

371.260 13

พ 639 ก

จ. 1

การเปรียบเทียบความเที่ยงตรงและความคลาดเคลื่อนมาตรฐานในการประมาณค่า
ความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีรูปแบบต่างกัน

ปริญญานิพนธ์

ของ

พชญา เกตุสกุล

19 S.ศ. 2539

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา

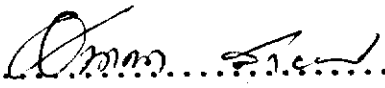
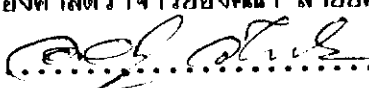
พฤษภาคม 2539

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

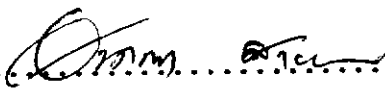
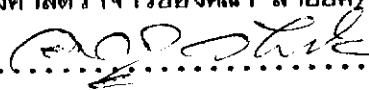
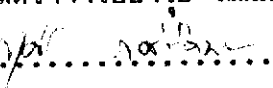
B. 52801

คณะกรรมการควบคุมและคณะกรรมการสอบได้พิจารณาปฏิญานพนธ์ฉบับนี้แล้ว
เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาการศึกษามหาบัณฑิต
วิชาเอกการวัดผลการศึกษา ของมหาวิทยาลัยศรีนครินทรวิโรฒได้

คณะกรรมการควบคุม

..........ประธาน
(รองศาสตราจารย์อังคณา สายยศ)
..........กรรมการ
(รองศาสตราจารย์อาวุธ วัฒนสิน)

คณะกรรมการสอบ

..........ประธาน
(รองศาสตราจารย์อังคณา สายยศ)
..........กรรมการ
(รองศาสตราจารย์อาวุธ วัฒนสิน)
..........กรรมการที่แต่งตั้งเพิ่มเติม
(รองศาสตราจารย์ชุตี วงศ์รัตนะ)

บัณฑิตวิทยาลัยอนุมัติให้รับปฏิญานพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา ของมหาวิทยาลัยศรีนครินทรวิโรฒ

..........คณบดีบัณฑิตวิทยาลัย
(ดร.ศิริภา นุลสุวรรณ)

วันที่ 10 เดือน พฤษภาคม พ.ศ. 2539

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงได้ด้วยดี เพราะได้รับความช่วยเหลือในการให้คำปรึกษา
ข้อคิดเห็น ตลอดจนคำแนะนำต่าง ๆ อย่างดียิ่ง จากรองศาสตราจารย์อังคณา สายยศ และรอง-
ศาสตราจารย์อาวุธ วัฒนสิน กรรมการที่ปรึกษาปริญญานิพนธ์ ผู้วิจัยขอกราบขอบพระคุณเป็นอย่าง
สูง และขอกราบขอบพระคุณรองศาสตราจารย์สุศรี วงศ์รัตน์ ที่ได้กรุณาให้ข้อเสนอแนะเพิ่มเติม
จึงทำให้ปริญญานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น

ขอขอบพระคุณอาจารย์รุ่งฤดี ทวีพันธ์ิ สำนึกทดสอบทางการศึกษา กรมวิชาการ
กระทรวงศึกษาธิการ ที่กรุณาให้ความอนุเคราะห์ในการวิเคราะห์ข้อมูลด้วยเครื่องคอมพิวเตอร์

ขอขอบพระคุณอาจารย์อรินทร์ น่วมถนอม อาจารย์ทองอยู่ สาระ อาจารย์สุนันท์ จิมวิชัย
อาจารย์กัญญา โทดำมา และอาจารย์เตือนใจ อุ่นอนันต์ ที่กรุณาเป็นผู้เชี่ยวชาญตรวจสอบเครื่อง
มือในการวิจัย

ขอขอบพระคุณผู้อำนวยการ ผู้ช่วยผู้อำนวยการฝ่ายวิชาการ และคณะครูโรงเรียนในกลุ่ม
ตัวอย่าง รวมทั้งนักเรียนทุกคนที่ได้ให้ความร่วมมือในการเก็บข้อมูลเป็นอย่างดี

ขอขอบคุณเพื่อน ๆ วัดพล 3๒ รุ่นพี่ และรุ่นน้อง ที่มีส่วนช่วยให้ปริญญานิพนธ์ฉบับนี้สำเร็จ
ลุล่วงได้ด้วยดี

สุดท้ายนี้ผู้วิจัยขอโน้มรำลึกถึงพระคุณของบิดา มารดา ครู-อาจารย์ ตลอดจนญาติพี่น้องที่
ได้อบรมสั่งสอน ให้ความช่วยเหลือ และเป็นกำลังใจแก่ผู้วิจัยตลอดมา

พินญา เกตุสกุล

สารบัญ

บทที่	หน้า
1 บทนำ	1
ภูมิหลัง	1
ความมุ่งหมายของการศึกษาค้นคว้า	6
ความสำคัญของการศึกษาค้นคว้า.....	6
ขอบเขตของการศึกษาค้นคว้า	7
คำนิยามศัพท์เฉพาะ	7
2 เอกสารและงานวิจัยที่เกี่ยวข้อง	10
การทดสอบแบบเทเลอร์	10
ทฤษฎีการตอบข้อสอบ	19
การประมาณค่าพารามิเตอร์ของทฤษฎีการตอบข้อสอบ.....	26
การวิเคราะห์ความสอดคล้องกับโมเดล	27
คุณภาพของแบบทดสอบ	29
ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ.....	31
งานวิจัยที่เกี่ยวข้อง	32
สมมติฐานของการศึกษาค้นคว้า	37
3 วิธีดำเนินการศึกษาค้นคว้า	38
ประชากร	38
กลุ่มตัวอย่าง	38
เครื่องมือที่ใช้ในการศึกษาค้นคว้า	39
วิธีดำเนินการสร้างแบบทดสอบ	40
วิธีดำเนินการเก็บรวบรวมข้อมูล	47
การวิเคราะห์ข้อมูลและสถิติที่ใช้	48

4 ผลการวิเคราะห์ข้อมูล.....	51
สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล	51
การเสนอผลการวิเคราะห์ข้อมูล	52
ผลการวิเคราะห์ข้อมูล	52
5 สรุปผล อภิปราย และข้อเสนอแนะ	59
ความมุ่งหมายของการศึกษาค้นคว้า	59
กลุ่มตัวอย่าง	59
เครื่องมือที่ใช้ในการศึกษาค้นคว้า	59
วิธีดำเนินการเก็บรวบรวมข้อมูล	60
การเสนอผลการวิเคราะห์ข้อมูล	60
สรุปผลการวิเคราะห์ข้อมูล	61
อภิปรายผล	61
ข้อเสนอแนะ	64
บรรณานุกรม	66
ภาคผนวก	74
ประวัติย่อของผู้วิจัย	116

บัญชีตาราง

ตาราง

หน้า

1	จำนวนกลุ่มตัวอย่างที่ใช้ในการวิจัย จำแนกตามขนาดโรงเรียน จำนวนห้องเรียน และจำนวนนักเรียน.....	39
2	ค่าสถิติพื้นฐานของแบบทดสอบแยกกลุ่มทั้ง 3 รูปแบบ.....	53
3	ค่าความเที่ยงตรงและผลการทดสอบความแตกต่างของค่าความเที่ยงตรง ในการประมาณค่าความสามารถของผู้สอบ จากการทำแบบทดสอบ ที่แตกต่างกัน 3 รูปแบบ.....	54
4	ผลการทดสอบความแตกต่างของค่าความเที่ยงตรงในการประมาณค่า ความสามารถของผู้สอบ จากการทำแบบทดสอบที่แตกต่างกัน 3 รูปแบบ โดยทดสอบเป็นรายคู่.....	55
5	ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของ ผู้สอบ.....	56
6	ผลการทดสอบความแตกต่างของค่าความคลาดเคลื่อนมาตรฐานในการ ประมาณค่าความสามารถของผู้สอบ จากการทำแบบทดสอบที่แตกต่าง กัน 3 รูปแบบ.....	57
7	ผลการทดสอบความแตกต่างของค่าความคลาดเคลื่อนมาตรฐานในการ ประมาณค่าความสามารถของผู้สอบ จากการทำแบบทดสอบที่แตกต่าง กัน 3 รูปแบบ โดยทดสอบเป็นรายคู่.....	58
8	ค่าไอเกน (มากกว่า 1) เพอร์เซ็นต์ของความแปรปรวน และ เพอร์เซ็นต์ของความแปรปรวนสะสม จากการนำแบบทดสอบ จำนวน 4 ฉบับ มาวิเคราะห์องค์ประกอบ.....	76
9	ผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (b) ค่าความ คลาดเคลื่อนมาตรฐานในการประมาณค่าความยากของข้อสอบ (SEE) และค่าความสอดคล้องกับโมเดล (t - test) ของแบบทดสอบ ฉบับที่ 1.....	80

10	ผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (b) ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความยากของข้อสอบ (SEE) และค่าความสอดคล้องกับโมเดล (t - test) ของแบบทดสอบฉบับที่ 2.....	82
11	ผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (b) ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความยากของข้อสอบ (SEE) และค่าความสอดคล้องกับโมเดล (t - test) ของแบบทดสอบฉบับที่ 3.....	84
12	ผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (b) ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความยากของข้อสอบ (SEE) และค่าความสอดคล้องกับโมเดล (t - test) ของแบบทดสอบฉบับที่ 4.....	86
13	คะแนนสอบวิชาคณิตศาสตร์ (y) กับผลการประมาณค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน (PRT) แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง (RRT) และแบบทดสอบแยกกลุ่มสองครั้ง (DRT).....	89
14	ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน (SEE _{PRT}) แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง (SEE _{RRT}) และแบบทดสอบแยกกลุ่มสองครั้ง (SEE _{DRT}).....	100
15	ตารางวิเคราะห์เนื้อหา พฤติกรรม และจุดประสงค์เชิงพฤติกรรมวิชาคณิตศาสตร์ (ค 204) ชั้นมัธยมศึกษาปีที่ 2 เรื่อง สมการและอสมการตามหลักสูตรมัธยมศึกษาตอนต้น พุทธศักราช 2521 (ฉบับปรับปรุง พ.ศ. 2533)	112

บัญชีภาพประกอบ

ภาพประกอบ

หน้า

1	แสดงโครงสร้างของแบบทดสอบสองขั้นตอน ที่มีแบบทดสอบแยกกลุ่มมีค่า ความยากของข้อสอบใกล้เคียงกัน.....	14
2	แสดงโครงสร้างของแบบทดสอบสองขั้นตอน ที่มีแบบทดสอบแยกกลุ่มมีค่า ความยากของข้อสอบกระจายในช่วงกว้าง.....	15
3	แสดงโครงสร้างของแบบทดสอบสองขั้นตอน ที่มีแบบทดสอบแยกกลุ่ม สองครั้ง.....	16
4	โด่งลักษณะข้อสอบของ ราล์ฟ โมเดล.....	24
5	แผนผังแสดงขั้นตอนการดำเนินการสร้างแบบทดสอบ.....	40

บทนำ

การวัดผลการศึกษา เป็นการวัดทางพฤติกรรมศาสตร์ หรือลักษณะทางจิตวิทยาของมนุษย์ สิ่งที่ต้องการวัดส่วนมากเป็นคุณลักษณะภายใน (Latent Traits) เช่น ความสามารถ ความถนัด และเจตคติ เป็นต้น สิ่งเหล่านี้ไม่สามารถวัดได้โดยตรงเหมือนกับการวัดทางกายภาพ ต้องอาศัยการวัดทางอ้อม โดยใช้ข้อคำถามในแบบทดสอบเป็นสื่อ เพื่อกระตุ้นให้บุคคลแสดงพฤติกรรมออกมาด้วยการตอบแบบทดสอบ แล้วจึงอนุมานผลการตอบจากคะแนนที่ได้ไปอธิบาย หรือทำนายคุณลักษณะภายในของบุคคล (สง. ลักษณะ. 2525 : 47 ; Lord. 1980 : 3) การวัดคุณลักษณะภายในโดยอาศัยการวัดทางอ้อมในลักษณะนี้ มักมีความคลาดเคลื่อนในการวัดรวมอยู่ด้วย ซึ่งทำให้ข้อสรุปที่ได้ขาดความถูกต้อง ดังนั้นนักวัดผลจึงได้พยายามสร้างเครื่องมือ และพัฒนาทฤษฎีการวัดต่าง ๆ ขึ้นมา เพื่อที่จะอธิบายหรือทำนายคุณลักษณะที่แท้จริงของบุคคล ให้มีความถูกต้อง และเชื่อถือได้มากที่สุด ทฤษฎีการวัดที่นิยมใช้กันอย่างแพร่หลายมาตั้งแต่อดีตจนถึงปัจจุบันก็คือทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory : CTT)

ทฤษฎีการทดสอบแบบดั้งเดิมเป็นทฤษฎีที่เน้นการประมาณค่าคะแนนจริงหรือความสามารถที่แท้จริงโดยอาศัยคะแนนของกลุ่มผู้สอบเป็นสำคัญ ทำให้ได้รับการวิจารณ์ว่าเป็นทฤษฎีที่ตั้งอยู่บนข้อตกลงเบื้องต้นที่อ่อนแอไม่มีเหตุผลเพียงพอในการประมาณค่าความสามารถที่แท้จริง ซึ่งมีจุดอ่อนหลายประการ ดังนี้ (Hambleton and Swaminathan. 1985 : 1 - 3)

ประการแรก ค่าสถิติของข้อสอบ ได้แก่ ค่าความยาก และค่าอำนาจจำแนกของข้อสอบ มีค่าไม่คงที่ จะแปรเปลี่ยนไปตามกลุ่มตัวอย่างของผู้สอบที่มีความสามารถแตกต่างกัน ดังนั้นการนำค่าสถิติของข้อสอบดังกล่าวมาใช้ประโยชน์จะมีข้อจำกัด โดยจะใช้ได้กับกลุ่มผู้สอบที่มีลักษณะเหมือนกับกลุ่มตัวอย่างที่นำมาใช้ในการประมาณค่าสถิติของข้อสอบเท่านั้น

ประการที่สอง การเปรียบเทียบความสามารถของผู้สอบแต่ละคน ถูกจำกัดอยู่กับสถานการณ์ของการทดสอบ กล่าวคือ จะต้องใช้แบบทดสอบที่มีข้อคำถามเหมือนกัน หรือเป็นแบบทดสอบคู่ขนานกันเท่านั้น ถ้าใช้ข้อคำถามไม่เหมือนกันจะไม่สามารถนำคะแนนมาเปรียบเทียบกันได้ เพราะคะแนนของผู้สอบแต่ละคนขึ้นอยู่กับค่าสถิติของข้อสอบ

ประการที่สาม ความเชื่อมั่นของแบบทดสอบ ที่นิยามในลักษณะของแบบทดสอบคู่ขนานนั้น เป็นไปได้ยากที่จะสัมฤทธิ์ผลในทางปฏิบัติ เนื่องจากไม่สามารถสร้างแบบทดสอบให้คู่ขนานกันได้อย่างแท้จริง และผลการตอบของผู้สอบอาจไม่เหมือนเดิมในการตอบครั้งที่สอง

ประการที่สี่ ไม่มีพื้นฐานในการที่จะบอกว่าผู้สอบคนหนึ่งมีความน่าจะเป็น ในการตอบข้อสอบข้อหนึ่งได้ถูกต้องเพียงใด ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมีประโยชน์ในการปรับปรุงแบบทดสอบให้เหมาะสมกับระดับความสามารถของผู้สอบ ซึ่งจะทำได้วัดความสามารถของผู้สอบได้อย่างแท้จริง

ประการที่ห้า การกำหนดข้อตกลงเบื้องต้น ที่เกี่ยวกับความแปรปรวนของความคลาดเคลื่อนในการวัดของผู้สอบทุกคนมีค่าเท่ากัน ซึ่งขัดแย้งกับลอร์ดและโนวิก (Lord and Novick. 1968 : 183) ที่พบว่า ความคลาดเคลื่อนมาตรฐานในการวัดสำหรับผู้สอบ ที่มีความสามารถปานกลางจะมีค่าน้อยกว่าความคลาดเคลื่อนมาตรฐานในการวัดสำหรับผู้สอบที่มีความสามารถสูงหรือต่ำ

นอกจากนี้ ทฤษฎีการทดสอบแบบดั้งเดิมยังไม่สามารถให้คำตอบที่ชัดเจน ในการแก้ปัญหาเกี่ยวกับการทดสอบอีกหลายอย่างเช่น การหาความลำเอียงของข้อสอบ (Item Bias) การเทียบคะแนนการสอบ (Test Score Equating) การจัดข้อสอบตามระดับความสามารถ (Tailored Testing) และการตัดสินสภาพการรอบรู้ (Mastery Testing) เป็นต้น จากปัญหาต่าง ๆ ดังที่กล่าวมาแล้วนี้ นักวัดผลจึงได้พยายามพัฒนาทฤษฎีการวัดขึ้นมาใหม่เพื่อที่จะแก้ปัญหาดังกล่าวทฤษฎีที่น่าสนใจก็คือทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) หรือทฤษฎีที่มีชื่อเรียกต่าง ๆ กัน เช่น ทฤษฎีคุณลักษณะภายใน (Latent Trait Theory) ทฤษฎีโค้งลักษณะข้อสอบ (Item Characteristic Curve Theory) หรือ ทฤษฎีการทดสอบแบบใหม่ (Modern Test Theory) เป็นต้น (ผจงจิต อินทสุวรรณ. 2525 : 54 ; Hambleton and others. 1978 : 468)

ทฤษฎีการตอบข้อสอบ เป็นทฤษฎีที่ใช้โมเดลทางคณิตศาสตร์อธิบายความสัมพันธ์ระหว่างการตอบข้อสอบ (Test Performance) กับระดับความสามารถของผู้สอบ (Ability) โดยใช้พารามิเตอร์ของข้อสอบ (Item Parameter) และพารามิเตอร์ของผู้สอบ (Examinee Parameter) เป็นตัวกำหนดลักษณะของข้อสอบเพื่อที่จะอธิบายหรือทำนายความน่าจะเป็นที่ผู้สอบคนหนึ่งจะตอบข้อสอบข้อหนึ่งได้ถูกต้อง โมเดลของทฤษฎีการตอบข้อสอบสามารถแสดงได้ในรูปของฟังก์ชันทางคณิตศาสตร์ที่เรียกว่าโค้งลักษณะข้อสอบ (Item Characteristic Curve) ซึ่งมีการพัฒนา

ชั้นหลายโมเดลด้วยกัน เช่น นอร์มัล ออจีฟ โมเดล (Normal Ogive Model) โลจิสติก โมเดล (Logistic Model) และ ราสช์ โมเดล (Rasch Model) เป็นต้น แต่ละโมเดลนี้จะมีจำนวนพารามิเตอร์ของข้อสอบแตกต่างกันไป กล่าวคือ แบบพารามิเตอร์หนึ่งตัวจะมีเฉพาะพารามิเตอร์ค่าความยาก (Difficulty) เท่านั้นแบบพารามิเตอร์สองตัวจะเพิ่มพารามิเตอร์ค่าอำนาจจำแนก (Discrimination) เข้าไปอีกตัวหนึ่ง และแบบพารามิเตอร์สามตัวจะมีทั้งพารามิเตอร์ค่าความยาก พารามิเตอร์ค่าอำนาจจำแนก และพารามิเตอร์ค่าการเดา (Guessing Parameter) สำหรับการเลือกโมเดลของทฤษฎีการตอบข้อสอบมาใช้นั้น จะต้องเลือกให้เหมาะสมกับสถานการณ์ต่าง ๆ ตามข้อตกลงเบื้องต้นของแต่ละโมเดล ถ้าเลือกใช้ได้ถูกต้องแล้วจะสามารถแก้ปัญหาที่เกิดจากทฤษฎีการทดสอบแบบดั้งเดิมได้ โดยทฤษฎีนี้มีจุดเด่นหลายประการ ดังนี้

ประการแรก ค่าพารามิเตอร์ของข้อสอบเป็นอิสระจากกลุ่มผู้สอบ ที่ใช้ในการประมาณค่า กล่าวคือ ค่าพารามิเตอร์ของข้อสอบแต่ละข้อ ได้แก่ ค่าความยาก ค่าอำนาจจำแนก และค่าการเดา เป็นค่าคงที่ ไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบ ถึงแม้ว่าผู้สอบต่างกลุ่มกัน ค่าพารามิเตอร์ของข้อสอบจะมีค่าคงเดิมเสมอ และในทำนองเดียวกัน ค่าพารามิเตอร์ความสามารถของผู้สอบเป็นอิสระจากข้อสอบที่ใช้ในการประมาณค่า จะไม่แปรเปลี่ยนไปตามค่าพารามิเตอร์ของข้อสอบ ไม่ว่าผู้สอบจะทำข้อสอบข้อใด หรือได้ทำข้อสอบเมื่อใดก็ตาม (Hambleton and Swaminathan. 1985 : 11)

ประการที่สอง การเปรียบเทียบความสามารถของผู้สอบ จะไม่ขึ้นอยู่กับข้อคำถามในแบบทดสอบ ถึงแม้ว่าผู้สอบจะใช้ข้อคำถามต่างกัน ก็สามารถนำค่าความสามารถมาเปรียบเทียบกันได้ ทั้งนี้เนื่องจากค่าความสามารถที่ประมาณค่ามาได้นั้น เป็นคะแนนโลจิท (Logit) ซึ่งอยู่ในมาตราวัดเดียวกัน (Common Ability Scale) (Hambleton and Cook. 1977 : 90 - 91)

ประการที่สาม การรายงานคุณภาพของข้อสอบในรูปค่าอินฟอร์เมชัน (Information) สามารถรายงานได้ทั้งเป็นรายข้อ และทั้งฉบับ ค่าอินฟอร์เมชันเป็นตัวบ่งชี้ถึงความถูกต้องแม่นยำ (Accuracy) ในการประมาณค่าสามารถ สามารถนำมาใช้แทนค่าความเชื่อมั่น และค่าความคลาดเคลื่อนมาตรฐานในการวัดได้ (Hambleton and Cook. 1977 : 84)

จากคุณลักษณะดังกล่าว จึงทำให้แนวคิดผลนำทฤษฎีการตอบข้อสอบไปประยุกต์ใช้ในการวัดผลทางการศึกษาและจิตวิทยาในหลาย ๆ เรื่องด้วยกัน เช่น การสร้างคลังข้อสอบ (Item Banking) การพัฒนาแบบทดสอบ (Test Development) การหาความลำเอียงของข้อสอบ

(Item Bias) การทดสอบแบบอิงเกณฑ์ (Criterion - Reference Testing) การเทียบคะแนนการสอบ (Test Score Equating) และการทดสอบแบบเจาะจง (Tailored Testing) เป็นต้น (Hambleton and others. 1978 : 492 - 499)

การทดสอบแบบเจาะจง เป็นผลจากการนำทฤษฎีการตอบข้อสอบมาประยุกต์ใช้ในการทดสอบ โดยนำเทคนิคทางด้านคอมพิวเตอร์เข้ามาดำเนินการสอบ เพื่อแก้ปัญหาการทดสอบแบบดั้งเดิม (Conventional Testing) ที่ใช้กันมาตั้งแต่ในอดีตจนถึงปัจจุบัน ซึ่งนักวิจัยเชื่อกันว่าวิธีการทดสอบแบบดั้งเดิม ไม่มีประสิทธิภาพไม่สามารถวัดสมรรถภาพที่แท้จริงได้ (สุธรรม จันทร์หอม. 2529 : 10) เนื่องจากการทดสอบแบบดั้งเดิมนั้น จะนำแบบทดสอบไปทดสอบกับผู้สอบเป็นกลุ่ม ผู้สอบทุกคนต้องทำข้อสอบทุกข้อ ในแบบทดสอบจะประกอบด้วยกลุ่มข้อสอบที่มีการจัดเรียงลำดับไว้แน่นอน และมีข้อสอบจำนวนมาก ทั้งนี้เพราะว่าข้อสอบที่มีความยาวเพิ่มขึ้นจะทำให้มีความเชื่อมั่นและความเที่ยงตรงสูง (อนันต์ ศรีโสภณ. 2525 : 99) ข้อสอบที่จัดไว้จะเป็นข้อสอบที่มีค่าความยากเหมาะสมกับกลุ่มผู้สอบที่มีความสามารถในระดับปานกลางเท่านั้น ไม่เหมาะสมกับกลุ่มผู้สอบที่มีความสามารถค่อนข้างสูงหรือต่ำ (ฉัตรภา พรหมมา. 2525 : 45) ดังนั้นผู้สอบที่มีความสามารถสูงหรือต่ำกว่าระดับปานกลางของกลุ่มมาก ๆ จะให้รายละเอียดเกี่ยวกับตัวผู้สอบต่ำ (Vale and Weiss. 1975 : 6) กล่าวคือ ถ้าผู้สอบคนใดได้ตอบข้อสอบที่มีค่าความยากเกินระดับความสามารถของเขาแล้ว ผู้สอบจะเกิดความท้อถอยมีแนวโน้มจะตอบข้อสอบโดยการเดา และถ้าผู้สอบคนใดได้ตอบข้อสอบที่มีค่าความยากต่ำกว่าระดับความสามารถของเขาแล้ว ผู้สอบจะเกิดความเบื่อหน่าย เพราะข้อสอบไม่ช่วยให้ผู้สอบได้ใช้ความสามารถอย่างเต็มที่ ผลกระทบเหล่านี้ทำให้การประมาณค่าความสามารถมีความคลาดเคลื่อน ความถูกต้องแม่นยำในการประมาณค่าความสามารถลดลง (Weiss. 1974 : 7) ดังนั้นนักวิจัยจึงได้พัฒนาการทดสอบแบบใหม่ขึ้นมา โดยการจัดข้อสอบให้มีค่าความยากเหมาะสมกับระดับความสามารถของผู้สอบเป็นรายบุคคล และเรียกวิธีการทดสอบในลักษณะนี้ว่า การทดสอบแบบเจาะจง (Tailored Testing) (Lord. 1980 : 151 ; citing Turnbull. 1951)

หลักการของการทดสอบแบบเจาะจงก็คือ ผู้สอบแต่ละคนไม่จำเป็นต้องทำข้อสอบเหมือนกันทุกข้อ จำนวนข้อไม่จำเป็นต้องเท่ากัน การตอบข้อสอบจะขึ้นอยู่กับรูปแบบและผลการตอบข้อสอบในข้อที่ผ่านมาของผู้สอบ (Weiss and Kingsbury. 1984 : 361) การคัดเลือกข้อสอบที่จะให้ผู้สอบได้ตอบในแต่ละข้อนั้นจะเป็นไปตามรูปแบบ และเงื่อนไขของการทดสอบแบบเจาะจงชนิด

ต่าง ๆ ซึ่งมียุทธวิธีในการสร้างและดำเนินการสอบ 2 ยุทธวิธีใหญ่ ๆ (Weiss. 1974 : 3) คือ ยุทธวิธีสองขั้นตอน (Two - Stage Strategies) และยุทธวิธีหลายขั้นตอน (Multi - Stage Strategies)

ยุทธวิธีสองขั้นตอนเป็นรูปแบบอย่างง่ายของการทดสอบแบบเทเลอร์จะใช้แบบทดสอบสองขั้นตอน (Two-Stage Test) ดำเนินการสอบทั้งประกอบด้วยแบบทดสอบสองชุด แบบทดสอบชุดแรกเรียกว่า แบบทดสอบแยกกลุ่ม (Routing Test) สำหรับผู้สอบในขั้นที่ 1 ซึ่งผู้สอบทุกคนต้องทำเหมือนกันเพื่อแยกระดับความสามารถไปทำข้อสอบในชุดที่สอง แบบทดสอบแยกกลุ่มนี้จะประกอบด้วยข้อสอบที่มีจำนวนข้อไม่มากนัก ถูกคัดเลือกขึ้นมาโดยยึดค่าความยากของข้อสอบเป็นหลัก ซึ่งมีด้วยกันหลายรูปแบบ เช่น แบบที่มีค่าความยากใกล้เคียงกัน (Peaked) แบบที่มีค่าความยากกระจายในช่วงกว้าง (Rectangular) และแบบที่มีแบบทดสอบแยกกลุ่มสองครั้ง (Double) เป็นต้น ส่วนแบบทดสอบชุดสองเป็นแบบทดสอบหลัก (Main Test) สำหรับผู้สอบในขั้นที่ 2 แบบทดสอบหลักนี้จะแยกเป็นแบบทดสอบฉบับย่อย ๆ (Subtest) ประมาณ 3 - 4 ฉบับ ตามระดับความยากของข้อสอบ ถ้าผู้สอบได้รับการประเมินจากแบบทดสอบแยกกลุ่มว่ามีความสามารถต่ำ ก็จะได้รับแบบทดสอบหลักฉบับที่ง่าย ถ้ามีความสามารถปานกลางก็จะได้รับแบบทดสอบหลักที่ยากปานกลาง และถ้ามีความสามารถสูงก็จะได้รับแบบทดสอบหลักฉบับที่ยาก (Lord. 1971 : 227)

ในปัจจุบันได้มีการนำเอาการทดสอบแบบเทเลอร์มาใช้ในการวัดผลการศึกษามากยิ่งขึ้น ทั้งนี้เชื่อกันว่ามีประสิทธิภาพสูงกว่าการทดสอบแบบดั้งเดิม ดังผลการวิจัยของ อรินทร์ น่วมถนอม (2537) แมคคอร์มิค (Mc Cormick D.J. 1978) ลินน์ และคนอื่น ๆ (Linn and others. 1969) เรคเคส (Reckase, M.D. 1978) ไวส์ส์ (Weiss. 1982) โมรีโน และคนอื่น ๆ (Moreno and others. 1984) พบว่าการทดสอบแบบเทเลอร์ให้ค่าความเที่ยงตรงและความเชื่อมั่นเท่ากับหรือมากกว่าการทดสอบแบบดั้งเดิม โดยที่ใช้จำนวนข้อสอบน้อยกว่าเดิม ทั้งนี้อาจเนื่องมาจากการทดสอบแบบเทเลอร์ สามารถกำจัดข้อสอบที่มีค่าความยากไม่เหมาะสมกับความสามารถของผู้สอบออกไป (Larkin and Weiss. 1974 : 1) ทำให้ลดจำนวนข้อสอบได้ถึงร้อยละ 50 (Weiss. 1974 : 362) เวลาที่ใช้ในการทดสอบก็น้อยลง (Spinetti and Hambleton. 1977 : 156) จึงส่งผลทำให้ประหยัดค่าใช้จ่ายในการดำเนินการทดสอบ (Urry. 1977 : 194) นอกจากนี้ยังทำให้ผู้สอบไม่เกิดความเบื่อหน่าย หรือเคร่งเครียดในการทำข้อสอบ ทำให้ผู้สอบมีทัศนคติที่ดีต่อการทดสอบแบบเทเลอร์ (Schmidt, Urry and Gugel. 1978 :

272)

จากที่กล่าวมาแล้วทั้งหมด จะเห็นว่าการทดสอบแบบเทเลอร์สามารถนำมาใช้ในการสอบ
แบบการทดสอบแบบดั้งเดิมได้เป็นอย่างดี แต่มีข้อยุ่งยากในการดำเนินการสอบเมื่อใช้กับการสอบ
โดยการเขียนตอบแบบเดิม (Papers and Pencil Test) ถึงแม้ว่าในปัจจุบันการวัดผลใน
เมืองไทยได้นำระบบคอมพิวเตอร์มาใช้บ้างแล้วก็ตาม โอกาสที่จะนำไปใช้อย่างแพร่หลายคงเป็น
เรื่องยาก แต่แบบทดสอบสองขั้นตอนเป็นรูปแบบอย่างง่ายของการทดสอบแบบเทเลอร์ มีวิธีการ
ไม่สลับซับซ้อนสามารถนำมาประยุกต์ใช้กับการเขียนตอบแบบเดิมได้ ผู้วิจัยจึงมีความสนใจที่จะนำ
แบบทดสอบสองขั้นตอนมาศึกษา โดยศึกษามุ่งเน้นในเรื่องแบบทดสอบแยกกลุ่มซึ่งมี 3 รูปแบบ คือ
แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายใน
ช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง การที่จะเลือกใช้แบบทดสอบแยกกลุ่มรูปแบบใดนั้น จะ
ต้องคำนึงถึงประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ ดังนั้นผู้วิจัยจึงได้ศึกษาการ
เปรียบเทียบความเที่ยงตรง และความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของ
ผู้สอบระหว่างแบบทดสอบแยกกลุ่มที่มีรูปแบบแตกต่างกัน โดยการศึกษาครั้งนี้ผู้วิจัยนำผลการทดสอบ
ไปประมาณค่าตามทฤษฎีการตอบข้อสอบ แบบราสช์ โมเดล (Rasch Model) โดยใช้โปรแกรม
ไบคาล (BICAL)

ความมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อเปรียบเทียบค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ
ระหว่างแบบทดสอบแยกกลุ่มที่มีรูปแบบต่างกัน
2. เพื่อเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ
ของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีรูปแบบต่างกัน

ความสำคัญของการศึกษาค้นคว้า

ผลการศึกษาในครั้งนี้ ทำให้ทราบถึงความเที่ยงตรงและความคลาดเคลื่อนมาตรฐานใน
การประมาณค่าความสามารถของผู้สอบ จากแบบทดสอบแยกกลุ่มที่มีรูปแบบต่าง ๆ กัน ซึ่งจะเป็น
ประโยชน์ในการพิจารณาเลือกรูปแบบของแบบทดสอบให้เหมาะสมกับระดับความสามารถของผู้สอบ
และยังช่วยทำให้ตัดสินใจเลือกใช้แบบทดสอบที่มีคุณภาพในการวัดผลสัมฤทธิ์ทางการเรียน เพื่อให้

ปรับปรุงการเรียนการสอนต่อไป

ขอบเขตของการศึกษาค้นคว้า

1. ประชากรที่ใช้ในการศึกษาค้นคว้านี้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ในภาคเรียนที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษาในจังหวัดลพบุรี จำนวน 23 โรงเรียน จำนวนห้องเรียน 147 ห้องเรียน จำนวนนักเรียน 6,024 คน

2. กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้านี้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ในภาคเรียนที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษาในจังหวัดลพบุรี จำนวน 7 โรงเรียน จำนวนห้องเรียน 11 ห้องเรียน จำนวนนักเรียน 419 คน ซึ่งได้จากวิธีการสุ่มแบบแบ่งชั้น (Stratified Random Sampling)

3. ตัวแปรที่ใช้ในการศึกษา

3.1 ตัวแปรอิสระ ได้แก่ รูปแบบของแบบทดสอบแยกกลุ่ม 3 รูปแบบ

3.1.1 แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน

3.1.2 แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง

3.1.3 แบบทดสอบแยกกลุ่มสองครึ่ง

3.2 ตัวแปรตาม ได้แก่

3.2.1 ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ

3.2.2 ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของ

ผู้สอบ

4. เนื้อหาที่ใช้ในการสร้างแบบทดสอบครั้งนี้ เป็นเนื้อหาวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 2 (ค 204) เรื่องสมการและอสมการ ตามหลักสูตรมัธยมศึกษาตอนต้น พุทธศักราช 2521 (ฉบับปรับปรุง พ.ศ. 2533)

คำนิยามศัพท์เฉพาะ

1. การทดสอบแบบเทเลอร์ (Tailored Testing) หมายถึง การทดสอบโดยใช้ชุดของข้อสอบ ที่มีการปรับความยากของข้อสอบให้เหมาะสมกับระดับความสามารถของผู้สอบเป็นรายบุคคล

2. แบบทดสอบสองขั้นตอน (Two - Stage Test) หมายถึง แบบทดสอบที่ประยุกต์มาจากแบบทดสอบดั้งเดิม แต่มีการจัดเรียงข้อสอบโดยยึดค่าความยากเป็นหลัก แบบทดสอบในรูปแบบนี้จะประกอบด้วยแบบทดสอบแยกกลุ่ม และแบบทดสอบหลัก ซึ่งมีรายละเอียดดังนี้

2.1 แบบทดสอบแยกกลุ่ม (Routing Test) เป็นแบบทดสอบขั้นแรกให้ผู้สอบทุกคนต้องสอบเหมือนกัน เพื่อที่จะจัดแยกระดับความสามารถของผู้สอบว่าผู้สอบคนใด จะได้ไปสอบแบบทดสอบหลักฉบับใดต่อไป ซึ่งแบบทดสอบแยกกลุ่มนี้จะประกอบด้วย

2.1.1 แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน มีค่าความยากอยู่ระหว่าง -0.50 และ 0.50

2.1.2 แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างมีค่าความยากอยู่ระหว่าง -2.00 และ 2.00

2.1.3 แบบทดสอบแยกกลุ่มสองครั้ง โดยในครั้งแรกจะมีค่าความยากอยู่ระหว่าง -0.50 และ 0.50 และครั้งที่สองจะมีค่าความยากอยู่สองช่วง ช่วงแรกมีค่าความยากอยู่ระหว่าง -0.90 และ -0.50 และช่วงที่สองมีค่าความยากอยู่ระหว่าง 0.50 และ 0.80

2.2 แบบทดสอบหลัก (Main Test) เป็นแบบทดสอบขั้นที่สองที่ผู้สอบจะต้องเลือกทำตามระดับความสามารถของผู้สอบ โดยมีแบบทดสอบฉบับย่อย ๆ ให้เลือกจำนวน 3 - 4 ฉบับ และจะจัดแยกฉบับตามค่าความยากของข้อสอบตั้งแต่ง่ายที่สุดจนถึงยากที่สุด

3. ค่าความสามารถของผู้สอบ (Examinee's Ability : θ) หมายถึง ระดับความสามารถของผู้สอบแต่ละคน ซึ่งวัดในหน่วย θ ที่มีค่าเฉลี่ยเท่ากับ 0 และ ค่าเบี่ยงเบนมาตรฐานเท่ากับ 1 สำหรับค่าความสามารถมีค่าเป็นไปได้ตั้งแต่ $-\infty$ ถึง $+\infty$

4. ค่าความยากของข้อสอบ (Item Difficulty : b) หมายถึง ค่าความสามารถตรงจุดเปลี่ยนโค้งคุณลักษณะข้อสอบแต่ละข้อ (Item Characteristic Curve) ซึ่งเป็นระดับความสามารถที่จะตอบข้อสอบข้อนั้นได้ถูกต้อง ด้วยความน่าจะเป็น .50 เมื่อไม่มีการเดา

5. โค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC) หมายถึง กราฟของฟังก์ชันการตอบข้อสอบที่แสดงความสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับความน่าจะเป็นที่ผู้สอบ ซึ่งมีความสามารถระดับนั้น จะตอบข้อสอบข้อนั้นได้ถูกต้อง

6. ฟังก์ชันอินฟอร์เมชันของข้อสอบ (Item Information Function : $I(\theta, U_i)$) หมายถึง ความแม่นยำในการประมาณค่าความสามารถจริงจากผลการตอบข้อสอบนั้นๆ ซึ่งมีค่าเท่า

กับอัตราส่วนของกำลังสองของความชันโค้งลักษณะข้อสอบ ต่อค่าความแปรปรวนของข้อสอบข้อนั้น

7. ฟังก์ชันอินฟอร์เมชันของแบบทดสอบ (Test Information Function : $I(\theta)$) หมายถึง ผลรวมของค่าฟังก์ชันอินฟอร์เมชันของข้อสอบทุก ๆ ข้อในแบบทดสอบนั้น ซึ่ง ณ ระดับความสามารถใดที่ฟังก์ชันอินฟอร์เมชันของแบบทดสอบมีค่าสูง แสดงว่า แบบทดสอบนั้นมีคุณภาพดีมากในการประมาณค่าความสามารถ ณ ระดับนั้น

8. ค่าความเที่ยงตรงในการประมาณค่าความสามารถ การวิจัยครั้งนี้เป็นการหาความเที่ยงตรงเชิงสภาพ (Concurrent Validity) หมายถึง ดัชนีที่บ่งถึงคุณภาพของแบบทดสอบ ในการประมาณค่าความสามารถของผู้สอบได้ถูกต้อง ตรงกับสภาพที่เป็นจริงในปัจจุบัน ซึ่งคำนวณได้จากความสัมพันธ์ระหว่าง ความสามารถที่ได้จากการประมาณค่า กับคะแนนสอบปลายภาคในวิชาคณิตศาสตร์ โดยใช้สูตรสัมประสิทธิ์สหสัมพันธ์เพียร์สัน

9. ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (Standard Error of Ability Estimation : $SEE(\theta)$) หมายถึง ค่าส่วนเบี่ยงเบนมาตรฐานของความคลาดเคลื่อนในการประมาณค่าความสามารถ ซึ่งมีค่าเท่ากับรากที่สองของส่วนกลับของค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบ

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการศึกษาค้นคว้าครั้งนี้ ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องในเรื่องต่าง ๆ ดังนี้

1. การทดสอบแบบเทเลอร์
 - 1.1 หลักการของการทดสอบแบบเทเลอร์
 - 1.2 ยุทธวิธีของการทดสอบแบบเทเลอร์
 - 1.3 องค์ประกอบของการทดสอบแบบเทเลอร์
2. ทฤษฎีการตอบข้อสอบ
 - 2.1 หลักการของทฤษฎีการตอบข้อสอบ
 - 2.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบ
 - 2.3 โมเดลของทฤษฎีการตอบข้อสอบ
3. การประมาณค่าพารามิเตอร์ของทฤษฎีการตอบข้อสอบ
4. การวิเคราะห์ความสอดคล้องกับโมเดล
5. คุณภาพของแบบทดสอบ
6. ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ
7. งานวิจัยที่เกี่ยวข้อง
 - 7.1 งานวิจัยภายในประเทศ
 - 7.2 งานวิจัยต่างประเทศ

การทดสอบแบบเทเลอร์

1. หลักการของการทดสอบแบบเทเลอร์

การทดสอบแบบเทเลอร์ (Tailored Testing) หรือชื่อที่เรียกต่าง ๆ กัน เช่น การทดสอบที่ปรับระดับความสามารถ (Adaptive Testing) การทดสอบโดยใช้คอมพิวเตอร์ (Computerized Testing) การทดสอบเป็นรายบุคคล (Individualized Testing) การทดสอบแบบโปรแกรม (Programmed Testing) หรือการทดสอบที่มีการจัดเรียงลำดับของข้อสอบ (Sequential Item Testing) เป็นต้น (Lord. 1980 : 150 - 151) หลักการของการ

ทดสอบแบบเทเลอร์ก็คือพยายามปรับระดับความยากของข้อสอบ ให้มีความเหมาะสมกับระดับความสามารถของผู้สอบเป็นรายบุคคล (Hambleton and Swaminathan. 1985 : 296 - 297) ผู้สอบแต่ละคนไม่จำเป็นต้องทำข้อสอบเหมือนๆ กันทุกข้อ หรือจำนวนข้อของข้อสอบก็ไม่จำเป็นต้องเท่ากัน ทั้งนี้ขึ้นอยู่กับรูปแบบและผลการตอบข้อสอบในการทดสอบแบบเทเลอร์นั้น ๆ (Weiss and Kingsbury. 1984 : 361) โดยทั่วไปแล้ว ผู้ดำเนินการสอบไม่สามารถจะทราบระดับความสามารถของผู้สอบก่อนที่จะเริ่มการสอบ ดังนั้นจึงให้ผู้สอบเริ่มทำข้อสอบที่มีระดับความยากปานกลางก่อน ถ้าผลการตอบข้อสอบถูกข้อต่อไปจะได้ตอบข้อสอบที่มีระดับความยากมากขึ้น แต่ถ้าตอบข้อสอบข้อนั้นผิด ข้อต่อไปจะได้ตอบข้อสอบที่มีระดับความยากลดลงและการตอบข้อสอบข้อต่อไป ก็จะดำเนินการในลักษณะนี้ไปจนสิ้นสุดการทดสอบ (Green and others. 1984 : 347) การตอบข้อสอบในลักษณะดังกล่าวนี้ เป็นการปรับระดับความยากของข้อสอบให้มีความเหมาะสมกับระดับความสามารถของผู้สอบนั่นเอง

2. ยุทธวิธีของการทดสอบแบบเทเลอร์

การเลือกข้อสอบเพื่อดำเนินการทดสอบของการทดสอบแบบเทเลอร์แบ่งออกเป็น 2 ยุทธวิธีใหญ่ ๆ ดังนี้ (Weiss. 1974 : 78-110)

2.1 ยุทธวิธีสองขั้นตอน (Two - Stage Strategies)

2.2 ยุทธวิธีหลายขั้นตอน (Multi - Stage Strategies)

2.2.1 รูปแบบแยกทางคงที่ (Fixed - Branching Model)

2.2.1.1 รูปปิรามิดแบบขนาดขั้นคงที่ (Constant Step Size Pyramid)

2.2.1.2 รูปปิรามิดแบบขนาดขั้นแปรผัน (Variable Step Size Pyramid)

2.2.1.3 รูปปิรามิดแบบข้างตัด (Truncated Pyramid)

2.2.1.4 รูปปิรามิดแบบมีหลายข้อในแต่ละขั้น (Multiple - Item Pyramid)

2.2.1.5 รูปปิรามิดแบบให้น้ำหนักแก่ตัวเลือกของข้อสอบเพื่อแยกทาง (Pyramid Using Differential Response Option Branching)

2.2.1.6 แบบทดสอบเฟล็กซิเลเวล (Flexilevel Test)

2.2.1.7 แบบทดสอบปรับระดับขึ้น (Stradaptive Test)

2.2.2 รูปแบบแยกทางแปรผัน (Variable - Branching Model)

2.2.2.1 ยุทธวิธีของเบส์ (Bayesian Strategies)

2.2.2.2 ยุทธวิธีความเป็นไปได้สูงสุด (Maximum Likelihood Strategies)

ในการศึกษาคั้งนี้ ผู้วิจัยได้นำยุทธวิธีสองขั้นตอนมาทำการศึกษา ดังนั้นผู้วิจัยขอแนะนำเสนอรายละเอียดเฉพาะยุทธวิธีสองขั้นตอนเท่านั้น

ยุทธวิธีสองขั้นตอน

แบบทดสอบสองขั้นตอน (Two - Stage Test) จัดเป็นยุทธวิธีสองขั้นตอนของการทดสอบแบบเทเลอร์อย่างง่าย ๆ ไม่สลับซับซ้อนสามารถนำมาประยุกต์ใช้กับแบบทดสอบแบบเขียนตอบได้เป็นอย่างดี ไม่จำเป็นต้องใช้คอมพิวเตอร์ ดังนั้นการสอบโดยปกติแล้วแบบทดสอบสองขั้นตอนจะประกอบด้วย แบบทดสอบ 2 ชุด ชุดแรกเป็นแบบทดสอบแยกกลุ่ม (Routing Test) ชุดที่สองเป็นแบบทดสอบหลัก (Main Test) (Lord. 1971 : 227 - 228) ซึ่งมีรายละเอียดดังนี้

1. แบบทดสอบแยกกลุ่มเป็นแบบทดสอบขั้นตอนแรกของแบบทดสอบสองขั้นตอน ซึ่งผู้สอบทุกคนจะต้องสอบเหมือนกัน สอบเพื่อแยกระดับความสามารถของผู้สอบ ก่อนที่จะไปทำแบบทดสอบหลักในขั้นตอนที่สองซึ่งมีด้วยกันหลายฉบับ ผู้สอบแต่ละคนจะทำแบบทดสอบหลักฉบับใดนั้นจะขึ้นอยู่กับผลการตอบข้อสอบในแบบทดสอบแยกกลุ่ม โดยปกติแล้วแบบทดสอบแยกกลุ่มจะมีลักษณะเป็นแบบทดสอบสั้น ๆ อาจมีจำนวนข้อเพียง 10 - 15 ข้อเท่านั้น เพื่อความสะดวกรวดเร็วในการตรวจให้คะแนนที่จะประมาณระดับความสามารถของผู้สอบเบื้องต้น คะแนนที่ได้จากแบบทดสอบแยกกลุ่มนี้จะใช้เป็นเกณฑ์ในการแยกกลุ่มซึ่งมีด้วยกันหลายรูปแบบ เช่น มีค่าความยากใกล้เคียงกัน (Peaked) มีค่าความยากกระจายในช่วงกว้าง (Rectangular) หรือเป็นแบบทดสอบแยกกลุ่มสองครึ่ง (Doudle) เป็นต้น

2. แบบทดสอบหลักเป็นแบบทดสอบขั้นตอนที่สองของแบบทดสอบสองขั้นตอน โดยที่ผู้สอบทุกคนจะต้องผ่านการทำแบบทดสอบแยกกลุ่มมาแล้ว แบบทดสอบหลักนี้จะประกอบไปด้วยแบบทดสอบฉบับย่อย ๆ ประมาณ 3 - 5 ฉบับ โดยปกติแล้วจะมีฉบับละประมาณ 20 - 30 ข้อ จัดเรียงตามค่าความยากของข้อสอบโดยฉบับแรกจะเป็นแบบทดสอบอย่างง่าย ๆ และฉบับต่อ ๆ ไปก็จะยากขึ้น

ตามลำดับ ผู้สอบแต่ละคนอาจจะทำแบบทดสอบหลักฉบับเดียวกัน หรือต่างฉบับกันก็ได้ซึ่งขึ้นอยู่กับผลการประเมินความสามารถในแบบทดสอบแยกกลุ่ม กล่าวคือ ผู้สอบคนใดมีความสามารถในระดับต่ำจะต้องไปทำแบบทดสอบหลักฉบับที่ง่าย และผู้สอบคนใดมีความสามารถในระดับปานกลางจะต้องไปทำแบบทดสอบหลักฉบับที่มีความยากปานกลาง ส่วนผู้สอบคนใดมีความสามารถในระดับสูงจะต้องไปทำแบบทดสอบหลักฉบับที่ยากขึ้นตามลำดับ

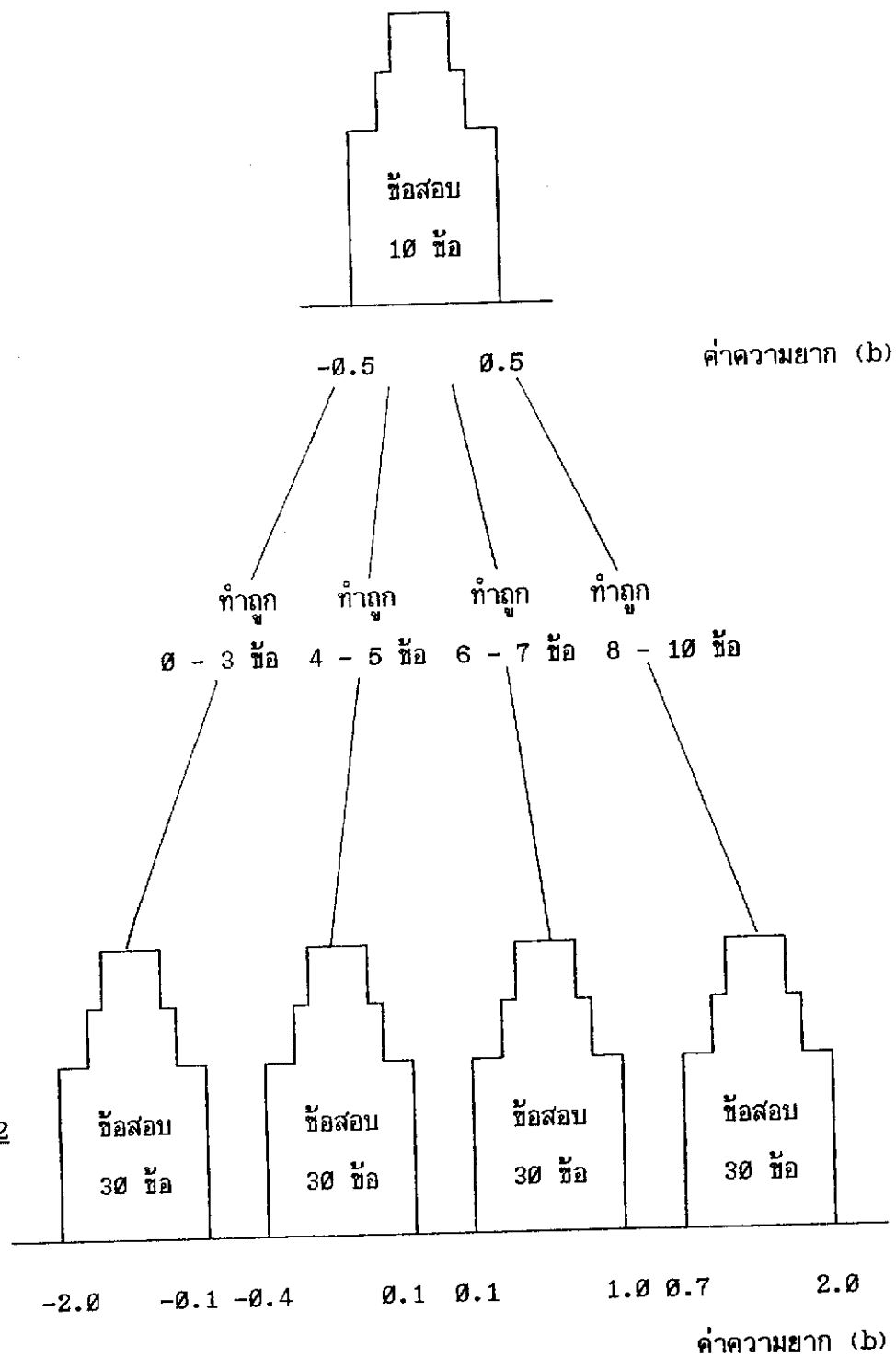
การสร้างแบบทดสอบฉบับย่อย ๆ ในแบบทดสอบหลักนั้น ลอร์ด (Lord. 1980 : 140-141) ได้เสนอแนะว่า แบบทดสอบหลักควรมีข้อสอบบางข้อที่มีค่าความยากของข้อสอบคาบเกี่ยวกันเป็นรายชื่อ (Overlap) เพื่อลดความคลาดเคลื่อนในการแยกระดับความสามารถของแบบทดสอบแยกกลุ่ม ซึ่งอาจเกิดขึ้นได้เมื่อผู้สอบได้คะแนนใกล้เคียงกับคะแนนจุดตัด สำหรับจำนวนข้อสอบที่มีค่าความยากคาบเกี่ยวกัน สามารถคำนวณได้โดยใช้สูตร ดังนี้ (Lord. 1980 : 140-141)

$$m = \frac{(n_1 L - n)}{(L - 1)}$$

เมื่อ m แทน จำนวนข้อสอบที่คาบเกี่ยวกัน
 L แทน จำนวนฉบับของแบบทดสอบหลัก
 n แทน จำนวนข้อสอบทั้งหมด
 n_1 แทน จำนวนข้อสอบแต่ละฉบับย่อย ๆ

โครงสร้างของแบบทดสอบสองขั้นตอน ที่มีรูปแบบต่าง ๆ กัน ดังที่กล่าวมาแล้ว มีลักษณะดังนี้ (Weiss. 1974 : 4, 6, 9)

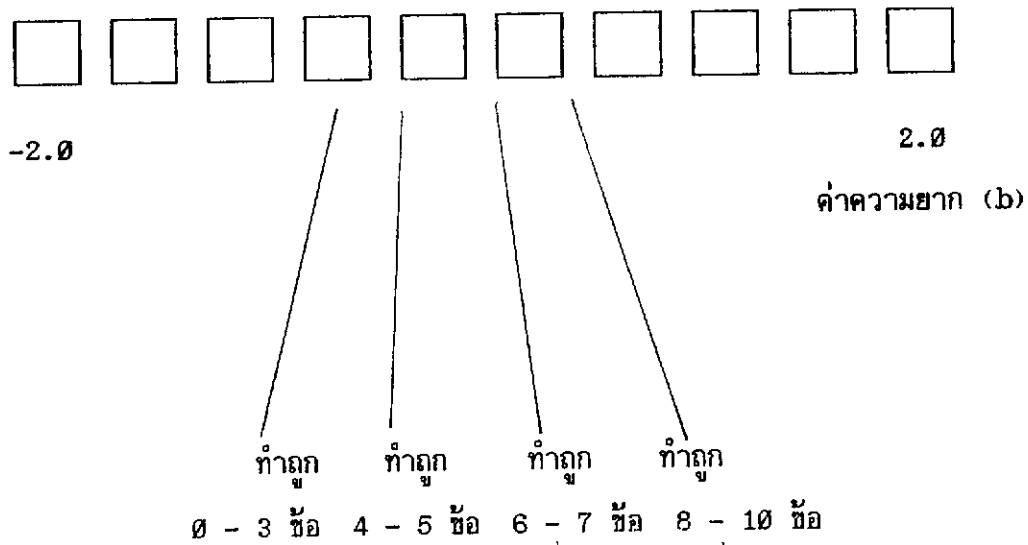
ขั้นที่ 1



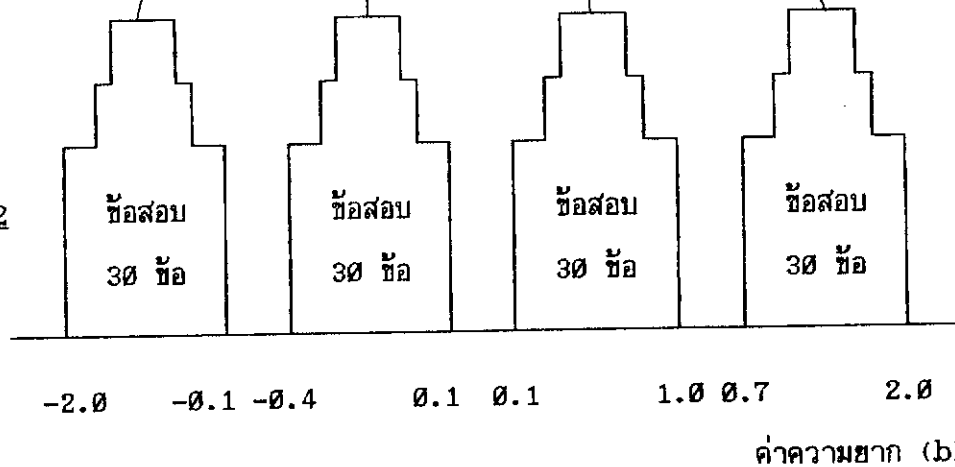
ภาพประกอบ 1 โครงสร้างของแบบทดสอบสองขั้นตอน ที่มีแบบทดสอบแยกกลุ่มมีค่าความยากใกล้เคียงกัน

ชั้นที่ 1

ข้อสอบ 10 ข้อ

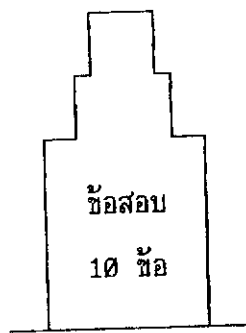


ชั้นที่ 2

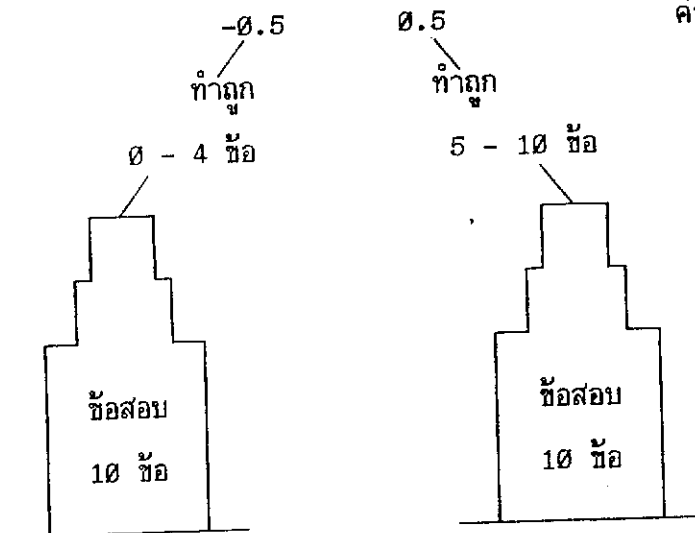


ภาพประกอบ 2 โครงสร้างของแบบทดสอบสองขั้นตอน ที่มีแบบทดสอบแยกกลุ่มมีค่าความยาก
กระจายในช่วงกว้าง

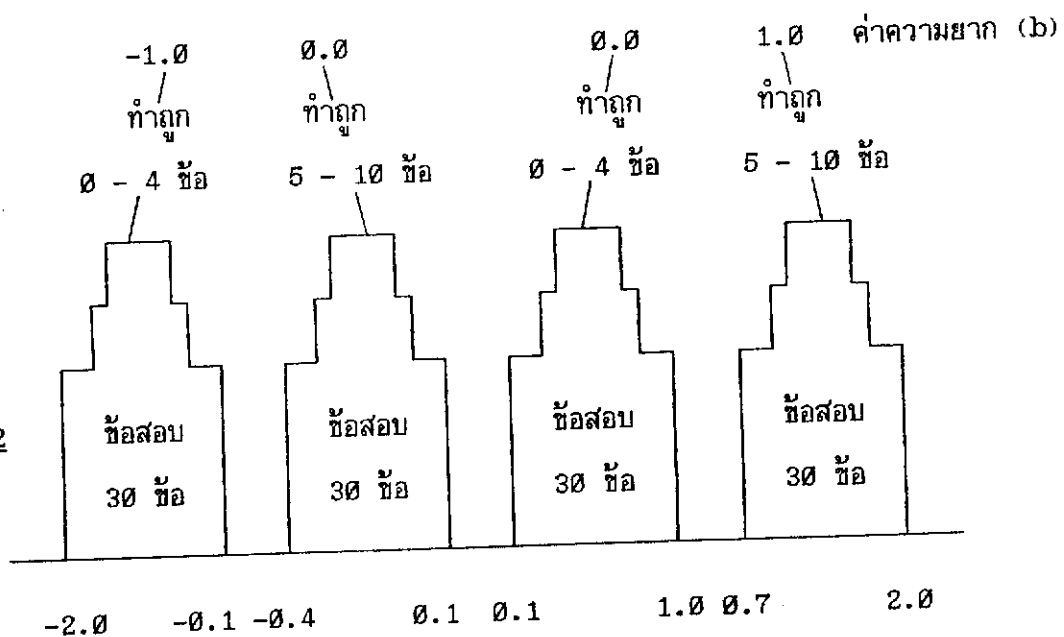
ชั้นที่ 1



ค่าความยาก (b)



ชั้นที่ 2



ค่าความยาก (b)

ภาพประกอบ 3 โครงสร้างของแบบทดสอบสองขั้นตอน ที่มีแบบทดสอบแยกกลุ่มสองครั้ง

3. องค์ประกอบของการทดสอบแบบเทเลอร์

องค์ประกอบต่าง ๆ ของการทดสอบแบบเทเลอร์ มีลักษณะดังนี้ (Weiss and Kingsbury. 1984 : 362 - 364)

3.1 โมเดลของการตอบข้อสอบ (Item Response Models) การทดสอบแบบเทเลอร์สามารถนำทฤษฎีการตอบข้อสอบที่มีพารามิเตอร์ 1, 2 หรือ 3 ตัว แบบโลจิสติก โมเดล (Logistic Model) หรือออร์มัล ออร์จีฟ โมเดล (Normal Ogive Model) มาประยุกต์ใช้ได้ การจะเลือกใช้โมเดลใดนั้นก็ควรพิจารณาธรรมชาติของข้อสอบ และควรมีความเหมาะสมกับข้อมูลที่ใช้ในการตอบข้อสอบจากการวิจัยของยูริ (Urry. 1977 : 184) พบว่าการทดสอบแบบเทเลอร์มีผลดีน้อยลงเมื่อใช้โมเดลที่มีพารามิเตอร์ไม่เพียงพอ การเลือกใช้พารามิเตอร์ 3 ตัวมีความเหมาะสมมากกว่าพารามิเตอร์ 2 ตัว หรือตัวเดียว

3.2 กลุ่มข้อสอบ (Item Pool) การทดสอบแบบเทเลอร์จะต้องใช้จำนวนข้อสอบมากพอ และค่าพารามิเตอร์ของข้อสอบที่ประมาณค่าตามทฤษฎีการตอบข้อสอบต้อง จัดให้อยู่ในมาตรวัดเดียวกัน และครอบคลุมในช่วงระดับความสามารถของประชากร จากการวิจัยของยูริ (Urry. 1977 : 184) พบว่า กลุ่มของข้อสอบที่มีคุณภาพนั้นจะต้องมีข้อสอบอย่างน้อย 100 ข้อ ค่าพารามิเตอร์ของข้อสอบที่ประกอบด้วย ค่าอำนาจจำแนกจะต้องมีค่ามากกว่า .8 ค่าความยากจะต้องอยู่ในช่วง -2.0 ถึง $+2.0$ และค่าการเดาจะต้องน้อยกว่า .3

3.3 ระดับในการเริ่มต้น (Entry Level) การทดสอบแบบเทเลอร์ ผู้สอบแต่ละคนอาจได้รับข้อสอบที่แตกต่างกัน ถ้าได้รับข้อสอบที่แตกต่างกันแล้วก็จะเริ่มทำข้อสอบที่มีระดับความยากต่างกันไปด้วย การเริ่มทำข้อสอบที่มีระดับความยากคลาดเคลื่อนไปจากระดับความสามารถจะไม่มีผลกระทบต่อประมาณค่าระดับความสามารถเท่าใดนัก แต่ถ้าผู้สอบได้ทำข้อสอบข้อแรกโดยตรงกับระดับความสามารถของเขา จะสามารถลดจำนวนข้อสอบที่จะใช้ในการทดสอบให้น้อยลง

3.4 กฎในการคัดเลือกข้อสอบ (Item Selection Rule) ในการทดสอบแบบเทเลอร์มีวิธีการคัดเลือกข้อสอบ 2 วิธี คือการคัดเลือกข้อสอบที่มีค่าอินฟอร์เมชันสูงสุด (Maximum Information Item Selection) และการคัดเลือกข้อสอบแบบเบย์ส์ (Bayesian Item Selection) การคัดเลือกข้อสอบที่มีค่าอินฟอร์เมชันสูงสุดจะเลือกใช้ข้อสอบที่ให้ค่าอินฟอร์เมชันสูงสุด ณ ระดับความสามารถของผู้สอบ และการคัดเลือกข้อสอบด้วยวิธีของเบย์ส์ นั้นสามารถลด

ความแปรปรวนภายหลังของการประมาณค่าความสามารถ ทั้งนี้เพราะว่าค่าอินโฟร์เมชันและความแปรปรวนภายหลังของเบส์เป็นฟังก์ชันที่มีความสัมพันธ์กัน

3.5 วิธีการให้คะแนน (Scoring Method) ในการประมาณค่าความสามารถของผู้สอบแต่ละคนจะใช้วิธีความเป็นไปได้สูงสุด (Maximum Likelihood Method) และวิธีของเบส์ (Bayesian Method) การประมาณค่าด้วยวิธีของเบส์เป็นที่นิยมใช้กันมาก ในการทดสอบแบบเทเลอร์ ทั้งนี้เพราะว่าสามารถประมาณค่าได้ค่อนข้างคงที่ถึงแม้ว่ามีข้อสอบเพียงข้อเดียวหรือในการที่ผู้สอบตอบข้อสอบถูกต้องทั้งหมด หรือผิดทั้งหมดก็สามารถประมาณค่าได้ แต่วิธีความเป็นไปได้สูงสุดไม่สามารถประมาณค่าในสถานการณ์เช่นนี้ได้ นอกจากนี้การให้คะแนนของการทดสอบแบบเทเลอร์ยังมีอีกหลายวิธี ดังนี้ (Weiss. 1974 : 7)

3.5.1 ให้คะแนนตามค่าความยากของข้อสอบที่ยากที่สุดที่ตอบถูก (The Most Difficulty of Item Answered Correctly) ข้อสอบข้อใดที่มีค่าความยากสูงสุดที่ผู้สอบได้ตอบถูก แสดงว่าเป็นค่าความสามารถของผู้สอบ

3.5.2 ให้คะแนนตามค่าเฉลี่ยความยากของข้อสอบที่ตอบถูก (Average Difficulty of All Item Answered Correctly) การให้คะแนนโดยวิธีนี้ต้องเอาค่าความยากของข้อสอบที่ผู้สอบตอบถูกต้องทั้งหมดในแบบทดสอบมาหาค่าเฉลี่ย ค่าเฉลี่ยความยากที่คำนวณได้ถือว่าเป็นความสามารถของผู้สอบ

3.5.3 ให้คะแนนตามค่าเฉลี่ยความยากของข้อสอบทุกข้อที่ตอบ (Average Difficulty) ค่าความยากเฉลี่ยของข้อสอบทุกข้อในแบบทดสอบที่ผู้สอบได้ตอบ แสดงว่าเป็นค่าความสามารถของผู้สอบ โดยไม่คำนึงว่าข้อสอบที่ตอบไปนั้นจะตอบถูกหรือผิด

3.5.4 ให้คะแนนตามค่าความยากของข้อสอบข้อสุดท้ายที่ตอบ (Difficulty of the Final Item) ค่าความยากของข้อสอบที่เป็นข้อสุดท้ายในแบบทดสอบซึ่งผู้สอบได้ตอบ แสดงว่าเป็นค่าความสามารถของผู้สอบ โดยไม่คำนึงว่าข้อสุดท้ายจะตอบถูกหรือผิด

3.5.5 ให้คะแนนตามค่าความยากของข้อสอบที่ต่อจากข้อสุดท้ายที่ตอบ (Difficulty of $(N + 1)^{th}$ Item) การให้คะแนนโดยวิธีนี้คล้ายกับวิธีที่ 3.5.4 แต่มีข้อแตกต่างตรงที่ถ้าข้อสุดท้ายที่ทำถูกคะแนนในข้อต่อไปจะเพิ่มขึ้น แต่ถ้าข้อสุดท้ายที่ทำผิดคะแนนในข้อต่อไปก็จะลดลง

ทฤษฎีการตอบข้อสอบ

1. หลักการของทฤษฎีการตอบข้อสอบ

ทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT) หรือทฤษฎีที่มีชื่อเรียกต่าง ๆ กัน เช่น ทฤษฎีคุณลักษณะภายใน (Latent Trait Theory) ทฤษฎีโค้งลักษณะข้อสอบ (Item Characteristic Curve Theory) หรือทฤษฎีการทดสอบแบบใหม่ (Modern Test Theory) เป็นต้น (ผจงจิต อินทสุวรรณ. 2525 : 54) หลักการของทฤษฎีการตอบข้อสอบก็คือ ใช้โมเดลทางคณิตศาสตร์อธิบายความสัมพันธ์ ระหว่างความสามารถที่มีอยู่ในตัวของผู้สอบ กับผลการสอบนั้น โดยที่ความสามารถของผู้สอบเป็นสิ่งที่เราไม่สามารถวัด หรือสังเกตได้โดยตรง แต่จะสามารถอธิบายหรือพยากรณ์ได้จากคะแนนการตอบของผู้สอบ ซึ่งเป็นสิ่งที่เราสามารถวัด หรือสังเกตได้ (Lord and Novick. 1968 : 358 ; Hambleton and Swaminathan. 1985 : 9) ความสัมพันธ์ระหว่างความสามารถกับผลการสอบดังที่กล่าวมาแล้วนี้ สามารถเขียนให้อยู่ในรูปของความสัมพันธ์ทั่ว ๆ ไปได้ดังนี้ (Lord. 1982 : 13)

$$P = f (U_1 / \theta_1, \theta_2, \theta_3, \dots, \theta_k ; \beta_1)$$

เมื่อ	P	แทน	ผลการสอบ
	f	แทน	ฟังก์ชัน
	U_1	แทน	ผลการตอบข้อสอบ (ตอบถูก $U_1 = 1$; ตอบผิด $U_1 = 0$)
	β_1	แทน	ค่าพารามิเตอร์ของข้อสอบที่ i
	$\theta_1, \theta_2, \dots, \theta_k$	แทน	ระดับความสามารถที่ 1, 2, ..., k

ความสัมพันธ์ดังกล่าวนี้เป็นความสัมพันธ์ในรูปทั่ว ๆ ไป นักวัดผลจึงได้พยายามกำหนดให้ออกมาเป็นฟังก์ชันทางคณิตศาสตร์ ที่มีชื่อเรียกว่า ฟังก์ชันลักษณะข้อสอบ (Item Response Function) หรือโค้งลักษณะข้อสอบ (Item Characteristic Curve) ซึ่งเป็นฟังก์ชันเพิ่มทางเดียว (Monotonically Increasing Function) (Hambleton and Swaminathan. 1985 : 13) ฟังก์ชันทางคณิตศาสตร์ดังกล่าวนี้มีด้วยกันหลายโมเดล ทั้งนี้จะขึ้นอยู่กับข้อตกลง

เบื้องต้นเกี่ยวกับทฤษฎีการตอบข้อสอบ ดังที่จะกล่าวต่อไปนี้

2. ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบ

ทฤษฎีการตอบข้อสอบ มีข้อตกลงเบื้องต้นทั่ว ๆ ไป 3 ประการ ดังนี้

2.1 ความเป็นมิติเดียว (Unidimension) ความสามารถ หรือคุณลักษณะของผู้สอบที่จะเป็นตัวกำหนดพฤติกรรมในการตอบข้อสอบข้อใด ๆ อาจมีได้หลายลักษณะ ลักษณะทั้งหมดรวมเรียกว่า ลาเท็นสเปส (Latent Space) โดยทั่วไปแล้วถ้าผลการตอบข้อสอบสามารถอธิบายได้เพียงความสามารถเดียวก็เพียงพอแล้ว เจอนไชในลักษณะนี้เรียกว่าความเป็นมิติเดียว (Hambleton and Swaminathan. 1985 : 16 - 17) และสามารถตรวจสอบความเป็นมิติเดียวของแบบทดสอบได้โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) เพื่อสังเกตค่าไอเกน (Eigen Value) ขององค์ประกอบที่มีค่าสูงสุด ว่าแตกต่างจากค่าอื่น ๆ อย่างชัดเจนหรือไม่ ในขณะที่ค่าไอเกนขององค์ประกอบที่เหลือมีค่าใกล้เคียงกัน (Warm. 1978 : 104 ; Lord. 1980 : 21) นอกจากนี้ยังมีวิธีง่าย ๆ ในการเลือกข้อสอบเพื่อทำให้เกิดความเป็นมิติเดียว ก็คือการเลือกข้อสอบที่มีค่าอำนาจจำแนกสูง ๆ (Green and others. 1984 : 349) และควรมีค่าอำนาจจำแนกตั้งแต่ 0.8 ขึ้นไป (Urry. 1977 : 184)

2.2 ความเป็นอิสระในการตอบข้อสอบ (Local Independence) หมายความว่า โอกาสในการตอบข้อสอบแต่ละข้อได้ถูกต้องเป็นอิสระจากกันแสดงว่า การตอบข้อสอบข้อใดข้อหนึ่งได้ถูกหรือผิด จะไม่มีผลต่อการตอบข้อสอบข้ออื่น ๆ (Hambleton and Swaminathan. 1985 : 23) สำหรับการตรวจสอบความเป็นอิสระในการตอบข้อสอบนั้น ให้สังเกตความเป็นมิติเดียวของแบบทดสอบ กล่าวคือ ถ้าแบบทดสอบฉบับใดมีความเป็นมิติเดียวแล้วแบบทดสอบฉบับนั้นจะมีความเป็นอิสระในการตอบข้อสอบด้วย (Warm. 1978 : 19 ; Lord. 1980 : 19)

2.3 โค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC) เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงความสัมพันธ์ ระหว่างโอกาสในการตอบข้อสอบข้อนั้นได้ถูกต้องกับระดับความสามารถที่วัดได้โดยใช้ชุดของข้อสอบหรือแบบทดสอบนั้น (Hambleton and Swaminathan. 1985 : 25) โค้งลักษณะข้อสอบมีด้วยกันหลายรูปแบบ แต่ละรูปแบบจะขึ้นอยู่กับโมเดลที่ใช้อธิบายความสัมพันธ์ เช่น ลาเท็นส์ดีซันทีนซ์ โมเดล (Latent Distance Model) โลจิสติก โมเดล (Logistic Model) และเกรด เรสพอน โมเดล (Graded Response Model) เป็นต้น

3. โมเดลของทฤษฎีการตอบข้อสอบ

โมเดลทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างความสามารถของผู้สอบกับผลการสอบ มีด้วยกันหลายโมเดล เมื่อแบ่งโมเดลตามลักษณะของข้อมูลหรือการตรวจให้คะแนนแล้ว สามารถแบ่งได้เป็น 3 ประเภทดังนี้ (Hambleton and Swaminathan. 1985 : 35)

3.1 โมเดลที่มีการตรวจให้คะแนนแบบสองค่า (Dichotomus) โดยถ้าตอบถูก ให้ 1 คะแนน และถ้าตอบผิดให้ 0 คะแนน ซึ่งแบ่งออกเป็นโมเดลต่าง ๆ ดังนี้

3.1.1 กัตต์แมน เพอร์เฟ็คท์ สเกล (Guttman Perfect Scale)

3.1.2 ลาเทนท์ ดีซ์ ทันทซ์ โมเดล (Latent Distance Model)

3.1.3 ลาเทนท์ ลิเนียร์ โมเดล (Latent Linear Model)

3.1.4 นอร์มัล ออจีฟ โมเดล ที่มีพารามิเตอร์ 1, 2, 3 ตัว

(One - , Two - , Three - Parameter Normal Ogive Model)

3.1.5 โลจิสติก โมเดล ที่มีพารามิเตอร์ 1, 2, 3, 4 ตัว (One - ,

Two - , Three - , Four - Parameter Logistic Model)

3.1.6 ราสช์ โมเดล (Rasch Model)

3.2 โมเดลที่มีการตรวจให้คะแนนแบบหลายค่า (Multi - Dichotomus) โดย ใช้กับข้อสอบที่ให้คะแนนหลายระดับ เช่น แบบทดสอบวัดเจตคติ แบบทดสอบวัดภาคปฏิบัติ และแบบ ทดสอบวัดบุคลิกภาพ เป็นต้น ซึ่งแบ่งออกเป็นโมเดลต่าง ๆ ดังนี้

3.2.1 นอมินอล เรสพอน โมเดล (Nominal Response Model)

3.2.2 เกรด เรสพอน โมเดล (Graded Response Model)

3.2.3 พาเชียล เครดิต โมเดล (Partial Credit Model)

3.3 โมเดลที่มีการตรวจให้คะแนนแบบต่อเนื่อง (Continuous) ได้แก่

คันทรีวอลล์ เรสพอน โมเดล (Continuous Response Model)

สำหรับในการศึกษาครั้งนี้ ผู้วิจัยได้เลือกใช้ ราสช์ โมเดล มาทำการวิเคราะห์ข้อมูล ดังนั้นผู้วิจัยจึงนำเสนอรายละเอียดเฉพาะ ราสช์ โมเดลเท่านั้น

ราล์ฟ โมเดล

ราล์ฟ โมเดล ถูกพัฒนาขึ้นโดยนักคณิตศาสตร์ชาวเดนมาร์ก ชื่อ จอร์จ ราล์ฟ (Georg Rasch) ราล์ฟได้มีแนวความคิดว่าควรจะมีโมเดลที่มีการวิเคราะห์ความเป็นอิสระจากแบบทดสอบ และจากสิ่งที่ถูกวัด (Husen and Postlethwaite, 1988 : 4177) หลักการของราล์ฟ ก็คือ ผู้สอบมีความสามารถจริง (Ability : β_v) แฝงอยู่ และข้อสอบก็มีความยาก (Difficulty : δ_i) แฝงอยู่ในแต่ละข้อ ดังนั้นพฤติกรรมคำตอบข้อสอบจึงขึ้นอยู่กับผลต่าง ระหว่างความสามารถของผู้สอบ (β_v) กับความยากของข้อสอบ (δ_i) ถ้าความสามารถของผู้สอบมีมากกว่าความยากของข้อสอบ ($\beta_v - \delta_i > 0$) โอกาสที่ผู้สอบจะตอบข้อสอบข้อนั้นได้ถูกต้อง จะมีค่ามากกว่า 0.5 ถ้าความสามารถของผู้สอบมีน้อยกว่าความยากของข้อสอบ ($\beta_v - \delta_i < 0$) โอกาสที่ผู้สอบจะตอบข้อสอบข้อนั้นได้ถูกต้อง จะมีค่าน้อยกว่า 0.5 หรือถ้าผู้สอบมีความสามารถพอ ๆ กับความยากของข้อสอบ ($\beta_v - \delta_i = 0$) โอกาสที่ผู้สอบจะตอบข้อสอบข้อนั้นได้ถูกต้องจะมีค่าครึ่งต่อครึ่ง (Wright and Stone, 1979 : 12) จากหลักการดังกล่าว แสดงว่า เครื่องมือและสิ่งที่ถูกวัดเป็นอิสระจากกัน ราล์ฟได้อธิบายโมเดลนี้ด้วยทฤษฎีความน่าจะเป็นที่มีการแจกแจงแบบปัวซอง (Poisson) ต่อมาได้มีการพัฒนาให้อยู่ในรูปโลจิสติก (Logistic) เพื่อทำให้ง่ายต่อการเข้าใจ และมีความสะดวกในการประมาณค่าพารามิเตอร์ต่าง ๆ จากหลักการดังกล่าวมาแล้ว สามารถเขียนเป็นโมเดลทางคณิตศาสตร์ได้ดังนี้ (Wright and Stone, 1979 : 15)

$$P[X_{v,i} = 1/\beta_v, \delta_i] = \frac{\exp(\beta_v - \delta_i)}{1 + \exp(\beta_v - \delta_i)}$$

เมื่อ $P[X_{v,i} = 1/\beta_v, \delta_i]$	แทน	ค่าความน่าจะเป็นที่ผู้สอบ ซึ่งมีระดับความสามารถ β จะตอบข้อสอบข้อที่ i ได้ถูกต้อง
β_v	แทน	ระดับความสามารถของผู้สอบคนที่ v
δ_i	แทน	ค่าความยากของข้อสอบข้อที่ i
exp	แทน	ค่า Exponential มีค่าเท่ากับ 2.7182818...

สำหรับ ราล์ช โมเดล มีลักษณะคล้ายกับ โลจิสติก โมเดล ที่มีพารามิเตอร์ 1 ตัว (One Parameter Logistic Model) ซึ่งพัฒนาขึ้นโดยเบิร์นบอม (Birnbbaum) ทั้งที่โมเดลทั้งสองนี้ถูกพัฒนาขึ้นมาอย่างอิสระจากกัน (Hambleton and Cook. 1977 : 82) โดยที่ราล์ช โมเดลนี้จะมีพารามิเตอร์ความยากของข้อสอบ (b) เพียงตัวเดียวเท่านั้น และโลจิสติก โมเดลที่มีพารามิเตอร์ 1 ตัวนั้นก็มีพารามิเตอร์ของข้อสอบที่ใช้อธิบายโมเดลเพียงตัวเดียวเหมือนกัน คือ พารามิเตอร์ความยากของข้อสอบ (b) ส่วนพารามิเตอร์อำนาจจำแนกของข้อสอบ (a) จะมีค่าเท่ากันทุกข้อ และพารามิเตอร์การเดาของข้อสอบ (c) มีค่าเป็นศูนย์ สามารถเขียนเป็นโมเดลทางคณิตศาสตร์ได้ดังนี้ (Hambleton and Swaminathan. 1985 : 47)

$$P_i(\theta) = \frac{\exp \frac{D\bar{a}(\theta - b_i)}{D\bar{a}(\theta - b_i)}}{1 + \exp \frac{D\bar{a}(\theta - b_i)}{D\bar{a}(\theta - b_i)}}$$

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

$\bar{a}$ แทน ค่าอำนาจจำแนกของข้อสอบ ซึ่งมีค่าเท่ากันทุกข้อ ($a = a_i$)

b_i แทน ค่าความยากของข้อสอบข้อที่ i

θ แทน ระดับความสามารถของผู้สอบ

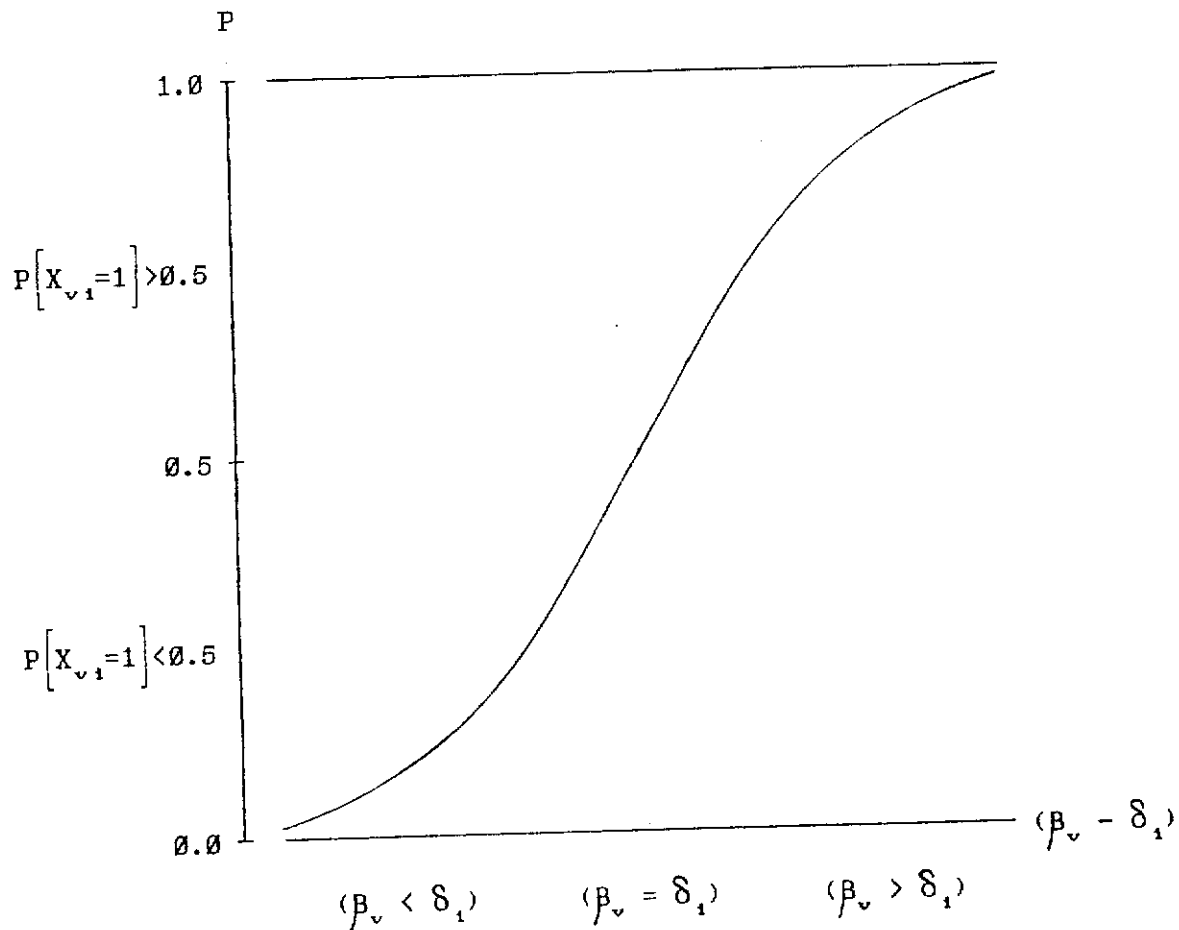
D แทน ค่าคงที่ (Scaling Constant) ซึ่งใช้ปรับโด่ง

ลักษณะของข้อสอบ มีค่าเท่ากับ 1.7

$\exp$ แทน ค่า Exponential มีค่าเท่ากับ 2.7182818...

เมื่อเปรียบเทียบระหว่าง ราล์ช โมเดล กับโลจิสติก โมเดล ที่มีพารามิเตอร์ 1 ตัว จะพบว่า ถ้าให้ $\beta_i = D\bar{a}\theta$ และ $\delta_i = D\bar{a}b_i$ แล้ว ราล์ชโมเดล จะเป็นโมเดลที่เหมือนกับโลจิสติก โมเดล นั่นเอง ดังนั้น ราล์ช โมเดล สามารถนำมาใช้แทน โลจิสติก โมเดล ที่มีพารามิเตอร์ 1 ตัวได้ สำหรับโด่งลักษณะข้อสอบของ ราล์ช โมเดล มีลักษณะดังนี้ (Wirght and Stone. 1979 : 14)

โอกาสที่จะตอบข้อสอบได้ถูกต้อง



ภาพประกอบ 4 โค้งลักษณะข้อสอบของ ราล์ช โมเดล

จากภาพประกอบ 4 จะเห็นว่าค่าความยากของข้อสอบ (δ_i) จะอยู่บนแกนระดับความสามารถของผู้สอบ (β_v) ตรงจุดของเส้นที่ลากมาจากจุดเปลี่ยนโค้ง ถ้า $\delta_i = .5$ และ $\beta_v = .5$ โอกาสที่ผู้สอบจะตอบข้อสอบได้ถูกต้องจะเท่ากับ 50 เปอร์เซ็นต์

ราล์ช โมเดล ได้รับการยอมรับจากนักวัดผลเป็นอย่างมาก ดังจะเห็นได้จากการพัฒนาโมเดลในตระกูลของราล์ชเพิ่มขึ้นอีกเพื่อประโยชน์ในการใช้งานให้มีความเหมาะสมกับลักษณะของข้อมูลที่แตกต่างกันออกไป ดังเช่นโมเดลต่อไปนี้ (Wright and Masters. 1982 : 38-59)

1. ไดโคโตมัสของราล์ช โมเดล (Rasch's Dichotomous Model)
2. พาเชียล เครดิต โมเดล (Partial Credit Model)
3. เรตติ้ง สเกล โมเดล (Rating Scale Model)
4. ไบโนเมียล ไตรอัลส์ (Binomial Trials Model)
5. ปัวซอง เคานท์ โมเดล (Poisson Counts Model)

ราล์ช โมเดล เป็นโมเดลที่มีผู้นิยมนำมาประยุกต์ใช้กันมาก เพราะโมเดลนี้มีจุดเด่นหลายประการ ดังนี้

1. เป็นโมเดลที่มีจำนวนพารามิเตอร์ของข้อสอบ น้อยกว่าโมเดลอื่น ๆ ของทฤษฎีการตอบข้อสอบ จึงทำให้ง่ายต่อการนำมาใช้ (Hambleton. 1989 : 157 ; citing Rasch. 1960, 1966)
2. ใช้ได้กับกลุ่มตัวอย่างที่มีขนาดเล็ก ประมาณ 100 คน ก็ประมาณค่าได้ (Wright. 1977 : 219)
3. ไม่จำเป็นต้องใช้กับข้อสอบที่มีการตรวจให้คะแนนแบบสองค่า ชนิดตอบถูกให้ 1 คะแนน และตอบผิดให้ 0 คะแนน อาจใช้กับการตรวจให้คะแนนแบบหลายค่า เช่น มาตราจัดอันดับ (Rating Scale) (Wright and Masters. 1982 : 48 - 49)
4. การประมาณค่าพารามิเตอร์ของข้อสอบ และพารามิเตอร์ของผู้สอบ มีความเป็นอิสระจากกัน กล่าวคือ ค่าความยากของข้อสอบจะไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง (Sample - Free) และค่าความสามารถของผู้สอบจะไม่แปรเปลี่ยนไปตามค่าความยากของข้อสอบ (Item - Free) (Wright and Panchapekesan. 1969 : 23)
5. การกระจายความสามารถของกลุ่มตัวอย่าง หรือความยากของข้อสอบ ไม่จำเป็นต้องเป็นโค้งปกติ เพียงแต่มั่นใจว่าจะเป็นโค้งปกติเท่านั้นก็ใช้ได้ และกลุ่มตัวอย่างไม่จำเป็นต้องได้มาจากการสุ่มก็ได้ (Wright and Stone. 1979 : 20)
6. ข้อสอบไม่จำเป็นต้องมีค่าอำนาจจำแนกเท่ากัน และไม่จำเป็นต้องคำนึงถึงการเดา (Wright and Panchapekesan. 1969 : 25)

การประมาณค่าพารามิเตอร์ของทฤษฎีการตอบข้อสอบ

การประมาณค่าพารามิเตอร์ของทฤษฎีการตอบข้อสอบ เพื่อการวิเคราะห์ข้อสอบสามารถประมาณค่าได้หลายวิธี ไรท์และมาสเตอร์ (Wright and Masters. 1982 : 60) ได้เสนอวิธีการประมาณค่าพารามิเตอร์ในราล์ส โมเดลไว้ 3 วิธี ดังนี้

1. วิธีพร็อกซ์ (The PROX Method) เป็นวิธีประมาณค่าพารามิเตอร์ตามวิธีของโคเฮน (Cohen's Approximation) ใช้กับแบบทดสอบที่มีการกระจายของค่าความสามารถของผู้สอบ และค่าความยากของข้อสอบมีลักษณะเป็นหรือมีแนวโน้มที่จะเป็นโค้งปกติ ลักษณะเด่นของการประมาณค่าโดยวิธีนี้คือ สามารถสะท้อนให้เห็นหลักการต่างๆ ของการวิเคราะห์ในราล์ส โมเดลได้เป็นอย่างดี ค่าที่ประมาณได้จะเป็นอิสระจากกลุ่มตัวอย่าง และความยาวของแบบทดสอบ โดยแปลงคะแนนให้อยู่ในรูปโลจิต (Logit)

2. วิธีแพร์ (The PAIR Method) เป็นวิธีการประมาณค่าพารามิเตอร์โดยใช้ข้อสอบจับคู่เป็นคู่ ๆ เช่น 1-2, 1-3, 1-4, ... เป็นต้น โดยนำจำนวนคนที่ตอบข้อสอบ 2 ข้อ แต่ตอบถูกเพียงข้อเดียวมาทำการวิเคราะห์ ดังนั้นจึงมีความเหมาะสมที่จะใช้ประมาณค่ากับแบบทดสอบที่มีจำนวนข้อไม่มากนัก การประมาณค่าโดยวิธีนี้ไม่ต้องอาศัยข้อตกลงเบื้องต้นเกี่ยวกับการกระจายของค่าความสามารถของผู้สอบ และค่าความยากของข้อสอบ ทั้งนี้เพราะสามารถจัดอิทธิพลพารามิเตอร์ของผู้สอบออกจากพารามิเตอร์ของข้อสอบได้ทั้งหมด ข้อดีของการประมาณค่าโดยวิธีแพร์คือ สามารถวิเคราะห์ข้อสอบที่มีข้อมูลไม่สมบูรณ์ได้ เช่น ผู้สอบบางคนทำแบบทดสอบไม่ครบทุกข้อ

3. วิธียูคอน (The UCON Method) เป็นวิธีการประมาณค่าพารามิเตอร์ด้วยวิธีความเป็นไปได้สูงสุดโดยไม่มีเงื่อนไข (Unconditional Maximum Likelihood Method) โดยใช้กับแบบทดสอบที่มีการกระจายของค่าความสามารถของผู้สอบ และค่าความยากของข้อสอบมีลักษณะเป็นโค้งปกติ จุดเด่นของการประมาณค่าโดยวิธีนี้คือ สามารถประมาณค่าพารามิเตอร์ของผู้สอบและข้อสอบได้พร้อมกัน การประมาณค่าโดยวิธียูคอนจะให้ผลใกล้เคียงกับวิธีแพร์

ในการศึกษาครั้งนี้ ผู้วิจัยวิเคราะห์ข้อมูลด้วยเครื่องคอมพิวเตอร์โดยใช้โปรแกรมไบคาล (BICAL) ซึ่งเป็นโปรแกรมสำเร็จรูป โปรแกรมนี้มีวิธีการประมาณค่าอยู่ 2 วิธี คือ วิธีพร็อกซ์ (PROX) และยูคอน (UCON) ผู้วิจัยได้เลือกใช้วิธียูคอน ประมาณค่าพารามิเตอร์

การวิเคราะห์ความสอดคล้องกับโมเดล

การประมาณค่าพารามิเตอร์ตามทฤษฎีการตอบข้อสอบ จำเป็นต้องมีการวิเคราะห์ความสอดคล้อง (Analysis Item Fit) เพื่อที่จะตรวจสอบว่าเหตุการณ์ที่ผู้สอบแต่ละคนตอบข้อสอบข้อหนึ่ง ๆ เป็นไปตามความคาดหมายหรือไม่เพียงใด การวิเคราะห์ความสอดคล้องกับโมเดลตามวิธีของราล์ฟ มีลักษณะ ดังนี้ (Wright and Stone. 1979 : 66 - 82)

1. การวิเคราะห์ความสอดคล้องของข้อสอบกับโมเดล (Analysis Item Fit)
การวิเคราะห์ในลักษณะนี้ เป็นการตรวจสอบการตอบข้อสอบในแต่ละข้อว่ามีความเหมาะสมกับค่าความหมายของการวัดตามโมเดลเพียงใด โดยมีวิธีการคำนวณ ดังนี้

1.1 คำนวณค่าคะแนนมาตรฐานส่วนที่เหลือ (Standard Residual : $Z_{v,i}^2$)

$$Z_{v,i}^2 = \exp(b - d) \text{ สำหรับ } X_{v,i} = 0 \text{ (ผู้สอบตอบข้อสอบผิด)}$$

$$Z_{v,i}^2 = \exp(d - b) \text{ สำหรับ } X_{v,i} = 1 \text{ (ผู้สอบตอบข้อสอบถูก)}$$

เมื่อ $Z_{v,i}^2$ แทน คะแนนมาตรฐานส่วนที่เหลือของผู้สอบคนที่ v ตอบข้อสอบข้อที่ i

b แทน ค่าความสามารถของผู้สอบ

d แทน ค่าความยากของข้อสอบ

$X_{v,i}$ แทน ค่าสังเกตของผู้สอบคนที่ v ตอบข้อสอบข้อที่ i

1.2 คำนวณค่า Fit Mean square (V_1) เพื่อประเมินค่าความสอดคล้องของข้อสอบข้อที่ i

$$V_1 = \sum_{v=1}^N Z_{v,i}^2 / (N - 1) \sim F_{(N-1), \infty}$$

เมื่อ N แทน จำนวนผู้สอบ

1.3 คำนวณค่า Fit Statistic (t_1) เพื่อประเมินความสอดคล้องของข้อสอบกับโมเดล

$$t_1 = [\ln(V_1) + (V_1 - 1)] \sqrt{\frac{(N - 1)}{8}} \sim N(0, 1)$$

สำหรับข้อสอบข้อใดจะมีความสอดคล้องกับโมเดล ก็ต่อเมื่อ ค่าเฉลี่ยของ t_1 มีค่าใกล้ 0 และค่าความเบี่ยงเบนมาตรฐานมีค่าใกล้ 1

2. การวิเคราะห์ความสอดคล้องของผู้ตอบกับโมเดล (Analysis Person Fit)
การวิเคราะห์ในลักษณะนี้มีประโยชน์ในการตรวจสอบค่าความเที่ยงตรง (Validity) ของการวัดในครั้งนั้น ๆ เมื่อผู้สอบตอบข้อสอบข้อหนึ่ง ๆ โดยมีวิธีการคำนวณ ดังนี้

2.1 คำนวณค่าคะแนนส่วนที่เหลือ (Standard Residual : Z_{v_1}) มีวิธีการคำนวณเหมือนกับการวิเคราะห์ความสอดคล้องของข้อสอบกับโมเดล

2.2 คำนวณค่า Fit Mean Square (V_v) เพื่อประเมินความสอดคล้องกับผู้ตอบคนที่ v

$$V_v = \sum_i^L Z_{v_1}^2 / (L - 1) \sim F_{(L - 1), \infty}$$

เมื่อ L แทน จำนวนข้อสอบ

2.3 คำนวณค่า Fit Statistic (t_v) เพื่อประเมินความสอดคล้องของผู้ตอบกับโมเดล

$$t_v = [\ln(V_v) + (V_v - 1)] \sqrt{\frac{(L - 1)}{8}} \sim N(0, 1)$$

สำหรับผู้สอบคนใดจะมีความสอดคล้องกับโมเดล ก็ต่อเมื่อ ค่าเฉลี่ยของ t_u มีค่าใกล้ 0 และค่าความเบี่ยงเบนมาตรฐานมีค่าใกล้ 1

ในการศึกษาครั้งนี้ผู้วิจัยคัดเลือกข้อสอบและผู้ตอบมีค่า Fit Statistic น้อยกว่า 3 (Wright and Stone. 1979 : 169)

คุณภาพของแบบทดสอบ

ในทฤษฎีการตอบข้อสอบ จะใช้ผลการตอบแบบทดสอบมาประมาณค่าความสามารถของผู้สอบ การประเมินคุณภาพของแบบทดสอบจะสังเกตได้จากความถูกต้องแม่นยำ (Precision) ในการประมาณค่าความสามารถของผู้สอบ ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบ (Test Information Function) จะเป็นดัชนีที่บ่งชี้ถึงความถูกต้องแม่นยำของการประมาณค่า (Birnbaum. 1968 : 418) ดังนั้น ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบสามารถนำมาตรวจสอบคุณภาพของแบบทดสอบแทนการหาความเชื่อมั่น (Reliability) และค่าความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement) ตามทฤษฎีการทดสอบแบบดั้งเดิมได้ (Hambleton and Swaminathan. 1985 : 236) ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบเป็นปริมาณที่เป็นส่วนกลับของกำลังสองของความกว้างของช่วงความเชื่อมั่น ที่เป็นผลมาจากการประมาณค่าความสามารถของผู้สอบ (Lord. 1980 : 65) ซึ่งเขียนในรูปสมการทางคณิตศาสตร์ได้ดังนี้ (Lord and Novick. 1968 : 454)

$$I(\theta) = \sum_{i=1}^n I(\theta, U_i)$$

เมื่อ $I(\theta)$ แทน ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบ

$I(\theta, U_i)$ แทน ค่าฟังก์ชันอินฟอร์เมชันของข้อสอบแต่ละข้อ

สำหรับฟังก์ชันอินฟอร์เมชันของข้อสอบ (Item Information Function) สามารถคำนวณได้จากสมการ ดังนี้

$$I(\theta, U_1) = \sum_{i=1}^n \frac{[P'_1(\theta)]^2}{P_1(\theta)Q_1(\theta)}$$

เมื่อ $I(\theta, U_1)$ แทน ค่าฟังก์ชันอินฟอร์เมชันของข้อสอบแต่ละข้อ

$P_1(\theta)$ แทน ค่าความน่าจะเป็นที่ผู้สอบซึ่งมีระดับความสามารถ θ จะตอบข้อสอบข้อที่ i ได้ถูกต้อง

$Q_1(\theta)$ แทน $1 - P_1(\theta)$

$P'_1(\theta)$ แทน ค่าความชันของโค้งลักษณะข้อสอบที่ระดับความสามารถ θ

จะเห็นได้ว่าผลรวมของฟังก์ชันอินฟอร์เมชันของข้อสอบแต่ละข้อในแบบทดสอบก็คือ ฟังก์ชันอินฟอร์เมชันของแบบทดสอบนั่นเอง ค่าฟังก์ชันอินฟอร์เมชันของข้อสอบ หรือแบบทดสอบเป็นดัชนีผสม (Composite Index) ที่สร้างขึ้นจากดัชนีบอกคุณลักษณะของข้อสอบหลาย ๆ ลักษณะ เช่น ดัชนีความยาก ดัชนีอำนาจจำแนก และดัชนีการเดา เป็นต้นแล้วรวมมาเป็นดัชนีเพียงตัวเดียวเพื่อบ่งชี้ถึงคุณภาพของแบบทดสอบ ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบฉบับหนึ่ง ๆ จะมีได้หลายค่าแตกต่างกันไปตามระดับของความสามารถ ที่ระดับความสามารถค่าใดค่าหนึ่งนั้นค่าฟังก์ชันอินฟอร์เมชันจะมีค่ามากหรือน้อยขึ้นอยู่กับ ค่าความคลาดเคลื่อนมาตรฐานในการวัดและความชันของโค้งลักษณะข้อสอบ ถ้ามีความชันเพิ่มมากขึ้นในขณะที่ความคลาดเคลื่อนมาตรฐานในการวัดมีค่าน้อยลง ค่าฟังก์ชันอินฟอร์เมชันจะมีค่าเพิ่มขึ้น (Lord. 1980 : 68) นอกจากนี้แล้วจำนวนข้อของแบบทดสอบยังส่งผลต่อค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบด้วย โดยถ้าแบบทดสอบฉบับใดมีจำนวนข้อสอบเพิ่มมากขึ้น ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบจะมีค่าเพิ่มมากขึ้นด้วย (Hambleton and Swaminathan. 1985 : 107)

ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ

ในการสอบทุกครั้งย่อมมีความคลาดเคลื่อนเกิดขึ้นแน่นอน อาจเกิดขึ้นน้อยหรือมากเท่า นั้น ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (Standard Error of Estimation : SEE (θ)) ตามทฤษฎีการตอบข้อสอบนี้จะอยู่ในรูปค่าเบี่ยงเบนมาตรฐานซึ่ง แตกต่างจากความคลาดเคลื่อนมาตรฐานในการวัด (Standard Error of Measurement : SE_{meas}) ตามทฤษฎีการทดสอบแบบดั้งเดิมตรงที่ความคลาดเคลื่อนมาตรฐานในการประมาณค่าจะ ใช้ความคลาดเคลื่อนที่เกิดจากการประมาณค่าคุณลักษณะของผู้ตอบ ที่คลาดเคลื่อนไปจากคุณลักษณะ จริง ส่วนความคลาดเคลื่อนมาตรฐานในการวัด จะใช้ความคลาดเคลื่อนที่เกิดจากคะแนนดิบ หรือ คะแนนที่ใช้ในการวัดที่คลาดเคลื่อนไปจากคะแนนจริง (Lord and Novick. 1968 : 66-67) สำหรับความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ คำนวณได้โดยอาศัย ค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบ (Test Information Function : $I(\theta)$) ดังนี้ (Hambleton and Cook. 1977 : 84 ; Warm. 1978 : 77)

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

ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบมีค่าแปรเปลี่ยน ไปตามสเกลความสามารถ และแปรผกผัน (Inverse) กับค่าฟังก์ชันอินฟอร์เมชันของแบบทดสอบ กล่าวคือ ถ้าแบบทดสอบฉบับใดมีค่าฟังก์ชันอินฟอร์เมชันสูง ณ ระดับความสามารถ θ แล้ว ความ คลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถจะมีค่าต่ำลง ณ ระดับความสามารถ θ นั้น แต่ถ้าแบบทดสอบฉบับใดมีค่าฟังก์ชันอินฟอร์เมชันต่ำ ณ ระดับความสามารถ θ แล้วความคลาด- เคลื่อนมาตรฐานในการประมาณค่าความสามารถจะมีค่าสูงขึ้น ณ ระดับความสามารถ θ นั้น นอก จากนี้แล้วค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ยังสามารถนำ ไปคำนวณหาค่าความเชื่อมั่นตามทฤษฎีการทดสอบแบบดั้งเดิมได้ โดยที่คะแนนเป็นคะแนนมาตรฐาน และมีค่าความเบี่ยงเบนในการวัดเท่ากับ 1 ให้สูตรในการคำนวณ ดังนี้ (Warm. 1978 : 77)

$$r_{tt} = 1 - [SEE(\theta)]^2$$

เมื่อ r_{cc} แทน ค่าความเชื่อมั่น

$SEE(\theta)$ แทน ค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานในการประมาณค่า
ความสามารถของผู้สอบ

งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวข้องกับการศึกษาความเที่ยงตรง และความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถ ของแบบทดสอบสองขั้นตอน มีดังต่อไปนี้

1. งานวิจัยภายในประเทศ

บรรค์ชัย คงเสน่ห์ (2530) ได้ศึกษาวิธีการสร้างแบบทดสอบสองขั้นตอนในวิชาฟิสิกส์ เรื่องการเคลื่อนที่และกฎการเคลื่อนที่ เป็นแบบทดสอบเลือกตอบชนิด 5 ตัวเลือก โดยสร้างจาก กลุ่มข้อสอบ 170 ข้อ แยกเป็นแบบทดสอบแยกกลุ่มที่มีค่าความยากอยู่ในช่วงใกล้เคียงกัน 20 ข้อ และแบบทดสอบหลัก 4 ฉบับ ๆ ละ 30 ข้อ โดยทำการทดสอบกับนักเรียนชั้นมัธยมศึกษาปีที่ 4 จำนวน 743 คน และยังได้ศึกษาความสัมพันธ์ระหว่างความสามารถทางการเรียนที่ใช้วิธีในการประมาณค่า 3 วิธี คือประมาณค่าโดยใช้โปรแกรมสำเร็จรูปโลจิสต์ ประมาณค่าโดยใช้ค่าความยากของข้อสอบที่ยากที่สุดที่นักเรียนตอบถูก และประมาณค่าโดยใช้ค่าความยากเฉลี่ยของข้อสอบที่นักเรียนตอบถูก ของแบบทดสอบหลักทั้ง 4 ฉบับ ผลการศึกษพบว่า

การประมาณค่าตามวิธีที่ 2 และวิธีที่ 3 มีความสัมพันธ์กันสูงในทางบวก กับแบบทดสอบหลักทั้ง 4 ฉบับ ส่วนวิธีที่ 1 กับวิธีที่ 3 มีความสัมพันธ์กันน้อยมากโดยเฉพาะในแบบทดสอบหลัก ฉบับที่ 4 มีความสัมพันธ์กันทางลบ คือ มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ -0.068 และการประมาณค่าตามวิธีที่ 1 และวิธีที่ 2 มีความสัมพันธ์กันปานกลางในทางบวก

นันทิยา พึงคำ (2531) ได้ศึกษาเปรียบเทียบคุณภาพของการทดสอบแบบ ซีเอที (CAT) กับการทดสอบแบบดั้งเดิม (Conventional) ในการวัดความสามารถด้านคำศัพท์ภาษาอังกฤษ โดยทำการศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 120 คน การทดสอบแบบ ซีเอที ใช้กลุ่มข้อสอบ (Item Pool) จำนวน 361 ข้อ โดยมีค่าความยากอยู่ระหว่าง -0.90552 ถึง 3.61860 มีค่าอำนาจจำแนกอยู่ระหว่าง 0.80334 ถึง 2.00000 และมีค่าการเดาอยู่ระหว่าง 0.04421 ถึง 0.29886 การประมาณค่าความสามารถของผู้สอบใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข (Conditional Maximum Likelihood) ส่วนการทดสอบแบบดั้งเดิมใช้ข้อสอบทั้งหมด 40 ข้อ

โดยมีค่าความยากอยู่ระหว่าง-0.76488 ถึง 2.97884 มีค่าอำนาจจำแนกอยู่ระหว่าง 0.94069 ถึง 2.0000 และมีค่าการเดาอยู่ระหว่าง 0.03462 ถึง 0.28620 ผลการศึกษพบว่า

การทดสอบแบบซีเอที่กับการทดสอบแบบดั้งเดิมมีความเที่ยงตรงเชิงเกณฑ์สัมพัทธ์ไม่แตกต่างกัน แต่การทดสอบแบบซีเอที่สามารถลดจำนวนข้อสอบลงได้มากกว่าครึ่งหนึ่ง

ความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ จากการทดสอบแบบซีเอที่มีค่าต่ำกว่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ จากการทดสอบแบบดั้งเดิมอย่างมีนัยสำคัญทางสถิติที่ระดับ .001

เกรียงศักดิ์ สุวรรณภาค (2532) ได้ศึกษาความสัมพันธ์ระหว่างรูปแบบของแบบทดสอบแยกกลุ่ม กับความสามารถทางการเรียนในวิชาคณิตศาสตร์ เรื่องเลขที่กำลัง โดยทำการศึกษาในตัวแปร 2 ตัว คือ รูปแบบของแบบทดสอบแยกกลุ่ม 3 แบบ คือ แบบที่มีค่าความยากอยู่ในช่วงใกล้เคียงกัน แบบที่มีค่าความยากกระจายตั้งแต่ง่ายที่สุดจนถึงยากที่สุดและแบบที่มีการแยกกลุ่มสองครั้งกับแบบทดสอบหลัก 4 ฉบับ และวิธีตรวจให้คะแนน 2 วิธี คือตรวจให้คะแนนตามค่าความยากของข้อสอบที่ยากที่สุดที่นักเรียนทำถูก และตรวจให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก โดยทำการศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 2,455 คนและประมาณค่าความสามารถทางการเรียนด้วยโปรแกรมสำเร็จรูปโลจิสต์ ผลการศึกษพบว่า

ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างผลการสอบที่ใช้แบบทดสอบแยกกลุ่มแต่ละฉบับ ซึ่งตรวจให้คะแนนในแต่ละวิธี กับความสามารถทางการเรียน มีค่าเป็นบวกอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกค่า

ค่าสัมประสิทธิ์สหสัมพันธ์ของคะแนน ที่ได้จากการตรวจให้คะแนนทั้ง 2 วิธี โดยแยกวิเคราะห์ตามแบบทดสอบแยกกลุ่มแต่ละแบบมีค่าเป็นบวกอย่างมีนัยสำคัญทางสถิติที่ระดับ.01 ทุกค่า

นริศ สวัสดิ์ (2533) ได้ศึกษาความสัมพันธ์ระหว่างคะแนนของแบบทดสอบสองขั้นตอนและคะแนนของแบบทดสอบรูปปริมาตร โดยใช้แบบทดสอบสองขั้นตอน ซึ่งแยกเป็นแบบทดสอบแยกกลุ่ม จำนวน 15 ข้อ และแบบทดสอบหลัก 4 ฉบับ ๆ ละ 20 ข้อ ส่วนแบบทดสอบรูปปริมาตรที่ใช้เป็นแบบทดสอบที่มีขนาดขั้นแปรผัน ขนาด 4 ขั้น 8 ขั้น และ 12 ขั้นอย่างละ 1 ฉบับ และใช้วิธีการให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก โดยทำการศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 2 จำนวน 815 คน ผลการศึกษพบว่า

ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนของแบบทดสอบสองขั้นตอน กับคะแนนของแบบทดสอบรูปปริมาตรทั้ง 3 ขนาด มีค่าเป็นบวกอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกค่า

ต่าย เชียงจี (2534) ได้ศึกษาเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) จากการทดสอบเทเลอ์รูปปริมาตรที่มีรูปแบบ จำนวนชั้น และวิธีการให้คะแนน ที่แตกต่างกัน โดยใช้กลุ่มทดลอง 32 กลุ่ม ๆ ละ 1,000 คน ในแต่ละกลุ่ม ทำการทดลองซ้ำ 40 ครั้ง หาค่าโอกาสในการตอบข้อสอบถูก ด้วยโมเดลโลจิสติก แบบหนึ่งพารามิเตอร์ โดยใช้การทดสอบเทเลอ์ที่มีลักษณะเป็นแบบทดสอบรูปปริมาตร แบบขนาดชั้นคงที่ และแบบขนาดชั้นแปรผันมีจำนวนชั้นของการทดสอบ 4 แบบ คือ แบบที่มีจำนวน 7 ชั้น 10 ชั้น 13 ชั้น และ 16 ชั้น การตรวจให้คะแนน 4 วิธี คือ ให้คะแนนตามค่าความยากของข้อที่ยากที่สุดที่ทำถูก ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำถูก ให้คะแนนโดยใช้วิธีของเบส์ และให้คะแนนโดยใช้วิธีความเป็นไปได้สูงสุดแบบมีเงื่อนไข ผลการศึกษพบว่า

การทดสอบรูปปริมาตรแบบขนาดชั้นแปรผันที่มี 16 ชั้น ให้คะแนนโดยใช้วิธีการของเบส์จะให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถต่ำสุด เมื่อเพิ่มจำนวนชั้นของรูปแบบทั้งสอง ทำให้ค่าความเชื่อมั่นและค่าความเที่ยงตรงเพิ่มขึ้น ค่าความคลาดเคลื่อนมาตรฐานในการวัดลดลง และยังพบว่าการใช้วิธีของเบส์ กับจำนวนชั้นตั้งแต่ 10 ชั้นขึ้นไปจะให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบต่ำสุด

อรินทร์ น่วมถนอม (2537) ได้ศึกษาความเที่ยงตรง และประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถของผู้สอบในวิชาคณิตศาสตร์ระหว่างแบบทดสอบดั้งเดิม แบบทดสอบสองขั้นตอน และแบบทดสอบแฟล็กชีเลเวล โดยศึกษากับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 441 คน ผลการศึกษพบว่า

ค่าความเที่ยงตรงของการประมาณค่าความสามารถ ระหว่างแบบทดสอบดั้งเดิมกับแบบทดสอบสองขั้นตอน มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนค่าความเที่ยงตรงของการประมาณค่าความสามารถ ระหว่างแบบทดสอบดั้งเดิมกับแบบทดสอบแฟล็กชีเลเวล และแบบทดสอบสองขั้นตอนกับแบบทดสอบแฟล็กชีเลเวล มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

ค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถระหว่างแบบทดสอบดั้งเดิมเทียบกับแบบทดสอบสองขั้นตอน และแบบทดสอบดั้งเดิมกับแบบทดสอบแฟล็กชีเลเวลมีค่าสูงกว่า 1 ในกลุ่มผู้สอบที่มีความสามารถระดับปานกลางแต่มีค่าต่ำกว่า 1 ในกลุ่มผู้สอบที่มีความสามารถระดับต่ำลงมา และระดับสูงขึ้นไป ส่วนค่าประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถ ระหว่างแบบทดสอบสองขั้นตอนเทียบกับแบบทดสอบแฟล็กชีเลเวล มีค่าสูงกว่า 1 ในกลุ่มผู้สอบที่มีความสามารถ

ระดับต่ำมาก ระดับปานกลางและระดับสูงมาก แต่มีค่าต่ำกว่า 1 ในกลุ่มผู้สอบที่มีความสามารถระดับต่ำและระดับสูง

2. งานวิจัยต่างประเทศ

ลินน์ และคนอื่น ๆ (Linn and others. 1969) ได้ศึกษาความแตกต่างของวิธีการแยกทาง 7 วิธี โดยในวิธีการแยกทาง 5 วิธีแรกเป็นแบบทดสอบสองขั้นตอน (Two - Stage Test) คือ แบบแยกกลุ่มสองครั้ง แบบแยกกลุ่มพิสัยกว้าง แบบแยกกลุ่มที่ใช้กลุ่มจำแนก แบบสุ่มข้อคำถามแบ่งตามลำดับ 4 กลุ่ม และแบบสุ่มข้อคำถามแบ่งตามลำดับ 3 กลุ่ม ส่วน 2 วิธีหลังเป็นแบบทดสอบรูปปิรามิด ซึ่งแบบที่หนึ่งเป็นชนิด 10 ขั้น และแบบที่สองเป็นชนิด 5 ขั้น โดยที่แต่ละขั้นมีข้อสอบ 5 ข้อ แบบทดสอบรูปปิรามิดทั้ง 2 แบบนี้ใช้วิธีแยกทางแบบเพิ่ม/ลดเท่ากัน และได้ศึกษากับนักเรียนจำนวน 4,885 คน โดยสร้างจากกลุ่มข้อสอบจำนวน 190 ข้อ ผลการศึกษพบว่า

แบบทดสอบรูปปิรามิดชนิด 10 ขั้น มีค่าสัมประสิทธิ์สหสัมพันธ์ .87 กับคะแนนรวมของแบบทดสอบ .95 กับแบบทดสอบรูปปิรามิดชนิด 5 ขั้น และ .89 ถึง .96 กับแบบทดสอบสองขั้นตอนทั้ง 5 แบบ

แบบทดสอบรูปปิรามิดชนิด 10 ขั้น มีความสัมพันธ์กับเกณฑ์การวัดสูงกว่าความสัมพันธ์กับแบบทดสอบดั้งเดิมที่มีความยากเท่ากัน และแบบทดสอบรูปปิรามิดชนิด 5 ขั้น มีความสัมพันธ์กับเกณฑ์การวัดสูงกว่าแบบทดสอบดั้งเดิมที่มีข้อสอบ 50 ข้อ นั่นคือแบบทดสอบรูปปิรามิดที่มีข้อสอบน้อยข้อกว่าจะมีความเที่ยงตรงมากกว่าเมื่อเทียบกับแบบทดสอบดั้งเดิม

ลาร์คิน และไวสส์ (Larkin and Weiss. 1975) ได้ศึกษาเปรียบเทียบแบบทดสอบสองขั้นตอนที่ประกอบด้วยแบบทดสอบแยกกลุ่ม จำนวน 10 ข้อ และแบบทดสอบหลัก 4 ฉบับ ๆ ละ 30 ข้อ กับแบบทดสอบรูปปิรามิดแบบที่ใช้ขนาดขั้นคงที่ชนิด 15 ขั้น และวิธีการตรวจให้คะแนน 4 วิธี คือ ให้คะแนนตามจำนวนข้อที่ทำถูก ให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ทำให้คะแนนตามค่าความยากเฉลี่ยของข้อสอบที่ตอบถูก และให้คะแนนตามค่าความยากของข้อสอบข้อสุดท้ายที่ทำ ดำเนินการทดสอบโดยใช้เครื่องคอมพิวเตอร์ทดสอบกับนักศึกษาจำนวน 111 คน ผลการศึกษพบว่า

คะแนนเฉลี่ยที่ได้รับจากการจัดลำดับก่อนหลังในการทดสอบ ระหว่างแบบทดสอบสองขั้นตอนกับแบบทดสอบรูปปิรามิด มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ นั่นคือ ในการจัดการสอบนั้นจะจัดให้ผู้สอบสอบแบบทดสอบชนิดใดก่อนหรือหลังไม่มีผลกระทบต่อคะแนนที่ได้

ค่าสัมประสิทธิ์สหสัมพันธ์ ระหว่างคะแนนที่ได้จากแบบทดสอบสองขั้นตอนกับแบบทดสอบรูป
ปิรามิดที่ใช้วิธีตรวจให้คะแนนทั้ง 4 วิธี มีค่าเท่ากับ .84, .81, .79 และ .83 ตามลำดับ

เมื่อพิจารณาค่าสัมประสิทธิ์สหสัมพันธ์ภายในของคะแนนที่ได้รับจากแบบทดสอบรูปปิรามิดที่
ใช้วิธีตรวจให้คะแนนทั้ง 4 วิธี ปรากฏว่า ความสัมพันธ์ของคะแนนที่ให้ตามวิธีที่ 2 กับวิธีที่ 3 มี
ค่าสูงสุดคือ .99 การให้คะแนนตามวิธีที่ 1 มีความสัมพันธ์ต่ำมากกับคะแนนที่ให้ตามวิธีที่ 2 และ
วิธีที่ 3 แต่มีความสัมพันธ์กันสูงกับคะแนนที่ให้ตามวิธีที่ 4 ส่วนคะแนนที่ให้ตามวิธีที่ 4 มีความ
สัมพันธ์ปานกลางกับคะแนนที่ให้ตามวิธีที่ 2 และวิธีที่ 3

เดอ กรูจิเตอร์ (De Gruijter. 1985) ได้ศึกษาความคลาดเคลื่อนมาตรฐานของ
นารามิเตอร์ค่าความยากของข้อสอบ โดยใช้ราล์ส โมเดล ผลการศึกษาพบว่า

ขนาดของความคลาดเคลื่อนมาตรฐานที่ได้จะมีค่าสอดคล้องกับผลการศึกษาของวิงเกอร์-
สกี และลอร์ด (1984) โดยขึ้นอยู่กับวิธีการที่ยึดเมตริกของความสามารถ กรณีของราล์ส โมเดล
การยึดเมตริกโดยกำหนดให้ค่าเฉลี่ยของค่าความยาก (b) เท่ากับศูนย์ และการกำหนดให้ค่าความ
ยาก (b) ทุกค่ามีค่าเท่ากับศูนย์ จะให้ผลที่แตกต่างกัน

เบย์รอฟฟ์ และซีรีย์ (Bayroff and Seeley. 1967) ได้ศึกษาความสัมพันธ์ระหว่าง
แบบทดสอบแยกกลุ่มกับแบบทดสอบดั้งเดิม โดยใช้แบบทดสอบแยกกลุ่มเพื่อการให้เหตุผลทางภาษา
และเหตุผลทางคณิตศาสตร์ ในการศึกษาครั้งนี้ทำการทดสอบกับกลุ่มตัวอย่าง จำนวน 120 คน
แบบทดสอบแยกกลุ่มประกอบด้วยข้อสอบจำนวน 8-9 ข้อ และแบบทดสอบดั้งเดิม 2 ฉบับ คือ แบบ
ทดสอบการให้เหตุผลทางภาษาจำนวน 50 ข้อ แบบทดสอบการให้เหตุผลทางคณิตศาสตร์ จำนวน
40 ข้อ ผลการศึกษาพบว่า

ความสัมพันธ์ระหว่างแบบทดสอบแยกกลุ่มกับแบบทดสอบดั้งเดิม ในด้านการให้เหตุผลของ
ภาษา มีค่าเท่ากับ 0.78 และความสัมพันธ์ระหว่างแบบทดสอบแยกกลุ่มกับแบบทดสอบดั้งเดิมในให้
ด้านการให้เหตุผลทางคณิตศาสตร์มีค่าเท่ากับ 0.74

แองกอฟฟ์ และฮัดเดิลสตัน (Angoff and Huddleston. 1958) ได้ทำการตรวจ
สอบแบบทดสอบสองขั้นตอน พบว่า แบบทดสอบสองขั้นตอนมีเทคนิคเหนือกว่าแบบทดสอบทั่ว ๆ ไป
แต่ขอบเขตของการเหนือกว่าของแบบทดสอบสองขั้นตอนนั้น ยังไม่เพียงพอที่จะขอมรับปัญหาเกี่ยว
กับการดำเนินการสอบ อย่างไรก็ตามทั้ง 2 คน มุ่งที่จะเพิ่มค่าความเชื่อมั่น และค่าความเที่ยงตรง
มากกว่าการลดเวลาในการดำเนินการสอบเมื่อยังคงค่าความเชื่อมั่น และค่าความเที่ยงตรงเท่า
เดิม

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องดังกล่าวมาแล้วนั้น จะพบว่า การทดสอบแบบเทเลอรันั้น จะมีประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบได้ดีกว่าการทดสอบแบบดั้งเดิม กล่าวคือ การทดสอบแบบเทเลอรัจะให้ค่าความเชื่อมั่นและความเที่ยงตรงสูงกว่าการทดสอบแบบดั้งเดิม ในขณะเดียวกัน ก็ให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถที่ต่ำกว่า นอกจากนี้ยังพบว่า การทดสอบแบบเทเลอรันั้น ใช้จำนวนข้อสอบ และเวลาในการสอบดำเนินการน้อยกว่าการทดสอบแบบทดสอบดั้งเดิม จากเหตุผลดังกล่าว ผู้วิจัยจึงมีความสนใจที่จะนำแบบทดสอบแยกกลุ่ม ซึ่งเป็นแบบทดสอบขั้นที่หนึ่งของแบบทดสอบสองขั้นตอน และเป็นรูปแบบหนึ่งของการทดสอบแบบเทเลอรัมาศึกษาทั้ง 3 รูปแบบ คือแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง โดยมุ่งศึกษาในเรื่องความเที่ยงตรง และความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ

สมมติฐานในการศึกษาค้นคว้า

1. ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง มีค่าแตกต่างกัน
2. ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง มีค่าแตกต่างกัน

บทที่ 3

วิธีดำเนินการศึกษาค้นคว้า

ประชากร

ประชากรที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษาในจังหวัดลพบุรี จำนวนโรงเรียน 23 โรงเรียน จำนวนห้องเรียน 147 ห้อง จำนวนนักเรียน 6,024 คน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษาในจังหวัดลพบุรี ซึ่งเลือกมาโดยการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) มีขนาดโรงเรียนเป็นชั้น (Strata) และมีห้องเรียนเป็นหน่วยการสุ่ม (Sampling Unit) ได้จำนวนโรงเรียน 7 โรงเรียน จำนวนห้องเรียน 11 ห้อง จำนวนนักเรียน 419 คน ดังรายละเอียดในตารางที่ 1

ตาราง 1 จำนวนกลุ่มตัวอย่างที่ใช้ในการวิจัย จำแนกตามขนาดโรงเรียน จำนวนห้องเรียน และ
จำนวนนักเรียน

โรงเรียน	จำนวนห้องเรียน	จำนวนนักเรียน
ขนาดใหญ่		
1. โศกยะเทียมวิทยาลัย	2	91
2. พระนารายณ์	2	90
ขนาดกลาง		
1. โศกตุมวิทยา	1	40
2. ดงตาลวิทยา	1	35
3. ท่าวังวิทยาคาร	2	68
ขนาดเล็ก		
1. บ้านเบิกวิทยาคม	2	55
2. บางกะปิปานเลิศวิทยา	1	40
รวม	11	419

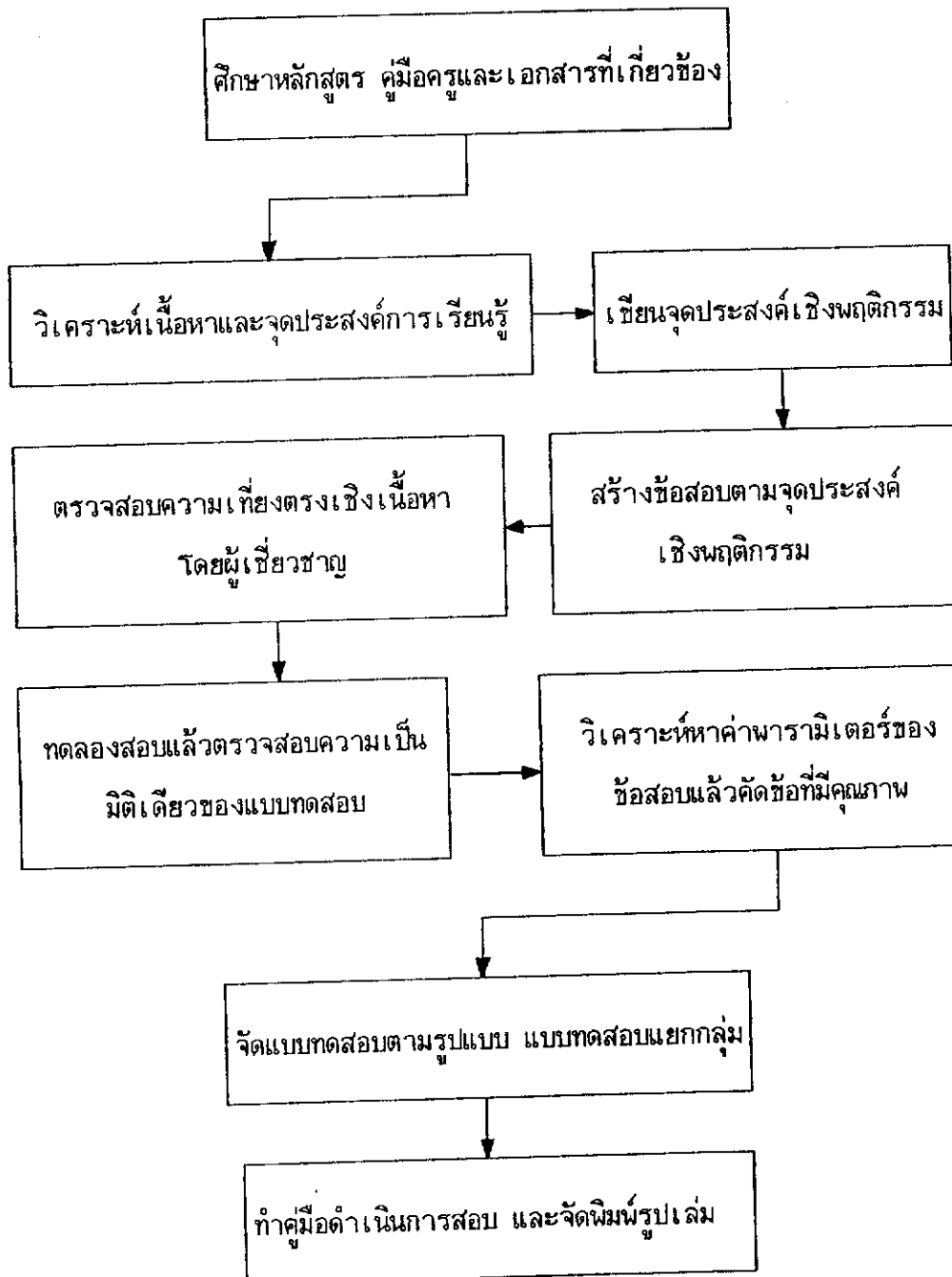
เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นแบบทดสอบเลือกตอบวัดผลสัมฤทธิ์ทางการเรียนชนิด
5 ตัวเลือก ซึ่งผู้วิจัยสร้างขึ้นจากเนื้อหาในวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 (ค 204) เรื่อง
สมการและอสมการ ตามหลักสูตรมัธยมศึกษาตอนต้น พุทธศักราช 2521 (ฉบับปรับปรุง พ.ศ.
2533) โดยจัดเป็นแบบทดสอบแยกกลุ่ม 3 รูปแบบ ดังนี้

- รูปแบบที่ 1 เป็นแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน จำนวน 20 ข้อ
- รูปแบบที่ 2 เป็นแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง จำนวน 20 ข้อ
- รูปแบบที่ 3 เป็นแบบทดสอบแยกกลุ่มสองครั้ง จำนวน 20 ข้อ

วิธีดำเนินการสร้างแบบทดสอบ

การสร้างแบบทดสอบผู้วิจัยได้ดำเนินการสร้างตามแผนผัง ดังนี้



ภาพประกอบ 5 แผนผังแสดงขั้นตอนการดำเนินการสร้างแบบทดสอบ

รายละเอียดในการสร้างแบบทดสอบ มีดังนี้

1. การสร้างกลุ่มข้อสอบ

1.1 ศึกษาเนื้อหาจากหลักสูตร คู่มือครู หนังสือแบบเรียนและเอกสารที่เกี่ยวข้อง
ในวิชาคณิตศาสตร์ (ค 204) ชั้นมัธยมศึกษาปีที่ 2 เรื่อง สมการและอสมการ

1.2 วิเคราะห์จุดประสงค์การเรียนรู้และเนื้อหา เพื่อนำมาเขียนเป็นจุดประสงค์
เชิงพฤติกรรม โดยเขียนได้ 14 จุดประสงค์

1.3 สร้างข้อสอบเลือกตอบ ชนิด 5 ตัวเลือก จำนวน 120 ข้อ โดยสร้างตาม
จุดประสงค์เชิงพฤติกรรม ซึ่งมีรายละเอียดดังนี้

จุดประสงค์ที่ 1	สร้างข้อสอบได้จำนวน 8 ข้อ
จุดประสงค์ที่ 2	สร้างข้อสอบได้จำนวน 4 ข้อ
จุดประสงค์ที่ 3	สร้างข้อสอบได้จำนวน 4 ข้อ
จุดประสงค์ที่ 4	สร้างข้อสอบได้จำนวน 4 ข้อ
จุดประสงค์ที่ 5	สร้างข้อสอบได้จำนวน 8 ข้อ
จุดประสงค์ที่ 6	สร้างข้อสอบได้จำนวน 12 ข้อ
จุดประสงค์ที่ 7	สร้างข้อสอบได้จำนวน 12 ข้อ
จุดประสงค์ที่ 8	สร้างข้อสอบได้จำนวน 4 ข้อ
จุดประสงค์ที่ 9	สร้างข้อสอบได้จำนวน 8 ข้อ
จุดประสงค์ที่ 10	สร้างข้อสอบได้จำนวน 8 ข้อ
จุดประสงค์ที่ 11	สร้างข้อสอบได้จำนวน 28 ข้อ
จุดประสงค์ที่ 12	สร้างข้อสอบได้จำนวน 4 ข้อ
จุดประสงค์ที่ 13	สร้างข้อสอบได้จำนวน 12 ข้อ
จุดประสงค์ที่ 14	สร้างข้อสอบได้จำนวน 4 ข้อ

1.4 นำข้อสอบที่สร้างขึ้น 120 ข้อ ไปให้ผู้เชี่ยวชาญซึ่งมีความรู้ในเนื้อหาและการ
เขียนข้อสอบจำนวน 5 คน เป็นผู้ตรวจสอบความเที่ยงตรงเชิงเนื้อหา (Content Validity)
โดยการพิจารณาความสอดคล้องระหว่าง ข้อสอบแต่ละข้อกับจุดประสงค์เชิงพฤติกรรม ข้อสอบที่ถือว่า
วัดได้ตรงตามจุดประสงค์เชิงพฤติกรรมจะต้องได้รับความเห็นชอบจากผู้เชี่ยวชาญไม่น้อยกว่า 3
คน สำหรับการประเมินความสอดคล้องมีแบบการประเมิน ดังตัวอย่างต่อไปนี้

ตัวอย่าง แบบประเมินความสอดคล้องระหว่างข้อสอบกับจุดประสงค์เชิงพฤติกรรม
คำชี้แจง ให้ท่านพิจารณาว่าข้อสอบต่อไปนี้วัดได้ตรงตามจุดประสงค์เชิงพฤติกรรมที่กำหนดไว้หรือไม่ โดยเขียนเครื่องหมาย ✓ ลงในช่องแสดงความคิดเห็น
 ซึ่งพิจารณาให้คะแนนตามเกณฑ์ดังนี้

ให้คะแนน +1 เมื่อท่านรู้สึกแน่ใจว่าข้อสอบวัดได้ตรงตามจุดประสงค์เชิงพฤติกรรม

ให้คะแนน 0 เมื่อท่านรู้สึกไม่แน่ใจว่าข้อสอบวัดได้ตรงตามจุดประสงค์เชิงพฤติกรรม

ให้คะแนน -1 เมื่อท่านรู้สึกแน่ใจว่าข้อสอบไม่ได้วัดตรงตามจุดประสงค์เชิงพฤติกรรม

จุดประสงค์เชิงพฤติกรรม	ข้อสอบ	คะแนนการพิจารณา		
		+1	0	-1
๑. เมื่อกำหนดสมการหรือสมการมาให้นักเรียนสามารถหาคำตอบของสมการหรือสมการได้	๑. ถ้า $5(x-4) = 3(x+2)$ แล้ว x มีค่าเท่าใด ก. 10 ข. 11 ค. 12 ง. 13 จ. 14	✓		
	๑๑. ถ้า $4x - 6 < 2$ แล้ว x มีค่าเท่าใด ก. $x > 2$ ข. $x < 2$ ค. $x > 4$ ง. $x < 4$ จ. $x < 8$		✓	

จากผลของการตรวจสอบความเที่ยงตรงเชิงเนื้อหาของผู้เชี่ยวชาญ ผลปรากฏว่าข้อสอบจำนวน 120 ข้อ ได้รับการพิจารณาว่ามีความสอดคล้องกับจุดประสงค์เชิงพฤติกรรมทุกข้อ

1.5 นำข้อสอบที่ได้ผ่านการพิจารณาจากผู้เชี่ยวชาญมาจัดเป็นแบบทดสอบ 4 ฉบับ ฉบับละ 30 ข้อ ซึ่งจัดโดยการสุ่มข้อสอบจากแต่ละจุดประสงค์เชิงพฤติกรรม

1.6 นำแบบทดสอบทั้ง 4 ฉบับ ไปทดสอบกับนักเรียนกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์หาค่าพารามิเตอร์ของข้อสอบซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ของโรงเรียนในสังกัดกรมสามัญศึกษาในจังหวัดลพบุรี โดยเป็นนักเรียนคนละกลุ่มกับกลุ่มตัวอย่างที่ใช้ในการวิจัย โดยมีจำนวนโรงเรียน 10 โรงเรียน จำนวนห้องเรียน 44 ห้องเรียนและจำนวนนักเรียน 1,679 คน ซึ่งได้มาจากการสุ่มอย่างง่าย (Simple Random Sampling) ดังรายละเอียดต่อไปนี้

โรงเรียนชัยบาดาลวิทยาคม	จำนวน 6 ห้องเรียน จำนวนนักเรียน 245 คน
โรงเรียนพระนารายณ์	จำนวน 6 ห้องเรียน จำนวนนักเรียน 290 คน
โรงเรียนพัฒนานิคม	จำนวน 6 ห้องเรียน จำนวนนักเรียน 225 คน
โรงเรียนโคกสำโรงวิทยา	จำนวน 5 ห้องเรียน จำนวนนักเรียน 200 คน
โรงเรียนหนองม่วงวิทยา	จำนวน 5 ห้องเรียน จำนวนนักเรียน 172 คน
โรงเรียนบ้านหมี่วิทยา	จำนวน 5 ห้องเรียน จำนวนนักเรียน 176 คน
โรงเรียนปิยะบุตร์	จำนวน 5 ห้องเรียน จำนวนนักเรียน 155 คน
โรงเรียนบ้านชีวิทยา	จำนวน 2 ห้องเรียน จำนวนนักเรียน 71 คน
โรงเรียนดงตาลวิทยา	จำนวน 2 ห้องเรียน จำนวนนักเรียน 70 คน
โรงเรียนโคกสกลวิทยา	จำนวน 2 ห้องเรียน จำนวนนักเรียน 75 คน

1.7 ผู้วิจัยได้สุ่มนักเรียนในห้องเรียนแต่ละคนเพื่อทำแบบทดสอบแต่ละฉบับ โดยการจัดการดำเนินการสอบให้เป็นระบบดังนี้ให้นักเรียนคนที่ 1, 2, 3 และ 4 ทำแบบทดสอบฉบับที่ 1, 2, 3 และ 4 ตามลำดับ และนักเรียนคนที่ 5, 6, 7 และ 8 จะได้ทำแบบทดสอบฉบับที่ 1, 2, 3 และ 4 ตามลำดับเหมือนเดิมสลับกันไปในลักษณะนี้ ซึ่งจะทำให้จำนวนนักเรียนที่ทำแบบทดสอบในแต่ละฉบับมีจำนวนใกล้เคียงกันคือ

- ฉบับที่ 1 มีนักเรียนทำแบบทดสอบ จำนวน 416 คน
- ฉบับที่ 2 มีนักเรียนทำแบบทดสอบ จำนวน 418 คน
- ฉบับที่ 3 มีนักเรียนทำแบบทดสอบ จำนวน 425 คน
- ฉบับที่ 4 มีนักเรียนทำแบบทดสอบ จำนวน 420 คน

1.8 นำผลการทดสอบมาตรวจให้คะแนนโดยข้อที่ตอบถูกให้ 1 คะแนน และข้อที่ตอบผิดให้ 0 คะแนน แล้วนำไปวิเคราะห์องค์ประกอบ (Factor Analysis) เพื่อตรวจสอบความเป็นมิติเดียวของแบบทดสอบโดยใช้โปรแกรมสำเร็จรูป SPSS-X ซึ่งเลือกสกัดองค์ประกอบด้วยวิธีวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis) แล้วหมุนแกนด้วยวิธีแวนริแมกซ์ (Varimax) ได้ผลดังตาราง 8 ในภาคผนวก ก ซึ่งปรากฏค่าไอเกน (Eigen Value) ของแต่ละฉบับดังนี้

ฉบับที่ 1 ค่าไอเกนอยู่ระหว่าง 1.03307 ถึง 3.99046 โดยที่ค่าไอเกนขององค์ประกอบที่ 1 มีค่า 3.99046 ซึ่งสูงกว่าค่าไอเกนขององค์ประกอบที่ 2 เป็นอย่างมาก และค่าไอเกนขององค์ประกอบที่ 2 กับองค์ประกอบที่เหลือมีค่าอยู่ระหว่าง 1.03307 ถึง 1.58862 ซึ่งไม่แตกต่างกันมากนัก

ฉบับที่ 2 ค่าไอเกนอยู่ระหว่าง 1.01029 ถึง 3.38590 โดยที่ค่าไอเกนขององค์ประกอบที่ 1 มีค่า 3.38590 ซึ่งสูงกว่าค่าไอเกนขององค์ประกอบที่ 2 เป็นอย่างมาก และค่าไอเกนขององค์ประกอบที่ 2 กับองค์ประกอบที่เหลือมีค่าอยู่ระหว่าง 1.01029 ถึง 1.43368 ซึ่งไม่แตกต่างกันมากนัก

ฉบับที่ 3 ค่าไอเกนอยู่ระหว่าง 1.01481 ถึง 4.48022 โดยที่ค่าไอเกนขององค์ประกอบที่ 1 มีค่า 4.48022 ซึ่งสูงกว่าค่าไอเกนขององค์ประกอบที่ 2 เป็นอย่างมาก และค่าไอเกนขององค์ประกอบที่ 2 กับองค์ประกอบที่เหลือมีค่าอยู่ระหว่าง 1.01481 ถึง 1.52096 ซึ่งไม่แตกต่างกันมากนัก

ฉบับที่ 4 ค่าไอเกนอยู่ระหว่าง 1.03941 ถึง 5.05963 โดยที่ค่าไอเกนขององค์ประกอบที่ 1 มีค่า 5.05963 ซึ่งสูงกว่าค่าไอเกนขององค์ประกอบที่ 2 เป็นอย่างมาก และค่าไอเกนขององค์ประกอบที่ 2 กับองค์ประกอบที่เหลือมีค่าอยู่ระหว่าง 1.03941 ถึง 1.58417 ซึ่งไม่แตกต่างกันมากนัก

ดังนั้นจึงกล่าวได้ว่าแบบทดสอบทั้ง 4 ฉบับ มีความเป็นมิติเดียว (Warm, 1978: 104 ; Lord, 1980 : 19) นั่นคือสามารถนำแบบทดสอบทั้ง 4 ฉบับไปวิเคราะห์ตามทฤษฎีการตอบข้อสอบได้

1.9 นำผลจากข้อ 1.8 ไปวิเคราะห์ข้อสอบรายข้อ เพื่อหาค่าพารามิเตอร์ความยากของข้อสอบตามทฤษฎีการตอบข้อสอบ แบบราล์ส โมเดล โดยใช้โปรแกรมสำเร็จรูป BICAL

ได้ผลการวิเคราะห์ที่มีค่าความยากของข้อสอบอยู่ในช่วง -2.23 ถึง 2.03 และค่าความสอดคล้องกับโมเดล (Fit Model : t - test) มีค่าอยู่ในช่วง -2.41 ถึง 3.32 ซึ่งมีรายละเอียดดังตาราง 9, 10, 11 และ 12 ในภาคผนวก ข

1.10 คัดเลือกข้อสอบที่มีความยากอยู่ในช่วง -2.00 ถึง 2.00 และมีค่าความสอดคล้องกับโมเดลอยู่ในช่วง -2.00 ถึง 2.00 ได้ผลดังนี้

ฉบับที่ 1 ได้ข้อสอบจำนวน 27 ข้อ มีค่าความยากอยู่ในช่วง -1.45 ถึง 1.55 และค่าความสอดคล้องอยู่ในช่วง -1.58 ถึง 0.84

ฉบับที่ 2 ได้ข้อสอบจำนวน 30 ข้อ มีค่าความยากอยู่ในช่วง -1.63 ถึง 1.23 และค่าความสอดคล้องอยู่ในช่วง -1.83 ถึง 1.15

ฉบับที่ 3 ได้ข้อสอบจำนวน 27 ข้อ มีค่าความยากอยู่ในช่วง -1.28 ถึง 1.76 และค่าความสอดคล้องอยู่ในช่วง -1.99 ถึง 1.39

ฉบับที่ 4 ได้ข้อสอบจำนวน 26 ข้อ มีค่าความยากอยู่ในช่วง -1.45 ถึง 1.25 และค่าความสอดคล้องอยู่ในช่วง -2.00 ถึง 1.79

1.11 นำข้อสอบในข้อ 1.10 จำนวน 110 ข้อ มาจัดเรียงข้อสอบจากข้อที่มีค่าความยากต่ำสุดไปหาข้อที่มีค่าความยากสูงสุด นั่นคือ จาก -1.63 ถึง 1.73 เพื่อเตรียมสร้างแบบทดสอบแยกกลุ่ม

2. การสร้างแบบทดสอบแยกกลุ่ม

แบบทดสอบแยกกลุ่มมี 3 ฉบับ ซึ่งสร้างโดยทำการสุ่มข้อสอบจากข้อ 1.11 ได้ผลดังนี้

ฉบับที่ 1 แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน มีค่าความยากอยู่ในช่วง -0.48 ถึง 0.46 จำนวน 20 ข้อ

ฉบับที่ 2 แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง มีค่าความยากอยู่ในช่วง -1.63 ถึง 1.76 จำนวน 20 ข้อ

ฉบับที่ 3 แบบทดสอบแยกกลุ่มสองครั้ง โดยครั้งแรกมีค่าความยากอยู่ในช่วง -0.46 ถึง 0.43 จำนวน 10 ข้อ และครั้งที่สองมีค่าความยาก 2 ระดับ คือ ค่าความยากใกล้เคียงกันในระดับที่ง่ายกว่าในช่วง -0.90 ถึง -0.55 จำนวน 10 ข้อ และค่าความยากใกล้เคียงกันในระดับที่ยากกว่าในช่วง 0.55 ถึง 0.80 จำนวน 10 ข้อ

3. จัดทำคู่มือดำเนินการสอบ และจัดพิมพ์รูปเล่ม

ลักษณะของแบบทดสอบที่ใช้ในการศึกษาค้นคว้า

แบบทดสอบที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นแบบทดสอบวัดผลสัมฤทธิ์ที่มีรูปแบบของแบบทดสอบแยกกลุ่มแตกต่างกัน ชนิดเลือกตอบ 5 ตัวเลือก โดยจัดเป็นแบบทดสอบแยกกลุ่ม 3 ฉบับ ฉบับละ 20 ข้อ ซึ่งมีรายละเอียดดังนี้

คำชี้แจง แบบทดสอบชุดนี้เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ เรื่อง สมการและอสมการ ซึ่งมีจำนวน 20 ข้อ ขอให้นักเรียนตั้งใจทำให้เต็มความสามารถ และควรใช้ความคิดอย่างรอบคอบ

ตัวอย่างข้อสอบ

(๘) พี่น้องสองคนมีอายรวมกันไม่เกิน 30 ปี เขียนเป็นประโยคสัญลักษณ์ได้ดังข้อใด

ก. $x - y < 30$

ข. $x - y = 30$

ค. $x + y < 30$

ง. $x + y = 30$

จ. $xy < 30$

คำตอบ คือ ข้อ ค

(๘๘) ถ้า $5(x - 4) = 3(x + 2)$ แล้ว x มีค่าเท่าใด

ก. 10

ข. 11

ค. 12

ง. 13

จ. 14

คำตอบ คือ ข้อ ง

การตรวจให้คะแนน

การตรวจให้คะแนนทั้ง 3 ฉบับ จะให้ 1 คะแนนสำหรับข้อที่ตอบถูก และให้ 0 คะแนนสำหรับข้อที่ตอบผิด

วิธีดำเนินการเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลในครั้งนี้ ผู้วิจัยดำเนินการตามขั้นตอน ดังนี้

1. วางแผนในการดำเนินการสอบ โดยติดต่อขอความร่วมมือในการนำแบบทดสอบไปทดสอบกับนักเรียนกลุ่มตัวอย่าง ตลอดจนกำหนดวัน และเวลาในการดำเนินการสอบ
2. นำข้อสอบวัดผลสัมฤทธิ์ทางการเรียนที่สร้างขึ้นจำนวน 120 ข้อ มาจัดทำเป็นแบบทดสอบ 4 ฉบับ ๆ ละ 30 ข้อ แล้วนำไปทดสอบกับกลุ่มตัวอย่างโดยเป็นนักเรียนคนละกลุ่มกับกลุ่มตัวอย่างที่ใช้ในการวิจัยจำนวน 1,679 คน จำนวน 44 ห้องเรียน โดยจัดการดำเนินการสอบเป็นระบบดังนี้คือ ให้นักเรียนคนที่ 1, 2, 3, 4 ทำแบบทดสอบฉบับที่ 1, 2, 3 และ 4 สลับกันไป ในลักษณะนี้จึงทำให้จำนวนนักเรียนที่ตอบแบบทดสอบในแต่ละฉบับมีจำนวนใกล้เคียงกัน
3. นำผลการสอบในข้อ 2 มาตรวจให้คะแนน แล้วนำไปวิเคราะห์องค์ประกอบและวิเคราะห์ข้อสอบรายข้อ แล้วจึงนำมาจัดกลุ่มข้อสอบตามค่าความยาก เพื่อเตรียมสร้างแบบทดสอบแยกกลุ่มที่มีรูปแบบแตกต่างกัน
4. นำแบบทดสอบในข้อ 3 ทั้ง 3 รูปแบบ คือ แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างและแบบทดสอบแยกกลุ่มสองครั้ง ไปทดสอบกับกลุ่มตัวอย่างจำนวน 419 คน โดยนักเรียนแต่ละคนต้องทำแบบทดสอบทั้ง 3 รูปแบบ สำหรับแบบทดสอบทั้ง 3 รูปแบบนั้น ผู้วิจัยสุ่มแต่ละรูปแบบไปทดสอบกับกลุ่มตัวอย่างจำนวน 3 ครั้ง โดยทั้งช่วงการทดสอบแต่ละครั้งเป็นเวลา 1 สัปดาห์
5. นำผลการสอบในข้อ 4 ไปวิเคราะห์ทางสถิติ และทดสอบสมมติฐานของการวิจัย

การวิเคราะห์ข้อมูลและสถิติที่ใช้

1. หาค่าสถิติพื้นฐาน คือ ค่าเฉลี่ย และค่าความเบี่ยงเบนมาตรฐาน
2. การตรวจสอบความเป็นมิติเดียว (Unidimensionality) ของแบบทดสอบ โดยนำผลการทดสอบกับกลุ่มตัวอย่างที่ใช้ในการทดลองสอบ ไปวิเคราะห์องค์ประกอบ (Factor Analysis) โดยใช้โปรแกรมสำเร็จรูป SPSS - X ซึ่งเลือกสกัดองค์ประกอบด้วยวิธีวิเคราะห์องค์ประกอบหลักและหมุนแกนด้วยวิธีแวนแมกซ์ แล้วสังเกตค่าไอเกน ขององค์ประกอบที่ 1 มีค่าสูงกว่าค่าไอเกนขององค์ประกอบที่ 2 อย่างมาก คือมีค่ามากกว่า 3 เท่าขององค์ประกอบที่ 2 และค่าไอเกนขององค์ประกอบที่ 2 มีค่าใกล้เคียงกับองค์ประกอบที่เหลือ แสดงว่า แบบทดสอบนั้น มีการวัดเพียงมิติเดียว (Warm. 1978 : 104 ; Lord. 1980 : 19)
3. ประเมินค่าพารามิเตอร์ความสามารถของผู้สอบ (θ) พารามิเตอร์ความยากของข้อสอบ (b) และค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) ตามทฤษฎีการตอบข้อสอบ แบบราสซัส โมเดล ด้วยโปรแกรมสำเร็จรูป BICAL
4. คำนวณค่าความเที่ยงตรงเชิงสภาพ (Concurrent Validity) ในการประมาณค่าความสามารถของผู้สอบ จากแบบทดสอบแยกกลุ่มที่แตกต่างกัน 3 รูปแบบ โดยการคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ระหว่าง ความสามารถที่ได้จากการประมาณค่าด้วยการทดสอบแบบทดสอบแต่ละรูปแบบ กับคะแนนสอบปลายภาคในวิชาคณิตศาสตร์ โดยใช้สูตรสัมประสิทธิ์สหสัมพันธ์เพียร์สัน (Pearson Product Moment Correlation Coefficient) (ล้วน สายยศ และอังคณา สายยศ. 2522 : 168)

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

- เมื่อ $r_{\theta Y}$ แทน ค่าความเที่ยงตรงเชิงสภาพของการประมาณค่าความสามารถ
- θ แทน ค่าความสามารถที่ได้จากการประมาณค่า
- Y แทน คะแนนสอบปลายภาควิชาคณิตศาสตร์
- N แทน จำนวนนักเรียนในกลุ่มตัวอย่าง

5. ทดสอบความแตกต่างของค่าความเที่ยงตรงเชิงสภาพ โดยนำค่าความเที่ยงตรงมาแปลงเป็นคะแนนมาตรฐาน (Z) ตามตารางของฟิชเชอร์ แล้วนำมาทดสอบโดยใช้สูตรไคสแควร์ (Chi Square : χ^2 ดังนี้ (Steel and Torrie. 1981 : 281)

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

เมื่อ χ^2 แทน ค่าจาก χ^2 distribution โดยที่ $df = K - 1$

Z แทน คะแนนมาตรฐานที่แปลงมาจากค่าความเที่ยงตรงเชิงสภาพตามตารางของฟิชเชอร์

N แทน จำนวนนักเรียนในกลุ่มตัวอย่าง

K แทน จำนวนค่าความเที่ยงตรงเชิงสภาพ

6. ถ้าพบความแตกต่างอย่างมีนัยสำคัญให้นำค่าความเที่ยงตรงเชิงสภาพมาทดสอบความแตกต่างเป็นรายคู่ โดยใช้สูตรดังนี้ (ลิวน สายยศ และอังคณา สายยศ. 2522 : 227)

$$Z = \frac{Zr_1 - Zr_2}{\sqrt{\frac{1}{(N_1 - 3)} + \frac{1}{(N_2 - 3)}}}$$

เมื่อ Z แทน คะแนนมาตรฐานจากโค้งปกติ

Zr_1, Zr_2 แทน คะแนนมาตรฐานที่แปลงมาจากค่าสัมประสิทธิ์สหสัมพันธ์ตามตารางของฟิชเชอร์

N_1, N_2 แทน จำนวนนักเรียนในกลุ่มตัวอย่างที่ทำแบบทดสอบแต่ละรูปแบบ

7. ทดสอบความแตกต่างระหว่าง ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ โดยใช้การวิเคราะห์ความแปรปรวนแบบการวัดซ้ำ (Analysis of variance for a repeated measures design) (Hays. 1994 : 569)

8. ถ้าพบความแตกต่างอย่างมีนัยสำคัญให้นำค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ มาทดสอบความแตกต่างเป็นรายคู่ โดยใช้วิธีของเชฟเฟ (Scheffe') และใช้สูตรดังนี้ (ลิวัน สายยศ และอังคณา สายยศ. 2536 : 119)

$$S_c = \sqrt{(K - 1) F(\alpha; df_1, df_2)} \sqrt{MS_E \left[\sum_{j=1}^k \frac{(C_j)^2}{n_j} \right]}$$

เมื่อ S_c	แทน	ค่าวิกฤตของ Scheffe'
K	แทน	จำนวนค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ
$F(\alpha; df_1, df_2)$	แทน	เปิดค่า F ในตารางการแจกแจงค่า F ที่ มีค่า .05 หรือ .01 และ df_1 คือ ค่า df_B และ df_2 คือ ค่า df_w
MS_E	แทน	ค่าความคลาดเคลื่อนของความแปรปรวนในตารางวิเคราะห์ความแปรปรวน
C_j	แทน	สัมประสิทธิ์ Contrast ซึ่งมีค่าเป็น 1, -1, 1, -1, ...
n_j	แทน	จำนวนคนในกลุ่มตัวอย่างแต่ละกลุ่ม

บทที่ 4

ผลการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้เพื่อให้เกิดความเข้าใจตรงกันในการนำเสนอและแปลผลการวิเคราะห์ข้อมูล ผู้วิจัยจึงได้กำหนดสัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูลดังนี้

สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล

$\bar{X}$	แทน	ค่าเฉลี่ย
S	แทน	ค่าความเบี่ยงเบนมาตรฐาน
N	แทน	จำนวนนักเรียนในกลุ่มตัวอย่าง
b	แทน	ค่าความยากของข้อสอบ
θ	แทน	คะแนนความสามารถของผู้สอบ
Y	แทน	คะแนนสอบวิชาคณิตศาสตร์
PRT	แทน	แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน
RRT	แทน	แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง
DRT	แทน	แบบทดสอบแยกกลุ่มสองครั้ง
$r_{\theta Y}$	แทน	ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ
Z	แทน	คะแนนมาตรฐานซึ่งแปลงมาจากค่าสัมประสิทธิ์สหสัมพันธ์ตามตารางของพิชเชอร์
k	แทน	จำนวนค่าของสัมประสิทธิ์สหสัมพันธ์
df	แทน	ขั้นของความเป็นอิสระ
χ^2	แทน	ค่าไคสแควร์
SEE_{PRT}	แทน	ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ จากแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน
SEE_{RRT}	แทน	ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ จากแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง

SEE_{DRT}	แทน	ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ จากแบบทดสอบแยกกลุ่มสองครั้ง
$\overline{SEE}(\theta)$	แทน	ค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ
F	แทน	ค่าที่ใช้พิจารณาการแจกแจง F
SS	แทน	ผลบวกกำลังสองของคะแนนแบบทดสอบ
MS	แทน	ความแปรปรวนของคะแนนแบบทดสอบ
S_c	แทน	ค่าวิกฤตของเชฟเฟ (Scheffe')

การเสนอผลการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลตามลำดับ ดังนี้

1. ผลการหาค่าสถิติพื้นฐานของ แบบทดสอบแยกกลุ่มที่มีความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง
2. ผลการหาค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ และการทดสอบความแตกต่างของค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง
3. ผลการหาค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ และการทดสอบความแตกต่าง ของค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง

ผลการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลตามลำดับ ดังนี้

1. ค่าสถิติพื้นฐานของแบบทดสอบแยกกลุ่มทั้ง 3 รูปแบบ
หลังจากจัดแบบทดสอบแยกกลุ่มทั้ง 3 รูปแบบ ๗ ละ 20 ข้อ และนำไปทดสอบกับกลุ่มตัวอย่างจำนวน 419 คน แล้วจึงได้นำมาวิเคราะห์ค่าสถิติพื้นฐาน ได้ผลดังตาราง 2

ตาราง 2 ค่าสถิติพื้นฐานของค่าความยากของข้อสอบและค่าความสามารถของผู้สอบ

แบบทดสอบ	b		e	
	$\bar{X}$	S	$\bar{X}$	S
PRT	-0.0030	0.2875	-0.3724	0.8328
RRT	-0.0580	1.1097	-0.5499	0.8541
DRT	-0.0170	0.6057	-0.3377	0.7445

จากตาราง 2 แสดงให้เห็นว่า ค่าความยากเฉลี่ยของข้อสอบมีค่าอยู่ระหว่าง -0.0580 ถึง -0.0030 โดยที่แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันมีค่าความยากเฉลี่ยของข้อสอบสูงสุด แบบทดสอบแยกกลุ่มสองครั้งมีค่าความยากเฉลี่ยของข้อสอบรองลงมา และแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างมีค่าความยากเฉลี่ยของข้อสอบต่ำสุด นั่นคือ แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างง่ายที่สุด แบบทดสอบแยกกลุ่มสองครั้งง่ายรองลงมา และแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันง่ายน้อยที่สุด ส่วนค่าความเบี่ยงเบนมาตรฐานของแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างมีค่าสูงสุดในขณะที่แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันมีค่าความเบี่ยงเบนมาตรฐานต่ำสุด

ส่วนค่าความสามารถเฉลี่ยของผู้สอบมีค่าอยู่ระหว่าง -0.5499 ถึง -0.3377 โดยที่แบบทดสอบแยกกลุ่มสองครั้ง มีค่าความสามารถเฉลี่ยของผู้สอบสูงสุด และแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างมีค่าความสามารถเฉลี่ยของผู้สอบต่ำสุด ส่วนค่าความเบี่ยงเบนมาตรฐานของแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างมีค่าสูงสุดในขณะที่แบบทดสอบแยกกลุ่มสองครั้งมีค่าความเบี่ยงเบนมาตรฐานต่ำสุด

2. ผลการหาค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ และการทดสอบความแตกต่างของค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ระหว่าง

แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้งนั้น ผู้วิจัยได้นำค่าประมาณความสามารถของผู้สอบที่ได้จากการทดสอบด้วย แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง จากนักเรียนจำนวน 419 คนไปคำนวณหาค่าความเที่ยงตรง โดยการคำนวณหาค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน ระหว่างค่าความสามารถของผู้สอบจากการทดสอบด้วยแบบทดสอบแต่ละรูปแบบ กับคะแนนสอบวิชาคณิตศาสตร์ แล้วทดสอบความแตกต่างของค่าความเที่ยงตรงระหว่างแบบทดสอบทั้ง 3 รูปแบบ ด้วยวิธีแปลงค่าสัมประสิทธิ์สหสัมพันธ์ ให้อยู่ในรูปคะแนนมาตรฐานตามตารางของฟิชเชอร์ โดยใช้ค่าไคสแควร์ ได้ผลดังตาราง 3

ตาราง 3 ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ และผลการทดสอบความแตกต่างของค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบจากการทำแบบทดสอบที่แตกต่างกัน 3 รูปแบบ

แบบทดสอบ	r_{oy}	Z	χ^2
PRT	0.5516	0.6198	9.3954**
RRT	0.6431	0.7647	
DRT	0.6783	0.8269	

** $P < .01$

จากตาราง 3 แสดงให้เห็นว่า ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ จากการทดสอบด้วยแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างและแบบทดสอบแยกกลุ่มสองครั้งมีค่าเท่ากับ 0.5516, 0.6431 และ 0.6783 ตามลำดับ ซึ่งค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบจากแบบ

ทดสอบแยกกลุ่มสองครั้งมีค่าสูงสุด รองลงมาเป็นแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันตามลำดับ แสดงว่าแบบทดสอบแยกกลุ่มสองครั้งวัดความสามารถของผู้สอบได้ตรงกับสภาพความเป็นจริงมากที่สุด และแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน วัดความสามารถของผู้สอบได้ตรงกับสภาพความเป็นจริงน้อยที่สุด และเมื่อนำมาทดสอบความแตกต่าง ของค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ปรากฏว่า แบบทดสอบทั้ง 3 รูปแบบมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่ามีค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ แตกต่างกันอย่างน้อย 1 คู่ ดังนั้น เพื่อให้ทราบค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบจากแบบทดสอบคู่ใดมีความแตกต่างกัน ผู้วิจัยจึงนำค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบจากแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง ไปทดสอบความแตกต่างทีละคู่ ได้ผลดังตาราง 4

ตาราง 4 ผลของการทดสอบความแตกต่างของค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ จากการทำแบบทดสอบทั้ง 3 รูปแบบ โดยทดสอบเป็นรายคู่

		PRT	RRT	DRT
แบบทดสอบ	Z	0.6198	0.7647	0.8269
PRT	0.6198	-	2.0909*	2.9885**
RRT	0.7647		-	0.8975
DRT	0.8269			-

** P < .01

* P < .05

จากตาราง 4 แสดงให้เห็นว่า ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน กับแบบทดสอบแยกกลุ่มสองครั้ง มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันกับแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างกับแบบทดสอบแยกกลุ่มสองครั้งมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

3. ผลการหาค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ และการทดสอบความแตกต่างของค่าความคลาดเคลื่อนมาตรฐาน ในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง

ผู้วิจัยได้นำผลการสอบ ที่ได้จากการทดสอบด้วยแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง จากนักเรียนจำนวน 419 คน ไปคำนวณค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถได้ผลดังตาราง 5 (ซึ่งมีรายละเอียดดังตาราง 14 ในภาคผนวก ง)

ตาราง 5 ค่าสถิติพื้นฐานของค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ

แบบทดสอบ	$\bar{X}$	S
PRT	0.5459	0.0525
RRT	0.5611	0.0544
DRT	0.5090	0.0809

จากตาราง 5 แสดงให้เห็นว่า ค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ จากการทดสอบด้วยแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง มีค่าเท่ากับ 0.5459, 0.5611 และ 0.5090 ตามลำดับโดยที่ค่าเฉลี่ยของความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบจากแบบทดสอบแยกกลุ่มสองครั้งมีค่าต่ำสุด และจากแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างมีค่าสูงสุด ส่วนค่าความเบี่ยงเบนมาตรฐานของแบบทดสอบแยกกลุ่มสองครั้งมีค่าสูงสุด ในขณะที่แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันมีค่าความเบี่ยงเบนมาตรฐานต่ำสุด และเพื่อให้ทราบค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบแตกต่างกันหรือไม่ ผู้วิจัยจึงนำค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบทั้ง 3 รูปแบบไปทดสอบความแตกต่างโดยใช้การวิเคราะห์ความแปรปรวนแบบการวัดซ้ำได้ผลดังตาราง 6

ตาราง 6 ผลการทดสอบความแตกต่างของค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ จากการทำแบบทดสอบที่แตกต่างกัน 3 รูปแบบ

แหล่งความแปรปรวน	SS	df	MS	F
ระหว่างภายในบุคคล	2.1100	418	-	64.8696**
ระหว่างรูปแบบของแบบทดสอบ	0.5967	2	0.2984	
ความคลาดเคลื่อนระหว่าง-บุคคลกับรูปแบบ	3.8757	836	0.0046	
รวม	6.5824	1256		

** $P < .01$

จากตาราง 6 แสดงให้เห็นว่า ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มทั้ง 3 รูปแบบ มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงว่า มีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบแตกต่างกันอย่างน้อย 1 คู่ ดังนั้นเพื่อให้ทราบค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบจากแบบทดสอบคู่ใดมีความแตกต่างกัน ผู้วิจัยจึงนำค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ จากแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างและแบบทดสอบแยกกลุ่มสองครั้งไปทดสอบความแตกต่างทีละคู่ ได้ผลดังตาราง 7

ตาราง 7 ผลของการทดสอบความแตกต่างของค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ จากการทำแบบทดสอบทั้ง 3 รูปแบบ โดยทดสอบเป็นรายคู่

แบบทดสอบ		PRT	RRT	DRT
	SEE(๑)	๐.5498	๐.5611	๐.5๐9๐
PRT	๐.5498	-	๐.๐152**	๐.๐369**
RRT	๐.5611		-	๐.๐521**
DRT	๐.5๐9๐			-

** $P < .01$

$$S_c = 0.0136$$

จากตาราง 7 แสดงให้เห็นว่า ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันกับแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายอยู่ในช่วงกว้าง แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันกับแบบทดสอบแยกกลุ่มสองครั้งและแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างกับแบบทดสอบแยกกลุ่มสองครั้ง มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกค่า

บทที่ 5

สรุป อภิปรายผล ข้อเสนอแนะ

ความมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อเปรียบเทียบค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีรูปแบบต่างกัน
2. เพื่อเปรียบเทียบค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีรูปแบบต่างกัน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้า เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษาในจังหวัดลพบุรี จำนวนโรงเรียน 7 โรงเรียน จำนวนห้องเรียน 11 ห้องเรียน จำนวนนักเรียน 419 คน ซึ่งได้มาจากการสุ่มแบบแบ่งชั้น (Stratified Random Sampling)

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้า เป็นแบบทดสอบเลือกตอบวัดผลสัมฤทธิ์ทางการเรียน ชนิด 5 ตัวเลือก ซึ่งผู้วิจัยสร้างขึ้นเองจากเนื้อหาในวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 2 (ค 204) เรื่องสมการและอสมการ ตามหลักสูตรมัธยมศึกษาตอนต้น พุทธศักราช 2521 (ฉบับปรับปรุง พ.ศ. 2533) โดยมี 3 รูปแบบ ดังนี้

1. แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน จำนวน 20 ข้อ
2. แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง จำนวน 20 ข้อ
3. แบบทดสอบแยกกลุ่มสองครั้ง จำนวน 20 ข้อ

วิธีดำเนินการเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลในครั้งนี้ ผู้วิจัยดำเนินการตามขั้นตอน ดังนี้

1. วางแผนในการดำเนินการสอบ โดยติดต่อขอความร่วมมือในการนำแบบทดสอบไปทดสอบกับนักเรียนกลุ่มตัวอย่าง ตลอดจนกำหนดวัน และเวลาในการดำเนินการสอบ
2. นำแบบทดสอบทั้ง 3 รูปแบบ คือ แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง ไปทดสอบกับกลุ่มตัวอย่างจำนวน 419 คน โดยนักเรียนแต่ละคนจะต้องทำแบบทดสอบทั้ง 3 รูปแบบ สำหรับแบบทดสอบทั้ง 3 รูปแบบนั้นผู้วิจัยจะสุ่มแต่ละรูปแบบไปทดสอบกับกลุ่มตัวอย่าง จำนวน 3 ครั้ง โดยทั้งช่วงการทดสอบแต่ละครั้งเป็นเวลา 1 สัปดาห์
3. นำผลการสอบในข้อ 2 มาจัดเตรียมข้อมูลสำหรับการประมาณค่าความสามารถของผู้สอบ และทดสอบสมมติฐานของการวิจัยต่อไป

การเสนอผลการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลตามลำดับ ดังนี้

1. ผลการหาค่าสถิติพื้นฐานของแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง
2. ผลการหาค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ และการทดสอบความแตกต่างของค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง
3. ผลการหาค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ และการทดสอบความแตกต่าง ของค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง

สรุปผลการวิเคราะห์ข้อมูล

1. ค่าสถิติพื้นฐานของแบบทดสอบแยกกลุ่มทั้ง 3 รูปแบบ พบว่า แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันมีค่าความยากเฉลี่ยของข้อสอบสูงสุด แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างมีค่าความยากเฉลี่ยของข้อสอบต่ำสุด ส่วนแบบทดสอบแยกกลุ่มสองครั้งมีค่าความสามารถเฉลี่ยของผู้สอบสูงสุด แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างมีค่าความสามารถเฉลี่ยของผู้สอบต่ำสุด และแบบทดสอบแยกกลุ่มสองครั้งมีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบต่ำสุด แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างมีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบสูงสุด

2. ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันกับแบบทดสอบแยกกลุ่มสองครั้ง และแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน กับแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 ตามลำดับ ส่วนค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างกับแบบทดสอบแยกกลุ่มสองครั้ง มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

3. ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน กับแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันกับแบบทดสอบแยกกลุ่มสองครั้ง และแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างกับแบบทดสอบแยกกลุ่มสองครั้ง มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกค่า

อภิปรายผล

1. จากการศึกษาเปรียบเทียบ ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง พบว่า แบบทดสอบแยกกลุ่มสองครั้งจะให้ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบสูงสุด รองลงมาได้แก่ แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน ตามลำดับ เมื่อนำค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบจากแบบทดสอบแยก

กลุ่มทั้งสามรูปแบบมาทดสอบความแตกต่าง ปรากฏว่า ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน กับแบบทดสอบแยกกลุ่มสองครั้งและแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันกับแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 ตามลำดับ ซึ่งเป็นไปตามสมมติฐานที่ตั้งไว้ในข้อ 1 และยังสอดคล้องกับผลการวิจัยของ เกรียงศักดิ์ สุวรรณภาค (2532 : 50) ที่พบว่าค่าสัมประสิทธิ์สหสัมพันธ์ ระหว่างผลการสอบที่ใช้แบบทดสอบแยกกลุ่มที่มีรูปแบบแตกต่างกันสามรูปแบบ คือ แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างและแบบทดสอบแยกกลุ่มสองครั้งกับความสามารถทางการเรียน มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01 ทุกค่า โดยที่ค่าสหสัมพันธ์ข้างต้นถูกจัดเป็นค่าความเที่ยงตรงของแบบทดสอบตามทฤษฎีการตอบข้อสอบ (งามนิตย์ ชาติทอง และปรีชา เจริญวรรณ. 2527 : 37 อ้างอิงมาจาก Lord and Novick. 1968) แสดงว่า แบบทดสอบแยกกลุ่มสองครั้งสามารถแยกระดับความสามารถของผู้สอบได้ดีกว่าแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน ทั้งนี้เนื่องจากแบบทดสอบแยกกลุ่มสองครั้ง ผู้สอบจะต้องทำข้อสอบถึงสองครั้ง เพื่อแยกระดับความสามารถของเขาให้ใกล้เคียงกับความสามารถที่แท้จริงให้มากที่สุด แล้วจะทำให้ผู้สอบได้ตอบข้อสอบที่มีค่าความยากตรงกับความสามารถของเขา ซึ่งจะส่งผลให้การประมาณค่าความสามารถของผู้สอบมีความถูกต้องแม่นยำยิ่งขึ้น ถึงแม้ว่าแบบทดสอบแยกกลุ่มสองครั้งกับแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน จะมีโครงสร้างการจัดค่าความยากของข้อสอบคล้ายกันก็ตาม แต่การที่จะนำแบบทดสอบแยกกลุ่มไปทดสอบเพียงครั้งเดียว โอกาสที่จะแยกระดับความสามารถของผู้สอบคลาดเคลื่อนจากความสามารถที่แท้จริงก็เป็นไปได้ จึงมีผลทำให้ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบจากแบบทดสอบแยกกลุ่มสองครั้ง สูงกว่าแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน ส่วนการเปรียบเทียบระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน กับแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง ปรากฏว่าแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างสามารถแยกระดับความสามารถของผู้สอบได้ดีกว่าแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน ทั้งนี้เนื่องจาก แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน เป็นแบบทดสอบที่มีการจัดข้อสอบให้มีค่าความยากอยู่ในระดับปานกลาง ซึ่งคล้ายกับแบบทดสอบดั้งเดิม ที่มีการจัดข้อสอบให้มีความเหมาะสมกับกลุ่มผู้สอบ เฉพาะความสามารถในระดับปานกลางเท่านั้น แต่จะไม่คำนึงถึงกลุ่มผู้สอบที่ความสามารถในระดับสูงหรือ

ระดับต่ำ ส่วนแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง เป็นแบบทดสอบที่มีการจัดข้อสอบให้มีความยากเรียงลำดับจากง่ายที่สุดไปจนถึงยากที่สุดซึ่งมีความเหมาะสมมากกว่าไม่ว่าผู้สอบจะมีความสามารถในระดับสูง ปานกลาง หรือต่ำ จึงมีผลทำให้การประมาณค่าความสามารถของผู้สอบ จากแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง มีความถูกต้องแม่นยำสูงกว่าแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน และเมื่อพิจารณาค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง กับแบบทดสอบแยกกลุ่มสองครั้ง ปรากฏว่า มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ซึ่งไม่เป็นไปตามสมมติฐานที่ตั้งไว้ในข้อ 1 และยังขัดแย้งกับผลการวิจัยของ เกรียงศักดิ์ สุวรรณภาค (2532 : 50) ดังที่กล่าวมาแล้วข้างต้น แสดงว่าแบบทดสอบแยกกลุ่มทั้งสองรูปแบบนี้สามารถแยกระดับความสามารถของผู้สอบได้ใกล้เคียงกัน นั่นคือ ถ้าคำนึงถึงคุณภาพในด้านความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้งแล้ว จะเลือกใช้แบบทดสอบแยกกลุ่มรูปแบบใดก็ได้

2. จากการศึกษาเปรียบเทียบ ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง พบว่า แบบทดสอบแยกกลุ่มสองครั้งจะให้ค่าความคลาดเคลื่อนมาตรฐานต่ำสุด สูงขึ้นไป ได้แก่ แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน และแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง ตามลำดับ เมื่อนำค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ จากแบบทดสอบแยกกลุ่มทั้งสามรูปแบบมาทดสอบความแตกต่าง ปรากฏว่า ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ระหว่าง แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกค่า ซึ่งเป็นไปตามสมมติฐานที่ตั้งไว้ในข้อ 2 โดยที่แบบทดสอบแยกกลุ่มสองครั้ง เป็นแบบทดสอบที่สามารถแยกระดับความสามารถของผู้สอบได้ดีที่สุด เมื่อเปรียบเทียบกับแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน และแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง เนื่องจากแบบทดสอบแยกกลุ่มสองครั้งผู้สอบต้องทำแบบทดสอบแยกกลุ่มถึงสองครั้ง เพื่อจะได้ตอบข้อสอบที่มีค่าความยากตรงกับระดับความสามารถของเขา ซึ่งจะส่งผลให้การประมาณค่าความสามารถมีความถูกต้องแม่นยำสูง ค่าความคลาดเคลื่อนมาตรฐาน

ในการประมาณค่าความสามารถของผู้สอบจะลดลง ซึ่งสอดคล้องกับผลการวิจัยของ แฮมเบิลตัน และกรูเจอร์ (Hambleton and Grujter. 1983 : 355 - 367) ที่พบว่าการใช้ข้อสอบที่มีระดับความยากพอเหมาะกับระดับความสามารถของผู้สอบแล้ว จะสามารถลดค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าลงได้ นั่นคือ แบบทดสอบแยกกลุ่มสองครั้งมีความเหมาะสมมากที่สุดที่จะนำมาใช้แยกระดับความสามารถของผู้สอบ เพื่อไปตอบแบบทดสอบหลักตามรูปแบบของแบบทดสอบสองขั้นตอน เนื่องจากมีความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบสูง แต่ให้ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบต่ำ ดังที่กล่าวมาแล้วข้างต้น ส่วนค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบจากแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างมีค่าสูงสุด ทั้งนี้อาจมีสาเหตุมาจากแบบทดสอบตามรูปแบบนี้จะจัดข้อสอบที่มีค่าความยากเรียงลำดับจากง่ายที่สุดไปจนถึงยากที่สุด แต่แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันนั้น มีข้อสอบที่มีค่าความยากระดับปานกลางเท่านั้น และจากการวิจัยของกรีน และคนอื่น ๆ (Green and others. 1984 : 352) พบว่า ข้อสอบที่สร้างขึ้นเพื่อให้ผู้สอบที่มีความสามารถระดับปานกลางจะมีความแปรปรวนคลาดเคลื่อนค่อนข้างต่ำและจะมีความแปรปรวนคลาดเคลื่อนค่อนข้างสูง เมื่อผู้สอบมีความสามารถระดับต่ำมากหรือสูงมาก ดังนั้นจึงสอดคล้องกับการวิจัยในครั้งนี้พบว่า แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างมีค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบสูงกว่าแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน

ข้อเสนอแนะ

1. ควรมีการศึกษาเพิ่มเติม โดยนำแบบทดสอบหลัก (Main Test) มาศึกษาร่วมกับแบบทดสอบแยกกลุ่ม (Routing Test) แล้วจึงนำคะแนนที่ได้จากแบบทดสอบทั้งสองประเภทไปพิจารณาค่าความสามารถของผู้สอบ
2. ควรมีการศึกษาเพิ่มเติมในเรื่องคุณภาพของแบบทดสอบ ที่เกี่ยวกับความเชื่อมั่น (Reliability) ตามรูปแบบของทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory)
3. ควรมีการศึกษาในลักษณะเดียวกันนี้ โดยเพิ่มวิธีการให้คะแนนที่แตกต่างกัน เช่น ให้คะแนนด้วยวิธีของเบย์ส (Bayesian Method) ให้คะแนนตามค่าเฉลี่ยความยากของข้อที่ทำถูก (Average Difficulty of All Items Answered Correctly) ให้คะแนนตามค่าความ

ยากของข้อที่ยากที่สุดทำถูก (The Most Difficulty of Item Answered Correctly) ให้คะแนนตามค่าเฉลี่ยความยากของทุกข้อที่ตอบ (Average Difficulty) เป็นต้น

4. ความรู้การศึกษาในลักษณะเดียวกันนี้ โดยใช้โมเดลที่มีความละเอียดซับซ้อนขึ้น เช่น โลจิสติก โมเดลที่มี 2 หรือ 3 พารามิเตอร์ เป็นต้น

บรรณานุกรม

- เกรียงศักดิ์ สุวรรณภาค. ความสัมพันธ์ระหว่างรูปแบบบางรูปแบบของแบบทดสอบแยกกลุ่มกับความ
ความสามารถทางการเรียน ในวิชาคณิตศาสตร์. ปรินซ์ตันนิพนธ์ กศ.ม. กรุงเทพฯ :
มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2532. อัดสำเนา.
- ขรรค์ชัย คงเสน่ห์. การสร้างแบบทดสอบสองขั้นตอนโดยใช้ทฤษฎีการตอบข้อคำถาม.
ปรินซ์ตันนิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร,
2530. อัดสำเนา.
- งามนิตย์ ชาททอง และปรีชา เครือวรรณ. "การให้คะแนนตัวเองในการสอบแบบทดสอบ
แฟล็กชีเลเวล," รายงานการวิจัย. ขอนแก่น : มหาวิทยาลัยขอนแก่น, 2527.
- ฉัตรนภา พรหมมา. "Tailored Testing : ผลสำเร็จของการประยุกต์ใช้ Latent Trait
Theory," วารสารการวัดผลการศึกษา. 4(2) : 45 - 53 ; กันยายน - ธันวาคม
2525.
- ด้าย เขียงฉี. การศึกษาเปรียบเทียบประสิทธิภาพในการประมาณค่าความสามารถของผู้สอบ
จากการทดสอบเทเลอร์รูปปิรามิด ที่มีรูปแบบ จำนวนชั้น และวิธีการให้คะแนนที่แตก
ต่างกันโดยใช้วิธีมอนติ คาร์โล. ปรินซ์ตันนิพนธ์ กศ.ด. กรุงเทพฯ : มหาวิทยาลัย-
ศรีนครินทรวิโรฒ ประสานมิตร, 2534. อัดสำเนา.
- นริศ สวัสดิ์. การศึกษาความสัมพันธ์ระหว่างคะแนนของแบบทดสอบสองขั้นตอนและคะแนนของ
แบบทดสอบรูปปิรามิด. ปรินซ์ตันนิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัย
ศรีนครินทรวิโรฒ ประสานมิตร, 2533. อัดสำเนา.
- นันทิยา พิงคำ. การเปรียบเทียบคุณภาพการทดสอบแบบ เอ ที ซี และแบบประเมินนิยมในการ
วัดความสามารถด้านคำศัพท์ภาษาอังกฤษของนักเรียนชั้นมัธยมศึกษาปีที่ 3. วิทยานิพนธ์
ค.ม. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2531. อัดสำเนา.
- ผจงจิต อินทสุพรรณ. "Latent Trait Theory," วารสารการวัดผลการศึกษา. 3(3) :
51 - 69 ; มกราคม - เมษายน 2525.
- ล้วน สายยศ และอังคณา สายยศ. สถิติวิทยาทางการศึกษา. กรุงเทพฯ : วัฒนาพานิช,
2522.

ล้วน สายยศ และอังคณา สายยศ. เทคนิคการวิจัยทางการศึกษา. กรุงเทพฯ : ศูนย์ส่งเสริมวิชาการ, 2536.

สงบ ลักษณะ. "ความสามารถเกี่ยวกับการตอบข้อสอบ," วารสารการวัดผลการศึกษา.

4(1) : 47-54 ; พฤษภาคม - สิงหาคม 2525.

สุธรรม จันทร์หอม. "แบบทดสอบต่อเนื่อง : Tailored Test," วารสารการวัดผลการศึกษา. 7(21) : 10 - 22 ; มกราคม - เมษายน 2529.

อนันต์ ศรีโสภ. การวัดผลการศึกษา. กรุงเทพฯ : ไทยวัฒนาพานิช, 2525.

อรินทร์ น่วมถนนอม. การศึกษาความเที่ยงตรงและประสิทธิภาพสัมพัทธ์ของการประมาณค่าความสามารถในวิชาคณิตศาสตร์ ระหว่างแบบทดสอบดั้งเดิม แบบทดสอบสองขั้นตอน และแบบทดสอบแฟล็กชีลเเวล. ปริญญาโท กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2537. อัดสำเนา.

Angoff, W.H. and E.M. Huddleston. "The Multiple - Level Experiment : A Study of a Two - Stage Test System for the College Board Scholastic Aptitude Test," Statistical Report 58 - 21. Princeton N.J. : Educational Testing Service, 1958.

Bayroff, A.G. and L.C. Seeley. "An Exploratory Study of Branching Tests," Technical Report. Note 188 ; June, 1967.

Birnbaum, Allan. "Some Latent Trait Model and Their Use in Inferring in Examinee's Ability," In F.M. Novick. Statistical Theories of Mental Test Scores. Reading, Massachusetts : Addison - Wesley, 1968.

- De Gruijter, D.N.M. "A Note on The Asymptotic Variance - Covariance Matrix of Item Parameter Estimates in The Rasch Model," Psychometrika. 50 : 247 - 249 ; 1985.
- Green, Bert F. and others. "Technical Guidelines for Assessing Computerized Adaptive Tests," Journal of Educational Measurement. 21(4) : 347 - 360 ; Winter, 1984.
- Hambleton, R.K. "Applications of Item Response Theory," International Journal of Educational Research. 13(2) : 121 - 220 ; 1989.
- Hambleton, R.K. and D.N.M. Gruijter "Application of Response Models to criterion - Referenced Test Item Selection," Journal of Educational Measurement. 20 : 355 - 367 ; 1983.
- Hambleton, R.K. and H. Swaminathan. Item Response Theory. Boston : Kluwer - Nijhoff Publishing, 1985.
- Hambleton, R.K. and L.L. Cook. "Latent Trait Models and Their Use in the Analysis of Educational Test Data," Journal of Education Measurement. 14(2) : 75 - 96 ; Summer, 1977.
- Hambleton, R.K. and others. "Developments in Latent Trait Theory : Model, Technical Issues, and Applications," Review of Educational Research. 48(4) : 467 - 510 ; 1978.

Hays L. William. Statistic. Fifth Edition. Harcourt Brace College Publishers. Copyright, 1994.

Husen, Torsten and t. Neville Postlethwaite. "Rasch Measurement Models," The International Encyclopedia of Education. Pergamon Press, 1985...v.

Larkin, K.C. and D.J. Weiss "An Empirical Investigation of Computer Administered Pyramidal Ability Testing," Research Report 74 - 3. Department of Psychology, University of Minnesota, 1974.

———. "An Empirical Comparison of Two - Stage and Pyramidal Adaptive Ability Testing," Research Report 75 - 1. Minneapolis : University of Minnesota, Department of Psychology, Psychometric Methods Program, 1975.

Linn, R.L. and others." The Development and Evaluation of Several Programmed Testing Methods," Educational and Psychological Measurement. 29(1) : 129 - 146 ; 1969.

Lord, F.M "A Theoretical Study of Two - Stage Testing," Psychometrika. 36(3) : 227 - 242 ; 1971.

———. Applications of Item Response Theory to Practical Testing Problem. Hillsdale, New Jersey : Lawrence Erlbaum Associates, 1980.

Lord, F.M. "Item Response Theory and Equating a Technical Summary,"
in Holland, Paul W. and Donald B. Rubin. Test Equating.
New York : Academic Press, 1982.

Lord, F.M and M.R. Novick. Statistical Theories of Mental Test Scores.
Massachusetts : Addison - Wesley, 1968.

Mc Cormick, D.J. "Tailor - APL : An Interactive Computer Program for
Individual Tailored Testing," technical Report. No. 5,
University of Southern California, Los Angeles, Department of
Psychology, March, 1978.

Moreno, K.E. and others. "Relationship Between Corresponding Armed
Services Vocational Aptitude Battery and Computerized Adaptive
Testing Subtests," Applied Psychological Measurement. 8(2) :
155 - 163 ; 1984.

Rasch, G. Probabilistic Model for some Intelligence and Attainment
Test. Copenhagen : the Danish Institute for Educational
Research, 1960.

Reckase, M.D. "Computerized Achievement Testing Using the Simple
Logistic Model," Research Report. Missouri University
Columbia, 1978.

- Schmidt, F.L., V.W. Urry and J.F. Gugel. "Computer Assisted Tailored Testing : Examinee Reactions and Evaluations," Educational and Psychological Measurement. 38(2) : 265 - 273 ; 1978.
- Spinetti, J.P. and R.K. Hambleton. "A Computer Simulation Study of Tailored Testing Strategies for Objective - Based Instructional Program," Educational and Psychological Measurement. 37(1) : 139 - 158 ; 1977.
- Steel, Robert G.D. and James H. Torrie. Principles and Procedures of Statistics. 2nd ed. Singapore : McGraw - Hill, Inc., 1981.
- Urry, V.W. "Tailored Testing : A Successful Application of Latent Trait Theory," Journal of Educational Measurement. 14(2) : 181 - 196 ; 1977.
- Vale, C.D. and D.J. Weiss. "A Simulation Study of Stradative Ability Testing," Research Report 75 - 6. Psychometric Method Program. Department of Psychology, University of Minnesota Minneapolis, 1975.
- Warm, T.A. "A Primer of Item Response Theory," Technical Report 941278. Oklahoma : U.S. Coast Guard Institute, 1978.
- Weiss, D.J. "Strategies of Adaptive Ability Measurement," Research Report 74 - 5. Minneapolis : University of Minnesota, Psychometric Methods Program, Department of Psychology, 1974.

- Weiss, D.J. "Improving Measurement Quality and Efficiency with Adaptive Testing," Applied Psychological Measurement. 6 : 473 - 492 ; 1982.
- Weiss, D.J. and G.G. Kingsbury. "Application of Computerized Adaptive Testing to Educational Problems." Journal of Educational Measurement. 21(4) : 361 - 375 ; 1984.
- Wright, B.D. "Misunderstanding the Rasch Model." Journal of Educational Measurement. 14 : 219 - 225 ; Fall, 1977.
- Wright, B.D. and M.H. Stone. Best Test Design. Chicago : MESA Press, 1979.
- Wright, B.D. and G.N. Masters Rating Scale Analysis. Chicago : MESA Press, 1982.
- Wright, B.D. and N. Panchapakesan, "A Procedure for Sample Free Item Analysis," Educational and Psychological Measurement. 29 : 23 - 48 ; 1969.

ภาคผนวก

ภาคผนวก ก

ผลการวิเคราะห์องค์ประกอบที่มีค่าไอเกนมากกว่า 1
เพื่อตรวจสอบความเป็นมิติเดียวของแบบทดสอบที่นำไปทดลองสอบ จำนวน 4 ฉบับ

ตาราง 8 ค่าไอเกน(มากกว่า 1)เปอร์เซ็นต์ของความแปรปรวน และเปอร์เซ็นต์ของความแปรปรวนสะสม จากการนำแบบทดสอบ จำนวน 4 ฉบับมาวิเคราะห์องค์ประกอบ

แบบทดสอบ ฉบับที่	องค์ประกอบที่	ค่าไอเกน	เปอร์เซ็นต์ของ ความแปรปรวน	เปอร์เซ็นต์ของความ แปรปรวนสะสม
1	1	3.99046	10.0	10.0
	2	1.58862	5.6	15.6
	3	1.48023	4.9	20.5
	4	1.38770	4.6	25.2
	5	1.32777	4.4	29.6
	6	1.25856	4.2	33.8
	7	1.20003	4.0	37.8
	8	1.17331	3.9	41.7
	9	1.12090	3.7	45.4
	10	1.07590	3.6	49.0
	11	1.03307	3.4	52.5
2	1	3.38590	11.3	11.3
	2	1.43368	4.8	16.1
	3	1.38308	4.6	20.7
	4	1.35252	4.5	25.2
	5	1.31155	4.4	29.6
	6	1.22196	4.1	33.6
	7	1.21019	4.0	37.7

ตาราง 8 (ต่อ)

แบบทดสอบ ฉบับที่	องค์ประกอบที่	ค่าไอเกน	เปอร์เซ็นต์ของ ความแปรปรวน	เปอร์เซ็นต์ของความ แปรปรวนสะสม
2	8	1.18682	4.0	41.6
	9	1.11059	3.7	45.3
	10	1.10186	3.7	49.0
	11	1.06580	3.6	52.5
	12	1.05193	3.5	56.1
	13	1.01029	3.4	59.4
3	1	4.48022	14.9	14.9
	2	1.52096	5.1	20.0
	3	1.45324	4.8	24.8
	4	1.27518	4.3	29.1
	5	1.24263	4.1	33.2
	6	1.21267	4.0	37.3
	7	1.14586	3.8	41.1
	8	1.11896	3.7	44.8
	9	1.09139	3.6	48.5
	10	1.02880	3.4	51.9
	11	1.01481	3.4	55.3

ตาราง 8 (ต่อ)

แบบทดสอบ ฉบับที่	องค์ประกอบที่	ค่าไอเกน	เปอร์เซ็นต์ของ ความแปรปรวน	เปอร์เซ็นต์ของความ แปรปรวนสะสม
4	1	5.05963	16.9	16.9
	2	1.58417	5.3	22.1
	3	1.38373	4.6	26.8
	4	1.27632	4.3	31.0
	5	1.23006	4.1	35.1
	6	1.14988	3.8	38.9
	7	1.09663	3.7	42.6
	8	1.05180	3.5	46.1
	9	1.03941	3.5	49.6

ภาคผนวก ข

ค่าประมาณพารามิเตอร์ความยากของข้อสอบ

ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความยากของข้อสอบ

และค่าความสอดคล้องกับโมเดล จากแบบทดสอบจำนวน 4 ฉบับนำไปทดลองสอบ

ตาราง 9 ผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (b) ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความยากของข้อสอบ (SEE) และค่าความสอดคล้องกับโมเดล (t - test) ของแบบทดสอบฉบับที่ 1

ข้อ	b	SEE	t
1	-2.19	0.13	-0.65
2	-0.01	0.11	-0.44
3	-0.82	0.10	-1.45
4	0.42	0.12	0.47
5	0.73	0.13	0.54
6	-1.63	0.11	-2.13
7	-1.45	0.11	-0.61
8	0.34	0.12	-1.58
9	-0.06	0.11	-0.67
10	0.21	0.11	0.35
11	-0.13	0.11	-0.11
12	-0.13	0.11	-0.45
13	0.97	0.14	-0.30
14	-0.01	0.11	-0.27
15	0.35	0.12	0.78
16	2.03	0.20	0.23
17	0.80	0.13	0.84
18	1.55	0.17	-0.51
19	-0.35	0.10	-1.08
20	0.90	0.13	0.48

ตาราง 9 (ต่อ)

ข้อ	b	SEE	t
21	-1.25	0.11	-0.58
22	-0.15	0.11	-0.58
23	0.55	0.12	-0.09
24	0.48	0.12	-0.05
25	-0.74	0.10	-0.71
26	-0.03	0.11	-0.33
27	-0.21	0.11	-0.38
28	-0.21	0.11	-0.57
29	0.67	0.13	0.61
30	-0.65	0.10	0.70

ตาราง 10 ผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (b) ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความยากของข้อสอบ (SEE) และค่าความสอดคล้องกับโมเดล (t - test) ของแบบทดสอบฉบับที่ 2

ข้อ	b	SEE	t
1	-1.63	0.11	-0.55
2	0.83	0.14	1.03
3	-1.23	0.10	-1.16
4	-0.48	0.10	-0.13
5	0.76	0.13	0.27
6	-0.75	0.10	-1.62
7	-0.10	0.11	0.33
8	0.61	0.13	-1.61
9	-0.15	0.11	-1.65
10	0.29	0.12	0.55
11	-0.09	0.11	-1.61
12	-0.18	0.11	-0.34
13	-0.45	0.10	-1.33
14	0.43	0.12	-0.34
15	-1.00	0.10	0.99
16	0.08	0.11	1.14
17	0.06	0.11	0.61
18	1.23	0.15	0.53
19	0.14	0.11	-1.42
20	0.13	0.11	-0.40

ตาราง 10 (ต่อ)

ข้อ	b	SEE	t
21	0.29	0.12	0.81
22	-0.14	0.11	-0.71
23	0.27	0.12	-1.22
24	-0.23	0.11	-1.83
25	0.63	0.13	1.15
26	0.06	0.11	-0.13
27	-0.09	0.11	-0.44
28	0.46	0.12	-0.74
29	-0.17	0.11	0.12
30	0.39	0.12	0.64

ตาราง 11 ผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (b) ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความยากของข้อสอบ (SEE) และค่าความสอดคล้องกับโมเดล (t - test) ของแบบทดสอบฉบับที่ 3

ข้อ	b	SEE	t
1	-0.21	0.11	-0.36
2	1.76	0.17	-0.02
3	-0.73	0.10	-0.88
4	-0.77	0.10	1.14
5	-1.19	0.11	-0.61
6	-0.55	0.10	-2.41
7	0.53	0.12	-0.50
8	-0.04	0.11	-1.93
9	0.59	0.12	2.52
10	0.03	0.11	-0.85
11	0.30	0.12	0.20
12	-0.73	0.10	-1.94
13	-0.03	0.11	0.45
14	-0.31	0.11	1.39
15	0.08	0.11	0.08
16	0.07	0.11	1.04
17	-0.07	0.11	0.71
18	0.41	0.12	2.23
19	-0.90	0.10	-1.99
20	-1.28	0.11	-1.47

ตาราง 11 (ต่อ)

ข้อ	b	SEE	t
21	0.12	0.11	0.50
22	0.78	0.13	-1.14
23	-0.94	0.10	-1.25
24	0.04	0.11	-1.76
25	-0.08	0.11	0.84
26	0.43	0.12	0.24
27	0.13	0.11	-1.12
28	0.50	0.12	-0.53
29	1.05	0.14	-0.64
30	1.01	0.14	0.51

ตาราง 12 ผลการประมาณค่าพารามิเตอร์ความยากของข้อสอบ (b) ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความยากของข้อสอบ (SEE) และค่าความสอดคล้องกับโมเดล (t - test) ของแบบทดสอบฉบับที่ 4

ข้อ	b	SEE	t
1	0.04	0.11	0.21
2	-2.23	0.14	-0.59
3	-0.48	0.11	-0.56
4	-0.46	0.11	1.64
5	0.84	0.12	-0.37
6	-1.45	0.11	-1.47
7	0.28	0.11	-0.15
8	0.32	0.11	-1.30
9	-0.06	0.11	-1.66
10	-0.34	0.11	-0.58
11	-0.33	0.11	-1.55
12	1.03	0.13	3.32
13	0.53	0.12	-1.46
14	-0.16	0.11	1.79
15	-0.95	0.11	-1.21
16	1.25	0.14	1.10
17	-0.18	0.11	-1.05
18	0.51	0.12	0.56
19	-0.65	0.11	-2.14
20	-1.23	0.11	-2.00

ตาราง 12 (ต่อ)

ข้อ	b	SEE	t
21	-0.60	0.11	-1.48
22	0.64	0.12	-0.99
23	0.89	0.12	-0.96
24	0.32	0.11	-0.50
25	0.87	0.12	1.42
26	-0.69	0.11	-0.98
27	1.14	0.13	0.56
28	-0.21	0.11	0.08
29	0.56	0.12	-0.55
30	0.80	0.12	2.47

ภาคผนวก ค

คะแนนสอบวิชาคณิตศาสตร์ และค่าประมาณพารามิเตอร์ความสามารถของผู้สอบ
ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน
แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง
และแบบทดสอบแยกกลุ่มสองครั้ง

ตาราง 13 คะแนนสอบวิชาคณิตศาสตร์ (Y) และค่าประมาณพารามิเตอร์ความสามารถของผู้สอบ(θ) ระหว่างแบบทดสอบแยกกลุ่มที่มีความยากใกล้เคียงกัน(θ_{PRT}) แบบทดสอบแยกกลุ่มที่มีความยากกระจายในช่วงกว้าง(θ_{RRT}) และแบบทดสอบแยกกลุ่มสองครั้ง(θ_{DRT})

คนที่	Y	θ			คนที่	Y	θ		
		θ_{PRT}	θ_{RRT}	θ_{DRT}			θ_{PRT}	θ_{RRT}	θ_{DRT}
1	19	-0.20	-1.39	-0.22	21	21	0.41	0.81	-0.14
2	18	-1.10	0.02	-0.42	22	17	-1.10	-1.74	-0.70
3	27	-0.62	0.02	-0.42	23	26	-0.20	0.28	-1.03
4	27	0.00	-0.24	-0.22	24	25	-0.62	-0.78	-1.03
5	23	0.00	0.54	-0.42	25	19	0.00	0.02	-0.22
6	19	-0.20	-0.24	-0.42	26	24	0.20	-0.24	0.65
7	30	-0.41	-0.24	-1.03	27	17	-0.62	0.02	-0.70
8	30	0.85	0.81	0.13	28	21	-0.20	-0.51	-0.22
9	22	-0.41	-0.51	-0.14	29	18	-0.20	0.02	-0.42
10	33	0.62	0.54	0.13	30	28	0.85	0.28	0.36
11	32	0.20	-0.24	-0.42	31	32	0.20	0.54	0.39
12	32	0.85	1.09	-0.14	32	16	0.62	0.54	-0.22
13	33	0.85	1.09	-0.14	33	30	1.10	0.02	0.92
14	38	0.62	1.09	0.65	34	28	-0.20	0.28	-0.14
15	26	0.20	-0.78	-0.71	35	25	-0.85	-0.24	0.25
16	21	0.20	0.02	0.25	36	33	0.41	-0.51	-0.14
17	29	-0.85	-0.51	-0.70	37	31	0.41	1.09	0.39
18	25	0.41	-0.51	-0.71	38	31	0.62	0.02	-0.42
19	23	0.00	-0.24	-0.42	39	34	0.62	0.02	-0.71
20	22	-0.20	-1.39	-0.42	40	22	-0.85	-0.78	-0.96

ตาราง 13 (ต่อ)

คนท	Y	θ			คนท	Y	θ		
		θ_{PRT}	θ_{RRT}	θ_{DRT}			θ_{PRT}	θ_{RRT}	θ_{DRT}
41	27	0.00	0.54	-0.14	63	34	0.62	0.24	-0.42
42	29	-0.62	-1.74	-0.71	64	16	-0.41	0.81	-0.22
43	27	0.20	-0.24	-0.42	65	16	0.20	0.81	1.39
44	29	-0.85	0.54	-1.03	66	18	0.20	-1.74	-1.95
45	23	-0.41	-0.24	-0.42	67	10	0.20	-2.15	-1.95
46	22	0.62	-0.24	-0.42	68	23	-0.62	0.28	0.36
47	22	0.00	0.81	-0.22	69	17	0.20	0.54	0.39
48	26	-0.20	0.81	0.39	70	15	-2.18	0.54	-0.22
49	12	-1.10	-1.74	-1.95	71	14	-0.85	0.02	0.92
50	10	-0.85	-2.15	-1.95	72	20	-0.20	0.28	-0.14
51	7	-0.41	-0.78	-2.44	73	20	0.20	-0.24	0.25
52	12	-1.39	-2.67	-1.95	74	10	-1.39	-0.51	-0.14
53	17	-0.41	-1.08	-0.96	75	10	-2.18	1.09	0.39
54	17	-2.18	-0.51	-0.96	76	11	-0.85	0.02	-0.42
55	18	-1.73	0.81	-0.96	77	11	-0.62	0.02	-0.71
56	8	-1.73	-1.08	-0.70	78	11	-0.85	0.20	-0.22
57	18	-1.10	-0.39	-0.96	79	10	-1.73	-0.51	-0.96
58	11	-0.85	-0.51	-0.96	80	30	0.62	1.09	0.13
59	15	0.00	0.81	0.13	81	17	0.62	0.28	0.39
60	11	-2.90	-0.78	-0.70	82	15	0.00	-0.51	-0.96
61	12	0.20	-0.78	-1.24	83	20	0.20	-0.54	-0.14
62	5	-0.20	-0.51	-0.96	84	8	0.20	-0.24	-0.96

ตาราง 13 (ต่อ)

คนปี	Y	θ			คนปี	Y	θ		
		θ_{PRT}	θ_{RRT}	θ_{DRT}			θ_{PRT}	θ_{RRT}	θ_{DRT}
85	12	-0.62	-1.08	-1.24	107	22	-0.85	-0.51	-0.42
96	14	-0.41	0.02	-1.24	108	22	-0.41	0.28	-1.38
87	24	0.20	-0.78	-0.45	109	29	0.62	0.28	0.39
88	19	0.41	0.81	0.13	110	19	-0.41	-0.51	-0.70
89	35	-0.41	-0.51	-0.14	111	17	-0.85	0.54	-0.14
90	20	-1.39	-0.51	-0.71	112	20	0.62	0.02	-0.42
91	24	-0.20	0.02	0.65	113	24	0.41	-0.78	-0.96
92	20	0.41	0.02	1.26	114	21	-0.85	0.02	0.25
93	19	-0.41	-0.51	-0.71	115	28	-0.20	-0.24	0.39
94	18	-0.85	-0.51	-0.96	116	14	0.41	-0.78	-0.14
95	24	0.20	0.28	-1.38	117	12	-2.90	-1.08	-1.56
96	19	-0.41	-0.24	-0.71	118	18	0.00	-0.28	0.01
97	31	1.73	-0.51	-0.39	119	28	-0.20	-0.24	0.65
98	18	-0.62	-0.24	-1.38	120	19	-1.39	-0.51	-0.71
99	24	-0.20	-0.24	-0.42	121	23	0.00	-0.02	0.72
100	25	-0.20	0.02	0.01	122	19	-0.62	-0.51	0.65
101	26	-0.20	0.02	-0.14	123	16	-1.39	-1.08	0.72
102	18	-0.41	-0.51	-0.70	124	18	-0.41	-0.78	-1.38
103	17	-1.39	0.51	-1.03	125	22	-0.41	0.02	0.48
104	20	2.20	0.28	-0.42	126	31	0.00	0.02	0.65
105	26	0.62	0.02	0.13	127	17	-0.20	-0.78	-0.45
106	25	0.62	-0.24	-0.71	128	15	-0.62	-1.74	0.72

ตาราง 13 (ต่อ)

คนที่	Y	θ			คนที่	Y	θ		
		θ_{PRT}	θ_{RRT}	θ_{DRT}			θ_{PRT}	θ_{RRT}	θ_{DRT}
129	26	-0.20	-0.51	0.72	151	14	-0.85	0.28	-0.42
130	21	-0.62	-0.78	0.71	152	13	-0.85	-1.39	-1.24
131	22	0.00	-0.78	-1.38	153	16	-0.41	0.54	0.96
132	33	-0.20	0.28	0.39	154	12	-0.20	-0.51	-0.45
133	22	-0.41	-0.51	0.65	155	22	-0.21	-0.78	-0.42
134	15	0.20	-0.51	0.25	156	22	-0.04	0.02	-0.71
135	21	0.20	0.02	-0.22	157	24	0.20	0.02	-0.14
136	19	-0.20	-0.78	-0.22	158	23	-0.85	-1.39	0.48
137	13	-0.62	-1.74	-0.42	159	16	0.41	-1.08	0.48
138	20	-0.62	-0.78	-0.45	160	20	-0.62	-0.24	0.25
139	23	0.62	0.28	-0.14	161	15	-0.85	-0.24	0.01
140	15	-0.62	-0.78	-0.45	162	18	-0.41	-0.78	-0.70
141	19	-0.62	-1.39	0.01	163	19	-0.62	-1.39	-0.22
142	14	-1.39	-0.78	-0.70	164	24	1.10	0.28	0.39
143	17	-0.62	-0.78	-1.56	165	19	-0.41	0.02	-0.42
144	9	-1.10	-0.51	-1.56	166	11	-0.85	-0.51	-0.22
145	17	-0.62	-0.51	0.13	167	14	-0.41	-1.39	-0.96
146	22	-0.41	0.28	0.01	168	20	-0.41	-0.24	-0.14
147	9	-0.85	-0.24	0.01	169	21	-0.62	-0.51	0.65
148	19	0.20	0.02	0.39	170	20	-0.41	0.02	0.01
149	24	0.00	-0.51	0.48	171	12	0.41	-0.51	-1.24
150	15	-0.41	-0.78	-0.96	172	16	0.20	-1.08	-1.56

ตาราง 13 (ต่อ)

คนก	Y	θ			คนก	Y	θ		
		θ_{PRT}	θ_{RRT}	θ_{DRT}			θ_{PRT}	θ_{RRT}	θ_{DRT}
173	16	-0.85	-0.24	-0.96	195	11	-0.85	-0.78	-0.96
174	18	-0.85	0.02	0.48	196	5	-1.73	-1.74	-1.56
175	20	0.00	-0.24	-0.45	197	12	-1.39	-0.78	-1.56
176	19	0.00	-1.39	0.48	198	13	-1.10	-0.78	-0.70
177	14	-0.20	-1.08	-0.22	199	15	-0.62	-0.78	-1.24
178	18	-1.39	-0.24	-0.42	200	13	-1.73	-0.51	-0.96
179	25	0.85	-0.24	0.65	201	17	-0.85	-1.39	-0.45
180	15	-0.41	-0.78	-0.42	202	18	-0.62	0.02	-0.22
181	26	0.41	1.09	0.65	203	15	-1.39	-0.24	-0.70
182	14	-2.18	-1.39	-0.70	204	24	-0.62	-0.24	0.01
183	13	-2.18	-0.51	-0.96	205	20	-0.41	-0.24	0.48
184	11	-1.39	-1.74	-0.70	206	23	-0.41	-0.78	-1.03
185	11	-0.85	-1.39	-0.96	207	19	-0.85	-1.08	-0.71
186	11	-0.85	-0.51	-0.70	208	18	-1.10	-1.39	0.96
187	13	-0.41	-1.08	-0.96	209	18	-1.10	-1.39	-0.96
188	19	-0.85	-0.24	-0.45	210	6	-0.85	-1.08	-0.96
189	15	-1.10	-1.08	-0.70	211	9	-0.85	-2.15	-0.96
190	16	-0.41	-0.78	-1.56	212	11	0.00	-1.74	-0.70
191	23	-0.85	-1.08	0.13	213	13	-0.85	-0.78	-0.70
192	16	-0.85	-0.78	-0.96	214	15	-0.62	-0.78	-0.70
193	14	-0.85	-1.08	-0.22	215	17	-0.20	-0.24	-1.22
194	16	-0.41	-1.74	-0.70	216	15	-1.73	0.02	-1.56

ตาราง 13 (ต่อ)

คนปี	Y	θ			คนปี	Y	θ		
		θ_{PRT}	θ_{RRT}	θ_{DRT}			θ_{PRT}	θ_{RRT}	θ_{DRT}
117	14	-1.10	-0.78	-1.24	239	19	-1.10	0.02	-0.96
118	10	-0.85	-1.08	-1.24	240	15	-0.62	-0.24	-1.56
119	16	0.00	-0.24	0.01	241	11	-1.39	-0.78	-1.24
120	12	-1.10	-0.51	-1.56	242	17	-1.39	-1.39	0.25
121	13	-1.10	-0.51	-1.24	243	22	-1.10	0.02	-0.96
122	11	-0.85	-1.08	-1.24	244	12	-1.73	-1.39	-0.22
123	10	0.85	-0.51	-1.56	245	15	-1.39	-0.24	-0.70
124	12	-0.85	-0.78	-0.96	246	18	-1.10	-0.78	-0.22
125	13	-0.62	-1.08	-1.24	247	16	-1.39	-0.51	-1.56
126	12	-0.85	-0.51	-0.96	248	18	-1.10	-0.78	-0.22
127	12	-1.39	-2.15	-1.56	249	23	0.20	0.54	0.25
128	11	-1.39	-0.78	-0.70	250	25	-0.41	0.54	0.01
129	11	-0.62	-0.78	-1.56	251	30	0.00	0.54	-1.03
130	11	-1.10	-0.78	-1.95	252	21	-0.85	-1.08	-0.45
131	14	-1.10	-0.78	-1.56	253	18	-0.62	-1.74	-0.70
132	16	-2.90	-1.39	-1.24	254	18	-1.10	-1.39	-0.71
133	10	-0.62	-1.08	-0.70	255	19	-1.10	-1.08	-0.45
134	10	-1.10	-1.39	-1.24	256	22	0.00	0.02	-1.24
135	8	-0.85	-0.51	-1.24	257	21	-0.85	-1.08	0.01
136	22	0.00	-1.08	0.39	258	12	-0.62	-1.74	-0.96
137	33	-0.20	0.28	-0.14	259	23	-1.39	-1.08	0.39
138	20	-1.39	0.02	-0.70	260	15	-0.85	-1.74	-0.96

ตาราง 13 (ต่อ)

คนปี	Y	θ			คนปี	Y	θ		
		θ_{PRT}	θ_{RRT}	θ_{DRT}			θ_{PRT}	θ_{RRT}	θ_{DRT}
261	16	-0.85	-1.08	0.25	283	17	-1.10	-0.78	-0.45
262	15	-1.10	-1.08	-0.70	284	13	-1.39	-1.08	-1.24
263	21	-0.41	-0.78	-0.45	285	16	-1.10	-1.39	-0.96
264	19	-0.62	-1.39	-0.70	286	20	-0.41	-0.51	-0.14
265	7	-1.10	-2.67	-0.96	287	15	-1.73	-1.08	-0.22
266	13	-0.62	-0.24	-1.38	288	11	-1.10	-1.39	-0.70
267	19	-0.62	-0.24	0.01	289	17	-1.10	-1.08	0.01
268	16	-1.10	-0.78	0.01	290	13	-0.85	-1.08	-0.22
269	10	-1.10	-2.67	-1.56	291	18	-0.41	-1.08	-0.96
270	21	-1.10	-0.51	0.01	292	12	-0.20	-1.39	-0.96
271	12	-0.85	-2.15	-0.22	293	9	0.00	-1.39	-1.24
272	15	-0.41	-1.08	-1.03	294	7	-1.39	-1.39	-0.70
273	16	-1.39	-0.51	-0.96	295	18	-0.41	-0.51	-0.70
274	14	-1.10	-1.08	-0.45	296	11	-0.20	-2.67	-2.44
275	18	-1.10	-1.74	-0.45	297	16	-0.62	-0.51	-0.70
276	18	-0.62	-0.78	-0.45	298	10	-0.85	-0.51	-0.96
277	14	-0.85	-1.08	0.01	299	19	-0.41	-0.24	0.01
278	12	-1.10	-1.08	-0.96	300	10	0.00	-0.51	-1.56
279	19	-0.62	-0.51	-0.45	301	10	-1.39	-0.51	-0.96
280	16	-0.85	-0.51	-0.22	302	12	-0.62	-0.78	-0.96
281	19	-1.10	-1.08	0.01	303	12	0.00	-1.08	-0.96
282	12	-0.41	-2.15	-1.24	304	7	-0.85	-0.78	-1.56

ตาราง 13 (ต่อ)

คนโท	Y	θ			คนโท	Y	θ		
		θ_{PRT}	θ_{RRT}	θ_{DRT}			θ_{PRT}	θ_{RRT}	θ_{DRT}
305	10	-1.10	-0.78	-0.96	327	11	-1.10	-0.78	-0.42
306	8	-0.85	-1.39	-2.44	328	19	-0.20	-0.78	-1.56
307	8	-0.41	-0.51	-0.70	329	10	0.20	-1.74	-0.45
308	9	-0.41	-0.78	-1.95	330	16	0.41	-0.24	-0.01
309	12	-0.41	-0.24	-1.24	331	17	0.62	-0.78	-0.22
310	10	-0.41	-0.78	-0.70	332	19	0.20	-0.78	-1.24
311	15	-0.85	-0.78	-0.45	333	12	0.41	-1.08	-0.22
312	19	-0.41	-0.78	-0.22	334	7	-0.62	-2.67	-0.96
313	8	-0.85	-0.51	-1.24	335	9	0.41	-1.39	-0.70
314	12	-0.41	-0.78	-1.38	336	9	-0.4	-1.08	-1.56
315	10	-0.85	-2.67	-0.45	337	15	0.62	-0.78	-0.45
316	13	-0.20	-1.74	0.01	338	19	0.41	-1.08	-0.96
317	7	-0.62	-1.74	-1.95	339	9	-0.41	-2.15	-1.95
318	10	-0.62	-2.15	-2.44	340	16	0.41	-0.78	-0.96
319	10	-0.41	-2.15	-0.70	341	15	0.62	-0.02	-0.22
320	13	-0.41	-0.78	-1.56	342	10	-0.41	-2.67	-0.70
321	12	0.41	-1.74	-0.96	343	13	0.20	-1.08	-1.56
322	8	-0.85	-2.15	-0.45	344	13	0.0	-0.78	-0.96
323	10	-0.20	-2.15	-1.56	345	15	-0.85	-0.78	-0.71
324	10	-1.39	-1.08	-1.95	346	19	-0.62	-0.51	0.25
325	5	0.41	-1.08	-1.24	347	15	-0.62	0.28	-0.45
326	10	-1.10	-0.78	-2.44	348	15	-0.41	-1.08	0.01

ตาราง 13 (ต่อ)

คนปี	Y	θ			คนปี	Y	θ		
		θ_{PRT}	θ_{RRT}	θ_{DRT}			θ_{PRT}	θ_{RRT}	θ_{DRT}
349	20	-1.10	-0.24	0.25	371	12	-1.39	-2.15	-1.56
350	16	-1.73	-1.39	-0.70	372	8	-1.10	-2.15	-1.95
351	25	-1.10	-0.78	0.01	373	10	-1.10	-0.78	-0.71
352	15	-1.39	-1.08	-0.22	374	11	-0.85	-1.74	-1.95
353	11	-0.41	-0.78	-0.22	375	14	-0.20	-1.08	-0.22
354	18	-0.41	-0.78	0.01	376	13	-0.41	0.02	-1.24
355	27	-0.41	0.20	0.48	377	9	-0.62	-1.39	-1.95
356	26	-0.85	-0.78	0.01	378	13	-0.20	-0.51	-0.45
357	10	0.00	0.02	-0.22	379	14	-0.41	-2.67	-1.56
358	10	-0.41	-0.78	-0.70	380	11	-0.41	-1.74	-1.24
359	16	-1.10	-1.39	-0.70	381	11	-1.39	-0.24	-0.96
360	18	-1.73	-1.08	-0.96	382	13	-0.20	-2.15	-1.56
361	22	-0.41	0.02	-0.22	383	17	0.00	-1.74	-0.70
362	21	-0.62	-0.51	0.01	384	9	-1.10	-1.39	-1.56
363	17	-0.62	-0.24	-0.22	385	37	1.10	0.54	0.39
364	12	0.41	-1.08	0.01	386	33	0.41	0.28	0.39
365	13	-0.20	-1.08	-1.24	387	23	0.41	0.28	0.13
366	9	-0.41	-1.39	-0.96	388	36	0.20	0.28	1.83
367	10	-1.10	-1.39	-1.56	389	23	0.00	0.81	1.83
368	14	-0.62	-1.08	-0.45	390	33	0.20	1.09	0.92
369	15	-0.62	-1.08	-1.95	391	30	1.39	1.40	1.50
370	14	-0.62	-1.08	-1.95	392	26	0.41	1.09	1.50

ตาราง 13 (ต่อ)

คนก ^๑	Y	θ			คนก ^๑	Y	θ		
		θ_{PRT}	θ_{RRT}	θ_{DRT}			θ_{PRT}	θ_{RRT}	θ_{DRT}
393	19	0.20	0.81	0.65	407	30	1.37	1.09	1.50
394	26	1.73	0.81	0.65	408	22	1.39	1.09	1.50
395	31	0.62	1.09	1.50	409	26	2.19	1.09	1.50
396	35	2.91	1.40	1.50	410	30	0.62	0.54	0.65
397	30	1.73	0.54	0.25	411	33	1.39	1.09	1.20
398	27	2.91	1.40	1.20	412	26	0.85	1.09	0.25
399	31	1.39	1.09	1.20	413	17	0.62	1.09	0.72
400	15	-0.20	0.81	0.39	414	23	1.39	1.09	0.92
401	18	1.39	1.09	0.13	415	28	2.91	1.40	1.50
402	26	1.10	1.09	0.48	416	29	1.10	0.54	0.92
403	30	1.73	1.09	1.20	417	20	1.10	0.54	1.50
404	28	2.19	0.28	1.83	418	30	1.39	1.09	1.50
405	17	1.73	1.09	0.39	419	17	0.00	0.02	0.65
406	31	1.73	1.09	1.50					

ภาคผนวก ง

ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ
ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน
แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง
และแบบทดสอบแยกกลุ่มสองครั้ง

ตาราง 14 ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ (SEE) ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน (SEE_{PRT}) แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง (SEE_{RRT}) และแบบทดสอบแยกกลุ่มสองครั้ง (SEE_{DRT})

คนที่	SEE_{PRT}	SEE_{RRT}	SEE_{DRT}	คนที่	SEE_{PRT}	SEE_{RRT}	SEE_{DRT}
1	0.46	0.59	0.50	21	0.47	0.54	0.54
2	0.53	0.52	0.55	22	0.53	0.63	0.52
3	0.48	0.52	0.55	23	0.46	0.52	0.59
4	0.46	0.53	0.50	24	0.48	0.55	0.59
5	0.46	0.53	0.55	25	0.46	0.52	0.50
6	0.46	0.53	0.55	26	0.46	0.53	0.53
7	0.47	0.53	0.59	27	0.48	0.52	0.52
8	0.50	0.54	0.53	28	0.46	0.53	0.50
9	0.47	0.53	0.54	29	0.46	0.52	0.55
10	0.48	0.53	0.53	30	0.50	0.52	0.53
11	0.46	0.53	0.55	31	0.46	0.53	0.50
12	0.50	0.56	0.54	32	0.48	0.53	0.54
13	0.50	0.56	0.54	33	0.53	0.52	0.54
14	0.48	0.56	0.53	34	0.46	0.52	0.49
15	0.46	0.55	0.57	35	0.50	0.53	0.54
16	0.46	0.52	0.49	36	0.47	0.53	0.53
17	0.50	0.53	0.57	37	0.47	0.56	0.55
18	0.47	0.53	0.55	38	0.48	0.52	0.57
19	0.46	0.53	0.55	39	0.48	0.52	0.57
20	0.46	0.59	0.55	40	0.50	0.55	0.54

ตาราง 14 (ต่อ)

คนท	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}	คนท	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}
41	0.46	0.53	0.54	61	0.46	0.55	0.56
42	0.48	0.63	0.57	62	0.46	0.53	0.54
43	0.46	0.53	0.55	63	0.48	0.56	0.53
44	0.50	0.53	0.59	64	0.47	0.53	0.49
45	0.47	0.53	0.50	65	0.46	0.55	0.54
46	0.48	0.53	0.55	66	0.46	0.53	0.50
47	0.46	0.54	0.50	67	0.46	0.53	0.54
48	0.46	0.54	0.53	68	0.48	0.53	0.50
49	0.53	0.63	0.68	69	0.46	0.54	0.53
50	0.50	0.70	0.68	70	0.76	0.53	0.52
51	0.47	0.55	0.79	71	0.50	0.54	0.68
52	0.57	0.81	0.68	72	0.46	0.63	0.61
53	0.47	0.57	0.54	73	0.46	0.53	0.54
54	0.76	0.53	0.54	74	0.57	0.81	0.50
55	0.64	0.54	0.54	75	0.76	0.59	0.54
56	0.64	0.57	0.52	76	0.50	0.53	0.61
57	0.53	0.59	0.54	77	0.48	0.81	0.56
58	0.50	0.53	0.54	78	0.50	0.52	0.50
59	0.46	0.54	0.53	79	0.64	0.53	0.54
60	1.05	0.55	0.56	80	0.48	0.56	0.53

ตาราง 14 (ต่อ)

คนที่	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}	คนที่	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}
81	0.48	0.52	0.53	101	0.46	0.52	0.54
82	0.46	0.53	0.54	102	0.47	0.53	0.52
83	0.46	0.53	0.54	103	0.57	0.53	0.59
84	0.46	0.53	0.54	104	0.46	0.52	0.55
85	0.48	0.57	0.56	105	0.48	0.52	0.53
86	0.47	0.52	0.56	106	0.48	0.53	0.57
87	0.46	0.55	0.50	107	0.50	0.53	0.55
88	0.47	0.54	0.53	108	0.47	0.52	0.63
89	0.47	0.53	0.54	109	0.48	0.52	0.53
90	0.57	0.53	0.57	110	0.47	0.53	0.52
91	0.46	0.52	0.53	111	0.50	0.53	0.54
92	0.47	0.52	0.56	112	0.48	0.52	0.55
93	0.47	0.53	0.57	113	0.47	0.55	0.54
94	0.50	0.53	0.54	114	0.50	0.52	0.49
95	0.46	0.52	0.63	115	0.46	0.53	0.53
96	0.47	0.53	0.57	116	0.47	0.55	0.54
97	0.64	0.53	0.53	117	1.05	0.57	0.61
98	0.48	0.53	0.63	118	0.46	0.52	0.49
99	0.46	0.53	0.55	119	0.46	0.53	0.53
100	0.46	0.52	0.49	120	0.57	0.53	0.57

ตาราง 14 (ต่อ)

คนที่	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}	คนที่	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}
121	0.46	0.52	0.51	141	0.48	0.59	0.49
122	0.48	0.53	0.53	142	0.57	0.55	0.52
123	0.57	0.57	0.51	143	0.48	0.55	0.61
124	0.47	0.55	0.63	144	0.53	0.53	0.61
125	0.47	0.52	0.50	145	0.48	0.53	0.53
126	0.46	0.52	0.53	146	0.47	0.52	0.49
127	0.46	0.55	0.50	147	0.50	0.53	0.49
128	0.48	0.63	0.51	148	0.46	0.52	0.53
129	0.46	0.53	0.51	149	0.46	0.53	0.50
130	0.84	0.55	0.57	150	0.47	0.55	0.54
131	0.46	0.55	0.63	151	0.50	0.52	0.55
132	0.46	0.52	0.53	152	0.50	0.59	0.56
133	0.47	0.53	0.53	153	0.47	0.53	0.54
134	0.46	0.53	0.49	154	0.46	0.53	0.50
135	0.46	0.52	0.50	155	0.47	0.55	0.55
136	0.46	0.55	0.50	156	0.46	0.52	0.57
137	0.48	0.63	0.55	157	0.46	0.52	0.54
138	0.48	0.55	0.50	158	0.50	0.59	0.50
139	0.48	0.52	0.54	159	0.47	0.57	0.50
140	0.48	0.55	0.50	160	0.48	0.53	0.49

ตาราง 14 (ต่อ)

คนปี	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}	คนปี	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}
161	0.50	0.53	0.49	181	0.47	0.56	0.53
162	0.47	0.55	0.52	182	0.76	0.59	0.52
163	0.48	0.59	0.50	183	0.76	0.53	0.54
164	0.53	0.52	0.53	184	0.57	0.63	0.52
165	0.47	0.52	0.55	185	0.50	0.59	0.54
166	0.50	0.53	0.50	186	0.50	0.53	0.52
167	0.47	0.59	0.54	187	0.47	0.57	0.54
168	0.47	0.53	0.54	188	0.50	0.53	0.50
169	0.48	0.53	0.53	189	0.53	0.57	0.52
170	0.47	0.52	0.49	190	0.47	0.55	0.61
171	0.47	0.53	0.56	191	0.50	0.57	0.53
172	0.46	0.57	0.61	192	0.50	0.55	0.54
173	0.50	0.53	0.54	193	0.50	0.57	0.50
174	0.50	0.52	0.50	194	0.47	0.63	0.52
175	0.46	0.53	0.50	195	0.50	0.55	0.54
176	0.46	0.59	0.50	196	0.64	0.63	0.61
177	0.46	0.57	0.50	197	0.57	0.55	0.61
178	0.57	0.53	0.55	198	0.53	0.55	0.52
179	0.50	0.53	0.53	199	0.48	0.55	0.56
180	0.47	0.55	0.55	200	0.64	0.53	0.54

ตาราง 14 (ต่อ)

คนก	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}	คนก	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}
201	0.50	0.59	0.50	221	0.53	0.53	0.56
202	0.48	0.52	0.50	222	0.50	0.57	0.56
203	0.57	0.53	0.52	223	0.50	0.53	0.61
204	0.48	0.53	0.49	224	0.50	0.55	0.54
205	0.47	0.53	0.50	225	0.48	0.57	0.56
206	0.47	0.55	0.59	226	0.50	0.53	0.54
207	0.50	0.57	0.57	227	0.57	0.70	0.61
208	0.53	0.59	0.54	228	0.57	0.55	0.52
209	0.53	0.59	0.54	229	0.48	0.55	0.61
210	0.50	0.57	0.44	230	0.53	0.55	0.68
211	0.50	0.70	0.54	231	0.53	0.55	0.61
212	0.46	0.63	0.52	232	1.05	0.59	0.56
213	0.50	0.55	0.52	233	0.48	0.57	0.52
214	0.48	0.55	0.52	234	0.53	0.59	0.56
215	0.46	0.53	0.50	235	0.50	0.53	0.56
216	0.64	0.52	0.61	236	0.46	0.57	0.53
217	0.53	0.55	0.56	237	0.46	0.52	0.54
218	0.50	0.57	0.56	238	0.57	0.52	0.52
219	0.46	0.53	0.49	239	0.53	0.52	0.54
220	0.53	0.53	0.61	240	0.48	0.53	0.61

ตาราง 14 (ต่อ)

คนที่	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}	คนที่	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}
241	0.57	0.55	0.56	261	0.50	0.57	0.49
242	0.57	0.59	0.49	262	0.53	0.57	0.52
243	0.53	0.52	0.54	263	0.48	0.55	0.50
244	0.64	0.59	0.50	264	0.48	0.59	0.52
245	0.57	0.53	0.52	265	0.53	0.81	0.54
246	0.53	0.55	0.50	266	0.48	0.53	0.63
247	0.57	0.53	0.61	267	0.48	0.53	0.49
248	0.53	0.55	0.50	268	0.53	0.55	0.49
249	0.46	0.53	0.49	269	0.53	0.81	0.61
250	0.47	0.53	0.49	270	0.53	0.53	0.49
251	0.46	0.53	0.59	271	0.50	0.70	0.50
252	0.50	0.57	0.50	272	0.47	0.57	0.59
253	0.48	0.63	0.52	273	0.57	0.53	0.54
254	0.53	0.59	0.57	274	0.53	0.57	0.50
255	0.53	0.57	0.50	275	0.53	0.63	0.50
256	0.46	0.52	0.56	276	0.48	0.55	0.50
257	0.50	0.57	0.49	277	0.50	0.57	0.49
258	0.48	0.63	0.54	278	0.53	0.57	0.54
259	0.57	0.57	0.53	279	0.48	0.53	0.50
260	0.50	0.63	0.54	280	0.50	0.53	0.50

ตาราง 14 (ต่อ)

คนที่	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}	คนที่	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}
281	0.53	0.57	0.49	301	0.57	0.53	0.54
282	0.47	0.70	0.56	302	0.48	0.55	0.54
283	0.53	0.55	0.50	303	0.46	0.57	0.54
284	0.57	0.57	0.56	304	0.50	0.55	0.61
285	0.53	0.59	0.54	305	0.53	0.55	0.54
286	0.47	0.53	0.54	306	0.50	0.59	0.79
287	0.64	0.57	0.50	307	0.47	0.53	0.52
288	0.53	0.59	0.52	308	0.47	0.55	0.68
289	0.53	0.57	0.49	309	0.47	0.53	0.56
290	0.50	0.57	0.50	310	0.47	0.55	0.52
291	0.47	0.57	0.54	311	0.50	0.55	0.50
292	0.46	0.59	0.54	312	0.47	0.55	0.50
293	0.46	0.59	0.56	313	0.50	0.53	0.56
294	0.57	0.59	0.52	314	0.47	0.55	0.63
295	0.47	0.53	0.52	315	0.50	0.81	0.50
296	0.46	0.81	0.79	316	0.46	0.63	0.49
297	0.48	0.53	0.52	317	0.48	0.63	0.68
298	0.50	0.53	0.54	318	0.48	0.70	0.79
299	0.47	0.53	0.49	319	0.47	0.70	0.52
300	0.46	0.53	0.61	320	0.47	0.55	0.61

ตาราง 14 (ต่อ)

คนที่	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}	คนที่	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}
321	0.47	0.63	0.54	341	0.48	0.52	0.50
322	0.50	0.70	0.50	342	0.47	0.81	0.52
323	0.46	0.70	0.61	343	0.46	0.57	0.61
324	0.57	0.57	0.68	344	0.46	0.55	0.54
325	0.47	0.57	0.56	345	0.50	0.55	0.57
326	0.53	0.55	0.79	346	0.48	0.53	0.49
327	0.47	0.55	0.55	347	0.48	0.52	0.50
328	0.46	0.55	0.61	348	0.47	0.57	0.49
329	0.46	0.63	0.50	349	0.53	0.53	0.49
330	0.47	0.53	0.49	350	0.64	0.59	0.52
331	0.48	0.55	0.50	351	0.53	0.55	0.49
332	0.46	0.55	0.56	352	0.57	0.57	0.50
333	0.47	0.57	0.50	353	0.47	0.55	0.50
334	0.48	0.81	0.54	354	0.47	0.55	0.49
335	0.47	0.59	0.52	355	0.47	0.52	0.50
336	0.47	0.57	0.61	356	0.50	0.55	0.49
337	0.48	0.55	0.50	357	0.46	0.52	0.50
338	0.47	0.57	0.54	358	0.47	0.55	0.52
339	0.47	0.70	0.68	359	0.53	0.59	0.52
340	0.47	0.55	0.54	360	0.64	0.57	0.54

ตาราง 14 (ต่อ)

คนที	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}	คนที	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}
361	0.47	0.52	0.50	381	0.47	0.53	0.54
362	0.48	0.53	0.49	382	0.46	0.70	0.61
363	0.48	0.53	0.50	383	0.46	0.63	0.52
364	0.47	0.57	0.49	384	0.53	0.59	0.61
365	0.46	0.57	0.56	385	0.53	0.53	0.53
366	0.47	0.59	0.54	386	0.47	0.52	0.53
367	0.53	0.59	0.61	387	0.47	0.52	0.53
368	0.48	0.57	0.50	388	0.46	0.52	0.62
369	0.48	0.57	0.68	389	0.46	0.54	0.62
370	0.48	0.57	0.68	390	0.46	0.56	0.54
371	0.57	0.70	0.61	391	0.57	0.58	0.58
372	0.53	0.70	0.68	392	0.47	0.56	0.58
373	0.53	0.55	0.57	393	0.46	0.54	0.53
374	0.50	0.63	0.68	394	0.64	0.54	0.53
375	0.46	0.57	0.50	395	0.48	0.56	0.58
376	0.47	0.52	0.56	396	1.04	0.58	0.58
377	0.48	0.59	0.68	397	0.46	0.53	0.49
378	0.46	0.53	0.50	398	1.04	0.58	0.55
379	0.47	0.81	0.61	399	0.57	0.56	0.55
380	0.47	0.63	0.56	400	0.64	0.54	0.53

ตาราง 14 (ต่อ)

คนที่	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}	คนที่	SEE _{PRT}	SEE _{RRT}	SEE _{DRT}
401	0.57	0.56	0.53	411	0.57	0.56	0.55
402	0.53	0.56	0.50	412	0.50	0.56	0.49
403	0.64	0.56	0.55	413	0.48	0.56	0.51
404	0.76	0.52	0.62	414	0.57	0.56	0.54
405	0.64	0.56	0.53	415	1.04	0.58	0.58
406	0.64	0.56	0.58	416	0.53	0.53	0.54
407	0.64	0.56	0.58	417	0.53	0.53	0.58
408	0.57	0.56	0.58	418	0.57	0.56	0.58
409	0.76	0.56	0.53	419	0.46	0.52	0.53
410	0.48	0.53	0.58				

ภาคผนวก จ

ตารางวิเคราะห์เนื้อหา พฤติกรรม และจุดประสงค์เชิงพฤติกรรม
วิชาคณิตศาสตร์ (ค 204) ชั้นมัธยมศึกษาปีที่ 2
เรื่อง สมการและอสมการ

ตาราง 15 ตารางการวิเคราะห์เนื้อหาและพฤติกรรม วิชาคณิตศาสตร์ (ค 204)
 ชั้นมัธยมศึกษา ปีที่ 2 เรื่อง สมการและอสมการ ตามหลักสูตรมัธยมศึกษาตอนต้น
 พุทธศักราช 2521 (ฉบับปรับปรุง พ.ศ. 2533)

เนื้อหา	พฤติกรรม	ความรู้- ความจำ	ความเข้าใจ	การนำไปใช้	การวิเคราะห์
สมการ และอสมการ					
1. ประโยคภาษาและประโยคสัญลักษณ์		✓	✓		
2. คำตอบของสมการ และคำตอบของอสมการ		✓	✓		
3. กราฟแสดงคำตอบ			✓		
4. การแก้สมการ					
4.1 สมบัติการเท่ากัน (สมบัติสมมาตร สมบัติการถ่ายทอด สมบัติการบวก สมบัติการคูณ)		✓	✓		
4.2 การแก้สมการโดยใช้ สมบัติสมมาตร สมบัติ การถ่ายทอด สมบัติ การบวก และสมบัติ การคูณ				✓	
5. โจทย์สมการ			✓	✓	✓

จุดประสงค์เชิงพฤติกรรม วิชาคณิตศาสตร์ (ค 204) ชั้นมัธยมศึกษาปีที่ 2
เรื่อง สมการและอสมการ ตามหลักสูตรมัธยมศึกษาตอนต้น พุทธศักราช 2521
(ฉบับปรับปรุง พ.ศ. 2533)

1. เนื้อหาย่อย ประโยคภาษาและประโยคสัญลักษณ์

พฤติกรรมความรู้ความจำ

- เมื่อกำหนดประโยคมาให้ นักเรียนสามารถบอกได้ว่า ประโยคใดเป็นสมการ
ประโยคใดเป็นอสมการ

พฤติกรรมความเข้าใจ

- เมื่อกำหนดประโยคภาษาหรือประโยคสัญลักษณ์มาให้ นักเรียนสามารถเปลี่ยน
ประโยคภาษาเป็นประโยคสัญลักษณ์ และเปลี่ยนประโยคสัญลักษณ์เป็นประโยค
ภาษาได้

2. เนื้อหาย่อย คำตอบของสมการ และคำตอบของอสมการ

พฤติกรรมความรู้ความจำ

- นักเรียนสามารถอธิบายความหมายของคำว่า "คำตอบของสมการ" หรือ
"คำตอบของอสมการ" ได้

พฤติกรรมความเข้าใจ

- เมื่อกำหนดจำนวนมาให้ นักเรียนสามารถบอกได้ว่าจำนวนนั้นเป็นคำตอบของ
สมการหรืออสมการหรือไม่
- เมื่อกำหนดสมการหรืออสมการมาให้ นักเรียนสามารถหาคำตอบของสมการ
หรือ อสมการได้

3. เนื้อหาย่อย กราฟแสดงคำตอบ

พฤติกรรมความเข้าใจ

- เมื่อกำหนดสมการหรืออสมการมาให้ นักเรียนสามารถเขียนกราฟแสดงคำตอบได้
- เมื่อกำหนดกราฟแสดงคำตอบของสมการ หรืออสมการมาให้ นักเรียนสามารถ
อ่านคำตอบของสมการ หรืออสมการจากกราฟได้

4. เนื่อหาย่อยการแก้สมการ

4.1 สมบัติการเท่ากัน (สมบัติสมมาตร สมบัติการถ่ายทอด สมบัติการบวก สมบัติการคูณ)

พฤติกรรมความรู้ความจำ

- นักเรียนสามารถอธิบาย ความหมายของสมบัติสมมาตร สมบัติการถ่ายทอด สมบัติการบวก และสมบัติการคูณได้

พฤติกรรมความเข้าใจ

- เมื่อกำหนดประโยคสัญลักษณ์ที่ใช้สมบัติการเท่ากันมาให้ นักเรียนสามารถบอกได้ว่าใช้สมบัติการเท่ากันข้อใด (สมมาตร การถ่ายทอด การบวก การคูณ)
- เมื่อกำหนดสมบัติการเท่ากันมาให้ (สมมาตร การถ่ายทอด การบวก การคูณ) นักเรียนสามารถหาประโยคสัญลักษณ์ที่ใช้สมบัติการเท่ากันนั้น ๆ ได้

4.2 การแก้สมการโดยใช้สมบัติสมมาตร สมบัติการถ่ายทอด สมบัติการบวก สมบัติการคูณ

พฤติกรรมการนำไปใช้

- เมื่อกำหนดสมการมาให้ นักเรียนสามารถใช้สมบัติสมมาตร สมบัติการถ่ายทอด สมบัติการบวก และสมบัติการคูณในการแก้สมการได้

5. เนื่อหาย่อยโจทย์สมการ

พฤติกรรมความเข้าใจ

- เมื่อกำหนดโจทย์ปัญหาเกี่ยวกับสมการมาให้ นักเรียนสามารถเขียนประโยคสัญลักษณ์ได้

พฤติกรรมการนำไปใช้

- เมื่อกำหนดโจทย์ปัญหาที่เกี่ยวข้องกับชีวิตประจำวันมาให้ นักเรียนสามารถแก้โจทย์ปัญหาโดยใช้ความรู้ เรื่อง สมการได้

พฤติกรรมวิเคราะห์

- เมื่อกำหนดโจทย์ปัญหาสมการมาให้ นักเรียนสามารถบอกได้ว่าโจทย์ปัญหานี้สมบรูณ์หรือไม่ และสามารถหาสิ่งที่ขาดหายไปได้

รายชื่อผู้เชี่ยวชาญ

1. อาจารย์อรินทร์ น่วมถนอม โรงเรียนบางกะปิ เขตบางกะปิ กรุงเทพมหานคร
2. อาจารย์ทองอยู่ สาระ โรงเรียนศรีอยุธยา เขตปทุมธานี กรุงเทพมหานคร
3. อาจารย์สุนันท์ จิมวัย โรงเรียนอัสสัมชัญ ลำโพง อำเภอเมือง สมุทรปราการ
4. อาจารย์กัญญา โตดำมา โรงเรียนอัสสัมชัญ ลำโพง อำเภอเมือง สมุทรปราการ
5. อาจารย์เตือนใจ อุ่นอินต์ โรงเรียนโคกสำโรงวิทยา อำเภอโคกสำโรง ลพบุรี

ประวัติย่อของผู้วิจัย

ชื่อ นางสาววิษญา ชื่อสกุล เกตุสกุล
 เกิดวันที่ 14 เดือนธันวาคม พุทธศักราช 2508
 สถานที่เกิด อำเภอเมือง จังหวัดลพบุรี
 สถานที่อยู่ปัจจุบัน บ้านเลขที่ 50/4 ซอยเทพสตรี 2 ตำบลทะเลชุบศร
 อำเภอเมือง จังหวัดลพบุรี โทร. 036 - 422029
 ตำแหน่งหน้าที่การงาน อาจารย์ 1 ระดับ 4
 สถานที่ทำงาน โรงเรียนพระนารายณ์ ตำบลท่าศาลา
 อำเภอเมือง จังหวัดลพบุรี โทร. 036 - 413111

ประวัติการศึกษา

พ.ศ. 2527	มัธยมศึกษาตอนปลาย (สายวิทย์-คณิต) จากโรงเรียนสตรีอ่างทอง
พ.ศ. 2531	ค.บ. (คณิตศาสตร์) จากวิทยาลัยครูพระนครศรีอยุธยา
พ.ศ. 2539	กศ.ม. (การวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

การเปรียบเทียบความเที่ยงตรงและความคลาดเคลื่อนมาตรฐานในการประมาณค่า
ความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีรูปแบบต่างกัน

บทคัดย่อ
ของ
พินิจา เกตุสกุล

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา

พฤษภาคม 2539

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อเปรียบเทียบความเที่ยงตรง และความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างและแบบทดสอบแยกกลุ่มสองครั้ง โดยศึกษากับกลุ่มตัวอย่างนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดลพบุรี จำนวน 419 คน ซึ่งได้มาจากการสุ่มแบบแบ่งชั้น เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาคณิตศาสตร์ที่ผู้วิจัยสร้างขึ้นเอง ซึ่งมี 3 รูปแบบ คือ แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน แบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง และแบบทดสอบแยกกลุ่มสองครั้ง การประมาณค่าความยากของข้อสอบ และค่าความสามารถของผู้สอบใช้โปรแกรมไบคาล ตามทฤษฎีการตอบข้อสอบ แบบราส์ช โมเดล

ผลการวิจัยสรุปได้ดังนี้

1. ค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันกับแบบทดสอบแยกกลุ่มสองครั้ง และแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันกับแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 ตามลำดับ ส่วนค่าความเที่ยงตรงในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างกับแบบทดสอบแยกกลุ่มสองครั้ง มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

2. ค่าความคลาดเคลื่อนมาตรฐานในการประมาณค่าความสามารถของผู้สอบ ระหว่างแบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกัน กับแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้าง แบบทดสอบแยกกลุ่มที่มีค่าความยากใกล้เคียงกันกับแบบทดสอบแยกกลุ่มสองครั้ง และแบบทดสอบแยกกลุ่มที่มีค่าความยากกระจายในช่วงกว้างกับแบบทดสอบแยกกลุ่มสองครั้ง มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกค่า

A COMPARISON OF VALIDITY AND STANDARD ERROR OF ABILITY
ESTIMATION BETWEEN DIFFERENT FORMS OF ROUTING TEST

AN ABSTRACT

BY

PHITCHAYA GETSAKUL

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

May 1996

The purpose of this research was to compare the validity and standard error of ability estimation among Peak Routing Test, Rectangular Routing Test and Double Routing Test. A sample of 419 Mathayomsuksa - 2 students of the Academic Year 1995 in Lopburi was selected by the stratified random sampling technique. The research instruments used were the mathematics achievement tests, constructed by the researcher, namely Peak Routing Test, Rectangular Routing Test and Double Routing Test. The Rasch Model in Item Response Theory by the BICAL Computer Program was used to estimate item difficulty and examinee's ability.

The results of the research were summarized as follows.

1. The validity of ability estimation between the Peak Routing Test and Double Routing Test, and Peak Routing Test and Rectangular Routing Test had statistically significant difference at the .01 level and .05 level respectively, while the validity of ability estimation between the Rectangular Routing Test and Double Routing Test had no statistically significant difference.

2. The standard error of ability estimation between the Peak Routing Test and Rectangular Routing Test, Peak Routing Test and Double Routing Test, Rectangular Routing Test and Double Routing Test had statistically significant difference at the .01 level.