

241.8.016

3 547

1.8

การสร้างแบบทดสอบและตารางเทียบระดับคะแนนตามแนวดิ่งในวิชาคณิตศาสตร์

มหาวิทยาลัยศรีนครินทรวิโรฒ
มหาวิทยาลัยศรีนครินทรวิโรฒ

ปริญญานิพนธ์

ของ

อาภรณ์ กาญจนกิจโสภณ

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

เพื่อเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาการศึกษามหาบัณฑิต

เมษายน 2531

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

177228

17 เม.ย. 2533

คณะกรรมการที่ปรึกษาประจำตัวนิสิตและคณะกรรมการสอบ ได้พิจารณาปริญญานิพนธ์ฉบับนี้แล้ว เห็นสมควรรับ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาการศึกษามหาบัณฑิตของมหาวิทยาลัยศรีนครินทรวิโรฒได้

คณะกรรมการที่ปรึกษา

.....ประธาน
.....กรรมการ
.....กรรมการ

คณะกรรมการสอบ

.....ประธาน
.....กรรมการ
.....กรรมการ

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้ สำนวจลุล่วงได้ด้วยความช่วยเหลือและแนะนำให้ข้อคิดเห็นและแก้ไขข้อบกพร่องต่าง ๆ จาก ดร.สงบ ลักษณะ และ ดร.วัน สังข์สะอาด ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอกราบขอบพระคุณ รองศาสตราจารย์ ดร.ส.วาสนา ประवालพฤษ์ ที่กรุณาให้ความรู้และแนะนำในการทำปริญญานิพนธ์นี้

ขอขอบพระคุณ ดร.สงบ ลักษณะ ที่ได้ให้ความรู้ คำแนะนำ และให้ใช้เครื่องคอมพิวเตอร์ โปรแกรม BICAL และคุณชาญศักดิ์ เรืองไตรรัตน์ พร้อมทั้งเจ้าหน้าที่ของฝ่ายสารสนเทศ กระทรวงศึกษาธิการ ที่ได้ช่วยเหลือในการวิเคราะห์ข้อมูล

ขอขอบพระคุณ ดร.สุธรรม จันทน์หอม และคุณอรรถพร คุณพันธ์ คุณนิรชร คุณพันธ์ ที่ได้ช่วยในการจัดส่งเอกสารมาให้ค้นคว้า

ขอขอบพระคุณคณะครูของโรงเรียนที่เป็นกลุ่มตัวอย่างและขอขอบใจนักเรียนทุกคนที่ให้ความร่วมมือในการเก็บข้อมูลเป็นอย่างดี

ขอขอบคุณเพื่อนวัดผลรุ่น 21 และน้องวัดผลรุ่น 22 ทุก ๆ คนที่เป็นกำลังใจ และช่วยเหลือในการทำปริญญานิพนธ์นี้

สุดท้ายนี้ ผู้วิจัยขอโน้มราลถึงพระคุณของบิดา-มารดา ครู-อาจารย์ พี่ น้อง และเพื่อน ๆ ที่เป็นกำลังใจในการทำปริญญานิพนธ์นี้

อาภรณ์ กาญจนกิจโสภณ

สารบัญ

บทที่	หน้า
1 บทนำ.....	1
ภูมิหลัง.....	1
จุดมุ่งหมายของการศึกษาค้นคว้า.....	4
ความสำคัญของการศึกษาค้นคว้า.....	5
ขอบเขตของการศึกษาค้นคว้า.....	5
นิยามศัพท์เฉพาะ.....	6
2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	8
ความหมายของการ เทียบระดับคะแนน.....	8
โมเดลการ เทียบระดับคะแนน โดยใช้ทฤษฎีการตอบข้อคำถาม.....	9
มโนทัศน์เกี่ยวกับราชคโ โมเดล.....	14
รูปแบบการ เชื่อมต่อของแบบทดสอบ.....	22
งานวิจัยที่เกี่ยวข้อง.....	28
งานวิจัยในต่างประเทศ.....	28
งานวิจัยในประเทศ.....	37
3 วิธีดำเนินการศึกษาค้นคว้า.....	41
ประชากร.....	41
กลุ่มตัวอย่าง.....	41
เครื่องมือที่ใช้ในการศึกษา.....	43
วิธีดำเนินการสร้างแบบทดสอบ เทียบคะแนนในแนวตั้ง.....	44
การสร้างแบบทดสอบ.....	44
การ เทียบระดับคะแนน.....	45

บทที่	หน้า
วิธีดำเนินการเก็บรวบรวมข้อมูล.....	46
การวิเคราะห์ข้อมูลและสถิติที่ใช้.....	47
4 ผลการวิเคราะห์ข้อมูล.....	49
สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล.....	49
การเสนอผลการวิเคราะห์ข้อมูล.....	50
ตอนที่ 1 เกี่ยวกับแบบทดสอบ.....	50
ตอนที่ 2 เกี่ยวกับความสามารถ.....	53
5 สรุปผล อภิปรายผล และข้อเสนอแนะ.....	66
จุดมุ่งหมายของการศึกษาค้นคว้า.....	66
กลุ่มตัวอย่าง.....	66
เครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	66
วิธีดำเนินการเก็บรวบรวมข้อมูล.....	67
สรุปผลการวิเคราะห์ข้อมูล.....	67
อภิปรายผล.....	69
ข้อเสนอแนะ.....	71
บรรณานุกรม.....	72
ภาคผนวก.....	75

บัญชีตาราง

ตาราง		หน้า
1	แสดงค่าความแตกต่างระหว่างความสามารถของคนกับความยากของข้อสอบ ในการทำนายผล.....	22
2	แสดงการนำรายค์โมเดลมาใช้ในการเทียบระดับคะแนน.....	25
3	รายชื่อโรงเรียนและจำนวนนักเรียนในกลุ่มตัวอย่างที่ใช้ในการทดลอง.....	42
4	ค่าสถิติพื้นฐานของแบบทดสอบ 3 ฉบับ.....	51
5	เปรียบเทียบค่าความยากข้อสอบรวม.....	52
6	ค่าสถิติพื้นฐานความยากของข้อสอบรวม.....	53
7	การเทียบระดับความสามารถของนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3.....	60
8	การเทียบระดับคะแนน.....	64

บัญชีภาพประกอบ

ภาพประกอบ		หน้า
1	แสดงการเทียบระดับคะแนนโดยใช้คะแนนจริงตามทฤษฎีการตอบข้อคำถาม.....	11
2	แสดงการเทียบระดับคะแนนโดยใช้คะแนนจริงจากแบบทดสอบร่วม.....	12
3	แสดงการเทียบระดับคะแนน 2 วิธี.....	23
4	การเชื่อมต่อของแบบทดสอบ 2 ฉบับ.....	24
5	แสดงการเชื่อมแบบ 2 ลูกโซ่.....	25
6	แสดงจุดเชื่อมต่อของการเชื่อมแบบ 2 ลูกโซ่.....	25
7	ตารางการเชื่อมของแบบทดสอบ 10 ฉบับ 7 ระดับ.....	26
8	ตาราง การเชื่อมต่อ โดยแบบทดสอบภายนอก.....	27
9	การเชื่อมรอยต่อของแบบทดสอบ.....	43
10	การเปรียบเทียบระหว่างคะแนนดิบกับค่าความสามารถและโค้งลักษณะ ข้อสอบของแบบทดสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1.....	54
11	การเปรียบเทียบระหว่างคะแนนดิบกับค่าความสามารถและโค้ง ลักษณะข้อสอบของแบบทดสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2.....	58
12	การเปรียบเทียบระหว่างคะแนนดิบกับค่าความสามารถและโค้งลักษณะ ข้อสอบของแบบทดสอบสำหรับชั้นมัธยมศึกษาปีที่ 3.....	58

บทที่ 1

บทนำ

ภูมิหลัง

การวัดผลการศึกษาจะเป็นเครื่องมืออันหนึ่งที่จะช่วยพัฒนาคุณภาพของการศึกษาในระดับต่าง ๆ เพราะผลจากการวัดจะเป็นพื้นฐานในการตัดสินใจของครูและนักการศึกษา เพื่อใช้ปรับปรุงวิธีสอน, การแนะแนว, การประเมินผลหลักสูตร, แบบเรียน การใช้อุปกรณ์การสอน ตลอดจนการจัดระบบบริหารทั่วไปของโรงเรียน และนอกจากนี้ยังช่วยปรับปรุงการเรียนของนักเรียนให้ดียิ่งขึ้น (อนันต์ ศรีโสภ 2520 : 1) โดยทำให้นักเรียนทราบสภาพการเรียนของตนเองว่า เก่ง-อ่อน บทร้องในเนื้อหาตอนใด มีความก้าวหน้ามาก-น้อยเพียงไร ซึ่งการวัดให้มีคุณค่าดังกล่าวนี้ต้องอาศัยเครื่องมือที่ดี นั่นคือ "แบบทดสอบ" (Test) นับว่าเป็นเครื่องมือที่ใช้กันแพร่หลายมากที่สุด ชวาล แพร์ตกุล (ชวาล แพร์ตกุล 2518 : 88) กล่าวว่า "การทดสอบเป็นวิธีวัดผลที่มีความสำคัญยิ่งและใช้มากที่สุดในการศึกษาจนนับได้ว่าเป็นหลักของการวัดผลในโรงเรียนก็ได้ การทดสอบที่ดีจะช่วยให้ครูทราบสภาพของนักเรียนและของครูเองได้ง่ายกว่าแบบอื่น ๆ มาก" เช่นเดียวกับ เดโช สนวนานท์ (เดโช สนวนานท์ 2503 : 76) และอนันต์ ศรีโสภ (อนันต์ ศรีโสภ 2520 : 2) ที่กล่าวว่า "เครื่องมือหรือวิธีการที่ดีที่สุดคือ การทดสอบหรือข้อสอบ"

ในปัจจุบันความต้องการที่สำคัญมากประการหนึ่งเกี่ยวกับการสร้างแบบทดสอบก็คือ การมีแบบทดสอบที่วัดคุณลักษณะเดียวกันหลาย ๆ ฉบับ โดยที่สามารถนำผลการสอบแต่ละฉบับมาเปรียบเทียบและแปลความหมายที่เป็นประโยชน์ในด้านการเรียน-การสอนได้ เช่น นักเรียนมีความสามารถเพิ่มขึ้นหรือไม่ คนใดควรไปเริ่มเรียนพื้นฐานมาก่อน คนใดควรได้รับการส่งเสริมให้ก้าวหน้าขึ้น เป็นต้น ความต้องการนี้นำไปสู่การสร้างแบบทดสอบที่เรียกว่า "แบบทดสอบคู่ขนาน" แต่ในการสร้างแบบทดสอบคู่ขนาน เป็นเรื่องที่ไม่สามารถสร้างได้อย่างแท้จริง จึงทำให้ไม่สามารถนำคะแนนที่ได้จากแบบทดสอบที่ว่าเป็นคู่ขนานนั้นมาเปรียบเทียบกันได้โดยตรง ปัญหานี้ก่อให้เกิดการเทียบระดับคะแนนจากแบบทดสอบที่วัดคุณลักษณะเดียวกัน แต่ต่างฉบับกัน นำมาเปรียบเทียบกันอย่างมีความหมายขึ้น วิธีดังกล่าวคือ

"การเทียบระดับคะแนน" (Equating) ซึ่งสามารถนำไปประยุกต์ใช้กับสถานการณ์ต่าง ๆ ที่เป็นประโยชน์ต่อการศึกษา ซึ่ง สบง ลักษณะ (สบง ลักษณะ 2525 : 23) ได้กล่าวถึงสถานการณ์ในการเทียบระดับคะแนนไว้ดังนี้

1. บุคคล 1 กลุ่มได้รับการทดสอบด้วยแบบทดสอบหนึ่งฉบับหรือมากกว่าในเวลาใกล้เคียงกัน แล้วต้องการจะเทียบระดับคะแนนระหว่างแบบทดสอบฉบับต่าง ๆ เหล่านั้น
2. แบบทดสอบฉบับเดียวกันแต่นำไปใช้ทดสอบบุคคลหลายกลุ่มที่อาจมีความสามารถแตกต่างกัน แล้วต้องการ เปรียบ เทียบบุคคลทั้งหมดทุกกลุ่มบนพื้นฐานของคะแนนจริงหรือความสามารถแท้
3. แบบทดสอบ 1 ฉบับแต่จัดแบ่งกลุ่มของข้อสอบใน 1 ฉบับ เป็นหลายระดับเรียงตามลำดับความง่ายไปหายากโดยมีข้อสอบร่วม (Anchor items) ระหว่างกลุ่มที่ติด ๆ กันเป็นบางข้อ ต้องการนำแบบทดสอบนี้ไปสอบกับบุคคลต่างระดับความสามารถกัน ผลคือคนต่างกลุ่มกันแต่ทำข้อสอบร่วมกัน และอีกหลายข้อที่คนต่างกลุ่มต่างทำอิสระแก่กันแล้วต้องการเทียบความสามารถหรือระดับคะแนนของผู้ตอบทุกกลุ่มในคราวเดียวกัน ซึ่งกรณีนี้เป็นกรณีสำคัญของการสร้างข้อสอบมาตรฐานชนิดหลายระดับ (Multi level) ไว้ในฉบับเดียวกัน นอกจากนี้ยังมีสถานการณ์ที่จำเป็นต้องใช้แบบทดสอบต่างฉบับกัน (วัดสิ่งเดียวกัน แต่อาจมี 3-4 ฉบับ) สอบกับเด็กกลุ่มใหญ่ที่แต่ละกลุ่มย่อยอาจใช้แบบทดสอบไม่เหมือนกัน แม้แต่เด็กที่นั่งใกล้กันก็อาจใช้แบบทดสอบคนละฉบับ เด็กคนละห้อง สอบคนละเวลา หรือเด็กมาสอบไม่พร้อมกัน ทั้งนี้ป้องกันการลอกคำตอบกัน และป้องกันการรู้ข้อสอบแล้วไปบอกเพื่อนที่ยังไม่สอบ จึงใช้แบบทดสอบหลายฉบับที่เทียบระดับคะแนนเข้าหากันเอาไว้ ทำให้เด็กทั้งหมด แม้จะสอบแบบทดสอบต่างกันก็อาจนำคะแนนมาเทียบในตารางเดียวกันได้ จากสถานการณ์ในการเทียบระดับคะแนนทำให้แนวคิดในเรื่องนี้แบ่งเป็น 2 ชนิดคือ การเทียบระดับคะแนนตามแนวนอน (Horizontal equation) เป็นกรณีที่มีข้อสอบ 2 ฉบับ ก และ ข มีความยากปานกลาง นำไปทดลองกับบุคคลประชากรเดียวกันแล้วต้องการว่า คะแนนของแบบทดสอบชุด ก จึงจะเทียบเท่ากับคะแนนของแบบทดสอบชุด ข กับการเทียบระดับคะแนนตามแนวตั้ง (Vertical equating) เป็นกรณีที่มีแบบทดสอบ 2 ชุด หรือมากกว่าซึ่งมีความยากต่างกันนำไปทดสอบกับประชากรต่างระดับความสามารถกัน เช่นกรณีแบบทดสอบมาตรฐานชนิด Multi-level

วิธีการที่ใช้ปรับคะแนนที่ได้จากแบบทดสอบแต่ละฉบับให้อยู่ในมาตราเดียวกัน และเทียบกันได้ นิยมใช้กันมากคือ (Petersen, Marco and Stewart. 1982 : 73)

1. การเทียบจากเปอร์เซนไทล์ที่เท่ากัน (Equi percentile equation) เป็นการแปลงคะแนนโดยใช้นิยามที่ว่า คะแนนสองตัวใด ๆ จะสามารถเทียบเท่ากันก็ต่อเมื่อคะแนนสองตัวนี้ต่างก็อยู่ในตำแหน่ง เปอร์เซนไทล์ที่เท่ากันของกลุ่มที่กำหนดให้

2. การเทียบเชิงเส้นตรง (Linear equating) เป็นวิธีการแปลงคะแนนผ่านคะแนนมาตรฐาน (Z-score) ซึ่งมีข้อตกลงว่า แบบทดสอบ 2 ฉบับนั้นจะต้องมีการกระจายของคะแนนเหมือนกัน

3. การเทียบโดยทฤษฎีการตอบข้อคำถาม เป็นวิธีการเทียบระดับคะแนนโดยใช้คะแนนจริง (True score) ของผู้สอบทั้งสองกลุ่ม

การเทียบระดับคะแนนวิธีที่ 1 และ 2 นั้น ใช้การวิเคราะห์โดยทฤษฎีดั้งเดิม (Classical theory) เป็นการประมาณค่าคะแนนจริง หรือความสามารถแท้โดยใช้คะแนนของกลุ่มผู้สอบเป็นสำคัญ ซึ่งถูกมองว่ามีปัญหาข้อโต้แย้งที่สำคัญ 3 ประการดังนี้ (สำเร็จ บุญเรืองรัตน์ 2529 : 41-42 อ้างอิงมาจาก Hambleton and others. 1978)

1. คุณสมบัติของข้อคำถามของแบบทดสอบไม่คงที่แปรเปลี่ยนไปตามกลุ่มผู้สอบคือ เมื่อวิเคราะห์ค่าความยาก ค่าอำนาจจำแนกของข้อสอบจะพบว่า ค่าเหล่านี้ไม่คงที่เปลี่ยนแปลงไปตามกลุ่มผู้สอบ นั่นก็หมายถึง แบบทดสอบมาตรฐานนั้นจะใช้ได้เฉพาะกับกลุ่มผู้สอบที่มีลักษณะเหมือน ๆ กันเท่านั้น

2. ในการเปรียบเทียบความสามารถของผู้สอบต่างกัน จะต้องใช้แบบทดสอบที่มีข้อคำถามเหมือนกัน หรือ เป็นแบบทดสอบคู่ขนาน

3. ข้อมูลที่ได้จากแบบทดสอบตามมาตรฐานดั้งเดิมนั้น ไม่สามารถบอกได้ว่าผู้สอบมีความสามารถทำอะไรได้ หรือบกพร่องตอนใด

การเทียบระดับคะแนนในวิธีที่ 3 เป็นทฤษฎีที่ใช้โมเดลทางคณิตศาสตร์แสดงความสัมพันธ์ระหว่างความสามารถแท้กับลักษณะการตอบข้อสอบ (Item response) ของแต่ละบุคคลได้อย่างชัดเจนและมีประสิทธิภาพสูง ซึ่งจุดเด่นของทฤษฎีนี้มี 1) ลักษณะเฉพาะของข้อสอบแต่ละข้อมีค่าแน่นอน (Invariant of item parameters) 2) จะใช้ข้อสอบกับใครก็ได้ (person-free) 3) จะสอบข้อใดก็ได้ (Item free) ซึ่งเป็นข้อที่สามารถตอบปัญหาข้อขัดแย้งจากการวิเคราะห์โดยทฤษฎีประมาณคุณลักษณะภายใน (Latent trait theory) หรือทฤษฎีโครงสร้างข้อสอบ (Item

characteristic curve theory) (สงขม ลักษณะ 2525 : 47) เป็นทฤษฎีที่มีประสิทธิภาพในการเทียบระดับคะแนนวิธีหนึ่ง

แบบทดสอบที่ใช้กันอยู่ทั่ว ๆ ไปในปัจจุบันนี้ส่วนมากเป็นแบบทดสอบที่นักเรียนทุกคนต้องทำเหมือนกันหมดและทำทุกข้อ ไม่มีการเทียบระดับความรู้-ความสามารถเดิมของนักเรียนหรือเทียบว่านักเรียนมีความรู้-ความสามารถเพิ่มขึ้นจากเดิมหลังจากมีการเรียน-การสอนหรือไม่ ดังนั้น จึงสมควรจะมีแบบทดสอบที่ในฉบับเดียวกันมีหลายระดับ หรือแบบทดสอบหลายฉบับที่มีข้อสอบรวมที่สามารถจะนำผลการสอบมาเทียบระดับคะแนนบอกความรู้-ความสามารถ จุดบกพร่อง ความก้าวหน้าของนักเรียนได้ผู้วิจัยมีความสนใจที่จะสร้างแบบทดสอบหลาย ๆ ฉบับ ที่จะวัดในสิ่งเดียวกันแต่มีความยากต่างกันและระหว่างฉบับจะมีข้อสอบรวมอยู่ นำไปสอบกับนักเรียนหลายระดับ ซึ่งผลจากการทำข้อสอบรวมจะนำมาใช้สร้างตารางเทียบระดับคะแนน

ในการศึกษาครั้งนี้จะสร้างข้อสอบวิชาคณิตศาสตร์ เรื่องสมการและอสมการของนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 ที่วิธีวิเคราะห์ตามทฤษฎีราชคโหมเดล (Rasch model) ซึ่งเป็นวิธีวิเคราะห์ที่ได้แบบทดสอบ และตารางเทียบระดับคะแนนที่มีประสิทธิภาพ นอกจากนั้นยังเป็นวิธีวิเคราะห์ที่ครู-อาจารย์ สามารถนำไปปฏิบัติได้จริงตามสภาพการเรียน-การสอนในปัจจุบันได้เป็นอย่างดีอีกด้วย เพราะสามารถใช้กับกลุ่มตัวอย่างที่มีขนาดเล็กและวิเคราะห์ด้วยตนเองได้โดยไม่ต้องอาศัยเครื่องคอมพิวเตอร์

จุดมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อสร้างแบบทดสอบวิชาคณิตศาสตร์ เรื่องสมการ และอสมการ สำหรับนักเรียนชั้นมัธยมศึกษาตอนต้นที่วิเคราะห์ด้วยราชคโหมเดล
2. เพื่อสร้างตารางเทียบระดับคะแนนตามแนวราชคโหมเดลสำหรับ เปรียบเทียบระดับคะแนนตามความสามารถในวิชาคณิตศาสตร์เรื่องสมการและอสมการระหว่างนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3

ความสำคัญของการศึกษาค้นคว้า

1. ทำให้ได้แบบทดสอบวิชาคณิตศาสตร์เรื่องสมการและอสมการ สำหรับนักเรียนชั้นมัธยมศึกษาตอนต้น

2. ทำให้สามารถสร้างตารางเทียบระดับคะแนนในการวัดต่างระดับได้ เช่นในการวัดผลวิชาคณิตศาสตร์ของชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 เรื่องสมการและอสมการ ซึ่งจะช่วยให้ได้ข้อมูลที่แสดงถึงการพัฒนาการของนักเรียนกลุ่มเดิมที่ผ่านขึ้นไปเรียนชั้นสูงขึ้นไปว่านักเรียนมีความก้าวหน้าเป็นไปตามปกติหรือเบี่ยงเบนไปจากปกติ จะสามารถทราบได้จากการใช้เทคนิคการวัดผลนี้

3. จะเป็นประโยชน์ในเรื่องการพิจารณาความต่อเนื่องของหลักสูตร ในเรื่องสมการและอสมการ ของชั้นมัธยมศึกษาตอนต้น ว่าหลักสูตรจัดประสบการณ์ให้นักเรียนได้พัฒนาความสามารถเพิ่มขึ้นตามลำดับชั้น ในอัตราส่วนเช่นใด นักเรียนได้พัฒนามากที่สุดในชั้นใด อ่อนหรือลดลงในชั้นใด ซึ่งผลของการศึกษานี้จะช่วยชี้แนะการปรับปรุงการพัฒนาหลักสูตรการเรียนการสอนให้ดีขึ้น

4. เป็นแนวทางในการวิจัยเพื่อพัฒนาการสร้างแบบทดสอบเทียบระดับคะแนนในแนวอื่น ๆ ซึ่งจะเป็นประโยชน์ในการวัดและประเมินผลต่อไป

5. เป็นแนวทางในการนำราซซ์โมเดล (Rasch model) มาใช้ในการสร้างแบบทดสอบให้เหมาะสมกับสภาพปัจจุบันและเก็บสะสมข้อสอบไว้เป็นธนาคารข้อสอบต่อไป

ขอบเขตของการศึกษาค้นคว้า

1. ประชากรที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นนักเรียนในระดับชั้นมัธยมศึกษาตอนต้น ภาคเรียนที่ 2 ปีการศึกษา 2530 ของโรงเรียนในสังกัดกรมสามัญศึกษาเขตท้องที่การศึกษา 2 กรุงเทพมหานคร จำนวนทั้งหมด 24, 648 คน จากจำนวน 495 ห้องเรียน 14 โรงเรียน ซึ่งแยกเป็น

ชั้นมัธยมศึกษาปีที่ 1 จำนวน 8,200 คน 164 ห้องเรียน

ชั้นมัธยมศึกษาปีที่ 2 จำนวน 8,376 คน 166 ห้องเรียน

ชั้นมัธยมศึกษาปีที่ 3 จำนวน 8,072 คน 165 ห้องเรียน

2. กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นนักเรียนชั้นมัธยมศึกษาตอนต้น ในภาคเรียนที่ 2 ปีการศึกษา 2530 ของโรงเรียนในสังกัดกรมสามัญศึกษา เขตท้องที่การศึกษา 2 กรุงเทพมหานคร

ซึ่งใช้วิธีสุ่มแบบแบ่งชั้น (Stratified random sampling) คือมีขนาดโรงเรียนเป็นชั้น (Strata) และมีห้องเรียน เป็นหน่วยการสุ่ม (Sampling unit) การสุ่มตัวอย่างนี้จะสุ่มมา 2 กลุ่ม โดยกลุ่มที่ 1 จะใช้ทดสอบหาค่าพารามิเตอร์ของข้อสอบ และค่าความสามารถ กลุ่มที่ 2 ใช้สำหรับนำผลมาสร้าง ตารางเทียบระดับคะแนน ในแต่ละ กลุ่มจะมีนักเรียน 1,200 คน จาก 24 ห้องเรียน แบ่งเป็นระดับ ละ 400 คน

คำนิยามศัพท์เฉพาะ

1. การเทียบระดับคะแนน (Equating) คือการแปลงคะแนนจากแบบฟอร์มหนึ่ง (x) ไปยังอีกแบบฟอร์มหนึ่ง (y) โดยถือเอาความสามารถในการทำข้อสอบร่วมเท่ากัน
2. แบบทดสอบเทียบระดับคะแนน หมายถึง แบบทดสอบ 3 ฉบับ ที่ผู้วิจัยสร้างขึ้นตามหลักสูตร ของกระทรวงศึกษาธิการ พุทธศักราช 2521 วิชาคณิตศาสตร์ เรื่อง สมการและอสมการ สำหรับชั้น มัธยมศึกษาตอนต้นที่วิเคราะห์ค่าพารามิเตอร์ โดยใช้ราซค์โมเดล (Rasch model)
3. ข้อสอบร่วม (Anchor items) หมายถึง กลุ่มข้อคำถามที่อยู่ในแบบทดสอบทั้งสองฉบับ
4. ความสอดคล้องกับโมเดล (Fit model) หมายถึง การพิจารณาความเป็นจริงตาม การกระจายของความน่าจะเป็น (probability) ของการตอบถูกต้องตามลักษณะโค้งของข้อสอบใน 2 กรณี คือ

กรณีที่ 1 เกี่ยวกับผู้สอบ คือผู้สอบแต่ละคนน่าจะทำข้อสอบข้อง่าย ๆ ได้ถูกมากกว่าทำข้อ ยาก ๆ ได้ถูก

กรณีที่ 2 เกี่ยวกับข้อสอบคือข้อสอบแต่ละข้อนั้น คนที่มีความสามารถสูงน่าจะทำข้อสอบได้ ถูกต้องมากกว่าคนที่มีความสามารถต่ำ

5. แบบทดสอบร่วมภายใน (Internal anchor test) หมายถึง แบบทดสอบฟอร์ม x และแบบทดสอบฟอร์ม y ต่างก็มีข้อสอบซ้ำกันอยู่จำนวนหนึ่ง
6. การเทียบระดับคะแนนในแนวดิ่ง (Vertical equating) หมายถึง การเทียบระดับ คะแนนที่ได้จากกลุ่มตัวอย่างที่ต่างระดับกัน และสอบแบบทดสอบที่ต่างกัน
7. ความยากของข้อสอบ (Difficulty) คือ ความสามารถแท้ของประชากร ณ จุดที่มี ความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องที่ .50

ค่าความยากมีค่าตั้งแต่ $-\alpha$ ถึง $+\alpha$ แต่ในทางปฏิบัติมีค่าอยู่ระหว่าง -2 ถึง $+2$ โดยค่า -2 หมายความว่า ข้อสอบง่าย ส่วน $+2$ หมายความว่า ข้อสอบยาก

8. ความสามารถของผู้สอบ (Examination's ability) คือระดับความสามารถแท้ที่คำนวณจากคะแนนจริง (True score) แต่ปรับหน่วยให้เป็นคะแนนมาตรฐาน (สงข ลักษณะ 2525 : 28) ในทางปฏิบัติมีค่าอยู่ระหว่าง -3 กับ $+3$ โดยค่า -3 หมายความว่า มีความสามารถต่ำและค่า $+3$ หมายความว่า มีความสามารถสูง สัญลักษณ์แทนค่าความสามารถที่นิยมใช้คือ θ หรือ β แต่ในการศึกษาครั้งนี้จะใช้ β

9. ความไม่ผันแปรตามกลุ่ม (Invariance across group) คือ ค่าพารามิเตอร์ต่าง ๆ มีค่าแน่นอนไม่ผันแปรตามกลุ่มของผู้สอบ นั่นคือ เมื่อนำข้อสอบชุดเดิมไปสอบกับคนต่างกลุ่มกันจะปรากฏค่าพารามิเตอร์ต่างกัน ทั้งนี้เป็นเพราะแต่ละกลุ่มมีจุดตั้งต้น (Origin) ของความสามารถแท้ (θ หรือ β) บนแกนราบต่างกัน ถ้ามีการปรับ (Transform) ค่าความสามารถแท้ให้มีสเกลเดียวกันแล้วค่าพารามิเตอร์เหล่านี้จะเท่ากัน

10. แบบทดสอบที่มีคามิติเดียวกัน (Unidimensional test) หมายถึง แบบทดสอบที่วัดความสามารถ หรือคุณลักษณะเดียวกัน ซึ่งข้อสอบดังกล่าวจะต้องมีความ เป็นเอกพันธ์กันสูงหรือเป็นข้อสอบที่สร้างมาจากเนื้อหาเดียวกัน

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การศึกษาค้นคว้าครั้งนี้ ผู้วิจัยได้ศึกษาในเรื่องต่าง ๆ ดังนี้

ความหมายของการ เทียบระดับคะแนน

โมเดลการ เทียบระดับคะแนน โดยใช้ทฤษฎีการตอบข้อคำถาม

โมเดลเกี่ยวกับราชคโสมเดล

รูปแบบการ เชื่อมต่อของแบบทดสอบ

งานวิจัยที่เกี่ยวข้อง

ความหมายของการ เทียบระดับคะแนน

กัลลิเซน (Gulliksen. 1950 : 298-304) ได้ให้ความหมายของการ เทียบระดับคะแนนว่าเป็นวิธีการทำคะแนนที่ได้จากแบบทดสอบสองชุดของวิชาเดียวกัน ให้เป็นคะแนนที่สมมูลกันเปรียบเทียบกันโดยตรง

ฟลานาแกน (Flanagan. 1951 : 747-748) ได้กล่าวว่า การ เทียบระดับคะแนนเป็นวิธีการที่ทำคะแนนจากแบบทดสอบต่างชุดกัน ให้มีคุณลักษณะที่เทียบกันได้

แองกอฟ (Angoff. "Summary and derivation of equation methods used of ETS" in test equating 1982 : 56) ได้ให้คำจำกัดความว่า "การ เทียบระดับคะแนนเป็นการแปลงระบบของหน่วยหนึ่งของแบบทดสอบไปสู่ระบบของหน่วยวัดอีกหน่วยหนึ่ง ซึ่งการเปลี่ยนแปลงนั้นเป็นเหมือนฟังก์ชันทางคณิตศาสตร์ที่สามารถนำไปประยุกต์ใช้กับกลุ่มอื่น ๆ ได้"

ลอร์ด (Lord. 1980 : 195) ได้กล่าวว่า การ เทียบระดับคะแนนคือวิธีการเทียบคะแนนจากแบบทดสอบ X และแบบทดสอบ Y ที่มีความยากต่างกันจากความสามารถเดียวกัน

ชูศักดิ์ ชัมภลลิขิต (ชูศักดิ์ ชัมภลลิขิต 2527 : 2) ได้สรุปเกี่ยวกับการเทียบระดับคะแนนว่าเป็นกระบวนการที่เกี่ยวข้องกับกิจกรรม 2 ประการ คือ

1. กระบวนการที่ทำให้แบบทดสอบ 2 ฉบับใด ๆ มีความเท่าเทียมกัน หรือเท่ากันในเชิงของโครงสร้าง
2. การใช้วิธีการทางสถิติเพื่อปรับ (Adjust) คะแนนที่ได้จากแบบทดสอบแต่ละฉบับให้อยู่ในมาตราเดียวกันและเทียบกันได้

สงข ลักษณ์ (สงข ลักษณ์ 2525 : 22) ได้กล่าวว่า คะแนนจากแบบทดสอบสองฉบับวัดสิ่งเดียวกัน แต่ไม่จำเป็นต้องเป็นข้อสอบคู่ขนาน จะถือว่าเทียบเท่ากันได้ถ้าคะแนนจากแบบทดสอบทั้ง 2 นั้นมาจากคะแนนจริง (True score) หรือความสามารถแท้ (True ability) ที่เท่ากัน

จากความหมายดังกล่าว พอสรุปได้ว่า การเทียบระดับคะแนน หมายถึงวิธีการนำคะแนนจากแบบทดสอบฉบับหนึ่งไปเทียบกับอีกฉบับหนึ่ง โดยถือเอาความสามารถในการทำข้อสอบรวมของบุคคลผู้สอบ เป็น เกณฑ์

โมเดลการเทียบระดับคะแนนโดยใช้ทฤษฎีการตอบข้อคำถาม

ลอร์ด (Lord 1980 : 199-203) ได้กล่าวว่า แบบทดสอบสองชุดใด ๆ ที่ใช้เทียบระดับคะแนนกันนั้นจะต้องเป็นแบบทดสอบที่มีมิติเดียวกัน และเทียบคะแนนจากความสามารถที่เท่ากัน ซึ่งจะมีข้อกำหนดที่สำคัญ 3 ประการดังนี้

1. ความเท่าเทียมกัน (Equity) คือ สำหรับทุกความสามารถของคน (θ หรือ β) จะมีการกระจายของคะแนนแปลง ($x(y)$) เหมือนกับเงื่อนไขการกระจายของคะแนนจากแบบทดสอบที่ต้องการเทียบ
2. ความไม่ผันแปรตามกลุ่ม (Invariance across groups) คือ คะแนนที่แปลงจะเหมือนเดิมไม่เปลี่ยนแปลงไปตามกลุ่มประชากรที่ต้องการเทียบระดับคะแนน
3. ความสมมาตร (Symmetry) คือ ผลจากการเทียบระดับคะแนนจะต้องเหมือนกันไม่ว่าจะเป็นการเทียบระดับคะแนนจากแบบทดสอบ X ไปแบบทดสอบ Y หรือเป็นการเทียบจากแบบทดสอบ Y ไปแบบทดสอบ X

นอกจากนี้ ลอร์ดได้แบ่งโมเดลการเทียบระดับคะแนนโดยใช้ทฤษฎีการตอบข้อคำถามออกเป็นหลายวิธีดังนี้

1. วิธีการเทียบระดับคะแนนโดยใช้คะแนนจริง (True-score equating)

แบบทดสอบที่มีความยากต่างกัน เมื่อนำไปสอบกับประชากรที่มีความสามารถเท่ากันหรือคุณลักษณะเดียวกันแล้ว คะแนนที่ได้จะมีลักษณะการกระจายต่างกัน ซึ่งถ้าเราสามารถหาค่าคะแนนจริงได้ ก็จะสามารถทำการเทียบระดับคะแนนระหว่างความสามารถ (β) กับคะแนนจริง (True-score) จากความสัมพันธ์เชิงคณิตศาสตร์ดังนี้

$$\xi = \xi(\beta) = \sum_{i=1}^n p_i(\beta) \quad \dots\dots\dots (1)$$

$$\eta = \eta(\beta) = \sum_{j=1}^m p_j(\beta) \quad \dots\dots\dots (2)$$

ξ คือ คะแนนจริงจากแบบทดสอบฟอร์ม X

η คือ คะแนนจริงจากแบบทดสอบฟอร์ม Y

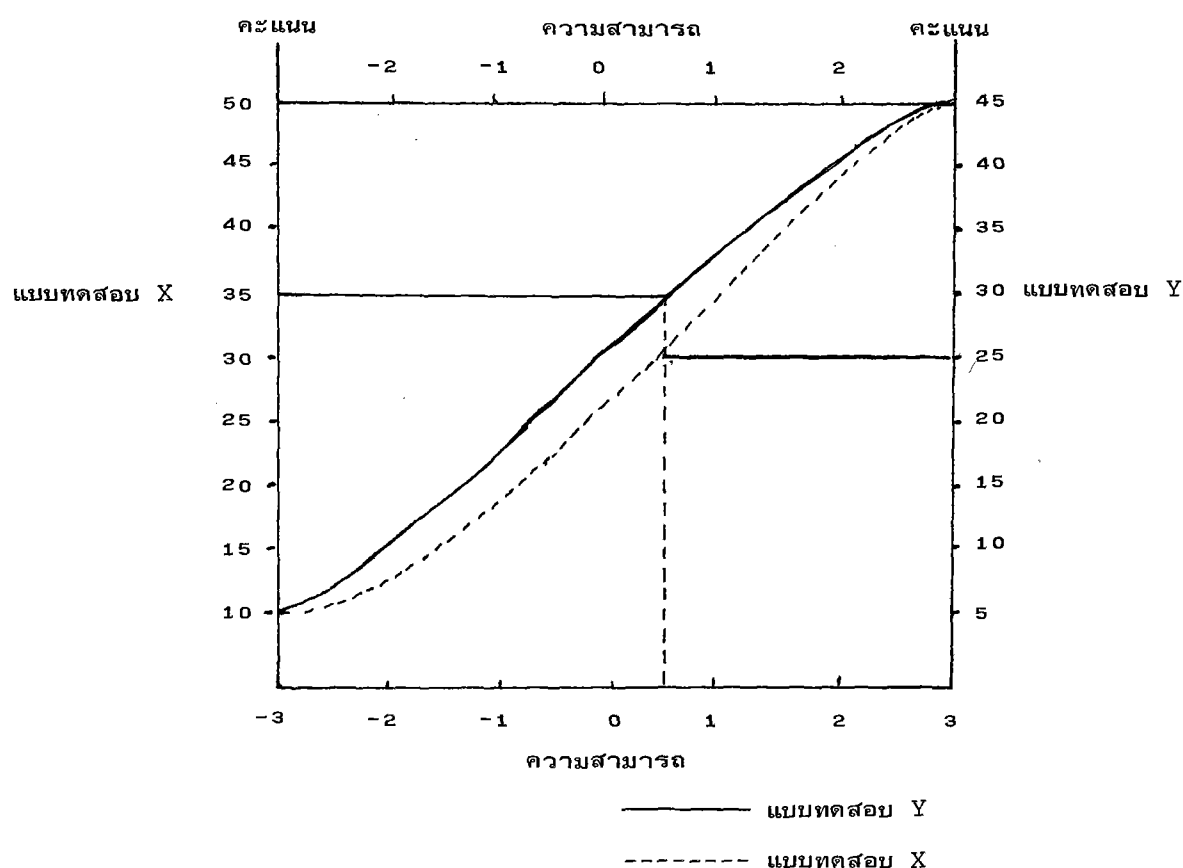
P คือ ความน่าจะเป็นของการตอบถูก

n คือ จำนวนข้อของแบบทดสอบฟอร์ม X

m คือ จำนวนข้อของแบบทดสอบฟอร์ม Y

สมการ (1) และ (2) เป็นสมการพารามेटริก (Parametric equations) ระหว่างคะแนนจริงของแบบทดสอบฟอร์ม X และฟอร์ม Y ดังนั้นหากแทนค่า β ใด ๆ ลงไปจะสามารถหาค่า ξ และ η ได้ และในทางตรงกันข้ามก็สามารถประเมินค่า β ได้ เมื่อทราบคะแนนจริงได้

ในทางปฏิบัติเราจะหาค่า $p_i(\beta)$ และ $p_j(\beta)$ ได้จากการประมาณค่าพารามิเตอร์ของข้อสอบ โดยเลือกที่จะใช้ชนิด one-parameter, two-parameter หรือ three-parameter จากนั้นนำคะแนนจริงของแบบทดสอบฟอร์ม X และแบบทดสอบฟอร์ม Y มาหาความสัมพันธ์กันโดยใช้ค่าความสามารถ (β) ที่ระดับเดียวกันดังภาพ

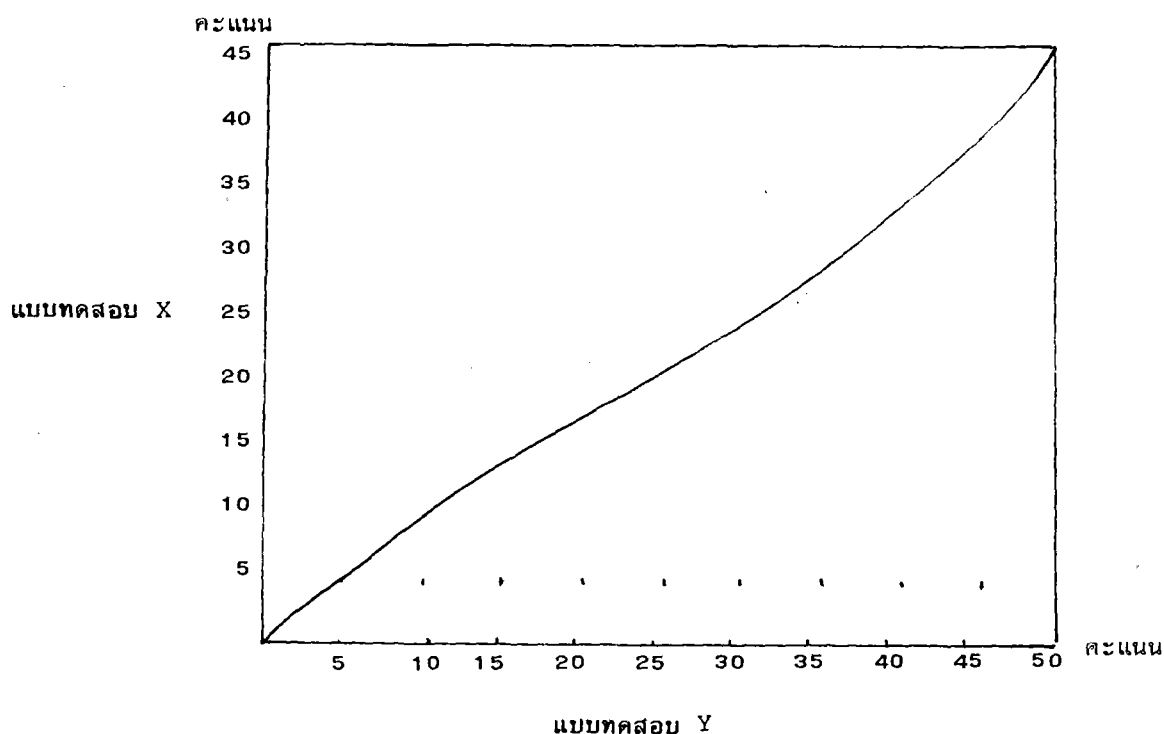


ภาพประกอบ 1 แสดงการเทียบระดับคะแนนโดยใช้คะแนนจริงตามทฤษฎีการตอบข้อคำถาม

2. วิธีการเทียบระดับคะแนนโดยใช้คะแนนจริงด้วยแบบทดสอบร่วม (True-score

equating with on anchor test) การใช้แบบสอบร่วมทำได้ 2 กรณี คือ ใช้แบบสอบร่วมรวมเข้าเป็นชุดเดียวกับแบบทดสอบที่ต้องการเทียบระดับคะแนน เรียกว่า แบบทดสอบร่วมภายใน (Internal anchor test) ส่วนในกรณีที่จัดแยกเป็นชุดแบบทดสอบต่างหากจากแบบสอบที่ต้องการเทียบระดับคะแนน เรียกว่า แบบสอบร่วมภายนอก (External anchor test) จุดมุ่งหมายของการใช้แบบสอบร่วม คือ เป็นการใช้แบบสอบร่วมเป็นตัวเชื่อมข้อมูลเข้าด้วยกัน เพื่อให้ค่าพารามิเตอร์ที่วิเคราะห์ออกมาอยู่ในระดับเดียวกัน ซึ่งถ้าขาดข้อสอบร่วมแล้ว จะไม่สามารถเทียบกันได้ นอกเสียจากว่ากลุ่มตัวอย่างที่สอบแบบทดสอบ X และ Y จะมีการกระจายของความสามารถเหมือนกัน

ในทางปฏิบัติจะหาค่าความสามารถ (β) ของกลุ่มตัวอย่างที่ทำแบบทดสอบร่วมไปเทียบระดับคะแนนกับแบบทดสอบฟอร์ม X และแบบทดสอบฟอร์ม Y ซึ่งมาเขียนกราฟได้ดังภาพ



ภาพที่ 2 แสดงการเทียบระดับคะแนนโดยใช้คะแนนจริงจากแบบทดสอบร่วม

3. วิธีการเทียบระดับคะแนนโดยใช้คะแนนดิบจากแบบทดสอบร่วม (Raw-score equating with an anchor)

การเทียบระดับคะแนนวิธีนี้เริ่มด้วยการประมาณลักษณะการกระจายของความสามารถในกลุ่มรวม ($\delta(\beta)$) โดยการวิเคราะห์มาจากการกระจายของคะแนนดิบในการทำข้อสอบร่วม ($\hat{\delta}(\beta)$) จากสมการ

$$\hat{\phi}_X(X) = \frac{1}{N} \sum_{a=1}^N \hat{\phi}_X(X|\hat{\beta}_a) \quad \dots\dots\dots (3)$$

เมื่อ $a = 1, 2, 3, \dots, N$ คือ จำนวนผู้เข้าสอบ

ถ้า $\hat{\delta}(\beta)$ มีลักษณะต่อเนื่อง (continuous) จะประมาณค่า $\phi(X)$ จากสมการ

$$\hat{\phi}_X(X) = \int_{-\alpha}^{\alpha} \hat{\phi}_X(X|\beta) \hat{\gamma}(\beta) d\beta \quad \dots\dots\dots (4)$$

เมื่อ $\hat{\phi}_x(X|\hat{\beta}_a)$ ได้มาจากการประมาณค่าพารามิเตอร์โดยใช้ทฤษฎีการตอบข้อคำถาม (Item parameter) ของข้อสอบในแบบทดสอบฟอร์ม X (Loard. 1980 : 45-46, 203-204) และสมการต่าง ๆ สามารถนำมาประยุกต์ใช้กับแบบทดสอบฟอร์ม Y ได้เช่นเดียวกัน

เนื่องจากแบบทดสอบฟอร์ม X และฟอร์ม Y เป็นอิสระจากกัน เมื่อความสามารถ (β) คงที่ การเทียบระดับคะแนนต้องคำนวณมาจากการกระจายร่วม (Joint distribution) ของคะแนนจากแบบทดสอบฟอร์ม X และฟอร์ม Y จากสมการ

$$\hat{\phi}(x,y) = \frac{1}{N} \sum_{a=1}^N \hat{\phi}_x(X|\hat{\beta}_a) \hat{\phi}_y(Y|\hat{\beta}_a) \dots\dots\dots (5)$$

หรือจากสมการ

$$\hat{\phi}(x,y) = \int_{-\alpha}^{\alpha} \hat{\phi}_x(X|\beta) \hat{\phi}_y(Y|\beta) \hat{\gamma}(\beta) d\beta \dots\dots\dots (6)$$

จากสมการข้างต้น จะเห็นบทบาทของแบบสอบร่วมได้ชัดเจน ทำให้สามารถประมาณการแจกแจงร่วม (Joint distribution) ของ X และ Y ได้ถึงแม้จะไม่มีผู้ใดทำแบบทดสอบทั้ง 2 ฉบับ

สมการ (6) เป็นการแจกแจงร่วมกันของตัวแปรสามตัวคือ β , X และ Y

เนื่องจาก β เป็นตัวระบุคะแนนจริงของ ξ และ η การแจกแจงนี้จึงนับว่าเป็นการแจกแจงร่วมกันของตัวแปร 4 ตัว คือ ξ , η , X และ Y สมการนี้ไม่มีทางอื่นจะประมาณได้ นอกจากจะจากความสัมพันธ์อ็อกซิเจนเปอร์เซ็นต์ระหว่าง X และ Y จาก 2 สมการสุดท้าย

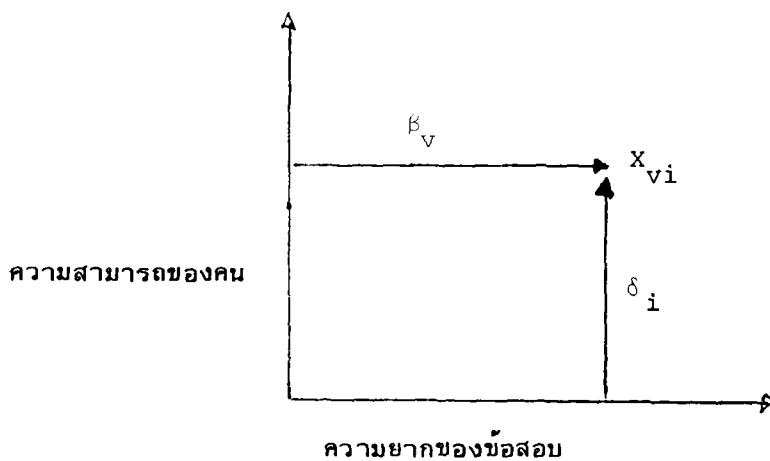
โมเดลเกี่ยวกับราชคโฆเด

เมื่อคน ๆ หนึ่งทำข้อสอบข้อหนึ่ง เหตุการณ์ที่เกิดขึ้นขณะนั้นคือคนพยายามใช้ความสามารถที่มีอยู่ เอาชนะข้อสอบด้วยการแก้ปัญหาโจทย์ข้อนั้น โอกาสที่คนจะทำข้อสอบถูกหรือไม่ขึ้นอยู่กับระดับความสามารถของตนเอง (Ability parameter : θ หรือ β) และระดับความยากของข้อสอบ

ถ้าให้ β_v เป็น ค่าของความสามารถของคน

δ_i เป็น ค่าความยากของข้อสอบ

เหตุการณ์ขณะที่บุคคลใช้ความสามารถแก้ปัญหาโจทย์ข้อหนึ่ง ๆ อาจเขียน เป็นรูปได้ดังนี้



ถ้าคนชนะข้อสอบ $X_{vi} = 1$

ถ้าคนแพ้ข้อสอบ $X_{vi} = 0$

ถ้าเปรียบเทียบความสามารถของคนกับความยากของข้อสอบในเชิงคณิตศาสตร์ได้ดังนี้

$$(\beta_v - \delta_i) > 0, \quad P\{X_{vi} = 1\} > .5 \quad (1)$$

$$(\beta_v - \delta_i) < 0, \quad P\{X_{vi} = 1\} < .5 \quad (2)$$

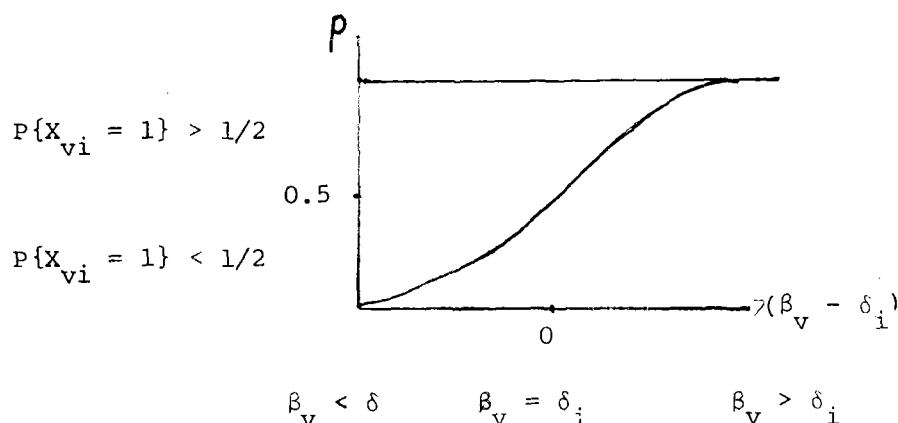
$$(\beta_v - \delta_i) = 0, \quad P\{X_{vi} = 1\} = .5 \quad (3)$$

จากสมการ (1) อธิบายได้ว่า เมื่อความสามารถของคน β_v มีค่ามากกว่าความยากของข้อสอบ δ_i ถ้านำความสามารถของคนลบด้วยความยากของข้อสอบแล้ว ผลลัพธ์จะมากกว่าศูนย์ โอกาสที่คนนั้นจะทำข้อสอบถูกจะมากกว่า .50

จากสมการ (2) อธิบายได้ว่า เมื่อความสามารถของคนลดยความยากของข้อสอบได้ ผลลัพธ์น้อยกว่าศูนย์แล้ว โอกาสที่คน ๆ นั้นจะทำข้อสอบได้ถูกก็จะน้อยกว่า .50

จากสมการ (3) อธิบายได้ว่าในเมื่อความสามารถของคนลดยความยากของข้อสอบได้ เท่ากับศูนย์พอดีแล้ว โอกาสที่คน ๆ นั้นจะทำข้อสอบนั้นได้คะแนนเท่ากับ .50 พอดี ซึ่งอาจจะทำถูกหรือ ผิดก็ได้

จากเงื่อนไขที่ได้อธิบายนี้ อาจนำมาเขียนเป็นรูปแสดงความน่าจะเป็นดังนี้



รูปแบบแสดงความสัมพันธ์ระหว่างผลการสอบและความสามารถของผู้สอบตามราชคโมเดล

เมื่อกำหนดให้ ความน่าจะเป็นของบุคคลที่ v

ความสามารถของบุคคล β_v

คะแนนของผู้สอบ X_{vi}

ความยากของข้อสอบ δ_i

จะได้สมการสำหรับใช้วิเคราะห์ความน่าจะเป็นในการทำข้อสอบที่เรียกว่า ราชคโมเดล

(Rasch Model) ได้ดังนี้

$$P\{X_{vi}/\beta_v, \delta_i\} = \exp[X_{vi}(\beta_v - \delta_i)]/[1 + \exp(\beta_v - \delta_i)] \dots \dots \dots (4)$$

โดยที่ x_{vi} จะมีค่า 2 ค่าคือ

$$x_{vi} = 0 \quad \text{เมื่อตอบผิด}$$

$$x_{vi} = 1 \quad \text{เมื่อตอบถูก}$$

ดังนั้นสามารถเขียนสมการสำหรับคำตอบที่ถูกใหม่ได้อีกว่า

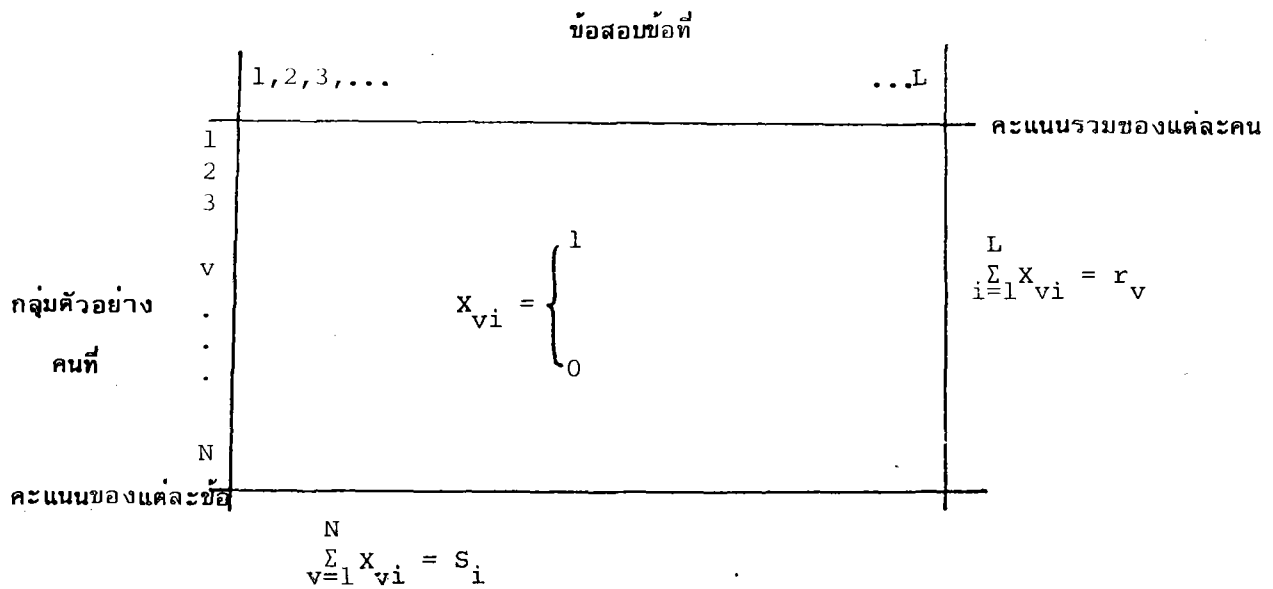
$$P\{x_{vi} = 1 | \beta_v, \delta_i\} = \exp(\beta_v - \delta_i) / [1 + \exp(\beta_v - \delta_i)] \dots\dots\dots (5)$$

และสำหรับคำตอบที่ผิด เขียนสมการใหม่ได้อีกว่า

$$P\{x_{vi} = 0 | \beta_v, \delta_i\} = 1 / [1 + \exp(\beta_v - \delta_i)] \dots\dots\dots (6)$$

ถ้านำสมการที่ (5) กับ (6) มารวมกันจะมีค่าเป็น 1 ทั้งนี้เพราะเป็นการรวมค่าความน่าจะเป็นของการตอบถูกและตอบผิด ซึ่งจะได้ผลลัพธ์เป็น 1 เสมอ และค่านี้จะเป็นตัวคูณค่า x_{vi} หรือคะแนนที่สังเกตได้นั่นเอง

เมื่อนำแบบทดสอบฉบับหนึ่งซึ่งมี L ข้อ มีจำนวนกลุ่มตัวอย่าง N คนนำผลการสอบมาเขียนอยู่ในรูปเมตริก (Matrix) ดังภาพ



เพราะฉะนั้น ถ้าบุคคล v แปรค่าตั้งแต่ 1 ถึง N และข้อสอบ i แปรค่าตั้งแต่ 1 ถึง L

จะได้ว่า

$$\text{คะแนนที่คนสอบได้ } r_v = \sum_{i=1}^L X_{vi}$$

$$\text{จำนวนคนที่ทำข้อสอบถูกในแต่ละข้อ } S_i = \sum_{v=1}^N X_{vi}$$

ถ้านำความสามารถของคนทุกคนและความยากของข้อทุกข้อ ไปแทนในสมการ (5) จะได้

$$P\{X_{vi}/\beta_v, \delta_i\} = \prod_v \prod_i \{ \exp(X_{vi}(\beta_v - \delta_i)) / (1 + \exp(\beta_v - \delta_i)) \} \dots\dots\dots (7)$$

ซึ่งค่าผลรวมของ $\prod_v$ และ $\prod_i$ กับค่า $\exp[X_{vi}(\beta_v - \delta_i)]$ ก็คือค่าของ Σ และ Σ

นั่นเอง

ดังนั้น

$$\prod_v \prod_i \exp[X_{vi}(\beta_v - \delta_i)] = \exp\left[\sum_{i=1}^L \sum_{v=1}^N X_{vi}(\beta_v - \delta_i)\right]$$

แต่เพราะว่า

$$\sum_{v=1}^N \sum_{i=1}^L x_{vi} \beta_v = \sum_v r_v \beta_v$$

และ

$$\sum_{v=1}^N \sum_{i=1}^L x_{vi} \delta_i = \sum_{i=1}^L s_i \delta_i$$

แทนค่าในสมการ (7) จะได้

$$P\{(X_{vi})/(\beta_v), (\delta_i)\} = \exp\left[\sum_{v=1}^N r_v \beta_v - \sum_{i=1}^L s_i \delta_i\right] / \prod_v \prod_i [\exp(\beta_v - \delta_i)] \dots (8)$$

จากสมการจะเห็นได้ว่า เราสามารถแยกความสามารถของคนออกจากคะแนนของข้อสอบได้ ทำให้วิเคราะห์ค่าความสามารถของคนโดยไม่ขึ้นอยู่กับข้อสอบ (person free) ได้ ในทำนองเดียวกัน เราอาจวิเคราะห์ค่าของคำถามโดยไม่ต้องขึ้นกับกลุ่มของคนได้ (Item free) ซึ่งเป็นลักษณะพิเศษของการวิเคราะห์แบบราชคส์ที่ทำให้สามารถวิเคราะห์ค่าความสามารถของคนและความยากของข้อสอบได้เป็นอิสระจากกัน

วิธีประมาณค่าพารามิเตอร์ในราชคส์โมเดล มีด้วยกันหลายวิธีคือ อวยพร วิบูลย์กัจจน์ 2526:

19-20 อ้างอิงจาก Wriht: after word in Rasch 1980: 188-192)

1. The LOG Method เป็นวิธีการประมาณค่าพารามิเตอร์ในรูป log ของความสำเร็จ ซึ่งเป็นวิธีที่เหมาะสมกับแบบทดสอบที่มีค่าอำนาจจำแนกเท่ากันคือ 1 และใช้จำนวนคนในการประมาณค่าพารามิเตอร์เป็นจำนวนมาก

2. The PAIR Method เป็นวิธีการประมาณค่าพารามิเตอร์ โดยใช้ข้อสอบจับกันเป็นคู่ ๆ เช่น 1-2, 1-3, 1-4, 2-4, ... เป็นต้น เอาจำนวนคนที่พยายามตอบ 2 ข้อแต่ถูกเพียงข้อเดียว มาวิเคราะห์เพื่อหา Sample free ดังนั้นจึงเหมาะที่จะใช้ประมาณค่าแบบทดสอบที่มีจำนวนข้อไม่มาก

3. The FCON Method เป็นวิธีประมาณค่าพารามิเตอร์โดยใช้ข้อสอบทุกข้อและทุกคนมาวิเคราะห์ เหมาะสำหรับแบบสอบที่มีจำนวนข้อสอบน้อยกว่า 30 ข้อ ถ้ามากกว่า 30 ข้อจะทำให้การประมาณค่าพารามิเตอร์มีความคลาดเคลื่อน แต่ภายหลังได้มีการปรับปรุงใหม่สามารถใช้กับแบบทดสอบที่มีจำนวนข้อ 60-70 ข้อได้

4. The UCON Method เป็นวิธีประมาณค่าพารามิเตอร์ด้วยวิธี Unconditional maximum likelihood procedure ใช้กับแบบทดสอบที่มีการกระจายของความสามารถของคนและความยากเป็นปกติ (Normal distribution) และใช้กับแบบทดสอบที่มีจำนวนข้อตั้งแต่ 25 ข้อขึ้นไปโดยมากจะใช้กับแบบทดสอบที่มี 1,000 ข้อขึ้นไป ถ้าจำนวนข้อน้อยจะใช้ FCON

5. The PROX Method เป็นวิธีประมาณค่าพารามิเตอร์ด้วยวิธีของ Cohen's approximation ใช้กับแบบทดสอบที่มีการกระจายความสามารถของคนและความยากของข้อสอบเป็นลักษณะ เบ้หรือมีแนวโน้ม เป็นปกติ โดยมาก PROX และ UCON จะให้ผลเหมือนกันจะต่างกันที่ค่าความคลาดเคลื่อนมาตรฐานเพียงเล็กน้อย

6. The UFORM Method ใช้กับแบบทดสอบที่มีการกระจายความยากของข้อสอบ เป็นแบบ Uniform

การทดสอบความสอดคล้อง (The analysis of fit)

ตามหลักการที่สำคัญของราชคโหมเดล คือ

1. คนที่มีความสามารถมากย่อมมีโอกาสที่จะทำข้อสอบถูกมากกว่าคนที่มีความสามารถน้อย
2. ภายในตัวคนเดียวโอกาสที่จะตอบถูกในข้อคำถามที่ง่าย ๆ ย่อมมีมากกว่าโอกาสที่จะตอบถูกในข้อคำถามยาก ๆ

ดังนั้น ในการวิเคราะห์ข้อสอบตามแนวของราชคโหม นั้น เมื่อวิเคราะห์เป็นรายข้อและรายคนได้แล้ว ยังต้องทดสอบดูความสอดคล้อง หรือความเป็นจริงตามเหตุตามผลโดยพิจารณาใน 2 กรณีคือ

กรณีที่ 1 พิจารณาความสมเหตุสมผลของผู้เข้าสอบ โดยดูว่าบุคคลนั้นทำแบบทดสอบข้อง่าย ๆ ถูกแล้วทำผิดในข้อยาก ๆ ซึ่งเป็นเหตุการณ์ที่สมเหตุสมผล แต่ถ้าบุคคลนั้นทำผิดในข้อง่าย ๆ แต่ทำถูกในข้อยาก ซึ่งเป็นเหตุการณ์ที่ไม่สมเหตุสมผล

กรณีที่ 2 พิจารณาว่าเหตุการณ์ที่ข้อสอบข้อหนึ่ง ๆ นั้น คนที่มีความสามารถมากจะทำข้อสอบได้ถูก ส่วนคนที่มีความสามารถน้อยจะทำข้อสอบผิด ซึ่งมีเหตุการณ์ที่ตรงข้ามกันจะเรียกว่า ข้อสอบนั้นมีความบกพร่องหรือข้อสอบนั้นไม่สอดคล้องกับโมเดล

วิธีวิเคราะห์ความสอดคล้อง

จากที่น่าจะเป็นของเหตุการณ์

$$P_{vi} = \exp(\beta_v - \delta_i) / (1 + \exp(\beta_v - \delta_i))$$

ค่าความแปรปรวนของที่น่าจะเป็นของเหตุการณ์

$$SD_p^2 = P_{vi}(1 - P_{vi})$$

ค่าความคลาดเคลื่อนมาตรฐานของที่น่าจะเป็นของเหตุการณ์

$$SD_p = (P_{vi}(1 - P_{vi}))^{1/2}$$

คะแนนมาตรฐานของเหตุการณ์ $Z_i = (X_i - P_{vi}) / (P_{vi}(1 - P_{vi}))^{1/2}$

ซึ่งคะแนนมาตรฐาน Z_{vi} จะประมาณว่ามีค่าเฉลี่ยเป็น 0 ค่าความเบี่ยงเบนมาตรฐานเป็น 1 หรือมีคุณสมบัติการแจกแจงเป็นโค้งปกตินั่นเอง

$$Z_{vi} \sim N(0,1)$$

ค่ากำลังสองของคะแนนมาตรฐานนี้ จะมีค่าใกล้เคียงกับค่าไคสแควร์ (chi-square) ที่เป็นความอิสระของตัวแปรเท่ากับ 1 หรือ

$$Z_{vi}^2 \sim \chi_1^2$$

การที่ให้ค่า Z_{vi}^2 มีค่าเท่ากับ χ^2 (chi-square) เพื่อให้เป็นหลักฐานอ้างอิงในการตัดสินใจว่าเหตุการณ์ใดสอดคล้องหรือไม่สอดคล้องกับค่าคาดหวังทางสถิติ ถ้าเราวิเคราะห์แล้วพบว่าค่ากำลังสอง

(Z_{vi}^2) ใหญ่มากเราก็จะเริ่มสงสัยได้แล้วว่ามีความผิดปกติเกิดขึ้นเพราะค่านี้คือตรรกะของความคลาดเคลื่อนอย่างหนึ่งนั่นเอง

เมื่อให้บุคคล v ทำข้อสอบ i คะแนนได้ จะแทนด้วย X_{vi}

$$X_{vi} = 0 \quad \text{ถ้าตอบผิด}$$

$$X_{vi} = 1 \quad \text{ถ้าตอบถูก}$$

ค่ากำลังสองของคะแนนมาตรฐานจะได้

$$Z_{vi}^2 = \exp(\beta - \delta) \quad \text{สำหรับ } X_{vi} = 0$$

$$Z_{vi}^2 = \exp(\delta - \beta) \quad \text{สำหรับ } X_{vi} = 1$$

การวิเคราะห์ความไม่สมจริงของผู้สอบ หรือกรณีที่ 1 วิเคราะห์โดยรวมค่าความคลาดเคลื่อนมาตรฐานกำลังสองจากข้อแรกถึงข้อสุดท้ายที่คนทำข้อสอบดังนี้

$$v = (\sum Z_{vi}^2) / (L-1) \quad L \text{ คือ จำนวนข้อ}$$

$$t_v = [\ln v + v - 1] [(L-1)/8]^{1/2}$$

การวิเคราะห์ความไม่สมจริงของข้อสอบรายข้อ หรือกรณีที่ 2 ให้รวมค่าความคลาดเคลื่อนมาตรฐานกำลังสองของรายข้อที่ผู้สอบทำข้อสอบข้อนั้นแล้วคำนวณดังนี้

$$v = (\sum Z_{vi}^2) / (N-1) \quad N \text{ คือ จำนวนคน}$$

$$t_i = [\ln(v) + v - 1] [(N-1)/8]^{1/2}$$

ประสิทธิภาพของข้อสอบนอกจากจะต้องสอดคล้องกับโมเดลแล้วยังต้องพิจารณาถึงความแตกต่างระหว่างความสามารถของคนกับความยากของข้อสอบ คือถ้าข้อสอบข้อใดมีความสามารถของกลุ่มตัวอย่างกับค่าความยากใกล้เคียงกัน ข้อสอบนั้นก็จะมีประสิทธิภาพในการใช้มาก แต่ถ้าความสามารถของคนกับความยาก

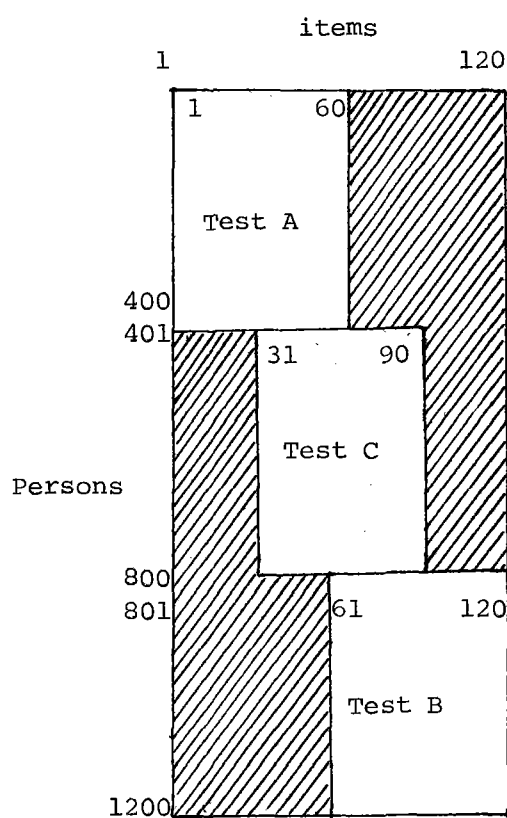
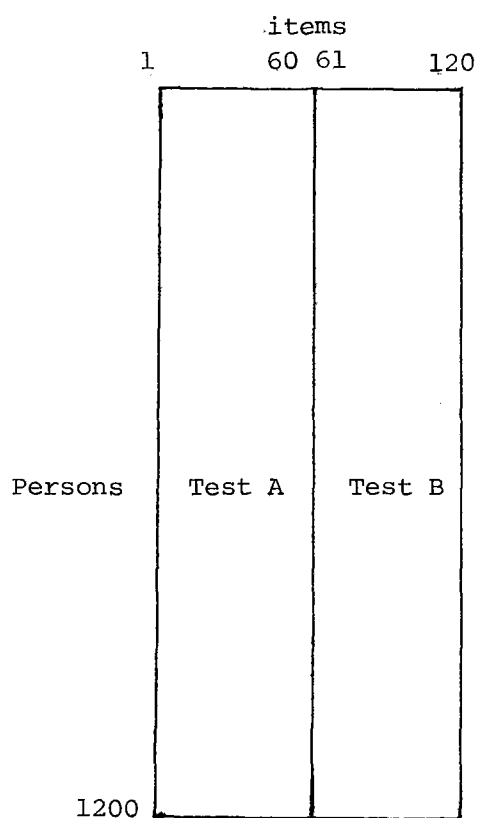
ของข้อสอบแตกต่างกันมาก ข้อสอบนั้นก็จะมึประสิทธิภาพค่า ไรท์ และสโตน (Wright and Stone 1979 : 75) ได้จัดระดับประสิทธิภาพของข้อสอบจากความแตกต่างระหว่างความสามารถของคนกับความยากของข้อสอบไว้ดังนี้

ตาราง 1 แสดงค่าความแตกต่างระหว่างความสามารถของคนกับความยากของข้อสอบในการทำนายผล

ความเหมาะสมกับจุดประสงค์	ค่าความแตกต่าง	ประสิทธิภาพของข้อสอบ
เหมาะสมมากที่สุด	$ \beta - \delta < 1$	มากกว่า 79%
เหมาะสมมาก	$1 < \beta - \delta < 2$	มากกว่า 45%
ไม่ค่อยเหมาะสม	$2 < \beta - \delta < 3$	น้อยกว่า 45%
ไม่เหมาะสม	$3 < \beta - \delta < 4$	น้อยกว่า 18%
ไม่ควรใช้	$4 < \beta - \delta $	น้อยกว่า 7%

รูปแบบการ เชื่อมต่อของแบบทดสอบ

การเทียบมาตรฐานจากเปอร์เซนไทล์ที่เท่าเทียม (Equal-percentile) หรือวิธีเชิงเส้นตรง เป็นวิธีที่สิ้นเปลืองทั้งเวลาและค่าใช้จ่ายเพราะต้องใช้กลุ่มตัวอย่างขนาดใหญ่เพื่อให้เกิดความมั่นใจในการสรุปผล และหน่วยตัวอย่างต้องทำข้อสอบทุกข้อ แต่การเทียบระดับคะแนนโดยใช้ราซค์โมเดล (Rasch model) จะช่วยประหยัดเวลาและค่าใช้จ่ายเพราะหน่วยตัวอย่างไม่ต้องทำข้อสอบทุกข้อ ต่างทำข้อสอบที่เป็นอิสระแล้วทำข้อสอบร่วมกันประมาณ 10-20 ข้อดังภาพประกอบ

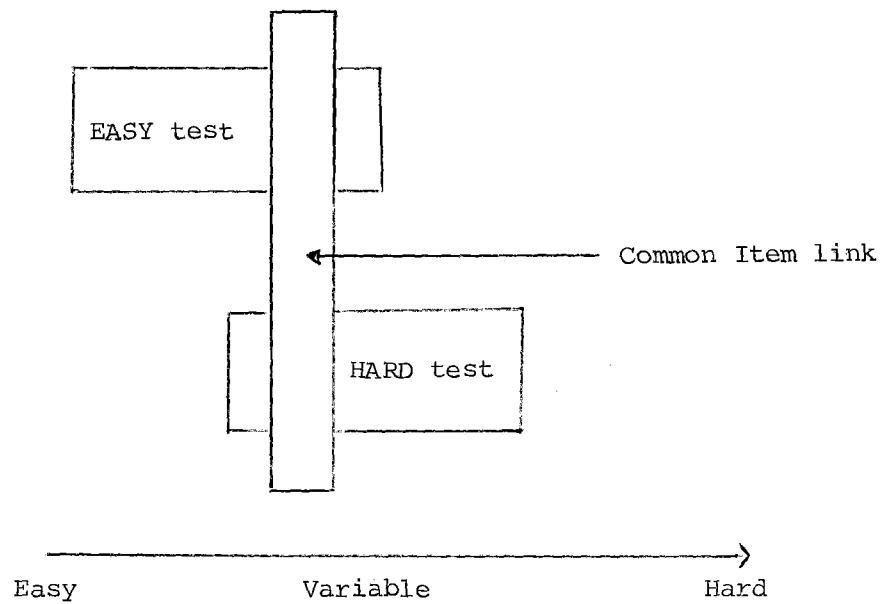
Tradition Equal-percentileRasch Common Item

ภาพประกอบ 3 แสดงการเทียบระดับคะแนน 2 วิธี

จากภาพประกอบ 3 จะเห็นว่าการเทียบระดับคะแนนโดยเทียบจากเปอร์เซ็นต์ไทล์ที่เท่าเทียมกัน ผู้สอบต้องทำข้อสอบ 120 ข้อทุกคน ส่วนการเทียบระดับคะแนนโดยวิธีราชคโหมเดล แบ่งหน่วยตัวอย่างเป็น 3 กลุ่ม ๆ ละ 400 คน แบ่งแบบทดสอบเป็น 3 ฉบับ (A, B, C) โดยฉบับ C เป็นข้อสอบรวมที่เป็นข้อสอบซ้ำกับฉบับ A 30 ข้อ ซ้ำกับฉบับ B 30 ข้อ ผู้สอบทุกคนทำข้อสอบคนละ 60 ข้อ

การเทียบระดับคะแนนโดยราชคโหมเดลทำให้เกิดแนวคิดในการสร้างธนาคารข้อสอบโดยการเพิ่มแบบทดสอบและตัวเชื่อมขยายออกไป จะทำให้เกิดแบบทดสอบหลายฉบับที่เชื่อมต่อกันและสามารถนำมาเทียบระดับคะแนนหลายรูปแบบ เช่น

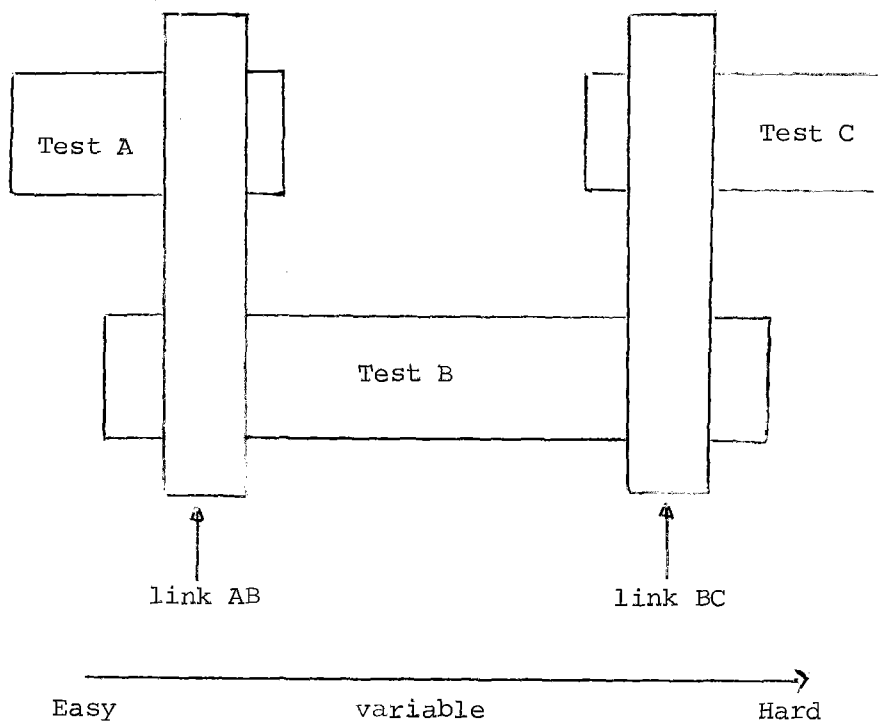
การเชื่อมต่อแบบธรรมดา



ภาพประกอบ 4 การเชื่อมต่อของแบบทดสอบ 2 ฉบับ

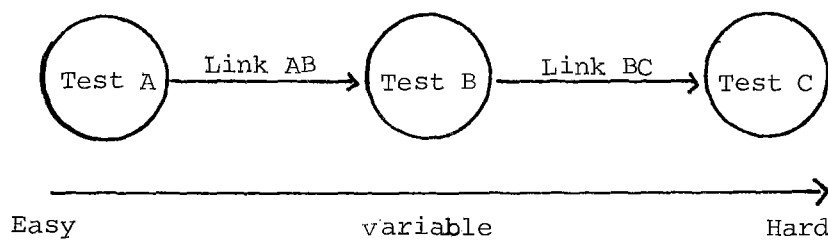
จากภาพเป็นการเชื่อมต่อของแบบทดสอบ 2 ฉบับที่มีฉบับหนึ่งง่ายอีกฉบับหนึ่งยากโดยมีข้อสอบร่วมที่ประกอบด้วยข้อยากในฉบับง่ายรวมกับข้อง่ายในฉบับยาก

การเชื่อมแบบ 2 จุดใช้



ภาพประกอบ 5 แสดงการเชื่อมแบบ 2 จุดใช้

จากภาพแสดงให้เห็นการเชื่อมต่อของแบบทดสอบ 3 ฉบับ ที่มีจุดเชื่อมต่อ 2 จุด ซึ่งสามารถแสดงให้เห็นจุดเชื่อมได้ชัดเจนดังภาพ

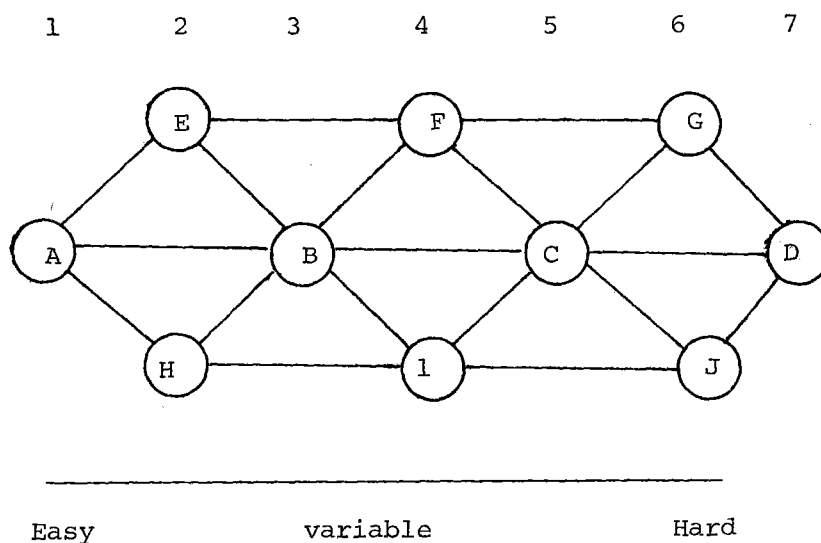


ภาพประกอบ 6 แสดงจุดเชื่อมต่อของการเชื่อมแบบ 2 จุดใช้

ในรูปแบบการเชื่อมแบบ 2 ลูกโซ่นี้ ถ้าเป็นการเทียบมาตราในเรื่องความยากของแบบทดสอบ จะเป็นการเทียบในแนวนอน (Horizontally) แต่ถ้าใช้กับกลุ่มตัวอย่างที่ต่างกันจะ เป็นการเทียบมาตราในแนวตั้ง (vertically)

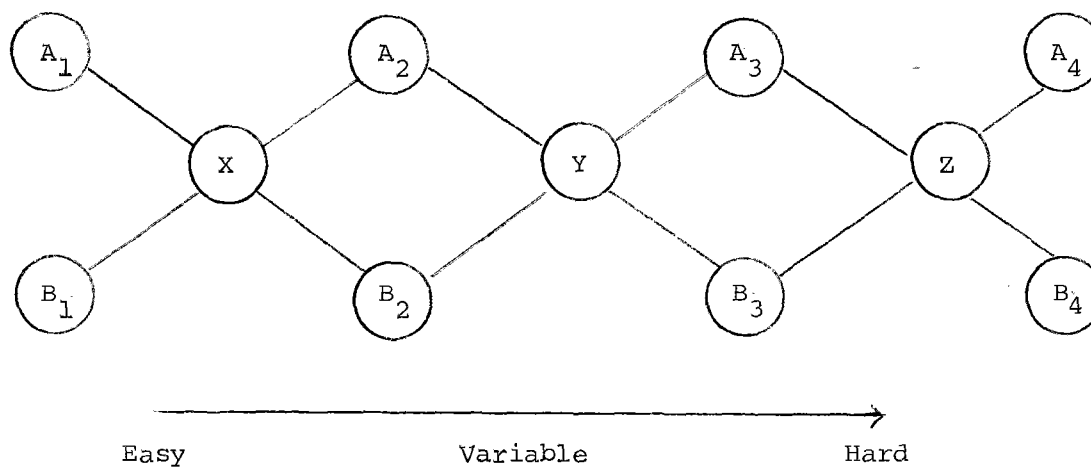
กรณีที่มีแบบทดสอบ 4 ฉบับ หรือมากกว่า การเชื่อมแบบทดสอบเหล่านี้จะทำในรูปตาข่าย เช่น มีแบบทดสอบ 10 ฉบับ ที่มีความยากแตกต่างกัน 7 ระดับ สามารถเชื่อมต่อกันเป็นตาข่ายดังภาพ

ระดับความยาก



ภาพประกอบ 7 ตาข่ายการเชื่อมของแบบทดสอบ 10 ฉบับ 7 ระดับ

จากภาพมีแบบทดสอบ 10 ฉบับ คือตั้งแต่ A ถึง J แต่ละฉบับมีข้อสอบ 60 ข้อ มีจุดเชื่อมต่อทั้งหมด 19 จุด แบบทดสอบแต่ละฉบับจะมีข้อสอบร่วมฉบับละ 10 ข้อ ดังนั้น ข้อสอบร่วมทั้งหมดเท่ากับ $19 \times 10 = 190$ ข้อ เหลือข้อสอบที่ไม่ใช่ข้อสอบร่วม $600 - 190 = 410$ ข้อ และถ้าให้นักเรียนทำแบบทดสอบฉบับละ 200 คน ต้องใช้นักเรียนทั้งหมด 2000 คน จะเห็นว่าการสร้างแบบทดสอบเชื่อมต่อดังกล่าวนี้นี้ สามารถนำไปสร้างธนาคารข้อสอบ หรือชุดของแบบทดสอบต่อไปได้อีกมากมาย ซึ่งในกรณีที่มีแบบทดสอบอยู่แล้วแต่ไม่มีข้อสอบในแต่ละฉบับ เป็นข้อสอบร่วมกันเลย การเชื่อมต่อจะอยู่ในลักษณะอีกแบบหนึ่งดังภาพต่อไปนี้



ภาพประกอบ 8 ตาข่ายการเชื่อมต่อโดยแบบทดสอบภายนอก

จากภาพแสดงการเชื่อมต่อของแบบทดสอบ 2 ชุด คือ ชุด 1 มีแบบทดสอบ 4 ฉบับ คือ A_1, A_2, A_3, A_4 ชุด 2 มีแบบทดสอบ 4 ฉบับ คือ B_1, B_2, B_3, B_4 แบบทดสอบที่เป็นตัวเชื่อม คือ แบบทดสอบ X, Y, Z

งานวิจัยที่เกี่ยวข้อง

งานวิจัยเกี่ยวกับการเทียบระดับคะแนนโดยใช้ราชคโหมเดลในต่างประเทศนี้ สเกล และลิซซิท (Skaggs Gary and Lissitz Robertw 1986 : 498-501) ได้สรุปเป็นตารางไว้ดังนี้

ตาราง 2 แสดงการนำราชคโหมเดลมาใช้ในการเทียบระดับคะแนน

ปีที่ศึกษา	ผู้ศึกษา	แบบทดสอบ	กลุ่มตัวอย่าง (จำนวน)	วิธีการ	ตัวแปรที่เทียบ ระดับคะแนน	วิธีประเมิน
1968	ไรท์	LSAT	นักเรียนกฎหมาย (980)	ราชค	แบบทดสอบฉบับง่าย กับฉบับยากและ ความสามารถของ คนกลุ่มต่ำกับกลุ่มสูง	- เปรียบเทียบความ แตกต่างมาตรฐาน ของแบบทดสอบ - ลักษณะของโค้ง - ความสัมพันธ์ b
1968	แอนดิสัน ด็ออร์เนย์ และอีเวอร์	แบบทดสอบวัด เชาว์ปัญญา	ทหาร (610,875)	ราชค	ความยากของแบบ ทดสอบจากกลุ่ม ตัวอย่าง 2 กลุ่ม	ความสัมพันธ์ b
1974	ไวทลีและ เดวิส	คำอุปมาอุปมัย	นักเรียนมัธยม (950)	ราชค	แบบทดสอบย่อย ฉบับยากและฉบับ ง่าย ข้อคู่กับข้อคี่ และแบบทดสอบที่สุ่ม	ความแตกต่าง มาตรฐานของ ความสามารถ
1975	ทินสเลย์	คำอุปมาอุปมัย	นักเรียนมัธยม, ลูกจ้างชั้นดี	ราชค	ความยากของข้อ สอบจากกลุ่ม ตัวอย่างที่ต่างกัน	ความสัมพันธ์

ปีที่ศึกษา	ผู้ศึกษา	แบบทดสอบ	กลุ่มตัวอย่าง (จำนวน)	วิธีการ	ตัวแปรที่เทียบ ระดับคะแนน	วิธีประเมิน
1977	เรนทและ มัลลฮอริ	แบบทดสอบ 7 ฉบับ	นักเรียนระดับ 4 - 6 (1300-2000)	ราชค	ธรรมชาติของลำดับ ขั้นการอ่านประเมิน เทียบมาตราการอ่าน 7 ฉบับ	ความ เบี่ยงเบน มาตรฐานของ b และความแตก ของกลุ่มตัวอย่าง ข้ามกลุ่ม ($\hat{\theta}$)
1978	สไลด์และ ลินด์	แบบทดสอบ วัดผลสัมฤทธิ์ ทาง คณิตศาสตร์	นักศึกษาปีที่ 1 (390-510)	ราชค	เทียบระดับในแนวตั้ง จากกลุ่มตัวอย่างย่อย ที่ความสามารถต่าง กัน	ความแตกต่าง มาตรฐานของ ความสามารถ
1979	สไลด์และ ลินด์	ความเข้าใจ ทางการอ่าน และคำศัพท์	นักเรียนระดับ 5 (510-570)	ราชค	เทียบระดับคะแนน แนวตั้งจากกลุ่ม ตัวอย่างย่อยที่ ความสามารถต่างกัน	ความแตกต่าง มาตรฐานของ ความสามารถ
1979	มาร์โค, พีเตอร์เซนด และสตีวาร์ท	SAT-ภาษา	นักเรียนมัธยม (ส่วนมาก) (1580)	ราชค, 3 พารามิเตอร์ อิกวิเปอร์ เซนไทล์, เส้นตรง	เทียบระดับคะแนนใน พารามิเตอร์แนวนอนกับแนวตั้ง, กลุ่มตัวอย่างที่สุ่มกับ ที่คล้ายกัน แบบทดสอบ ร่วมภายนอกกับ ภายใน ความยาก ของแบบทดสอบร่วม IRT กับอิกวิเปอร์ เซนไทล์	ค่าความคลาด- เคลื่อนกำลังสอง และดัชนีความ ลำเอียง

ตาราง 2 (ต่อ)

ปีที่ศึกษา	ผู้ศึกษา	แบบทดสอบ	กลุ่มตัวอย่าง (จำนวน)	วิธีการ	ตัวแปรที่เทียบ ระดับคะแนน	วิธีประเมิน
1980	ลอยด์และ ชูเวอร์	ITBS : ความเข้าใจ ทางคณิต- ศาสตร์	นักเรียนระดับ 6-8 (245-300)	ราชค	การเทียบระดับ คะแนนในแนวตั้ง จากตัวอย่างที่มี ระดับกัน	เปรียบเทียบ จากกราฟ
1981	กัสติย์	ITBS : ความเข้าใจ ทางการอ่าน	นักเรียนที่มีความ สามารถสูงของ ระดับ 6-8 (5500)	ราชค	การเทียบระดับ คะแนนในแนวตั้ง โดยเปรียบเทียบ ระดับความสามารถ กับระดับผลการเรียน	เปรียบเทียบโดย ดูการเปลี่ยนแปลง ของกราฟ
1981 (a)	ดิฟกี	MAT : การอ่าน	นักเรียนระดับ 6-8	ราชค	การเทียบระดับ คะแนนในแนวตั้งของ แบบทดสอบฉบับย่อย	ค่าความแตกต่าง มาตรฐานฟังก์ชัน ของความล่าเอียง คะแนนรวม
1981	โดเลน	ITED : คำศัพท์ เกี่ยวกับ ความยาว ของสระ	นักเรียนระดับ 9-12 (1580-1925)	ราชค, 2 และ 3 พารามิเตอร์ อิกวิเปอร์ เซนต์ไทล์ เส้นตรง	การเทียบระดับ คะแนนในแนวนอน กับแนวตั้ง, เปรียบเทียบโมเดล การเทียบระดับ คะแนน, คำศัพท์กับ ความยาวของสระ	cross-vali- dation ค่าเฉลี่ย กำลังสองของ ความคลาดเคลื่อน

ตาราง 2 (ต่อ)

ปีการศึกษา	ผู้ศึกษา	แบบทดสอบ	กลุ่มตัวอย่าง (จำนวน)	วิธีการ	ตัวแปรที่เทียบ ระดับคะแนน	วิธีประเมิน
1981	เพชันท	ITED- วิธีเขียน	นักเรียนระดับ 9-12 (1000)	ราชค, 2- 3 พารามิ- เตอร์, อิกวิเปอร์ เซนดไทล์	เปรียบเทียบการ เทียบระดับคะแนน ในแนวดิ่ง	ความสัมพันธ์ ระหว่างมาตรา คะแนนของผู้สอบ ด้วยจุดเริ่มต้น- เกณฑ์บนพื้นฐานวิธี เทียบระดับคะแนน แบบอิกวิเปอร์ เซนดไทล์กับ คะแนนรวม
1982	โสม	CTBS	นักเรียนระดับ3 (2400-4300)	ราชค- การเลือก ข้อสอบให้ เหมาะสมกับ โมเดล	เปรียบเทียบการ ประเมินความ สามารถและเทียบ ระดับคะแนนความ ที่ต่างกัน	ความแตกต่าง มาตรฐานไป ประยุกต์ใช้กับ cross- valrination
1983	ลอยด์	ITBS- นิยามทาง คณิตศาสตร์	นักเรียนระดับ 3-8 (250)	ราชค- การเลือก ข้อสอบให้ เหมาะสมกับ โมเดล	เปรียบเทียบวิธีการ เทียบระดับคะแนน แบบทดสอบร่วม ภายนอกกับภายใน ของการเทียบระดับ คะแนนในแนวดิ่ง	การเปลี่ยนแปลง ของฟังก์ชันจาก กราฟ

ในปี 1978 สไลด์ และลินด์ (Slinde and Line.1978 : 23-55) ได้ศึกษาเกี่ยวกับการเทียบระดับคะแนนในแนวตั้งที่วิเคราะห์โดยราล์มโมเคล กับคะแนนการสอบเข้าเรียนในมหาวิทยาลัยอิลินอยในวิชาคณิตศาสตร์จำนวน 50 ข้อ กลุ่มตัวอย่าง 1,365 คน ผู้วิจัยได้ตัดข้อ 37 - 50 ออก เพราะมีสัดส่วนของการตอบถูกของนักเรียนน้อยกว่า .20 หรือมากกว่า .80 นำข้อสอบที่เหลือ 36 ข้อ มาแบ่งเป็น 2 ฉบับ คือ ฉบับยาก กับฉบับง่าย จำนวนฉบับละ 18 ข้อ ส่วนนักเรียนที่เป็นกลุ่มตัวอย่างถูกคัดออก 58 คน เพราะได้คะแนน 0 หรือ 18 และที่เหลือผู้วิจัยได้นำมาแบ่งเป็น 3 กลุ่ม โดยใช้คะแนนจากแบบทดสอบฉบับง่ายเป็นเกณฑ์ คือผู้ได้คะแนนตั้งแต่ 14 ขึ้นไปเป็นกลุ่มสูงมีจำนวน 506 คน ผู้ได้คะแนนต่ำกว่า 10 ลงมาเป็นกลุ่มต่ำมีจำนวน 390 คน กลุ่มกลางคือผู้ได้คะแนนตั้งแต่ 11 ถึง 13 มีจำนวน 411 คน ผู้วิจัยได้นำแบบทดสอบ 2 ฉบับ และกลุ่มตัวอย่าง 3 กลุ่มมาศึกษาการเทียบระดับคะแนนได้ผลดังนี้

การเปรียบเทียบแบบทดสอบฉบับง่าย กับฉบับยาก โดยใช้กลุ่มตัวอย่างทั้งหมดพบว่าค่าเฉลี่ยของแบบทดสอบฉบับง่าย 11.975 ฉบับยาก 6.514 ซึ่งต่างกัน 5.461 แต่เมื่อเปรียบเทียบค่าเฉลี่ยของความสามารถพบว่า ค่าเฉลี่ยของแบบทดสอบฉบับง่าย 0.144 ฉบับยากมีค่า 0.069 แตกต่าง 0.045 ซึ่งแสดงว่าความสามารถของกลุ่มตัวอย่างไม่ขึ้นอยู่กับแบบทดสอบนอกจากนี้ยังพบว่าค่าเฉลี่ยและค่าความเบี่ยงเบนของความแตกต่างมาตรฐานมีค่า -0.023 และ 1.039 ซึ่งมีค่าใกล้เคียงศูนย์และหนึ่งตามความคาดหวัง

ผลการเปรียบเทียบแบบทดสอบฉบับง่าย ฉบับยาก กับกลุ่มสูง กลุ่มต่ำพบว่าในกลุ่มสูงมีค่าเฉลี่ยของแบบทดสอบฉบับง่าย 15.353 ฉบับยาก 8.642 แตกต่างกัน 6.711 ในกลุ่มต่ำค่าเฉลี่ยของแบบทดสอบฉบับง่าย 7.533 ฉบับยาก 4.258 แตกต่างกัน 3.274 เมื่อเปรียบเทียบค่าเฉลี่ยของความสามารถได้ว่า ค่าเฉลี่ยความสามารถของกลุ่มสูงในแบบทดสอบฉบับยาก 0.981 ฉบับง่าย 0.921 แตกต่างกัน -0.060 กลุ่มต่ำในแบบทดสอบฉบับง่าย -0.870 ฉบับยาก -0.865 แตกต่างกัน -0.005 ซึ่งแสดงว่าค่าเฉลี่ยของความสามารถนี้สอดคล้องกันน้อยกว่าการเปรียบเทียบจากตัวอย่างทั้งหมดแต่ก็ยังเป็นค่าที่มีประสิทธิภาพอยู่

เมื่อนำค่าความสามารถของกลุ่มสูงกับกลุ่มต่ำมาเทียบระดับคะแนนโดยการพล็อตกราฟโดยให้แกนนอน เป็นความสามารถของกลุ่มสูงแกนตั้ง เป็นความสามารถของกลุ่มต่ำพบว่าถ้าใช้เส้นตรงที่ทำมุม 45 องศา กับแกนนอนหรือเส้นตรงที่มีค่าความชันเป็น 1 มาเป็นหลักในการเปรียบเทียบจะได้ว่าแบบทดสอบฉบับง่ายจะมีแนว เป็นเส้นตรงอยู่เหนือเส้นหลัก แบบทดสอบฉบับยากก็มีแนว เป็นเส้นตรง

เหมือนกันแต่อยู่ต่างเส้นหลัก แต่ทั้งหมดก็ยังเป็นเส้นตรงที่เกือบจะขนานกับเส้นหลักนอกจากบางจุดที่แบบทดสอบฉบับง่ายให้ผลสูงกว่าการประเมินหรือแบบทดสอบฉบับยากให้ค่าต่ำกว่าการประเมิน ซึ่งความแตกต่างระหว่างเส้นหลักจะประมาณ 0.5 หรือ 0.6 หน่วย ซึ่งเป็นการแสดงให้เห็นว่าความสามารถของแต่ละบุคคลไม่ขึ้นอยู่กับกลุ่มซึ่งเป็นคุณสมบัติที่เหมาะสมในการจะนำมาใช้ในการเทียบระดับคะแนน ส่วนการเปรียบเทียบแบบทดสอบฉบับยาก ฉบับง่าย กับกลุ่มที่ความสามารถปานกลางโดยประเมินจากกลุ่มที่มีความสามารถสูง กับกลุ่มที่มีความสามารถต่ำ พบว่าผลที่ได้ยังไม่เป็นที่พอใจมากนัก สไลด์และลีนด์ ได้เสนอแนะว่า การเทียบระดับคะแนนในแนวตั้งโดยใช้ราชคโหมเดล ควรใช้วิธีที่มีแบบทดสอบร่วมและควรเลือกข้อสอบที่เหมาะสมกับโมเดล

สไลด์และลีนด์ (Slide and Line, 1979 : 159-165) ได้ศึกษาการเทียบระดับคะแนนในแนวตั้งซ้ำอีกครั้ง โดยใช้แบบทดสอบความเข้าใจในการอ่านและคำศัพท์ (SRA reading comprehension and vocabulary) ที่มีความยากแตกต่างกันจำนวน 60 ข้อ โดยการแบ่งเป็น 2 ฉบับ คือ ฉบับยาก 30 ข้อ เป็นข้อที่มีค่าเฉลี่ยสัดส่วนของการตอบถูก .41 ฉบับง่าย 30 ข้อ เป็นข้อที่มีค่าเฉลี่ยสัดส่วนของการตอบถูก .65 ข้อสอบทั้งหมดผ่านการวิเคราะห์โดย Factor analysis กลุ่มตัวอย่างที่ใช้ศึกษามีจำนวน 1638 คน การวิเคราะห์โดยใช้คอมพิวเตอร์โปรแกรม BICAL ได้คัดกลุ่มตัวอย่างที่ได้ศูนย์หรือได้คะแนนเต็มออก 58 คน ที่เหลือนำมาแบ่งเป็น 3 กลุ่ม เป็นกลุ่มสูง 506 คน กลุ่มกลาง 509 คน และกลุ่มต่ำ 565 คน นำมาเปรียบเทียบกันได้ผลดังนี้

เมื่อเปรียบเทียบแบบทดสอบฉบับยาก ฉบับง่ายกับกลุ่มตัวอย่างทั้ง 3 กลุ่มพบว่าค่าเฉลี่ยของความสามารถและของแต่ละกลุ่มในแบบทดสอบแต่ละฉบับมีค่าใกล้เคียงกัน คือกลุ่มสูงมีค่าความสามารถเฉลี่ยของแบบทดสอบฉบับยาก 1.583 ฉบับง่าย 1.644 กลุ่มกลางมีค่าความสามารถเฉลี่ยในแบบทดสอบฉบับยาก -.766 ฉบับง่าย -.782 ซึ่งสอดคล้องกับเหตุผลที่ว่า ข้อสอบต่างก็มีอิสระในการวัดคือค่าความยากของแบบทดสอบไม่ได้ขึ้นอยู่กับกลุ่มตัวอย่าง ซึ่งเป็นคำตอบที่สนับสนุนในการนำเอาราชคโหมเดลมาใช้ในการเทียบระดับคะแนนในแนวตั้ง เมื่อประเมินความสามารถจากกลุ่มหนึ่งไปใช้อีกกลุ่มหนึ่ง เช่น กลุ่มกลางไปใช้กับกลุ่มสูง กลุ่มสูงไปใช้กับกลุ่มต่ำ ผลที่ได้พบว่าค่าความสามารถเฉลี่ยเท่าเทียมกัน 2 กลุ่มเท่านั้น คือกลุ่มสูงไปใช้กับกลุ่มกลาง กับกลุ่มกลางไปใช้กับกลุ่มสูง นอกนั้นมีค่าต่างกัน และเมื่อเทียบความสัมพันธ์ระหว่างความยาก และอำนาจจำแนกระหว่างกลุ่มสูง กลุ่มกลางกลุ่มต่ำได้

-.13, -.38, -.68 ตามลำดับ ซึ่งการมีความสัมพันธ์กันนี้เป็นปัญหาในการนำการวิเคราะห์แบบ
 ราชคโหมเดลมาใช้เปรียบเทียบกับระดับคะแนนในแนวดิ่ง ซึ่งการจะให้ผลการเทียบระดับคะแนนที่ดี
 จะต้องไม่มีความสัมพันธ์กัน

กัฟตราสัน (Gustafsson. 1979 : 153-158) ได้ศึกษาการเทียบระดับคะแนนในแนวดิ่ง
 โดยราชคโหมเดลของสโลดและลินด์ โดยนำมาศึกษาใหม่กับกลุ่มตัวอย่าง 1,300 คน ซึ่งมีการกระจาย
 เป็นปกติมีค่าเฉลี่ยเป็นศูนย์ ค่าความเบี่ยงเบนมาตรฐานเป็นหนึ่ง แบบทดสอบที่ใช้มี 36 ข้อ มีค่าความ
 น่าจะเป็นของการตอบถูกเท่ากับ -1.5, -.5, .5 และ 1.5 นำแบบทดสอบมาแบ่งเป็น 2 ฉบับ
 คือ ฉบับง่าย 18 ข้อ ฉบับยาก 18 ข้อ แบ่งกลุ่มตัวอย่างเป็น 3 กลุ่ม โดยใช้คะแนนจากแบบทดสอบ
 ฉบับง่าย คือคะแนนตั้งแต่ 15 ขึ้นไปเป็นกลุ่มสูง (จำนวน 443 คน) ตั้งแต่ 11 คะแนนลงมาเป็นกลุ่ม
 ต่ำ (จำนวน 440 คน) นอกนั้นเป็นกลุ่มกลาง (จำนวน 417 คน) ใช้คอมพิวเตอร์คำนวณค่าพารามิเตอร์
 จากเงื่อนไข Maximum likelihood (CML) ซึ่งเป็นการวิเคราะห์ตามวิธีราชคโหมเดล
 ได้ความยากของข้อสอบจากกลุ่มตัวอย่างทั้งหมด และตามกลุ่มย่อย (กลาง, ต่ำ, สูง) นำค่าความยาก
 มาเปรียบเทียบกัน พบว่า กลุ่มต่ำประเมินค่าความยากของแบบทดสอบฉบับยากสูงกว่ากลุ่มรวมและ
 ประเมินความยากของแบบทดสอบฉบับง่ายต่ำกว่ากลุ่มรวม ในกลุ่มสูงจะประเมินตรงข้ามกันซึ่งกัฟตราสัน
 ได้สรุปว่า เป็นสิ่งรบกวนเพียงเล็กน้อย และยังได้แนะนำให้ใช้เงื่อนไข maximum likelihood
 (CML) ในการวิเคราะห์ว่า เป็นสิ่งที่เหมาะสมที่สุด การเทียบระดับคะแนนในแนวดิ่ง โดยใช้
 ราชคโหมเดล ควรเป็นข้อสอบชนิดเลือกตอบและควรใช้กับกลุ่มที่มีการเดาต่ำ

ลอร์ด และฮูเวอร์ (Loyd and Hoover. 1980 : 179-193) ได้ศึกษาเกี่ยวกับการเทียบ
 ระดับคะแนนในแนวดิ่งที่วิเคราะห์ตามราชคโหมเดล กลุ่มตัวอย่างที่ใช้ศึกษาเป็นนักเรียน 3 ระดับ คือ
 ระดับ 6, 7, 8, ในรัฐไอโอวา จำนวน 1,956 คน โดยในแต่ละระดับได้หาค่ามัธยฐานของแต่ละระดับ
 ได้ว่า ระดับ 6 มีมัธยฐานที่ 12, 13 ระดับ 7 มีมัธยฐานที่ 12, 13, 14 ระดับ 8 มีมัธยฐานที่ 12, 13,
 14 แบบทดสอบที่ศึกษาเป็นแบบทดสอบวัดทักษะเบื้องต้นทางคณิตศาสตร์ การวิเคราะห์ใช้โปรแกรม
 LOGIST โดยใช้ one-parameter แล้วนำคะแนนดิบมาเทียบระดับคะแนนโดยใช้ความสามารถในการ
 ทำข้อสอบรวมเป็นตัวเชื่อมโดยแบ่งการเทียบระดับคะแนนใน 2 ลักษณะ คือ 1) การเทียบระดับคะแนน
 ในระดับที่ติดกัน (ระดับ 12 กับระดับ 13, ระดับ 13 กับระดับ 14) มีข้อสอบรวมภายใน 30 ข้อ 2)
 การเทียบระดับคะแนนในระดับที่ไม่ติดกัน (ระดับ 12 กับระดับ 14) มีข้อสอบรวม 15 ข้อ ได้ค่าความ

ยากเฉลี่ยมีค่าตั้งแต่ -1.65 ถึง $.23$ ค่าเบี่ยงเบนมาตรฐานตั้งแต่ $.86$ ถึง 1.62 ค่าความยากเฉลี่ยของข้อสอบรวม 30 ข้อ มีค่าตั้งแต่ -1.33 ถึง $.33$ ค่าเบี่ยงเบนมาตรฐานมีค่าตั้งแต่ $.63$ ถึง 1.30 ความยากเฉลี่ยของแบบทดสอบรวม 15 ข้อ มีค่าตั้งแต่ -1.42 ถึง $-.44$ ค่าเบี่ยงเบนมาตรฐานมีค่าตั้งแต่ $.50$ ถึง 1.61 ผู้วิจัยได้นำคะแนนดิบที่เทียบจากความสามารถที่เท่ากันมาสร้างกราฟเทียบคะแนนของระดับ 7 เทียบระหว่างมัธยมศึกษา 13 กับ 13 ระดับ 8 เทียบระหว่างมัธยมศึกษา 13 กับ 14 ระดับ 6 เทียบมัธยมศึกษา 12 กับ 13 ระดับ 7 เทียบมาตรฐาน 12 กับ 13 ระดับ 8 เทียบมัธยมศึกษา 12 กับ 13 ส่วนการเทียบระดับคะแนนระดับที่ไม่ติดกันสร้างกราฟเทียบระดับคะแนน ระดับ 7 เทียบมัธยมศึกษา 12 กับ 14 ระดับ 8 เทียบมัธยมศึกษา 12 กับ 14 ระดับ 7 เทียบมัธยมศึกษา 12, 13, 14 ระดับ 8 เทียบมัธยมศึกษา 12, 13, 14 และเทียบข้ามระดับ คือระดับ 6 มัธยมศึกษา 12 กับ ระดับ 7 ที่มัธยมศึกษา 12 ระดับ 8 ที่มัธยมศึกษา 12 แล้วได้สร้างตารางเปรียบเทียบคะแนนดิบในแต่ละระดับ

โคเลน (Kolen, 1981 : 1-11) ได้ศึกษารูปแบบของการเทียบระดับคะแนนระหว่างรูปแบบดั้งเดิมสองวิธีคือ รูปแบบเชิงเส้นตรงและรูปแบบอิกวิเพอร์เซนไทล์ กับรูปแบบการเทียบระดับคะแนนโดยใช้โค้งลักษณะข้อสอบที่ศึกษาทั้ง 1, 2, 3 พารามิเตอร์ แต่ละวิธีเทียบด้วยค่าประมาณจากคะแนนจริง (Estimated true score equating) และเทียบด้วยการประมาณจากค่าสังเกต (Estimated observed score equating) ใช้ข้อมูลจากโครงการ IOWA Test of Education Development (ITED) ค.ศ. 1978 ศึกษา 2 รูปแบบ คือ เทียบจากฟอร์มที่มีความยากเท่าเทียมกันและเทียบจากฟอร์มที่มีความยากต่างกัน กลุ่มตัวอย่างเป็นนักเรียนระดับที่ 1 คือระดับเกรด 9 และ 10 ระดับ 2 คือระดับเกรด 11 และ 12 จำนวนทั้งหมด 10,728 คน จากโรงเรียนในรัฐไอโอว่า 34 แห่ง การศึกษาที่ใช้เกณฑ์จาก Cross-validation criterion ผลการวิจัยพบว่า การเทียบระดับคะแนนทั้งหมดแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ $.01$ และ $.05$ การเทียบระดับคะแนนโดยใช้วิธีอิกวิเพอร์เซนไทล์เหมาะสมที่สุดกับแบบทดสอบที่มีความยากต่างกัน ส่วนการเทียบระดับคะแนนโดยใช้โค้งลักษณะข้อสอบพบว่า กรณิ์ใช้พารามิเตอร์ตัวเดียวยังเป็นวิธีที่ไม่ค่อยถูกต้องสำหรับแบบทดสอบที่ต่างกัน ซึ่งเนื่องจากอาจมีการเดาเกิดขึ้นมาก ส่วนกรณิ์ใช้พารามิเตอร์ 3 ตัว มีปัญหาที่เกิดขึ้นอย่างหลีกเลี่ยงไม่ได้คือ การกะประมาณค่าไม่ถูกต้องของ Lower asymptote parameter

โคเลน และ ไวน์เนย์ (Kolen and Whitnuy, 1982 : 279-292) ได้เปรียบเทียบความถูกต้องของการเทียบระดับคะแนนโดยใช้แบบทดสอบรวม 4 วิธี คือ วิธีอิกวิเพอร์เซนไทล์ เส้นตรง

พารามิเตอร์ตัวเดียว พารามิเตอร์ 3 ตัว ใช้แบบสอบ General Educational Development (GED) ซึ่งเป็นแบบทดสอบวัดผลสัมฤทธิ์เพื่อใช้ในการตัดสินใจให้ประกาศนียบัตรแก่ผู้ที่ต้องการเทียบความรู้ระดับเตรียมอุดมศึกษาทั่วประเทศ แบบทดสอบมี 12 ชุด โดยมีฟอร์มที่ 12 เป็นแบบสอบร่วมที่เหลืออีก 11 ฟอร์มเป็นแบบทดสอบที่ใช้เทียบระดับคะแนนผู้เข้าสอบแต่ละคนต้องทำแบบทดสอบ 2 ฟอร์ม ในจำนวน 11 ฟอร์ม ส่วนฟอร์มที่ 12 ต้องทำทุกคน กลุ่มตัวอย่างมาจากการสอบ ปี 1980 จำนวนมากกว่า 800,000 คน ทำการสุ่มแบบแบ่งชั้นภูมิหลายชั้นจากประเภทของโรงเรียน เขตทางภูมิศาสตร์ สถานภาพทางสังคม สุดท้ายได้นักเรียนที่เป็นตัวอย่างโรงเรียนละ 22 คน จำนวนคนในแต่ละชุดของแบบทดสอบประมาณ 200 คน ประมาณค่าพารามิเตอร์โดยใช้โปรแกรม LOGIST แล้วสร้างคะแนนสมมูลระหว่างแบบสอบอีก 11 ชุด การเทียบผลการวิเคราะห์โดยใช้ Cross-Validation ผลการศึกษาพบว่า วิธีการเทียบระดับคะแนนโดยใช้อิควิเปอร์เซนไทล์และการใช้โด่งลักษณะข้อสอบด้วย 3 พารามิเตอร์ ได้ผลไม่เป็นที่ยอมรับ ขณะเดียวกันวิธีการเทียบระดับคะแนนโดยใช้พารามิเตอร์ตัวเดียวให้ผลเทียบเท่าเพียงพอกับรูปแบบ 3 พารามิเตอร์ ในการประเมินผลการศึกษาค้างนี้มีการเปรียบเทียบอื่น ๆ ซึ่งสรุปได้ว่า ความถูกต้องของวิธีการเทียบระดับคะแนนขึ้นอยู่กับองค์ประกอบหลายอย่าง เช่น ลักษณะของแบบทดสอบ รูปแบบของการเทียบ ขนาดของกลุ่มตัวอย่าง เป็นต้น

โฮมส์ (Holmes, 1982 : 139-147) ได้ศึกษาความเหมาะสมของการเทียบระดับคะแนนในแนวทางโดยใช้ราชคโมเดล กับกลุ่มตัวอย่างที่เป็นนักเรียนเกรด 3 จำนวน 5,310 คน เกรด 4 จำนวน 3,393 คน แบบทดสอบที่ใช้เป็นชุด Comprehensive tests of basic skills (1973) ฟอร์ม S จำนวน 183 ข้อ ซึ่งได้รับการวิเคราะห์ว่าเป็นมิติเดียวกัน ขึ้นแรกจากจุดประสงค์เดียวกัน แล้วนำมาวิเคราะห์โดย Factor analysis เลือกข้อที่มีค่า Load สูงกว่า 3.00 ได้ข้อสอบที่แสดงว่าเป็นมิติเดียวกัน 32 ข้อ นำมาคำนวณหาค่าพารามิเตอร์โดยใช้โปรแกรม BICAL แยกข้อสอบออกเป็น 3 ฉบับ ฉบับง่ายมีค่าความยากตั้งแต่ -0.48 ถึง 48 ฉบับ ปานกลางมี 8 ข้อ มีระดับความยากตั้งแต่ 1.25 ถึง .73 ฉบับยากมี 12 ข้อ ระดับความยากตั้งแต่ 0.53 ถึง 1.27 แบ่งกลุ่มตัวอย่างมาระดับละ 1,000 คน เพื่อตรวจสอบโดย Cross Validation ผลของการเทียบระดับคะแนนพบว่า การเทียบระดับคะแนนโดยใช้ราชคโมเดลยังมีผลที่ไม่น่าพอใจ ซึ่งสาเหตุอาจเนื่องจากกลุ่มตัวอย่างที่ใช้เมื่อประเมินความสามารถแล้วพบว่ามีความสามารถสูงกว่าแบบทดสอบ ซึ่งการแก้ไขก็ต้องพยายามเลือกแบบทดสอบให้เหมาะกับโมเดล

ฟิลลิปส์ (Phillip 1986 : 107-118) ได้ศึกษาเกี่ยวกับระดับคะแนนในแนวตั้งที่วิเคราะห์ โดยราชคโมเดล เพื่อหาคำตอบที่ว่า ถ้านำสิ่งที่ไม่สอดคล้องกับโมเดลออกไปจะมีส่วนทำให้เกิด 1) การปรับปรุงข้อมูลให้เหมาะสมกับโมเดลขึ้น 2) มีอิทธิพลต่อการเทียบระดับคะแนนของแบบทดสอบ 2 ฉบับ 3) เปลี่ยนผลของการเทียบระดับคะแนน ผู้ศึกษาได้ใช้แบบทดสอบการอ่านและคณิตศาสตร์ทดสอบกับกลุ่มตัวอย่างที่เป็นนักเรียนเกรด 4 และเกรด 5 จำนวน 400-500 คน ตามลำดับ ดังนั้น 1) ใช้ โปรแกรม BICAL หาค่าพารามิเตอร์ และ เปรียบเทียบค่าก่อนและหลังเมื่อนำสิ่งที่ไม่เหมาะสมกับโมเดล ออก 2) เปรียบเทียบค่าเฉลี่ยค่าความ เบี่ยงเบนมาตรฐาน ค่าความยาก ค่าอำนาจจำแนก และค่า ความเชื่อมั่นของแบบทดสอบโดยใช้สูตร KR_{20} 3) เปรียบเทียบกับการเทียบระดับคะแนนแบบอิควิเปอร์ เซนไทล์ ซึ่งผลการศึกษาพบว่า การนำสิ่งที่ไม่เหมาะสมกับโมเดลออกไปไม่ได้มีส่วนทำให้ค่าจากการประเมิน การเทียบระดับคะแนนเปลี่ยนแปลง

งานวิจัยในประเทศ

งานวิจัยที่เกี่ยวกับคุณภาพของพารามิเตอร์ตัวเดียว

ปณิตดา วิทยวัฒน์ (ปณิตดา วิทยวัฒน์ 2528 : 62-71) ได้ศึกษาผลการวิเคราะห์ข้อสอบ โดยแบบเดิมกับวิธีโลจิสติกโมเดลทั้ง 3 พารามิเตอร์ คือ พารามิเตอร์ตัวเดียว พารามิเตอร์ 2 ตัว และพารามิเตอร์ 3 ตัว กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2527 ของจังหวัด เชียงใหม่ จำนวน 1,687 คน ที่สอบแบบทดสอบคณิตศาสตร์ (ค101) ชนิดเลือกตอบ 5 ตัว จำนวน 50 ข้อ ผู้วิจัยได้วิเคราะห์ข้อมูลโดยใช้โปรแกรม ซีไอเอ (Classical item analysis) โปรแกรม BICAL โปรแกรม LOGIST แล้วนำผลการวิเคราะห์มาเปรียบเทียบกันได้ว่า จำนวนข้อที่ได้รับการ คัดเลือกไว้จากวิธีวิเคราะห์แบบเดิม กับวิธีโลจิสติกโมเดลทั้ง 3 พารามิเตอร์ต่างกัน แต่จำนวนข้อ ที่วิเคราะห์จากพารามิเตอร์ตัวเดียว สองตัว สามตัว ไม่แตกต่างกัน ค่าความเชื่อมั่นของพารามิเตอร์ ตัวเดียว สองตัว สามตัว เมื่อปรับจำนวนข้อให้เท่ากันแล้วมีค่าไม่แตกต่างกัน และ เมื่อปรับจำนวนข้อ ให้เท่ากันค่าความ เชื่อมั่นของการวิเคราะห์แบบเดิมกับวิธีวิเคราะห์แบบโลจิสติกไม่ต่างกัน ถ้าพิจารณา ความสอดคล้องของข้อสอบจะพบว่า การคัดเลือกข้อสอบจากการวิเคราะห์พารามิเตอร์ตัวเดียวจะมีความ สอดคล้องกับการวิเคราะห์ข้อสอบด้วยวิธีอื่นอย่างไม่มีนัยสำคัญที่ระดับ .05

งานวิจัยที่เกี่ยวกับการเทียบระดับคะแนน

ชูชีพ พงษ์สมบุญ (ชูชีพ พงษ์สมบุญ 2528 : 112-118) ได้ศึกษาเปรียบเทียบรูปแบบการเทียบระดับคะแนน 3 วิธี คือ การเทียบเชิงเส้นตรง การเทียบแบบอิกวิเปอร์เซนไทล์ และการเทียบโดยใช้โค้งลักษณะข้อสอบ จากรูปแบบที่ใช้แบบทดสอบร่วมกับรูปแบบที่ใช้ผู้สอบร่วม เพื่อต้องการเปรียบเทียบประสิทธิภาพและความคงที่ของการเทียบระดับคะแนนทั้ง 3 วิธี ใช้กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ทั่วประเทศสุ่มมาจำนวน 3,721 คน ที่เข้าสอบโครงการตรวจคุณภาพการศึกษาของกรมวิชาการ กระทรวงศึกษาธิการ ในจำนวน 60 ข้อ แบ่งเป็น 2 ฉบับ ฉบับละ 38 ข้อ มีแบบสอบร่วมชนิดแบบสอบร่วมภายใน จำนวน 20 ข้อ ผลการเปรียบเทียบได้ว่า ประสิทธิภาพของการเทียบระดับคะแนนระหว่างรูปแบบทดสอบร่วมกับรูปแบบที่ใช้ผู้สอบร่วมในแต่ละวิธีประสิทธิภาพไม่แตกต่างกัน ความคงที่ของวิธีการเทียบระดับคะแนนโดยรูปแบบที่ใช้ผู้สอบร่วมได้ผลว่า วิธีการเทียบโดยใช้อิกวิเปอร์เซนไทล์มีความคงที่มากกว่าการเทียบเชิงเส้นตรงแต่มีความคงที่พอ ๆ กับการเทียบโดยใช้โค้งลักษณะข้อสอบ ความคงที่ของวิธีการเทียบระดับคะแนนในรูปแบบที่ใช้แบบสอบร่วม ได้ผลเหมือนกับรูปแบบที่ใช้ผู้สอบร่วม

ภาวิณี ศรีสุขวัฒนานันท์ (ภาวิณี ศรีสุขวัฒนานันท์ 2529 : 155-170) ได้เปรียบเทียบผลของการใช้รูปแบบการเทียบระดับคะแนนต่างกัน 3 รูปแบบ คือ รูปแบบอิกวิเปอร์เซนไทล์ รูปแบบเชิงเส้นตรง รูปแบบอิงทฤษฎีการตอบข้อสอบแบบพารามิเตอร์ โดยใช้แบบสอบร่วมภายในต่างกัน 3 ขนาด คือ ขนาดร้อยละ 20 (7 ข้อ) ร้อยละ 40 (14 ข้อ) ร้อยละ 60 (21 ข้อ) โดยใช้ข้อสอบทั้งหมด 35 ข้อ การประเมินผลของการเทียบระดับคะแนน 2 ลักษณะ คือ 1) ค่าความคลาดเคลื่อนมาตรฐานของการเทียบระดับคะแนน (SEE) ซึ่งเป็นการวิเคราะห์ในเชิงทฤษฎีของรูปแบบ 2) ดัชนีการเปรียบเทียบความแตกต่าง (Index C) ที่ได้จากการวิเคราะห์กลุ่มตัวอย่างสอบทานผล ซึ่งเป็นการวิเคราะห์ความเพียงพอของวิธีการเทียบระดับคะแนนโดยตรงกับผลการเทียบสุดท้าย ส่วนประชากรและกลุ่มตัวอย่างแยกเป็น 2 กรณี 1) กรณีสอบเลือก ประชากรคือผลการสอบรายบุคคล ผู้มีคุณสมบัติขั้นต่ำตามกำหนดระเบียบการสอบคัดเลือกที่กำหนดไว้ 15,875 สุ่มตัวอย่างสำหรับการเทียบระดับคะแนน 2 กลุ่ม กลุ่มละ 1,500 คน สำหรับการสอบทานผลกลุ่มละ 1,500 คน 2) กรณีแบบสอบวัดผลสัมฤทธิ์ ประชากร คือ ผลการสอบรายบุคคลในวิชาที่กำหนดในระดับอุดมศึกษา จำนวน 12,383 คน สุ่มตัวอย่างเหมือนกรณีแรก แบบทดสอบมีกรณีละ 2 ชุด โดยมีชุดละ 35 ข้อ มีข้อสอบร่วม 21 ข้อ,

14 ข้อ และ 7 ข้อ โดยยึดหลักแบบทดสอบร่วมมีค่าความเชื่อมั่นเท่าเทียมกัน ในแต่ละคู่ใช้รูปแบบการเทียบระดับคะแนน 3 รูปแบบ สร้างตารางแปลงคะแนนสมมูลจากแบบทดสอบชุดที่ 1 ไปยังแบบทดสอบชุดที่ 2 แล้ววิเคราะห์ความคลาดเคลื่อนมาตรฐานของการเทียบระดับคะแนน ค่าดัชนี ประสิทธิภาพสัมพันธภาพ การวิเคราะห์กลุ่มสอบทานผล โดยคำนวณค่าดัชนีเปรียบเทียบความแตกต่าง (C) ผลการศึกษาพบว่า การเทียบระดับคะแนน 9 รูปแบบ ได้ผลเป็นที่ยอมรับว่า ความยาวของแบบทดสอบร่วมมีผลต่อความแม่นยำ และความเพียงพอของการเทียบระดับคะแนน แต่ในกรณีของแบบสอบคัดเลือกรูปแบบไม่ชัดเจนในด้านมิติเกี่ยวกับรูปแบบการเทียบระดับคะแนนพบว่า กรณีสอบคัดเลือกรูปแบบการเทียบระดับคะแนนที่ได้ผลดีตามลำดับคือ รูปแบบอิกวิเปอร์เซนไทล์ รูปแบบทฤษฎีการตอบข้อสอบ และเชิงเส้นตรง กรณีการสอบวัดผลสัมฤทธิ์ รูปแบบที่ให้ผลดีคือ รูปแบบเชิงเส้นตรง รูปแบบอิกวิเปอร์เซนไทล์ และรูปแบบทฤษฎีการตอบข้อสอบ

เรวดี อินทสระ (เรวดี อินทสระ 2530 : 64-69) ได้เปรียบเทียบความคลาดเคลื่อนของการเทียบระดับคะแนนและความเพียงพอจากการใช้รูปแบบอิงทฤษฎีการตอบข้อสอบกับรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ จากกลุ่มตัวอย่างจำนวน 2,823 คน ใช้แบบทดสอบวัดผลสัมฤทธิ์ทางคณิตศาสตร์ จำนวน 150 ข้อ วิเคราะห์โดยใช้โปรแกรม LOGIST 5 นำค่าพารามิเตอร์แยกเป็น 2 ฉบับ ฉบับละ 45 ข้อ ทั้งสองมีข้อสอบร่วม 15 ข้อ ไปทดสอบกับกลุ่มตัวอย่างแล้วนำมาเทียบระดับคะแนนและทานผลการเทียบระดับคะแนนได้ผลว่า เมื่อเทียบระดับคะแนนจากฉบับที่ 2 ไปฉบับที่ 1 เปรียบเทียบความสมมูลของคะแนนทั้งสองได้ว่า ช่วงคะแนน 1-18 รูปแบบอิงทฤษฎีการตอบข้อสอบมีคะแนนสมมูลสูงกว่ารูปแบบเทคนิคการวิเคราะห์องค์ประกอบ ช่วงคะแนน 19-23 ทั้งสองรูปแบบมีคะแนนสมมูลใกล้เคียงกันช่วงคะแนน 24-45 รูปแบบเทคนิคการวิเคราะห์องค์ประกอบมีคะแนนสมมูลสูงกว่ารูปแบบอิงทฤษฎีการตอบข้อสอบ เมื่อตรวจสอบผลได้ว่า การทานผลของการเทียบระดับคะแนนแบบอิงทฤษฎีการตอบข้อสอบมีดัชนีความเพียงพออยู่ในระดับความพอใจ ส่วนรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบดัชนีความเพียงพออยู่ในระดับปานกลาง และรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบมีส่วนความคลาดเคลื่อนมาตรฐานทั้งฉบับสูงกว่ารูปแบบอิงทฤษฎีการตอบข้อคำถาม

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องทั้งในประเทศไทย และต่างประเทศแล้ว จะเห็นได้ว่า รูปแบบการสร้างแบบทดสอบสำหรับเทียบระดับคะแนนจะให้มีความสมบัติตามที่ต้องการขึ้นอยู่กับ

กับปัจจัยหลายประการคือ มีประสิทธิภาพในเรื่องความถูกต้อง ความสะดวกในการเก็บข้อมูล ขนาดของกลุ่มตัวอย่าง ค่าใช้จ่าย ประโยชน์ที่ได้รับ และที่สำคัญคือความเป็นไปได้ในการนำไปปฏิบัติจริง ดังนั้น ทฤษฎีที่จะวิเคราะห์ข้อสอบ เพื่อสร้างแบบทดสอบสำหรับ เทียบระดับคะแนนที่เหมาะสมมากที่สุด คือการวิเคราะห์โดยใช้ราชคโหมเดล ซึ่งปัจจุบันนี้ยังไม่เป็นที่แพร่หลายในประเทศไทย ในอนาคตการวัดผล การศึกษาคงจะนำวิธีการวิเคราะห์แบบราชคโหมเดลมาใช้กันอย่างแพร่หลายในสภาพการเรียนการสอน ปัจจุบัน เพราะสามารถใช้ได้ทั้งในโรงเรียนเล็ก ๆ โรงเรียนใหญ่ ๆ หรือกลุ่มโรงเรียน และนอกจากนี้ยังสามารถวิเคราะห์ได้โดยไม่ต้องอาศัย เครื่องคอมพิวเตอร์ช่วย หรือใช้วิเคราะห์ด้วย เครื่องคอมพิวเตอร์ ขนาดไมโครได้ด้วย

บทที่ 3

วิธีดำเนินการศึกษาค้นคว้า

ประชากร

ประชากรที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นนักเรียนในระดับชั้นมัธยมศึกษาตอนต้น ในภาคเรียนที่ 2 ปีการศึกษา 2530 ของโรงเรียนสังกัดกรมสามัญศึกษา เขตท้องที่การศึกษา 2 กรุงเทพมหานคร จำนวน 14 โรงเรียน โดยแยกตามลำดับชั้นได้ดังนี้

ชั้น	จำนวนห้องเรียน	จำนวนนักเรียน (คน)
ชั้นมัธยมศึกษาปีที่ 1	164	8,208
ชั้นมัธยมศึกษาปีที่ 2	166	8,376
ชั้นมัธยมศึกษาปีที่ 3	165	8,072
รวม	495	24,656

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นนักเรียนชั้นมัธยมศึกษาตอนต้น ภาคเรียนที่ 2 ปีการศึกษา 2530 ของโรงเรียนในสังกัดกรมสามัญศึกษา เขตท้องที่การศึกษา 2 กรุงเทพมหานคร ซึ่งเลือกมาโดยการสุ่มแบบแบ่งชั้น (Stratified random sampling) มีขนาดของโรงเรียนเป็นชั้น (Strat) คือขนาดใหญ่พิเศษ ขนาดใหญ่ และมีห้องเรียนเป็นหน่วยการสุ่ม (Sampling unit) ซึ่งมีขั้นตอนการสุ่มดังนี้

1. ประมวลกลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้ โดยการแบ่งกลุ่มตัวอย่างเป็น 2 กลุ่ม เป็นกลุ่มที่ใช้คำนวณค่าพารามิเตอร์และกลุ่มที่ใช้เทียบระดับคะแนน กลุ่มละ 1,200 คน แบ่งย่อยเป็น

ระดับชั้นละ 400 คน (Rackase.1977 อ้างจาก Skaggs and Lissity.1986 : 516, Wright.1977 : 107, Whitelly.1977 : 229-230, Phillips.1986 : 109)

2. สุ่มโรงเรียนขนาดใหญ่พิเศษ ขนาดใหญ่ มาประมาณ 60 เปอร์เซ็นต์ โดยวิธีสุ่มอย่างง่าย (Simple random sampling) ได้จำนวน 8 โรงเรียน

3. สุ่มห้องเรียนจากโรงเรียนที่สุ่มในขั้นที่ 2 ซึ่งในแต่ละโรงเรียนจะมีนักเรียนประมาณ ห้องละ 50 คน โดยวิธีสุ่มอย่างง่ายได้กลุ่มตัวอย่างดังตารางต่อไปนี้

ตารางที่ 3 รายชื่อโรงเรียนและจำนวนนักเรียนในกลุ่มตัวอย่างที่ใช้ในการทดลอง

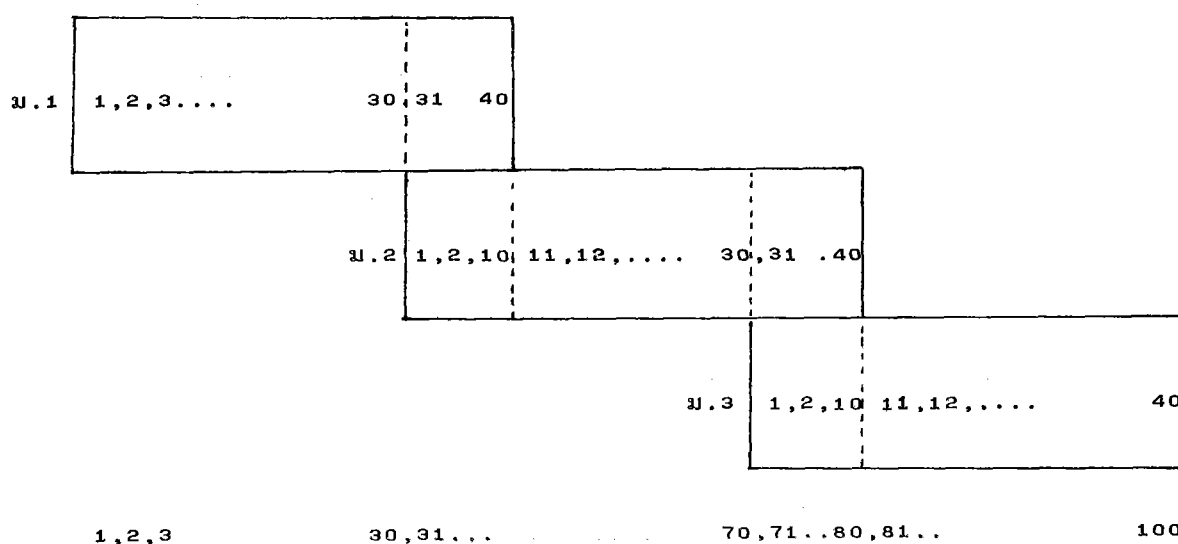
ขนาดโรงเรียน	โรงเรียน	จำนวนนักเรียน					
		การสร้าง เครื่องมือ			การ เทียบระดับคะแนน		
		ม.1	ม.2	ม.3	ม.1	ม.2	ม.3
ใหญ่พิเศษ ใหญ่	เทพศิรินทร์	50	50	50	70	80	70
	สายป้อนยา	50	50	50	60	48	70
	พุทธจักรวิทยา	35	40	40	50	24	70
	สตรีมหาพฤฒาราม	50	50	50	70	98	70
	สุทธิวราราม	50	50	50	70	87	55
	สันติราษฎร์วิทยาลัย	40	50	50	50	30	30
	วัดสระเกษ	44	50	50	50	25	30
	ไตรมิตรวิทยาลัย	40	50	34	70	90	60
	รวม	359	390	374	490	482	415

เครื่องมือที่ใช้ในการศึกษา

เครื่องมือที่ใช้การศึกษาครั้งนี้เป็นแบบทดสอบที่ผู้วิจัยสร้างขึ้นในเนื้อหาวิชาคณิตศาสตร์ เรื่องสมการ และอสมการ ระดับชั้นมัธยมศึกษาตอนต้น แบ่งเป็น 3 ระดับ ตามจุดประสงค์ และเนื้อหาในหลักสูตร ระดับละ 30 ข้อ ดังนี้

1. แบบทดสอบวิชาคณิตศาสตร์ ระดับมัธยมศึกษาปีที่ 1
2. แบบทดสอบวิชาคณิตศาสตร์ ระดับมัธยมศึกษาปีที่ 2
3. แบบทดสอบวิชาคณิตศาสตร์ ระดับมัธยมศึกษาปีที่ 3

แบบทดสอบแต่ละฉบับจะมีข้อสอบรวมภายใน (Internal anchor) จำนวน 10 ข้อ (Holms. 1982 : 140, Wright. 1977 : 10, Wright. 1977 : 107) ในระดับที่ติดกันดังภาพประกอบต่อไปนี้



ภาพประกอบ 9 การเชื่อมรอยต่อของแบบทดสอบ

สรุป จากภาพประกอบที่ 9 จะเห็นว่า แบบทดสอบชั้น ม.1 มีจำนวน 40 ข้อ โดยข้อ 31-40 เป็นข้อสอบร่วมระหว่างชั้น ม.1 กับ ม.2 แบบทดสอบชั้น ม.2 มีจำนวน 40 ข้อ โดยข้อ 1-10 เป็นข้อสอบร่วมระหว่างชั้น ม.1 กับ ม.2 และข้อ 31-40 เป็นข้อสอบร่วมระหว่างชั้น ม.2 กับ ม.3 แบบทดสอบชั้น ม.3 มีจำนวน 40 ข้อ โดยข้อ 1-10 เป็นข้อสอบร่วมระหว่างชั้น ม.2 กับ ม.3

วิธีดำเนินการสร้างแบบทดสอบเทียบคะแนนในแนวตั้ง

การสร้างแบบทดสอบเทียบระดับคะแนนในแนวตั้ง ที่วิเคราะห์โดยราชคโหมเดล แบ่งการสร้างออกเป็น 2 ตอน คือ ขั้นตอนการสร้างแบบทดสอบและขั้นตอนการเทียบระดับคะแนน

1. การสร้างแบบทดสอบ แบบทดสอบที่สร้างขึ้นเป็นแบบทดสอบชนิดเลือกตอบมี 5 ตัวเลือก ในวิชาคณิตศาสตร์ เรื่องสมการและอสมการ มีจำนวน 6 ฉบับ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 ระดับละ 2 ฉบับ โดยมีรายละเอียดการสร้างดังนี้

1.1 ศึกษาเนื้อหา จุดมุ่งหมายการสอนและจุดประสงค์เชิงพฤติกรรม จากหนังสือแบบเรียน คู่มือครู และเอกสารของกลุ่มโรงเรียน ในเรื่องสมการและอสมการ สำหรับนักเรียนชั้นมัธยมศึกษาตอนต้น

1.2 สร้างข้อสอบตามจุดประสงค์เชิงพฤติกรรมโดยหนึ่งจุดประสงค์จะสร้างข้อสอบระดับละประมาณ 15-20 ข้อ ได้ข้อสอบระดับละ 2 ฉบับ ฉบับละ 45 ข้อ รวมเป็นข้อสอบระดับละ 90 ข้อ

1.3 นำข้อสอบที่สร้างขึ้นในข้อ 1.2 ไปทดลองกับนักเรียนในกลุ่มตัวอย่างที่ 1 โดยนักเรียนแต่ละชั้นจะทำข้อสอบในระดับของตนทั้ง 2 ฉบับ

1.4 ตรวจให้คะแนน โดยให้คะแนนข้อที่ตอบถูก 1 คะแนน ข้อที่ตอบผิดให้ 0 คะแนน คัดเลือกกระดาษคำตอบที่ถูกทุกข้อ หรือผิดทุกข้อออก และตัดข้อสอบที่มีผู้ถูกหมดหรือผิดหมดออก

1.5 นำผลจากข้อ 1.4 ทั้งหมดมาวิเคราะห์เพื่อหาค่าพารามิเตอร์ของข้อสอบ คือ ค่าความยากของข้อสอบ (b หรือ δ) และค่าความสามารถ (θ หรือ β) โดยใช้โปรแกรมสำเร็จรูป BICAL ด้วยวิธี Unconditional maximum likelihood procedure (UCON)

1.6 คัดเลือกข้อสอบที่มีค่า ERR IMPAC น้อยกว่า .05 และมีค่า Item characteristic curve (ICC) สอดคล้องกับกลุ่มตัวอย่างเมื่อแบ่งกลุ่มตัวอย่างตามลำดับคะแนนเป็น 6 กลุ่ม

1.7 นำข้อสอบทั้งหมดมาพิจารณาคัดเลือก โดยพิจารณาจากข้อที่มีค่าสมรรถนะของความแตกต่างระหว่างค่าความสามารถกับค่าความยากน้อยกว่าหนึ่ง คือ $|\beta - \delta| < 1$ (Wright and Stone. 1979 : 75) ในที่นี้ใช้ค่าความสามารถที่เป็นค่าเฉลี่ยของแต่ละระดับชั้น

1.8 คัดเลือกข้อสอบจากข้อ 1.7 โดยพิจารณาแยกเป็น 2 กรณี

ก. ข้อสอบรวม แยกเป็น 2 ระดับ

- ระดับชั้นมัธยมศึกษาปีที่ 1 กับ 2 พิจารณาจากข้อสอบที่มีค่าความยากสูงสุดของระดับชั้นมัธยมศึกษาปีที่ 1 จำนวน 10 ข้อ ได้ค่าความยากตั้งแต่ -1.36 ถึง $-.84$

- ระดับชั้นมัธยมศึกษาปีที่ 2 กับ 3 พิจารณาจากข้อสอบที่มีค่าความยากสูงสุดของระดับชั้นมัธยมศึกษาปีที่ 2 จำนวน 10 ข้อ ได้ค่าความยากตั้งแต่ -1.19 ถึง $-.72$

ข. ข้อสอบเฉพาะแต่ละระดับ พิจารณาจากการกระจายของข้อสอบกับจุดประสงค์ โดยพิจารณาเลือกข้อสอบสำหรับชั้นมัธยมศึกษาปีที่ 1 จำนวน 30 ข้อ มัธยมศึกษาปีที่ 2 จำนวน 20 ข้อ มัธยมศึกษาปีที่ 3 จำนวน 30 ข้อ

1.9 นำข้อสอบที่คัดเลือกจากข้อ 1.8 มาเรียงตามลำดับแผนภูมิได้จำนวนข้อดังนี้

ก. ชั้นมัธยมศึกษาปีที่ 1 มีข้อสอบเฉพาะ 30 ข้อ ข้อสอบรวมกับมัธยมศึกษาปีที่ 2 จำนวน 10 ข้อ รวมเป็น 40 ข้อ

ข. ชั้นมัธยมศึกษาปีที่ 2 มีข้อสอบเฉพาะ 20 ข้อ ข้อสอบรวมกับมัธยมศึกษาปีที่ 1 จำนวน 10 ข้อ ข้อสอบรวมกับมัธยมศึกษาปีที่ 3 จำนวน 10 ข้อ รวมเป็น 40 ข้อ

ค. ชั้นมัธยมศึกษาปีที่ 3 มีข้อสอบเฉพาะ 30 ข้อ ข้อสอบรวมกับมัธยมศึกษาปีที่ 2 จำนวน 10 ข้อ รวมเป็น 40 ข้อ

2. การเทียบระดับคะแนน

2.1 นำแบบทดสอบที่สร้างเสร็จแล้วจากข้อ 1 ไปสอบกับกลุ่มตัวอย่างเพื่อจะนำผลมาสร้างตารางเทียบระดับคะแนน โดยนักเรียนชั้น ม.1 จะทำแบบทดสอบฉบับของ ม.1 นักเรียนชั้น ม.2 จะทำแบบทดสอบฉบับของ ม.2 และนักเรียนชั้น ม.3 ทำแบบทดสอบฉบับของ ม.3

2.2 ตรวจให้คะแนนแบบทดสอบ โดยตอบถูกให้ 1 คะแนน ตอบผิดหรือเว้นว่างให้ 0 คะแนน

2.3 นำคะแนนจากข้อ 2.2 มาวิเคราะห์เพื่อหาพารามิเตอร์ของข้อสอบ คือค่าความยาก (δ) และค่าความสามารถ โดยใช้โปรแกรมสำเร็จรูป BICAL

2.4 ปรับค่าความสามารถ (Transformation) ของนักเรียนชั้น ม.1, ม.2 และ ม.3 ให้อยู่ใน Scale เดียวกันโดยใช้หลัก Linear equation ตามสูตรของลอร์ดและสวเวอร์ (Loyd and Hoover. 1980 : 186) ด้วยวิธีการดังนี้

2.4.1 ปรับค่าความสามารถ (Transformation) ของนักเรียนชั้น ม.1 ให้อยู่ใน Scale เดียวกับ ม.2 โดยการนำค่าความยากเฉลี่ยของข้อสอบร่วมระหว่างนักเรียนชั้น ม.1 กับ ม.2 มาวิเคราะห์ ซึ่งจะได้อัตราที่นำไปปรับค่าความสามารถในการทำข้อสอบของนักเรียนชั้น ม.1 ให้อยู่ใน Scale เดียวกับ ม.2

2.4.2 ปรับค่าความสามารถ (Transformation) ของนักเรียนชั้น ม.3 ให้อยู่ใน Scale เดียวกับชั้น ม.2 โดยการนำค่าความยากเฉลี่ยข้อสอบร่วมระหว่างนักเรียนชั้น ม.2 กับ ม.3 มาวิเคราะห์ ซึ่งจะได้อัตราที่นำไปปรับค่าความสามารถในการทำข้อสอบของนักเรียนชั้น ม.3 ให้อยู่ใน Scale เดียวกับ ม.2

2.5 สร้างตารางเทียบความสามารถของนักเรียนชั้น ม.1, ม.2 และ ม.3 โดยนำค่าความสามารถของ ม.1 กับ ม.3 ที่ถูก Transformation ให้อยู่ใน Scale เดียวกับ ม.2 แล้วมาเทียบหาค่าที่ใกล้เคียงกันมากที่สุด

2.6 สร้างตารางเทียบระดับคะแนนของนักเรียนชั้น ม.1, ม.2 และ ม.3 โดยยึดความสามารถจากข้อ 2.5 เป็นหลัก

2.7 จัดพิมพ์คู่มือการใช้แบบทดสอบ พร้อมทั้งตารางเทียบระดับคะแนน

วิธีดำเนินการเก็บรวบรวมข้อมูล

1. วางแผนในการดำเนินการสอบโดยติดต่อขอความร่วมมือในการนำแบบทดสอบไปทดลองกับนักเรียนในกลุ่มตัวอย่าง ตลอดจนกำหนดวันและเวลาในการสอน

2. สร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน เรื่องสมการและอสมการสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1 จำนวน 90 ข้อ ชั้นมัธยมศึกษาปีที่ 2 จำนวน 90 ข้อ ชั้นมัธยมศึกษาปีที่ 3 จำนวน 90 ข้อ นำไปทดสอบกับนักเรียนตามระดับนั้น ๆ ระดับละ 400 คน

3. คัดเลือกแบบทดสอบที่นำไปทดลองแล้วให้มีคุณภาพตามต้องการ ระดับละ 40 ข้อ

4. นำแบบทดสอบของแต่ละระดับที่คัดเลือกไว้แล้วไปสอบกับนักเรียนในกลุ่มตัวอย่างที่ต้องการเทียบระดับคะแนนตามระดับนั้น ๆ ระดับละประมาณ 400 คน

การวิเคราะห์ข้อมูลและสถิติที่ใช้

1. หาค่าสถิติพื้นฐาน คือ ค่าเฉลี่ยและส่วน เบี่ยงเบนมาตรฐาน
2. วิเคราะห์หาค่าความยากของข้อสอบ, ค่า ERR IMPAC, ค่า Item characteristic curve (ICC) ของข้อสอบแต่ละข้อตามหลักการตอบข้อคำถามของราซค์โมเดล ด้วยโปรแกรมสำเร็จรูป BICAL
3. หาค่าความสามารถทางการเรียน (β) ของนักเรียนแต่ละคน เน้นที่ตอบถูกตามหลักการตอบข้อคำถามของราซค์โมเดลด้วยโปรแกรมสำเร็จรูป BICAL
4. ทดสอบความแตกต่างของคะแนนเฉลี่ยของข้อสอบรวม โดยใช้สูตร Z-test (ล้วน สายยศ และอังคณา สายยศ 2528 : 83)

$$Z = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

เมื่อ	$\bar{X}_1$	คือ	คะแนน เฉลี่ยของกลุ่มที่ 1
	$\bar{X}_2$	คือ	คะแนน เฉลี่ยของกลุ่มที่ 2
	S_1^2	คือ	ความแปรปรวนของกลุ่มที่ 1
	S_2^2	คือ	ความแปรปรวนของกลุ่มที่ 2
	n_1	คือ	จำนวนคนในกลุ่มที่ 1
	n_2	คือ	จำนวนคนในกลุ่มที่ 2

5. Transformation ค่าความสามารถของแต่ละระดับให้อยู่ใน Scale เดียวกันโดย

ใช้หลัก Linear equation (Loyd and Hoover, 1980 : 186)

$$\beta_j = \frac{a_j}{a_k} \beta_k + \delta_i - \frac{a_i}{a_k} \delta_k$$

เมื่อ	β_j	แทนค่าความสามารถของนักเรียนระดับ
	β_k	แทนค่าความสามารถของนักเรียนระดับ
	a_j	แทนค่าอำนาจจำแนกของข้อสอบรวมในระดับ
	a_k	แทนค่าอำนาจจำแนกของข้อสอบรวมในระดับ
	δ_j	แทนค่าความยาก เฉลี่ยของข้อสอบรวมในระดับ j
	δ_k	แทนค่าความยาก เฉลี่ยของข้อสอบรวมในระดับ k

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลและการแปรข้อมูลในการศึกษาครั้งนี้ ผู้วิจัยได้กำหนดสัญลักษณ์และอักษร
อักษรย่อต่าง ๆ ไว้ดังนี้

สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล

N	แทน	จำนวนนักเรียนในกลุ่มตัวอย่าง
K	แทน	จำนวนข้อของแบบทดสอบ
$\bar{X}$	แทน	คะแนนเฉลี่ย
S.D.	แทน	ค่า เบี่ยงเบนมาตรฐาน
β	แทน	ความสามารถทางการเรียน
β_1	แทน	ความสามารถทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 1
β_2	แทน	ความสามารถทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 2
β_3	แทน	ความสามารถทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 3
β_{12}	แทน	ความสามารถทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ที่ถูก Transformation ให้อยู่ใน Scale เดียวกับชั้นมัธยมศึกษาปีที่ 2
β_{32}	แทน	ความสามารถทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่ถูก Transformation ให้อยู่ใน Scale เดียวกับชั้นมัธยมศึกษาปีที่ 2
δ	แทน	ความยากของข้อสอบ
δ_1	แทน	ความยากของข้อสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1
δ_2	แทน	ความยากของข้อสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2
δ_3	แทน	ความยากของข้อสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3

- δ_{12} แทน ความยากของข้อสอบร่วมระหว่างชั้นมัธยมศึกษาปีที่ 1 กับ 2
- δ_{23} แทน ความยากของข้อสอบร่วมระหว่างชั้นมัธยมศึกษาปีที่ 2 กับ 3
- $\bar{\delta}_{12}$ แทน ความยากเฉลี่ยของข้อสอบร่วมระหว่างชั้นมัธยมศึกษาปีที่ 1 กับ 2
- $\bar{\delta}_{23}$ แทน ความยากเฉลี่ยของข้อสอบร่วมระหว่างชั้นมัธยมศึกษาปีที่ 2 กับ 3
- X_1 แทน คะแนนดิบของแบบทดสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1
- X_2 แทน คะแนนดิบของแบบทดสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2
- X_3 แทน คะแนนดิบของแบบทดสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3

การเสนอผลการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้ ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลเป็นตอน ๆ ดังนี้

ตอนที่ 1 เกี่ยวกับแบบทดสอบ

- 1.1 ค่าสถิติพื้นฐานของแบบทดสอบ
- 1.2 เปรียบเทียบความยากข้อสอบร่วม เป็นรายข้อ
- 1.3 ค่าสถิติพื้นฐานความยากของข้อสอบร่วม

ตอนที่ 2 เกี่ยวกับค่าความสามารถ

- 2.1 ค่าความสามารถกับคะแนนของนักเรียนชั้นมัธยมศึกษาปีที่ 1
- 2.2 ค่าความสามารถกับคะแนนของนักเรียนชั้นมัธยมศึกษาปีที่ 2
- 2.3 ค่าความสามารถกับคะแนนของนักเรียนชั้นมัธยมศึกษาปีที่ 3
- 2.4 การเทียบระดับความสามารถ
- 2.5 การเทียบระดับคะแนน

ผลการวิเคราะห์ข้อมูล

ตอนที่ 1 เกี่ยวกับแบบทดสอบ

- 1.1 ค่าสถิติพื้นฐานของแบบทดสอบ จากการนำแบบทดสอบ 3 ฉบับ ไปทดลองกับ

กลุ่มตัวอย่าง ได้ค่าสถิติพื้นฐานของคะแนนดังแสดงในตาราง 4

ตาราง 4 ค่าสถิติพื้นฐานของแบบทดสอบ 3 ฉบับ

ระดับชั้น	N	K	คะแนนสูงสุด	คะแนนต่ำสุด	$\bar{X}$	S.D.
ม.1	490	40	38	3	17.5653	7.3849
ม.2	482	40	30	3	14.4564	5.5679
ม.3	415	40	37	6	17.8843	7.1489

ผลการวิเคราะห์ข้อมูลจากตาราง 4 แสดงให้เห็นว่า คะแนนสูงสุดของทั้ง 3 ฉบับ ไม่เท่ากับคะแนนเต็ม และคะแนนต่ำสุดไม่เท่ากับศูนย์ เป็นผลให้การวิเคราะห์หาค่าพารามิเตอร์ ขึ้นต่อไปสามารถใช้กลุ่มตัวอย่างนี้ได้ทั้งหมดไม่ต้องตัดออก ค่าเฉลี่ยและค่าการกระจายของแบบทดสอบ สำหรับชั้น ม.1 กับ ม.2 มีค่าใกล้เคียงกันและสูงกว่าชั้น ม.2 แต่ทั้ง 3 ฉบับมีค่าเฉลี่ยต่ำกว่าครึ่งของ คะแนนเต็ม แสดงว่าแบบทดสอบทั้งสามฉบับเป็นแบบทดสอบที่ค่อนข้างยาก

1.2 เปรียบเทียบความยากข้อสอบรวมเป็นรายข้อ

ผู้วิจัยได้นำแบบทดสอบทั้ง 3 ฉบับไปวิเคราะห์ด้วยโปรแกรมสำเร็จรูป เพื่อ หาค่าความยากของข้อสอบรวมในแต่ละระดับ ปรากฏค่าความยากของข้อสอบรวมดังแสดงใน ตาราง 5

ตาราง 5 เปรียบเทียบค่าความยากของข้อสอบรวม

δ_{12}				δ_{23}			
ข้อที่	ม.1	ข้อที่	ม.2	ข้อที่	ม.2	ข้อที่	ม.3
31	0.574	1	-0.534	31	1.297	1	-1.644
32	0.507	2	-0.142	32	0.562	2	0.458
33	0.893	3	-0.092	33	0.245	3	-0.690
34	1.256	4	0.549	34	1.707	4	-0.431
35	0.735	5	-0.485	35	1.120	5	1.393
36	0.642	6	0.190	36	1.103	6	-0.453
37	0.893	7	-0.534	37	0.799	7	0.275
38	1.074	8	-0.051	38	0.562	8	-0.308
39	0.654	9	-0.872	39	1.171	9	-0.185
40	0.918	10	-0.212	40	0.428	10	-1.819

ผลการวิเคราะห์ข้อมูลจากตาราง 5 แสดงให้เห็นว่า ข้อสอบรวมระหว่างชั้น ม.1 กับ ม.2 นั้น เมื่อให้นักเรียนชั้น ม.1 ทำข้อสอบแล้วค่าความยากของทุกข้อจะมีค่าสูงกว่าที่นักเรียนชั้น ม.2 ทำ นั่นคือ ข้อสอบรวมระหว่างชั้น ม.1 กับ ม.2 ทุกข้อจะเป็นข้อสอบที่ยากสำหรับนักเรียนชั้น ม.1 แต่จะเป็นข้อสอบที่ง่ายสำหรับนักเรียนชั้น ม.2 ส่วนข้อสอบรวมระหว่างชั้น ม.2 กับ ม.3 ถ้าให้นักเรียนชั้น ม.2 ทำจะมีค่าสูงกว่านักเรียนชั้น ม.3 ทำ ยกเว้นข้อ 35 ในชั้น ม.2 หรือข้อ 5 ในชั้น ม.3 ซึ่งโจทย์เป็นเรื่องสมการที่ต้องใช้เงื่อนไขว่าตัวหารไม่เป็นศูนย์มาช่วยหาคำตอบ ซึ่งเป็นเรื่องที่อยู่ในหลักสูตรชั้น ม.2 จึงทำให้นักเรียนชั้น ม.2 สามารถหาคำตอบได้ถูกต้อง ส่วนนักเรียนชั้น ม.3 เป็นเรื่องที่ได้เรียนมานานแล้วก็อาจลืมไม่ระวังเงื่อนไขนี้ไป ส่วนข้ออื่น ๆ อีก 9 ข้อ ทุกข้อที่นักเรียนชั้น ม.2 ทำยังมีค่าความยากสูงกว่าที่นักเรียนชั้น ม.3 ทำ ซึ่งก็ยิ่งแสดงว่าข้อสอบรวมระหว่างชั้น ม.2 กับ ม.3 นั้น เป็นข้อสอบที่ยากสำหรับชั้น ม.2 แต่เป็นข้อสอบที่ง่ายสำหรับนักเรียนชั้น ม.3

1.3 ค่าสถิติพื้นฐานความยากของข้อสอบรวม

ผู้วิจัยได้นำแบบทดสอบรวมมาวิเคราะห์หาค่าสถิติพื้นฐาน ปรากฏผลดังแสดงใน

ตาราง 6 ค่าสถิติพื้นฐานความยากของข้อสอบรวม

ระดับชั้น	δ_{12}		δ_{23}	
	$\bar{\delta}_{12}$	S.D.	$\bar{\delta}_{23}$	S.D.
ม.1	0.8146	0.2356	-	-
ม.2	-0.2183	0.4087	0.8994	0.4549
ม.3	-	-	-0.3404	0.9472
ความแตกต่างของค่าเฉลี่ย	1.0329**		1.2398**	

** มีนัยสำคัญทางสถิติที่ระดับ .01

ผลการวิเคราะห์ข้อมูลจากตาราง 6 แสดงให้เห็นว่า ข้อสอบรวมระหว่างชั้น ม.1 กับ ม.2 นั้น จะมีค่าความยากเฉลี่ยในระดับชั้น ม.1 สูงกว่าค่าความยากเฉลี่ยในชั้น ม.2 และทั้งสองค่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.1 ส่วนข้อสอบรวมระหว่างชั้น ม.2 กับ ม.3 ค่าความยากเฉลี่ยของระดับชั้น ม.2 สูงกว่าค่าความยากเฉลี่ยของชั้น ม.3 และมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ตอนที่ 2 เกี่ยวกับความสามารถ

2.1 ค่าความสามารถกับคะแนนของนักเรียนชั้นมัธยมศึกษาปีที่ 1

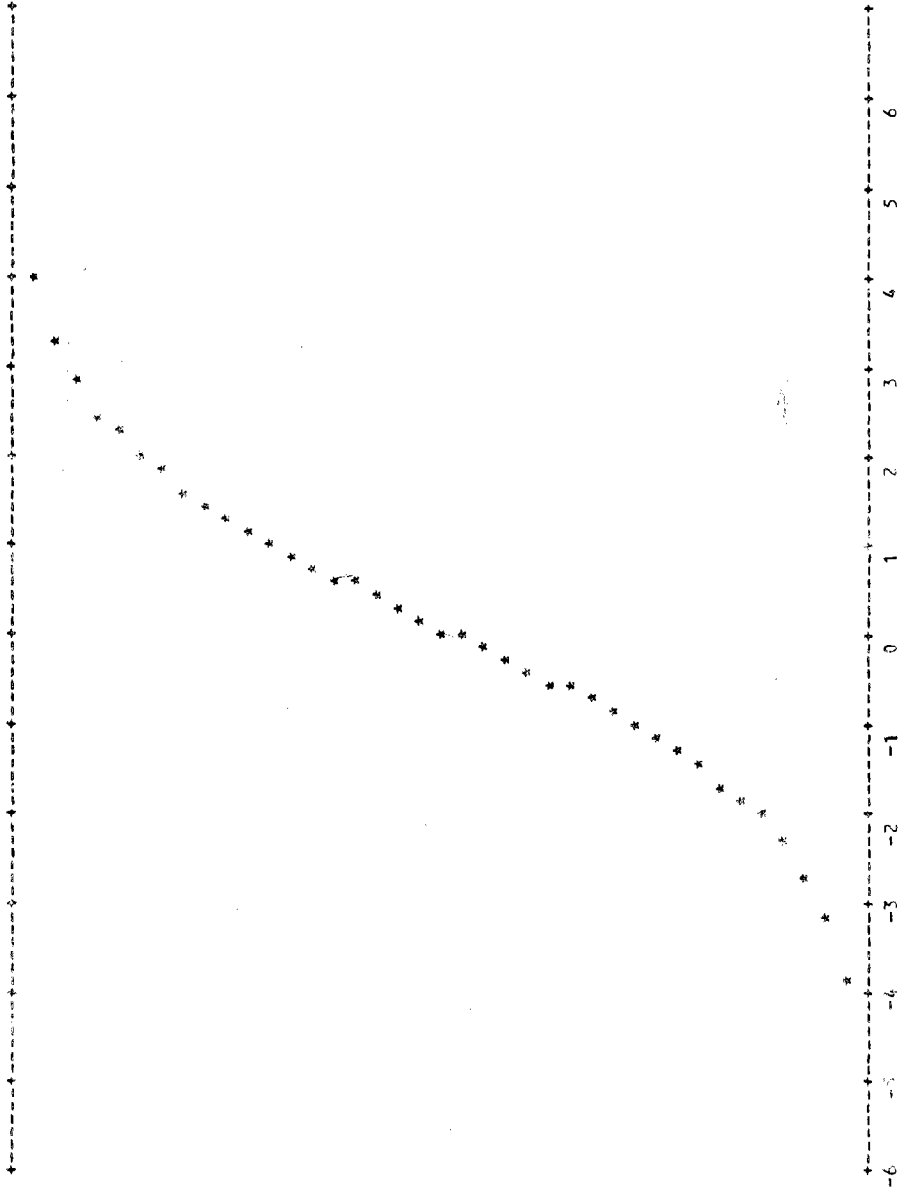
ผู้วิจัยได้นำแบบทดสอบสำหรับชั้น ม.1 ไปทดลองกับนักเรียนชั้น ม.1 จำนวน 490 คน แล้วมาวิเคราะห์ด้วยโปรแกรมสำเร็จรูป BICAL ได้ค่าความสามารถของนักเรียน ดังภาพประกอบ 10

COMPLETE SCORE EQUIVALENCE TABLE

RAW SCORE COUNT LOG STANDARD

RAW SCORE	COUNT	LOG ABILITY	STANDARD ERRORS
39	0	3.93	1.04
38	1	3.20	.75
37	3	2.75	.63
36	0	2.44	.55
35	4	2.15	.50
34	2	1.92	.47
33	3	1.72	.44
32	4	1.54	.42
31	6	1.36	.40
30	10	1.22	.39
29	8	1.08	.38
28	14	.94	.37
27	13	.81	.36
26	11	.69	.36
25	22	.56	.35
24	16	.45	.35
23	19	.33	.34
22	20	.21	.34
21	13	.10	.34
20	20	-.01	.34
19	16	-.13	.34
18	17	-.24	.34
17	23	-.35	.34
16	24	-.47	.35
15	23	-.59	.35
14	22	-.71	.35
13	28	-.83	.36
12	25	-.96	.37
11	25	-1.09	.38
10	22	-1.23	.39
9	31	-1.36	.40
8	21	-1.54	.41
7	12	-1.72	.47
6	7	-1.91	.46
5	3	-2.13	.45
4	1	-2.39	.54
3	1	-2.72	.62
2	0	-3.15	.74
1	0	-3.87	1.02

TEST CHARACTERISTIC CURVE



PERSON SEPARABILITY INDEX .84 (EQUIVALENT TO KR20)

40 ITEMS CALIBRATED ON 406 PERSONS

490 MEASURABLE PERSONS WITH MEAN ABILITY = -.31 AND STD. DEV. = .87

ภาพประกอบ 10 การเปรียบเทียบระหว่างคะแนนดิบกับค่าความสามารุและโค้งลักษณะข้อสอบของแบบทดสอบ

สำหรับนักเรียนมัธยมศึกษาปีที่ 1

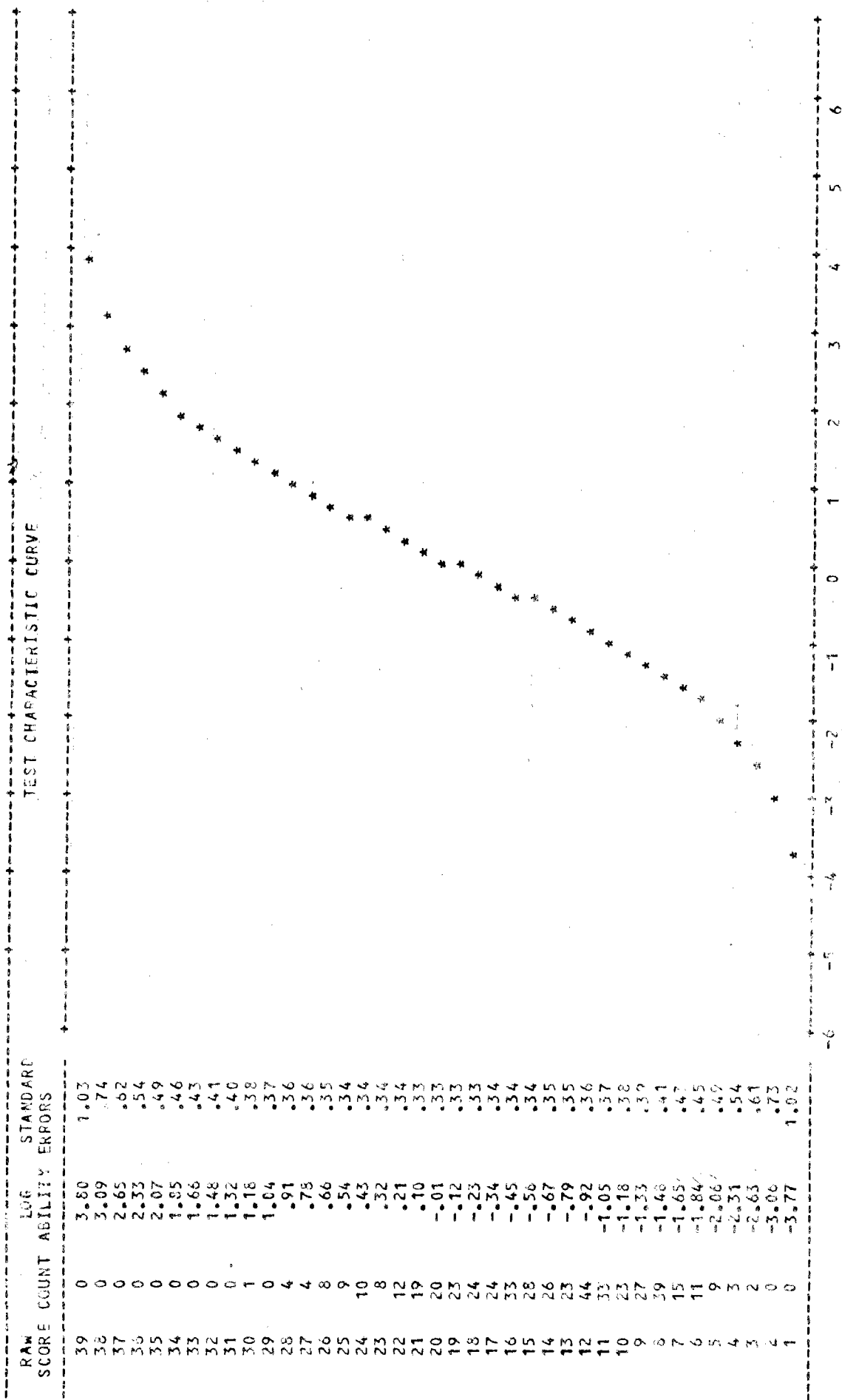
จากภาพประกอบ 10 เป็นการวิเคราะห์ผลการสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ซึ่งมีจำนวน 490 คน ทำข้อสอบ 40 ข้อ ในภาพจะแสดงโค้งลักษณะข้อสอบทั้งฉบับ ผลการสอบที่เรียงคะแนนจากสูงที่สุดไปต่ำสุดพร้อมทั้งจำนวนนักเรียนที่สอบได้ ค่าความสามารถของผู้สอบแต่ละระดับคะแนน ค่าความคลาดเคลื่อนมาตรฐาน และค่าความเชื่อมั่นของแบบทดสอบฉบับนี้ซึ่งมีค่าเท่ากับ .84

เมื่อพิจารณารายละเอียดพบว่า นักเรียนที่ได้คะแนนต่ำสุด 3 คะแนน มีค่าล้อยความสามารถเท่ากับ -2.72 จำนวน 1 คน คะแนนสูงสุด 38 คะแนน มีค่าล้อยความสามารถเท่ากับ 3.20 จำนวน 1 คน

2.2 ค่าความสามารถกับคะแนนของนักเรียนชั้นมัธยมศึกษาปีที่ 2

ผู้วิจัยได้นำแบบสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2 ไปทดลองกับกลุ่มตัวอย่าง 482 คน นำผลการสอบมาวิเคราะห์ด้วยโปรแกรมสำเร็จรูป BICAL ได้ผลดังภาพประกอบที่ 11

COMPLETE SCORE EQUIVALENCE TABLE



PERSON SEPARABILITY INDEX .72 (EQUIVALENT TO KR20)

40 ITEMS CALIBRATED ON 442 PERSONS

482 MEASURABLE PERSONS WITH MEAN ABILITY = -.67 AND STD. DEV. = .58

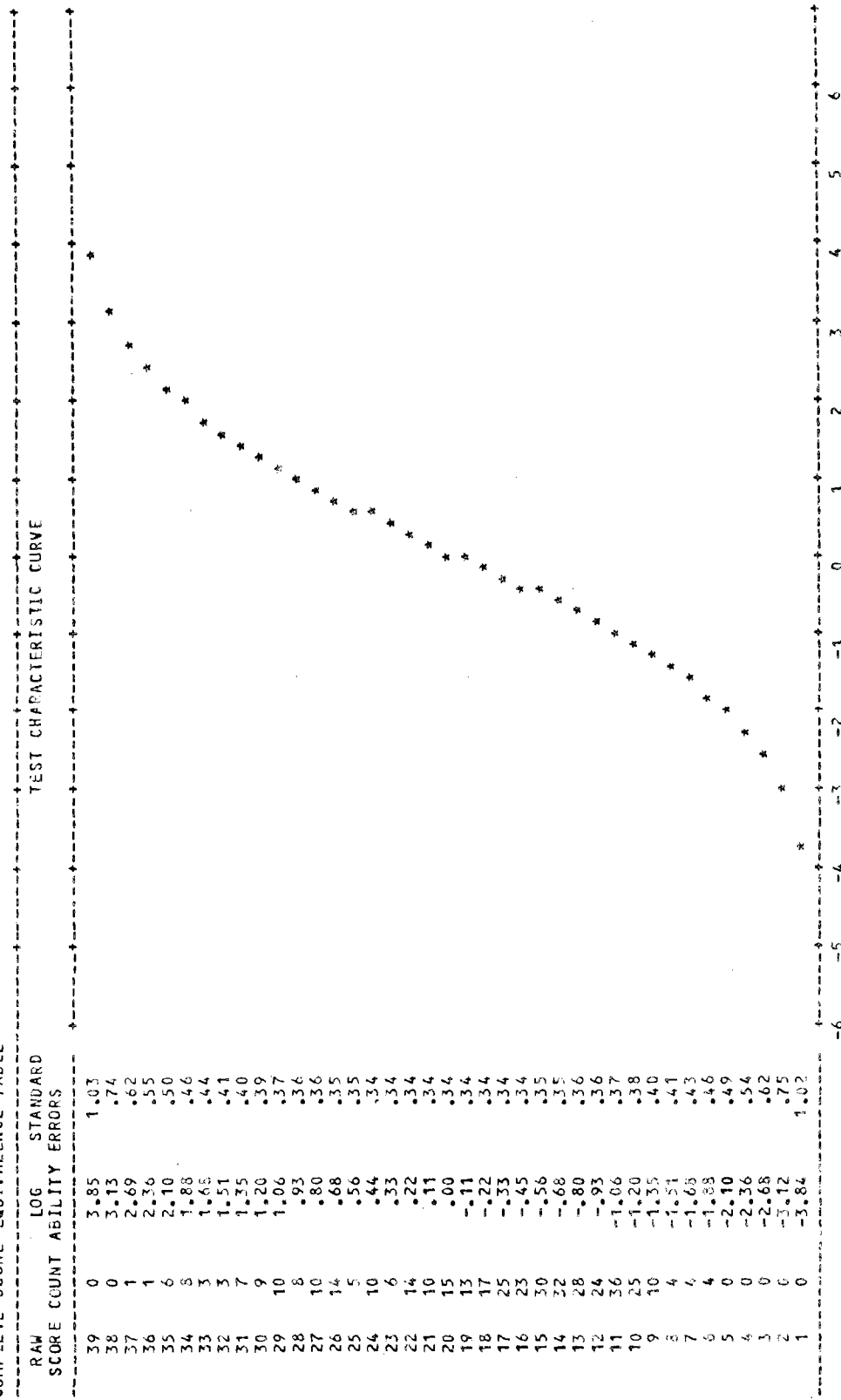
จากภาพประกอบ 11 เป็นการวิเคราะห์ผลการสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 2 ซึ่งมีจำนวน 482 คน ทำข้อสอบ 40 ข้อ ในภาพจะแสดงถึงโค้งลักษณะข้อสอบทั้งฉบับ ผลการสอบที่เรียงคะแนนจากสูงสุดไปต่ำสุดพร้อมทั้งจำนวนนักเรียนที่สอบได้ ค่าความสามารถของผู้สอบแต่ละระดับคะแนน ค่าความคลาดเคลื่อนมาตรฐาน และค่าความเชื่อมั่นของแบบทดสอบทั้งฉบับซึ่งมีค่าเท่ากับ .72

เมื่อพิจารณารายละเอียดพบว่า นักเรียนที่ได้คะแนนต่ำสุด 3 คะแนนซึ่งมีค่าล้นความสามารถเท่ากับ -2.63 จำนวน 2 คน คะแนนสูงสุด 30 คะแนน มีค่าล้นความสามารถเท่ากับ 1.18 จำนวน 1 คน

2.3 ค่าความสามารถกับคะแนนของนักเรียนชั้นมัธยมศึกษาปีที่ 3

ผู้วิจัยได้นำแบบสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 ไปทดลองกับกลุ่มตัวอย่าง 415 คน นำผลการวิเคราะห์ด้วยโปรแกรมสำเร็จรูป BICAL ได้ผลดังภาพประกอบที่ 12

COMPLETE SCORE EQUIVALENCE TABLE



PERSON SEPARABILITY INDEX .83 (EQUIVALENT TO KR20)

40 ITEMS CALIBRATED ON 40 PERSONS

415 MEASURABLE PERSONS WITH MEAN ABILITY = -.24 AND STD. DEV. = .80

ภาพประกอบ 12 การเปรียบเทียบระหว่างคะแนนเดิมกับค่าความสามารถ และโค้งลักษณะข้อสอบของแบบทดสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3

จากภาพประกอบ 12 เป็นการวิเคราะห์ผลการสอบของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ซึ่งมีจำนวน 415 คน ทำข้อสอบ 40 ข้อ ในภาพจะแสดงถึงโค้งลักษณะข้อสอบทั้งฉบับ ผลการสอบที่เรียงคะแนนจากสูงสุดไปต่ำสุด พร้อมทั้งจำนวนนักเรียนที่สอบได้ ค่าความสามารถของผู้สอบแต่ละระดับคะแนน ค่าความคลาดเคลื่อนมาตรฐานและค่าความเชื่อมั่นของแบบทดสอบฉบับนี้ซึ่งมีค่าเท่ากับ .83

เมื่อพิจารณารายละเอียดพบว่า นักเรียนที่ได้คะแนนต่ำสุด 6 คะแนน มีค่าลัดความสามารถเท่ากับ -1.88 จำนวน 4 คน คะแนนสูงสุด 37 คะแนนมีค่าลัดความสามารถเท่ากับ 2.69 จำนวน 1 คน

2.4 การเทียบระดับความสามารถ

ผู้วิจัยได้นำผลการวิเคราะห์หาค่าความยากเฉลี่ยของข้อสอบรวมไปใช้กับสูตรของ ลอร์ด และฮูเวอร์ (Lloyd and Hoover 1980: 186) ตามหลัก Linear equation ได้ตัวคงที่ในการนำไปปรับ (Transformation) ค่าความสามารถของนักเรียนชั้นมัธยมศึกษาปีที่ 1 กับมัธยมศึกษาปีที่ 3 ให้อยู่ใน Scale เดียวกับชั้นมัธยมศึกษาปีที่ 2 แล้วนำค่าความสามารถที่ได้ปรับ (Transformation) ให้อยู่ใน Scale เดียวกันแล้วมาเทียบหาค่าใกล้เคียงกันมากที่สุด ได้ผลดังตาราง 7

ตาราง 7 การเทียบระดับความสามารถของนักเรียนมัธยมศึกษาปีที่ 1, 2 และ 3

$\beta_{12} (\beta)$	β_2	$\beta_{32} (\beta_3)$
$\begin{cases} -3.4229 (-2.39) \\ -3.1629 (-2.13) \\ -2.9429 (-1.91) \end{cases}$	$\begin{cases} -3.77 \\ -3.06 \end{cases}$	
$\begin{cases} -2.7529 (-1.72) \\ -2.5729 (-1.54) \\ -2.4129 (-1.38) \\ -2.2629 (-1.83) \end{cases}$	$\begin{cases} -2.31 \\ -2.63 \end{cases}$	$-2.6002 (-3.84)$
$\begin{cases} -2.1229 (-1.09) \\ -1.9929 (-.96) \\ -1.8629 (-.83) \end{cases}$	$\begin{cases} -2.06 \\ -1.84 \end{cases}$	$-1.8802 (-3.12)$
$\begin{cases} -1.7429 (-.71) \\ -1.6229 (-.59) \\ -1.5029 (-.47) \\ -1.3829 (-.35) \end{cases}$	$\begin{cases} -1.65 \\ -1.48 \\ -1.33 \end{cases}$	$-1.4402 (-2.68)$
$\begin{cases} -1.2729 (-.24) \\ -1.1629 (-.13) \\ -1.0429 (-.01) \end{cases}$	$\begin{cases} -1.05 \\ -1.18 \end{cases}$	$-1.1211 (-2.36)$
$\begin{cases} -0.9329 (0.10) \\ -0.8229 (0.21) \end{cases}$	$\begin{cases} -0.79 \\ -0.92 \end{cases}$	$-0.8602 (-2.10)$
$\begin{cases} -0.7029 (0.33) \\ -0.5829 (0.45) \end{cases}$	$\begin{cases} -0.67 \\ -0.56 \end{cases}$	$-0.6402 (-1.88)$

ตาราง 7 (ต่อ)

$\beta_{12}(\beta_1)$	β_2	$\beta_{32}(\beta_3)$
$\begin{cases} -0.4729 & (0.56) \\ -0.3429 & (0.69) \end{cases}$	$\begin{cases} -0.45 \\ -0.34 \end{cases}$	$-0.4402 \quad (-1.68)$
$-0.2229 \quad (0.81)$	-0.23	$-0.2702 \quad (-1.51)$
$-0.0923 \quad (0.94)$	-0.12	$-0.1102 \quad (-1.35)$
$0.0471 \quad (1.08)$	$\begin{cases} 0.10 \\ -0.01 \end{cases}$	$0.0398 \quad (-1.20)$
$0.1871 \quad (1.22)$	$\begin{cases} 0.21 \\ 0.32 \end{cases}$	$\begin{cases} 0.1798 \quad (-1.06) \\ 0.3098 \quad (-0.93) \end{cases}$
$0.3471 \quad (1.38)$	0.43	$0.4398 \quad (-0.80)$
$0.5071 \quad (1.54)$	0.54	$0.5598 \quad (-0.68)$
$0.6871 \quad (1.72)$	$\begin{cases} 0.66 \\ 0.78 \end{cases}$	$\begin{cases} 0.6798 \quad (-0.56) \\ 0.7879 \quad (-0.45) \end{cases}$
$0.8871 \quad (1.92)$	0.91	$0.9098 \quad (-0.33)$
$1.1802 \quad (2.15)$	$\begin{cases} 1.04 \\ 1.18 \end{cases}$	$\begin{cases} 1.0198 \quad (-0.22) \\ 1.1298 \quad (-0.11) \end{cases}$
$1.3871 \quad (2.42)$	$\begin{cases} 1.32 \\ 1.48 \end{cases}$	$\begin{cases} 1.3498 \quad (0.11) \\ 1.4598 \quad (0.22) \\ 1.5698 \quad (0.33) \end{cases}$
$1.7171 \quad (2.75)$	$\begin{cases} 1.65 \\ 1.85 \end{cases}$	$\begin{cases} 1.6798 \quad (0.44) \\ 1.7998 \quad (0.56) \\ 1.9198 \quad (0.68) \end{cases}$

ตาราง 7 (ต่อ)

$\beta_{12}(\beta_1)$	β_2	$\beta_{32}(\beta_3)$
2.1671 (3.20)	$\begin{cases} 2.07 \\ 2.33 \end{cases}$	$\begin{cases} 2.0398 (0.80) \\ 2.1698 (0.93) \\ 2.2998 (1.06) \\ 2.4398 (1.20) \end{cases}$
		$\begin{cases} 2.5898 (1.35) \\ 2.7498 (1.51) \\ 2.9198 (1.68) \\ 3.1198 (1.88) \end{cases}$
		$\begin{cases} 3.3398 (2.10) \\ 3.5998 (2.36) \end{cases}$
2.8971 (3.93)	$\begin{cases} 2.65 \\ 3.09 \end{cases}$	
	3.80	

ตาราง 7 แสดงผลการวิเคราะห์ค่าความสามารถของนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 ที่เรียงลำดับจากความสามารถต่ำสุดไปหาความสามารถสูงสุดที่ได้ปรับ (Transformation) ให้อยู่ใน Scale เดียวกันทั้งหมด และค่าความสามารถตาม Scale เดิมของนักเรียนชั้นมัธยมศึกษาปีที่ 1 และ 2

เมื่อพิจารณารายละเอียดพบว่าค่าความสามารถตาม Scale ของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ที่จะนำมาปรับ (Transformation) และเทียบกับค่าความสามารถของนักเรียนชั้นมัธยมศึกษาปีที่ 2 ได้ จะเริ่มจากค่า -2.29 ถึง 3.93 โดยค่าความสามารถของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ที่เท่ากับ -2.39 เมื่อนำมาปรับ (Transformation) ให้อยู่ใน Scale เดียวกับชั้นมัธยมศึกษาปีที่ 2 จะมีค่าเท่ากับ -3.4229 ซึ่งจะเท่ากับค่าความสามารถ -3.77 ของนักเรียนชั้นมัธยมศึกษาปีที่ 2 ค่าความสามารถของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่จะนำมาปรับ (Transformation) ให้อยู่ใน

Scale เดียวกับชั้นมัธยมศึกษาปีที่ 2 จะมีค่าเท่ากับ -3.4229 ซึ่งจะเท่ากับค่าความสามารถ -3.77 ของนักเรียนชั้นมัธยมศึกษาปีที่ 2 ค่าความสามารถของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่จะนำมาปรับ (Transformation) และเทียบกับความสามารถในชั้นมัธยมศึกษาปีที่ 2 ได้จะเริ่มจาก -3.84 ถึง 2.36 โดยค่าความสามารถของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ที่เท่ากับ -3.84 เมื่อนำมาปรับ (Transformation) ให้อยู่ใน Scale เดียวกับชั้นมัธยมศึกษาปีที่ 2 จะมีค่าเท่ากับ -2.6002 และเท่ากับค่าความสามารถ -2.31 และ -2.63 ของนักเรียนชั้นมัธยมศึกษาปีที่ 2

ถ้าพิจารณาเปรียบเทียบค่าความสามารถของนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 พร้อมกัน จะพบว่าค่าความสามารถของแต่ละระดับชั้นเมื่อปรับ (Transformation) ให้อยู่ใน Scale เดียวกัน แล้วจะมีค่าใกล้เคียงกันหลายค่าเช่น ความสามารถของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 เท่ากับ 2.75 มีค่าเท่ากับ 1.85 และ 1.65 ของความสามารถชั้นมัธยมศึกษาปีที่ 2 และเท่ากับ 0.68 , 0.56 , 0.44 ของความสามารถชั้นมัธยมศึกษาปีที่ 3

2.5 การเทียบระดับคะแนน

ผู้วิจัยได้นำตาราง 7 ที่เป็นตารางเทียบความสามารถที่เท่ากันของแต่ละระดับชั้นไปเทียบหาค่าคะแนนดิบ ของแต่ละระดับชั้นจากภาพประกอบที่ 10-12 ได้ผลดังตาราง 8

ตาราง 8 การเทียบระดับคะแนน

X_1	X_2	X_3
4	1	
5-6	2	
7-10	2-4	1
11-13	5-6	2
14-17	7-9	3
18-20	10-11	4
21-22	12-13	5
23-24	14-15	6
25-26	16-17	7
27	18	8
28	19	9
29	20-21	10
30	22-23	11-12
31	24	13
32	25	14
33	26-27	15-16
34	28	17
35	29-30	18-19
36	31-32	20-23
37	33-34	24-26
38	35-36	27-30
39-40	37-38	31-34
	39-40	35

ตาราง 7 แสดงการเทียบระดับคะแนนของนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 ที่เรียงลำดับจากคะแนนต่ำสุดไปหาคะแนนสูงสุด โดยเทียบมาจากความสามารถที่เท่ากัน เมื่อพิจารณารายละเอียดจะพบว่า คะแนนของชั้นมัธยมศึกษาปีที่ 1 เริ่มตั้งแต่ 4 คะแนน ถึง 40 คะแนน ชั้นมัธยมศึกษาปีที่ 2 เริ่มจาก 1 คะแนน ถึง 40 คะแนน ชั้นมัธยมศึกษาปีที่ 3 เริ่มที่ 1 คะแนน ถึง 35 คะแนน โดยที่ 4 คะแนนของชั้นมัธยมศึกษาปีที่ 1 มีค่าเท่ากับ 1 คะแนนของชั้นมัธยมศึกษาปีที่ 2 แต่จะไม่เท่ากับคะแนนใดเลยในชั้นมัธยมศึกษาปีที่ 3 ส่วน 7-10 คะแนนของชั้นมัธยมศึกษาปีที่ 1 จะเท่ากับ 2-4 ของคะแนนในชั้นมัธยมศึกษาปีที่ 2 และเท่ากับ 1 คะแนน ของชั้นมัธยมศึกษาปีที่ 3

บทที่ 5

สรุปผล อภิปรายผล และข้อเสนอแนะ

จุดมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อสร้างแบบทดสอบวิชาคณิตศาสตร์ เรื่องสมการและอสมการสำหรับนักเรียนชั้นมัธยมศึกษาตอนต้นที่วิเคราะห์ด้วยรายค์โมเดล
2. เพื่อสร้างตารางเทียบระดับคะแนนตามแนวรายค์โมเดล สำหรับเปรียบเทียบคะแนนตามความสามารถในวิชาคณิตศาสตร์ เรื่องสมการและอสมการระหว่างนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นนักเรียนชั้นมัธยมศึกษาตอนต้นในภาคเรียนที่ 2 ปีการศึกษา 2530 ของโรงเรียนในสังกัดกรมสามัญศึกษา เขตท้องที่การศึกษา 2 กรุงเทพมหานคร จำนวน 2,510 คน ซึ่งเลือกมาโดยการสุ่มแบบแบ่งชั้น (Stratified random sampling)

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล ในการศึกษาค้นคว้าครั้งนี้ เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนเรื่องสมการและอสมการ จำนวน 3 ฉบับ ดังนี้

1. แบบทดสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1 จำนวน 40 ข้อ ซึ่งประกอบด้วยข้อสอบเฉพาะของชั้นมัธยมศึกษาปีที่ 1 จำนวน 30 ข้อ ข้อสอบร่วมกับชั้นมัธยมศึกษาปีที่ 2 จำนวน 10 ข้อ
2. แบบทดสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2 จำนวน 40 ข้อ ซึ่งประกอบด้วยข้อสอบเฉพาะของชั้นมัธยมศึกษาปีที่ 2 จำนวน 20 ข้อ ข้อสอบร่วมกับชั้นมัธยมศึกษาปีที่ 1 จำนวน 10 ข้อ และข้อสอบร่วมกับชั้นมัธยมศึกษาปีที่ 3 จำนวน 10 ข้อ

3. แบบทดสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 40 ข้อ ซึ่งประกอบด้วย
ข้อสอบเฉพาะของชั้นมัธยมศึกษาปีที่ 3 จำนวน 30 ข้อ ข้อสอบร่วมกับชั้นมัธยมศึกษาปีที่ 2 จำนวน
10 ข้อ

วิธีดำเนินการเก็บรวบรวมข้อมูล

ขั้นตอนในการดำเนินการเก็บรวบรวมข้อมูลในการศึกษาค้นคว้ามีดังนี้

1. วางแผนในการดำเนินการสอบ โดยติดต่อขอความร่วมมือในการนำเครื่องมือหรือ
แบบทดสอบไปทดลองกับนักเรียนกลุ่มตัวอย่าง ตลอดจนกำหนดวันและเวลาในการสอบ
2. นำแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนที่สร้างขึ้น 6 ฉบับ ระดับละ 2 ฉบับ ไปทดลอง
กับกลุ่มตัวอย่าง จำนวน 1,123 คน โดยนักเรียนแต่ละระดับชั้นจะทำแบบทดสอบคนละ 2 ฉบับ
3. นำแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนที่ได้คัดเลือกไว้อย่างดีแล้วไปทดสอบกับ
กลุ่มตัวอย่างที่ต้องการเทียบระดับคะแนนตามระดับนั้น ๆ จำนวน 1,387 คน

สรุปผลการวิเคราะห์ข้อมูล

จากการวิเคราะห์ข้อมูลแล้วสามารถสรุปผลการวิเคราะห์ได้ดังนี้

1. การคัดเลือกข้อสอบแต่ละระดับ จากการทดลองครั้งที่ 1 ปรากฏผลดังนี้
 - 1.1 แบบทดสอบสำหรับชั้นมัธยมศึกษาปีที่ 1 จำนวน 90 ข้อ มีค่าความยากตั้งแต่
-2.40 ถึง 3.43 มีค่า ERR IMPAC มากกว่า .05 จำนวน 9 ข้อ ค่าเฉลี่ยของความสามารถ
เท่ากับ 0.48 ดังนั้น ข้อสอบที่จะคัดเลือกไว้ต้องเป็นข้อสอบที่มีค่าความยากอยู่ระหว่าง -0.52 กับ
1.48 จึงเหลือข้อสอบที่จะนำมาพิจารณาค่าความสอดคล้องของค่า ICC และการกระจายของ
จุดประสงค์ 64 ข้อ ได้คัดเลือกข้อสอบที่เหมาะสมที่สุดไว้ 40 ข้อ
 - 1.2 แบบทดสอบสำหรับชั้นมัธยมศึกษาปีที่ 2 จำนวน 90 ข้อ มีค่าความยากตั้งแต่ -2.41
กับ 2.87 มีค่า ERR IMPAC มากกว่า .05 จำนวน 10 ข้อ ค่าเฉลี่ยของความสามารถเท่ากับ
-0.24 ดังนั้น ข้อสอบที่จะคัดเลือกไว้ต้องเป็นข้อสอบที่มีค่าความยากอยู่ระหว่าง -1.24 กับ 0.76

เหลือข้อสอบที่จะนำมาพิจารณาความสอดคล้องของค่า ICC และการกระจายของจุดประสงค์

58 ข้อ ได้คัดเลือกข้อสอบที่เหมาะสมที่สุดไว้ 30 ข้อ

1.3 แบบทดสอบสำหรับชั้นมัธยมศึกษาปีที่ 3 จำนวน 90 ข้อ มีค่าความยากตั้งแต่ -2.38 ถึง 2.98 มีค่า ERR IMPAC มากกว่า .05 จำนวน 9 ข้อ ค่าเฉลี่ยของความสามารถเท่ากับ -.35 ดังนั้น ข้อสอบที่คัดเลือกไว้ต้องเป็นข้อสอบที่มีค่าความยากอยู่ระหว่าง -1.35 กับ 0.65 เหลือข้อสอบที่จะนำมาพิจารณาความสอดคล้องของค่า ICC และการกระจายของจุดประสงค์ 64 ข้อ ได้คัดเลือกข้อสอบที่เหมาะสมที่สุดไว้ 30 ข้อ

2. การคัดเลือกข้อสอบรวม ได้พิจารณาข้อสอบที่มีค่าความยากสูงสุดของข้อสอบชั้นมัธยมศึกษาปีที่ 1 จำนวน 10 ข้อ เป็นข้อสอบร่วมระหว่างชั้นมัธยมศึกษาปีที่ 1 กับ 2 คัดข้อสอบที่มีค่าความยากสูงสุดของข้อสอบชั้นมัธยมศึกษาปีที่ 2 จำนวน 10 ข้อ เป็นข้อสอบร่วมระหว่างชั้นมัธยมศึกษาปีที่ 2 กับ 3

3. การจัดเรียงข้อสอบ ได้นำข้อสอบที่คัดเลือกไว้แล้วมาจัดเรียงดังนี้

3.1 แบบทดสอบสำหรับชั้นมัธยมศึกษาปีที่ 1 มีข้อ 1-30 เป็นข้อสอบเฉพาะข้อ 31-40 เป็นข้อสอบร่วมกับชั้นมัธยมศึกษาปีที่ 2

3.2 แบบทดสอบสำหรับชั้นมัธยมศึกษาปีที่ 2 มีข้อ 1-10 เป็นข้อสอบร่วมกับชั้นมัธยมศึกษาปีที่ 1 ข้อ 11-30 เป็นข้อสอบเฉพาะ ข้อ 31-40 เป็นข้อสอบร่วมกับชั้นมัธยมศึกษาปีที่ 3

3.3 แบบทดสอบสำหรับชั้นมัธยมศึกษาปีที่ 3 มีข้อ 1-10 เป็นข้อสอบร่วมกับชั้นมัธยมศึกษาปีที่ 2 ข้อ 11-40 เป็นข้อสอบเฉพาะ

4. ค่าสถิติพื้นฐานของแบบทดสอบทั้ง 3 ฉบับ พบว่า ค่าเฉลี่ยและการกระจายของคะแนนแบบทดสอบสำหรับชั้นมัธยมศึกษาปีที่ 1 กับปีที่ 3 มีค่าใกล้เคียงกัน และใกล้เคียงกับครึ่งของคะแนนเต็ม ส่วนแบบทดสอบสำหรับชั้นมัธยมศึกษาปีที่ 2 มีค่าเฉลี่ยและการกระจายของคะแนนต่ำกว่าทั้ง 2 ฉบับ

5. คุณภาพของข้อสอบรวม เมื่อพิจารณาเป็นรายข้อจะเห็นว่าข้อสอบของทั้ง 2 ระดับ เป็นข้อสอบที่มีค่าความยากสูง เมื่อสอบในระดับต่ำ และค่าความยากต่ำเมื่อนำไปสอบกับนักเรียนในระดับสูง และค่าความยากเฉลี่ยของข้อสอบรวมในแต่ละระดับมีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งสอดคล้องกับงานวิจัยของ สไลด์ และลินด์ (Slinde and Line 1978 : 23-55, 1979: 159-165)

6. การเทียบระดับความสามารถ เมื่อนำค่ายากเฉลี่ยของแต่ละระดับ ระดับตัวคงที่ที่จะนำไป Transformation ความสามารถของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ให้มี Scale เดียวกับชั้นมัธยมศึกษาปีที่ 2 คือ -1.0329 และค่าตัวคงที่ที่จะนำไป Transformation ความสามารถทางการเรียนของนักเรียนชั้นมัธยมศึกษาปีที่ 3 ให้มี Scale เดียวกับชั้นมัธยมศึกษาปีที่ 2 คือ $1,2398$ ได้ค่าความสามารถสูงสุดของนักเรียนชั้นมัธยมศึกษาปีที่ 1 คือ 3.93 มีค่าความสามารถเท่ากับ 3.09 กับ 2.65 ในชั้นมัธยมศึกษาปีที่ 2 และเท่ากับ 1.88 , 1.68 , 1.51 , และ 1.35 ของความสามารถในชั้นมัธยมศึกษาปีที่ 3 ความสามารถต่ำสุดของชั้นมัธยมศึกษาปีที่ 3 คือ -3.84 จะมีค่าเท่าความสามารถ -2.31 กับ -2.63 ในชั้นมัธยมศึกษาปีที่ 2 และเท่ากับความสามารถ -1.23 , -1.38 , -1.54 และ -1.72 ในชั้นมัธยมศึกษาปีที่ 1

7. การเทียบระดับคะแนน จะเรียงจากคะแนนต่ำสุดไปหาคะแนนมากที่สุด โดยคะแนนบางตอนจะอยู่ในรูป ช่วงคะแนน เริ่มจากคะแนนต่ำสุดของชั้นมัธยมศึกษาปีที่ 1 คือ ตั้งแต่ 4 คะแนน ถึง 40 คะแนน ชั้นมัธยมศึกษาปีที่ 2 เริ่มจาก 1 คะแนน ถึง 40 คะแนน ชั้นมัธยมศึกษาปีที่ 3 เริ่มจาก 1 คะแนน ถึง 35 คะแนน

ผลการวิเคราะห์ข้อมูลทั้งหมดจะได้ข้อสรุปว่า การสร้างแบบทดสอบสำหรับเทียบระดับคะแนนในแนวตั้ง (Vertical Equating) ในการศึกษาครั้งนี้ สามารถนำไปสร้างตารางเทียบระดับคะแนนเพื่อพัฒนาการของนักเรียนกลุ่ม เดิมที่ผ่านขึ้นไป เรียนชั้นสูงขึ้นว่านักเรียนมีความก้าวหน้าเป็นไปตามปกติหรือ เบี่ยงเบนไปตามปกติได้

อภิปรายผล

จากการศึกษาค้นคว้าครั้งนี้ จุดมุ่งหมายที่สำคัญคือต้องการสร้างตารางเทียบระดับคะแนนตามแนวตั้ง ซึ่งสามารถอภิปรายผลได้ดังนี้

1. คุณภาพของข้อสอบร่วม ผลจากการวิเคราะห์ค่าความยากของข้อสอบร่วมพบว่า ข้อสอบร่วมในชั้นมัธยมศึกษาปีที่ 1 กับมัธยมศึกษาปีที่ 2 มีค่าความยากรายข้อและค่าแตกต่างของค่าเฉลี่ยสอดคล้องกันอย่างมีนัยสำคัญทางสถิติที่ระดับ $.01$ ข้อสอบร่วมของระดับชั้นมัธยมศึกษาปีที่ 2 กับมัธยมศึกษาปีที่ 3 มีค่าความแตกต่างของค่าเฉลี่ยสอดคล้องกันอย่างมีนัยสำคัญทางสถิติที่ระดับ $.01$ แต่เมื่อพิจารณาค่า

ความยากง่ายข้อพบว่ามีข้อสอบ 1 ข้อที่มีค่าความยากในระดับต่ำกว่าค่าความยากในระดับสูง ซึ่งเมื่อพิจารณาโจทย์ปัญหาแล้วพบว่าเป็นเรื่องการแก้สมการ ที่มีเงื่อนไขว่าตัวหารต้องไม่เป็นศูนย์ ซึ่งเป็นบทเรียนในเรื่องระบบจำนวนเต็มของชั้นมัธยมศึกษาปีที่ 2 ซึ่งนักเรียนชั้นมัธยมศึกษาปีที่ 3 ได้เรียนมานาน ทำให้ไม่ระวังหรือลืมคำนึงถึงเงื่อนไขข้อนี้ แต่ข้อสอบรวมทั้งหมดก็มีค่าความยากต่างกัน ตามความต้องการคือมีค่าความยากที่สูงสำหรับนักเรียนชั้นต่ำและมีค่าความยากที่ต่ำในกลุ่มนักเรียนระดับสูง ซึ่งแสดงว่าการสร้างข้อสอบร่วมโดยตัดข้อสอบที่ยากสำหรับชั้นต่ำมา เป็นข้อสอบร่วม เป็นวิธีที่ใช้ได้สำหรับข้อสอบที่มีเนื้อหาต่อเนื่องกัน และในกรณีที่มีข้อสอบไม่เหมาะสมกับโมเดลอยู่เพียง 1 ข้อ นั้นก็ไม่มีผลทำให้การเทียบระดับคะแนนนั้นเปลี่ยนแปลงไป ซึ่งสอดคล้องกับงานวิจัยของ ฟิลลิปส์ (Phillip. 1986: 107-118)

2. ตารางเทียบระดับความสามารถของผู้สอบ ผลการวิเคราะห์ในครั้งนี้พบว่าค่าความสามารถของนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 มีค่าค่อนข้างจะใกล้เคียงกัน ซึ่งสาเหตุจากเนื้อหาในเรื่องสมการและอสมการเป็นเรื่องที่ไม่แตกต่างกัน ในแต่ละชั้นยังต้องใช้หลักในการแก้สมการเดียวกัน เพียงแต่ความซับซ้อนของโจทย์มากขึ้นเท่านั้น และเรื่องนี้ เป็นเรื่องที่มีพื้นฐานซึ่งได้ศึกษามาตั้งแต่ชั้นประถมศึกษาไปแล้ว จึงทำให้ค่าความสามารถของทั้ง 3 ระดับใกล้เคียงกัน แต่ค่าความสามารถที่ได้จากตารางยังเป็นตัวเลขที่บอกได้ว่านักเรียนชั้นมัธยมศึกษาปีที่ 1 ต้องใช้ความสามารถสูงกว่านักเรียนชั้นมัธยมศึกษาปีที่ 2 ในการแก้โจทย์ปัญหาเรื่องสมการและอสมการ นักเรียนชั้นมัธยมศึกษาปีที่ 2 ต้องใช้ความสามารถสูงกว่านักเรียนชั้นมัธยมศึกษาปีที่ 3 แก้โจทย์ปัญหาเรื่องสมการและอสมการ หรือโจทย์สมการและอสมการนั้น นักเรียนชั้นมัธยมศึกษาปีที่ 1 ต้องใช้ความสามารถมากกว่านักเรียนชั้นมัธยมศึกษาปีที่ 2 และ 3 ส่วนนักเรียนชั้นมัธยมศึกษาปีที่ 2 ก็จะใช้ความสามารถมากกว่านักเรียนชั้นมัธยมศึกษาปีที่ 3 ซึ่งสอดคล้องกับผลงานวิจัยของ มาร์โค (Marco. 1977 : 155)

3. ตารางเทียบระดับคะแนน ผลการวิเคราะห์พบว่า คะแนนของนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 ที่เทียบจากความสามารถเดียวกันมีค่าใกล้เคียงกันเนื่องจากเนื้อหาของเรื่องที่ศึกษาเป็นเรื่องที่ต่อเนื่อง แต่ค่าของคะแนนก็สอดคล้องกับผลงานวิจัยของ มาร์โค (Marco. 1977 : 155) ลอร์ดและฮูเวอร์ (Loyd and Hoover. 1980 : 186-187)

ข้อเสนอแนะ

จากการศึกษาค้นคว้าครั้งนี้ ทำให้ผู้วิจัยได้แนวคิดที่จะเสนอแนะ ดังนี้

1. ควรจะมีการนำวิธีการสร้างแบบทดสอบและตารางเทียบระดับคะแนนนี้ไปใช้ในการสร้างวิชาอื่น ๆ เพื่อช่วยให้นักเรียนและอาจารย์ผู้สอนทราบพัฒนาการในการเรียนการสอนของตน
2. การสร้างแบบทดสอบและตารางเทียบระดับคะแนนในเรื่องที่เนื้อหาไม่ต่อเนื่องกัน หรือใช้หลักการเดียวกัน การสร้างข้อสอบร่วมควรได้มีการทดลองกับกลุ่มตัวอย่างทั้งสองกลุ่ม
3. ควรจะมีการนำวิธีการสร้างตารางเทียบระดับคะแนนไปใช้เป็นแนวทางในการศึกษา และสร้างแบบทดสอบที่มีตารางเทียบระดับคะแนนในแนวอื่นที่เป็นชุดหรือมีหลายฉบับในเนื้อหาเดียวกัน เพื่อที่จะนำไปใช้ในการสอน เป็นการป้องกันการทุจริตได้
4. ควรจะมีการศึกษา เปรียบเทียบรูปแบบการเทียบระดับคะแนนโดยวิธีนี้กับการเทียบระดับคะแนนที่เทียบจากการปรับค่าความยากของข้อสอบที่ ไวท์ และ สโตน ได้เสนอไว้ (Wright and Stone. 1979:: 110-117) ว่ารูปแบบใดมีความเหมาะสมมากที่สุด

บรรณานุกรม

- ชวาล แพร์ตกุล เทคนิคการวัดผล พิมพ์ครั้งที่ 6 วัฒนาพานิช 2518, 432 หน้า
- ชูชีพ พงษ์สมบุญ การเปรียบเทียบการเทียบมาตราระหว่างรูปแบบที่ใช้ผู้สอบร่วมกับรูปแบบที่ใช้แบบสอบร่วม วิทยานิพนธ์คุรุศาสตรมหาบัณฑิต จุฬาลงกรณ์มหาวิทยาลัย 2518, 118 หน้า
- ✓ ชุศักดิ์ ชัมภลชิต Test Equating procedures เอกสารประกอบคำบรรยายในโครงการสัมมนาระดับชาติ ครั้งที่ 4 โครงการพัฒนาแบบทดสอบ โครงการพัฒนาการศึกษาอาเซียน 2527, 118 หน้า
- ชูศรี วงศ์รัตน์ สถิติเพื่อการวิจัย คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร 2525, 370 หน้า
- เดโช สนวนานนท์ จิตวิทยาสำหรับครู มงคลการพิมพ์ 2523, 452 หน้า
- ปนัดดา วัชวณะ การศึกษาผลวิเคราะห์ข้อสอบโดยใช้วิธีโลจิสติกโมเดลกับวิธีแบบเดิม (คลาสสิกคอลโมเดล) ในวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 วิทยานิพนธ์ ศศ.ม มหาวิทยาลัยเชียงใหม่ 2528, 71 หน้า
- ภาวิณี ศรีวัฒนานันท์ การเปรียบเทียบผลจากการใช้รูปแบบการเทียบมาตราที่ต่างกับเมื่อแบบทดสอบรวมมีความยาวเท่ากัน วิทยานิพนธ์ ด.ด จุฬาลงกรณ์มหาวิทยาลัย 2529, 147 หน้า
- เรวดี อินทสระ การเปรียบเทียบความคลาดเคลื่อนของการเทียบมาตราระหว่างรูปแบบอิงทฤษฎีการตอบข้อสอบกับรูปแบบการใช้เทคนิคการวิเคราะห์องค์ประกอบ วิทยานิพนธ์ ศศ.ม มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร 2530, 69 หน้า 371.26013 57690.
- ล้วน สายยศ และ อังคณา สายยศ หลักการวิจัยทางการศึกษา ทวีกิจการพิมพ์ 2524, 287 หน้า
- สงวน ลักษณะ "ความสามารถเกี่ยวกับการตอบข้อสอบ (Item Response Theory)" วารสารการวัดผลการศึกษา 4(1) : 47-54, 88-94 พฤษภาคม-สิงหาคม 2525

สงข ลักษณ์ "การเทียบระดับคะแนนระหว่างแบบทดสอบ (Equating)" วารสารการวัดผลการศึกษา 4(2) : 21-32 กันยายน-ธันวาคม 2525

สำเร็จ บุญเรืองรัตน์ "การพัฒนาทฤษฎีเลเท็น เทรทเพื่อวิเคราะห์ข้อสอบ" วารสารการวัดผลการศึกษา 7(21) : 41-68 มกราคม-เมษายน 2529

สุธรรม จันทร์ทอม การวิเคราะห์ข้อสอบแบบราชคโหมเดล คณะศึกษาศาสตร์ มหาวิทยาลัยเชียงใหม่ 2528, 38 หน้า (อัดสำเนา)

_____ "วิเคราะห์ผลการสอบแบบราชคโหมเดล" วารสารการวัดผลการศึกษา 4(3) : 45-61 มกราคม-เมษายน 2526.

อนันต์ ศรีโสภะ การวัดและประเมินผลการศึกษา พิมพ์ครั้งที่ 2 ไทยวัฒนาพานิช 2520, 250 หน้า

อวยพร วิบูลย์กาญจน์ การเปรียบเทียบผลการวิเคราะห์แบบสอบอุปมาอุปไมยด้วยคลาสสิกคอลโมเดลกับราชคโหมเดล วิทยานิพนธ์ ค.ม. จุฬาลงกรณ์มหาวิทยาลัย 2525, 61 หน้า

Angoff, W.H. "Summary and Derivation of Equating Methods Used at ETS" in Test Equating p 55-69, ed. by P.H. Holland and D.B. Rubin. New York, Academic Press, 1982

Flanagan T.C. Educational Measurement Washington D.C. American Council on Education 1951

Gustafsson Jan-Eric "The Rasch Model in Vertical Equating of Tests:A Critique of Slinde and Linn" Journal of Educational Measurement 16(3) : 153-158 FALL 1979

Gulliksen, Harold Theory of Mental Tests New York Willey 1950, 486 p.

Holmes S.E. "Unidimensionality and Vertical Equating with the Rasch Model" Journal of Educational Measurement 19(2) 139-147 Summer 1982

Kolen M.J. "Comparison of Traditional and Item Response Theory Methods for Equating Tests" Journal of Educational Measurement 18(1) : 2-11 Spring 1981

Kolen M.J. and Writney D.R. "Comparison of Four Procedures for Equating the "Tests of General Educational Development." Journal of Educational Measurement 19(4) : 279-293 Winter 1982

- Laksana Sangob A Comparison of an ANOVA Approach and ICC Approach for Assending Item Bias in an Achievement Test IOWA Testing Programs Occasional Papers Number 29 September 1980, 37 p.
- Lord F.M. Applications of Item Response Theory to Practical Testing Problem. Hillsdale New Jersey : Erlbaum 1980, 247 p.
- Loyd B.H. and Hoover H.D. "Vertical Equating Using the Rasch Model" Journal of Educational Measurement 17(3) : 179-193 Fall 1980
- Marco G.L. "Item Characteristic Curve Solutions to Three Intractable Testing Problems" Journal of Educational Measurement 14(2) : 139-160 Summer 1977
- Nancy S.P. Gary L.M and Elizabeth I.S "A Test of the Adequacy of Linear Score Equating Model" in Test Equating p 73-135. ed. by P.H. Holland and D.B. Rubin, New York, Academic Press, 1982
- Skaggs Gary and Lissitz R.W. "An Exploration of the Robustness of Four Test Equating Model" Applied Psychological Measurement 10(3) : 303-317 September 1986
- Slinde J.A. and Linn R.L. "An Exploration of Adequacy of the Rasch Model for the Problem of Vertical Equation" Journal of Educational Measurement 1(15) : 23-35 Spring 1987
- ____ "An Note on Vertical Via the Rasch Model Oups of Quite Different Ability and Tests of Quits Different Difficulty" Journal of Education Measurement 16(3) : 159-165 Fall 1979
- Petersen, Marco and Steward "A Test of Adequacy of Linear Score Equating Model" In Test Equating p 71-135 ed. by P.H. Holland and D.B. Rubin, New York, Academic Press, 1982
- Phillips S.E. "The Effects of the Deletion of Misfitting Persons on Vertical Equating Via the Rasch Model" Journal of Education Measurement 23(2) : 107-118 Summer, 1986
- Whitely S.E. "Models Meanings and Misunderstandings Some Issures Applying Rasch's Theory" "Journal of Education Measurement"14(3) : 228-235 Number 1977
- Wright B.D. "Solving Measurement Problems with the Rasch Model" Journal of Education Measurement 14(2) : 97-116 Summer 1977
- Wright B.D. and Stone M.H. Best Test Design MESA Press Chicago 1979, 223 p

ภาคผนวก ก.

คู่มือการสอบ แบบทดสอบ และตารางเทียบระดับคะแนนในแนวดิ่ง

ความมุ่งหมายของแบบทดสอบ

แบบทดสอบและตารางเทียบระดับคะแนนในแนวดิ่ง วิชาคณิตศาสตร์ เรื่องสมการและอสมการนี้ สร้างขึ้นเพื่อใช้สอบวัดการพัฒนาความสามารถทางการเรียนเรื่องสมการและอสมการของนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 ว่านักเรียนมีการพัฒนาไปในทางที่ก้าวหน้าหรือเบี่ยงเบนไปจากปกติ ซึ่งจะเป็นประโยชน์ต่อนักเรียนและอาจารย์ผู้สอนที่จะได้หาทางปรับปรุงแก้ไขต่อไป

โครงสร้างของแบบทดสอบ

แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ เรื่องสมการและอสมการนี้เป็นแบบทดสอบชนิดเลือกตอบมี 5 ตัวเลือก จำนวน 3 ฉบับ คือ

1. แบบทดสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1 จำนวน 40 ข้อ โดยที่ข้อ 31-40 เป็นข้อสอบร่วมกับระดับชั้นมัธยมศึกษาปีที่ 2
2. แบบทดสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2 จำนวน 40 ข้อ โดยที่ข้อ 1-10 เป็นข้อสอบร่วมกับระดับชั้นมัธยมศึกษาปีที่ 1 ข้อ 31-40 เป็นข้อสอบร่วมกับชั้นมัธยมศึกษาปีที่ 3
3. แบบทดสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 3 จำนวน 40 ข้อ โดยที่ข้อ 1-10 เป็นข้อสอบร่วมกับระดับชั้นมัธยมศึกษาปีที่ 2

การพัฒนาแบบทดสอบ

ในการสร้างแบบทดสอบนี้ ได้เริ่มดำเนินการสร้างตั้งแต่เดือนธันวาคม 2530 ถึง เมษายน 2531 จึงสำเร็จตามต้องการ โดยดำเนินการสอบปรับปรุงคัดเลือกและจัดแบบทดสอบ ดังนี้

การทดสอบครั้งที่ 1 ทดลองกับนักเรียนชั้นมัธยมศึกษาตอนต้น ปีการศึกษา 2530 จำนวน 1,123 คน โดยแยกเป็นชั้นมัธยมศึกษาปีที่ 1 จำนวน 359 คน มัธยมศึกษาปีที่ 2 จำนวน 390 คน มัธยมศึกษาปีที่ 3 จำนวน 374 คน โดยใช้แบบทดสอบ 6 ฉบับ ๆ ละ 45 ข้อ เป็นข้อสอบระดับละ

2 ฉบับ รวม 90 ข้อ แล้วมาวิเคราะห์ตามแนวราชคโหมเดล เพื่อหาค่าความยากและค่าความสามารถ โดยใช้โปรแกรมสำเร็จรูป BICAL แล้วคัดเลือกข้อสอบที่มีค่า ERR IMPAC น้อยกว่า .05 ค่า ITEM CHARACTERISTIC CURVE (ICC) สอดคล้องกับกลุ่มตัวอย่างเมื่อแบ่งกลุ่มตัวอย่างตามลำดับคะแนนเป็น 6 กลุ่ม มีค่าสัมบูรณ์ของความแตกต่างระหว่างค่าความสามารถกับค่าความยากน้อยกว่า 1 ($|\beta - \delta| < 1$) และพิจารณาการกระจายของจุดประสงค์ได้ข้อสอบสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1 จำนวน 40 ข้อ ได้นำข้อที่มีค่าความยากสูงสุดมาเป็นข้อสอบร่วมระหว่างชั้นมัธยมศึกษาปีที่ 1 กับ 2 จำนวน 10 ส่วนข้อสอบสำหรับชั้นมัธยมศึกษาปีที่ 2 จำนวน 30 ข้อ นำข้อที่มีค่าความยากสูงสุดของระดับจำนวน 10 ข้อ เป็นข้อสอบร่วมระหว่างชั้นมัธยมศึกษาปีที่ 2 กับ 3 และข้อสอบสำหรับชั้นมัธยมศึกษาปีที่ 3 จำนวน 30 ข้อ เป็นข้อสอบ 3 ระดับ ระดับละ 1 ฉบับ ฉบับละ 40 ข้อ

การทดสอบครั้งที่ 2 ทดสอบกับนักเรียนชั้นมัธยมศึกษาตอนต้น ปีการศึกษา 2530 จำนวน 1,387 คน แบ่งเป็นชั้นมัธยมศึกษาปีที่ 1 จำนวน 490 คน มัธยมศึกษาปีที่ 2 จำนวน 482 คน มัธยมศึกษาปีที่ 3 จำนวน 415 คน นำผลการสอบมาวิเคราะห์ด้วยโปรแกรมสำเร็จรูป BICAL เพื่อหาค่าความยากของข้อสอบร่วมและความสามารถของแต่ละระดับคะแนน แล้วนำค่าความยากเฉลี่ยของข้อสอบร่วมแต่ละระดับมาวิเคราะห์โดยใช้หลัก Liner equation และสูตรของ ลอยด์ และ ฮูเวอร์ (Loyd and Hoover 1980 : 186) เพื่อ transformation ความสามารถของนักเรียนชั้นมัธยมศึกษาปีที่ 1 กับ 3 ให้มี Scale เดียวกับชั้นมัธยมศึกษาปีที่ 2

คุณภาพของข้อสอบรวม

คุณภาพของข้อสอบแต่ละข้อนั้นจะคัดเลือกจากข้อที่มีค่า ERR IMPAC ต่ำกว่า .05, ค่า ICC สอดคล้องกับกลุ่มตัวอย่างที่แบ่งตามลำดับคะแนน, ค่า $|\beta - \delta| < 1$, การกระจายของจุดประสงค์ และความเชื่อมั่น (Reliability) ของแบบทดสอบทั้งฉบับ โดยที่ชั้นมัธยมศึกษาปีที่ 1 มีค่าความเชื่อมั่นเท่ากับ .84 ชั้นมัธยมศึกษาปีที่ 2 มีค่าเท่ากับ .72 ชั้นมัธยมศึกษาปีที่ 3 มีค่าเท่ากับ .83 ส่วนในการ Transformation ของแต่ละระดับจะต้องคำนึงถึงคุณภาพของข้อสอบรวม ซึ่งมีรายละเอียดดังนี้

ตารางที่ 1 ค่าสถิติพื้นฐานความยากของข้อสอบร่วม

ระดับชั้น	ข้อสอบร่วมระหว่างชั้น ม.1 กับ ม.2		ข้อสอบร่วมระหว่างชั้น ม.2 กับ ม.3	
	$\bar{X}$	SD	$\bar{X}$	SD
ม.1	0.8146	0.2356	-	-
ม.2	-0.2183	0.4087	0.8994	0.4549
ม.3	-	-	-0.3404	0.9472
ความแตกต่างของ ค่าเฉลี่ย	1.0329**		1.2398**	

** มีนัยสำคัญทางสถิติที่ระดับ .01

วิธีดำเนินการสอน

การดำเนินการสอนแบบทดสอบมีลำดับขั้นดังนี้

1. การเตรียมตัวก่อนสอบ ให้ปฏิบัติดังนี้

1.1 กำหนดวัน เวลา สถานที่สอบ จัดกรรมการคุมสอบ และผู้ตรวจคำตอบ

1.2 จัดเตรียมอุปกรณ์การสอบ ได้แก่ แบบทดสอบ กระดาษคำตอบ เฉลย ตารางเทียบระดับคะแนน ให้มีจำนวนเพียงพอสำหรับผู้เข้าสอบและผู้ตรวจ

1.3 ผู้ดำเนินการสอบต้องศึกษา คำชี้แจง วิธีดำเนินการสอบ การตรวจกระดาษคำตอบ การเทียบระดับคะแนนให้เข้าใจถูกต้องก่อน

2. วิธีดำเนินการสอบให้ปฏิบัติดังนี้

2.1 แจกกระดาษคำตอบพร้อมแบบทดสอบให้ตรงกับระดับชั้นที่ต้องการทดสอบ ให้นักเรียนกรอกรายการ

2.2 อ่านคำชี้แจงที่ปกหน้า แบบทดสอบให้นักเรียนฟัง แล้วให้นักเรียนลงมือทำ ใช้เวลาตามที่ปรากฏในปกหน้าข้อสอบ

3. การเตือนเวลา ให้ปฏิบัติดังนี้ เตือนเวลาครั้งแรกเมื่อเวลาผ่านไปครึ่งหนึ่งของแบบทดสอบ เตือนครั้งที่ 2 เมื่อเหลือเวลาอีก 5 นาทีจะหมดเวลา

4. วิธีปฏิบัติเมื่อหมดเวลา ให้ผู้สอบวางดินสอหรือปากกาแล้วเก็บกระดาษคำตอบและแบบทดสอบ

วิธีการตรวจให้คะแนน

คำตอบที่ถูกให้ 1 คะแนน ตอบผิดหรือเว้นว่างหรือใส่เครื่องหมายมากกว่า 1 ที่ให้ 0 คะแนน แล้วรวมคะแนนทั้งฉบับ

ตารางเทียบคะแนน

ตารางเทียบระดับคะแนนของแบบทดสอบ 3 ฉบับนี้ เป็นตารางที่เทียบจากความสามารถที่เท่าเทียมกัน กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาตอนต้น ปีการศึกษา 2530 เขตท้องที่การศึกษา 2 กรุงเทพมหานคร จำนวน 1,387 คน

ตารางที่ 2 ตาราง เทียบระดับคะแนน

ม.1	ม.2	ม.3
4	1	
5-6	2	
7-10	3-4	1
11-13	5-6	2
14-17	7-9	3
18-20	10-11	4
21-22	12-13	5
23-24	14-15	6
25-26	16-17	7
27	18	8
28	19	9
29	20-21	10
30	22-23	11-12
31	24	13
32	25	14
33	26-27	15-16
34	28	17
35	29-30	18-19
36	31-32	20-23
37	33-34	24-26
38	35-36	27-30
39-40	37-38	31-34
	39-40	35

การแปลผล

คะแนนที่อยู่ในแนวเดียวกัน แสดงว่านักเรียนมีความสามารถเท่ากัน เช่น เด็กชาย ก. ได้คะแนนสอบแบบทดสอบสำหรับชั้น ม.1 ได้ 28 คะแนน ได้คะแนนแบบทดสอบสำหรับ ม.2 ได้ 19 คะแนน และแบบทดสอบสำหรับชั้น ม.3 ได้ 9 คะแนน แสดงว่าเด็กชาย ก. ไม่มีความก้าวหน้าในเรื่องสมการและอสมการเลยเพราะเด็กชาย ก. มีความสามารถเท่าเดิม

เด็กชาย ข. ได้คะแนนสอบแบบทดสอบสำหรับชั้น ม.1 ได้ 28 คะแนน แบบทดสอบสำหรับชั้น ม.2 ได้ 24 คะแนน และแบบทดสอบสำหรับชั้น ม.3 ได้ 14 คะแนน แสดงว่าเด็กชาย ข. มีความสามารถในเรื่องสมการและอสมการก้าวหน้าขึ้น

เด็กชาย ค. ได้คะแนนสอบแบบทดสอบสำหรับชั้น ม.1 จำนวน 28 คะแนน แต่ทำแบบทดสอบสำหรับชั้น ม.2 ได้ 17 คะแนน แสดงว่าเด็กชาย ค. มีความบ่่ายเบนที่ผิดปกติในเรื่องสมการและอสมการ ต้องหาทางศึกษาหาสาเหตุเพื่อแก้ไขต่อไป

$$\begin{array}{r}
 18 \\
 22 \quad 54 \\
 \hline
 90 \\
 18 \quad 6 \\
 \hline
 108 \quad 0
 \end{array}$$

ภาคผนวก ข.

ตาราง 3 แสดงจุดประสงค์การเรียนรู้ จุดประสงค์เชิงพฤติกรรม และเนื้อหาของหลักสูตรมัธยมศึกษา
ตอนต้น

จุดประสงค์การเรียนรู้	จุดประสงค์เชิงพฤติกรรม	เนื้อหา
นักเรียนสามารถแก้สมการ และตรวจสอบคำตอบได้	ม.1 นักเรียนสามารถ 1. หาคำตอบของสมการอย่างง่าย ๆ โดย การลองค่าตัวแปรในสมการ 2. บอกได้ว่าจำนวนที่กำหนดให้เป็นคำตอบ ของสมการหรือไม่ 3. แก้สมการโดยนำคุณสมบัติการเท่ากันมา ใช้หาคำตอบและตรวจสอบคำตอบได้	ม.1 สมการอยู่ในรูปง่าย คือ $ax+b = c$ เมื่อ a, b, c เป็น จำนวนเต็ม x เป็น ตัวแปร
	ม.2 1. หาคำตอบของสมการอย่างง่าย ๆ โดย ลองแทนค่าตัวแปรด้วยจำนวนต่าง ๆ 2. บอกได้ว่าจำนวนที่กำหนดให้เป็นคำตอบ ของสมการ 3. เขียนกราฟแสดงคำตอบของสมการได้ 4. อ่านคำตอบของสมการจากกราฟได้ 5. ใช้คุณสมบัติการเท่ากันแก้สมการได้ 6. แก้สมการที่กำหนดให้ได้อย่างถูกต้อง และรวดเร็ว	ม.2 สมการอยู่ในรูป $ax+b = c$ $ax+bx+cx = d$ $a(x+b) = c$ เมื่อ a, b, c, d เป็น จำนวนจริง x เป็น ตัวแปร
	ม.3 นักเรียนสามารถ 1. แก้สมการเชิงเส้นตัวแปรเดียวได้ 2. บอกได้ว่าจำนวนที่กำหนดให้เป็น	ม.3 สมการอยู่ในรูปที่ซับซ้อน มีวงเล็บ, เศษส่วน แต่ค่าตัวแปรเป็น

จุดประสงค์การเรียนรู้	จุดประสงค์เชิงพฤติกรรม	เนื้อหา
ใช้สมการแก้โจทย์ปัญหาต่าง ๆ ได้อย่างถูกต้องและรวดเร็ว	คำตอบของสมการ ม. 1 - นักเรียนสามารถเขียนสมการแทนข้อความที่กำหนดให้ได้ - นักเรียนสามารถใช้สมการแก้โจทย์ปัญหาง่าย ๆ ได้อย่างถูกต้อง ม. 2 นักเรียนสามารถ - เปลี่ยนประโยคภาษาเป็นประโยคสัญลักษณ์ได้ - เปลี่ยนประโยคสัญลักษณ์เป็นประโยคภาษาได้ - สร้างสมการจากโจทย์ที่กำหนดให้และหาคำตอบตามโจทย์ที่ต้องการได้อย่างถูกต้อง ม. 3 นักเรียนสามารถหาคำตอบของโจทย์ปัญหาที่เป็นสมการเชิงเส้นตัวแปรเดียวได้	คิกรหนึ่ง
นักเรียนสามารถแก้สมการได้อย่างถูกต้อง	ม. 2 นักเรียนสามารถ - บอกได้ว่าประโยคใดเป็นสมการและประโยคใดเป็นสมการ - หาคำตอบของสมการได้โดยการแทนค่าตัวแปรด้วยจำนวนต่าง ๆ	ม. 2 สมการในรูปตัวแปร เป็นกำลังและมี สัมประสิทธิ์เป็นหนึ่ง $x+b > c$

จุดประสงค์การเรียนรู้	จุดประสงค์เชิงพฤติกรรม	เนื้อหา
	3. เขียนกราฟแสดงคำตอบของสมการได้ 4. อ่านคำตอบของสมการได้ 5. แก้อสมการได้อย่างถูกต้องและรวดเร็ว โดยใช้คุณสมบัติการบวกของการไม่เท่ากัน ม.3	ม.3 อสมการในรูปตัวแปร มีสัมประสิทธิ์ไม่ใช่ หนึ่งต้องใช้คุณสมบัติ การคูณ

ตาราง 4 แสดงจุดประสงค์เชิงพฤติกรรมของแต่ละระดับชั้น

จุดประสงค์	ระดับชั้น		
	ม.1	ม.2	ม.3
นักเรียนสามารถ			
1. บอกคุณสมบัติการเท่ากันได้	✓		
2. แก้อสมการหาคำตอบได้ถูกต้อง	✓	✓	✓
3. บอกได้ว่าจำนวนที่กำหนดให้เป็นคำตอบของสมการ	✓	✓	✓
4. เขียนกราฟแสดงคำตอบได้		✓	
5. เปลี่ยนประโยคภาษาเป็นประโยคสัญลักษณ์ได้	✓	✓	✓
6. เปลี่ยนประโยคสัญลักษณ์เป็นประโยคภาษาได้	✓	✓	✓
7. หาคำตอบของสมการจากโจทย์ปัญหาได้	✓	✓	✓
8. แก้อสมการหาคำตอบได้ถูกต้อง		✓	✓
9. เขียนกราฟแสดงคำตอบของสมการได้		✓	
10. หาคำตอบของสมการจากโจทย์ปัญหาได้		✓	✓

การสร้างแบบทดสอบและตารางเทียบระดับคะแนนแนวดิ่งในวิชาคณิตศาสตร์

บทคัดย่อ

ของ

อาภรณ์ กาญจนกิจโสภณ

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

เพื่อเป็นส่วนหนึ่งของการศึกษาค้นคว้าตามหลักสูตร

ปริญญาการศึกษามหาบัณฑิต

เมษายน 2531

การศึกษาครั้งนี้มีจุดมุ่งหมายเพื่อสร้างแบบทดสอบและตารางเทียบระดับคะแนนแนวตั้ง
 ในวิชาคณิตศาสตร์ เรื่องสมการและอสมการที่วิเคราะห์ด้วยราชคโหมเดล กลุ่มตัวอย่างเป็นนักเรียน
 ชั้นมัธยมศึกษาปีที่ 1, 2 และ 3 ปีการศึกษา 2530 ของโรงเรียนสังกัดกรมสามัญศึกษา ในเขต
 ท้องที่การศึกษา 2 กรุงเทพมหานคร จำนวน 2,510 คน โดยใช้วิธีสุ่มแบบแบ่งชั้น วิธีการศึกษา
 ประกอบด้วยการทดสอบครั้งที่ 1 เพื่อประมาณค่าพารามิเตอร์ของข้อสอบ และค่าความสามารถ
 ตามแบบราชคโหมเดลที่วิเคราะห์โดยโปรแกรมสำเร็จรูป BICAL แล้วคัดเลือกข้อสอบที่มีค่า
 ERR IMPACT น้อยกว่า .05 มีค่า ICC (Item characteristic curve) สอดคล้องกับกลุ่ม
 ตัวอย่าง เมื่อแบ่งกลุ่มตัวอย่างตามระดับคะแนนเป็น 6 กลุ่ม และพิจารณาค่าสัมบูรณ์ของความแตกต่าง
 ระหว่างความยากกับความสามารถเฉลี่ยน้อยกว่า 1 ($|\beta - \delta| < 1$) ได้แบบทดสอบ 3 ฉบับ ๆ ละ
 40 ข้อ โดยแบบทดสอบสำหรับชั้นมัธยมศึกษาปีที่ 1 กับมัธยมศึกษาปีที่ 2 มีข้อสอบร่วมกัน 10 ข้อ
 ชั้นมัธยมศึกษาปีที่ 2 กับมัธยมศึกษาปีที่ 3 มีข้อสอบร่วมกัน 10 ข้อ การทดสอบครั้งที่ 2 เพื่อวิเคราะห์
 หาค่าความสามารถและค่าความยากเฉลี่ยของข้อสอบร่วม เพื่อนำผลมาปรับค่าความสามารถของ
 นักเรียนชั้นมัธยมศึกษาปีที่ 1 และ 3 ให้อยู่ใน Scale เดียวกับชั้นมัธยมศึกษาปีที่ 2 โดยใช้หลัก
 Linear equation หลังจากนั้นผู้วิจัยได้สร้างตารางเทียบระดับคะแนนของนักเรียนชั้นมัธยมศึกษา
 ปีที่ 1, 2 และ 3 จากความสามารถที่เท่ากัน

ผลการศึกษาจะได้ตารางเทียบระดับคะแนนของนักเรียนชั้นมัธยมศึกษาปีที่ 1, 2 และ 3
 ซึ่งเทียบจากความสามารถที่เท่ากัน เช่น ความสามารถ .54 ของนักเรียนชั้นมัธยมศึกษาปีที่ 2
 ซึ่งมีคะแนน 25 คะแนน จะเท่ากับ 32 คะแนน ในชั้นมัธยมศึกษาปีที่ 1 และ 14 คะแนนในชั้น
 มัธยมศึกษาปีที่ 3 ซึ่งสามารถนำตารางเทียบระดับคะแนนที่ไปใช้ในการศึกษาความก้าวหน้าทาง
 การเรียนการสอนเรื่องสมการและอสมการของนักเรียนได้

A CONSTRUCTION TEST AND TABLE VERTICAL
EQUATING IN MATHEMATICS

AN ABSTRACT

BY

APORN KANJANAKITSOPON

Presented in partial fulfillment of the requirements
for the Master of Education degree
at Srinakharinwirot University
April 1988

The purpose of this studying was to construct a set of test items and vertical equating table addressing to the constant of equation and inequation in Mathematics. Estimates of item difficulty parameter and estimates of examinee's ability was obtained from the BICAL computer program utilizing the options which permit estimates under the Rasch model. The samples of the study, three level, all 2510 Mathayom Suksa I, II and III students in academic year 1987 from school's in Bangkok. These sample were randomized stratifically. The phases of the test development consist of 2 testings. The first one was to estimate item difficulty, ability mean of an examinee. All test of the three levels were then selected in terms of content, appropriate difficulty ($|\beta - \delta| < 1$), err impact less than .05 and item characteristic curve. There are three levels of items. Each level is composed of 40 items: adjacent levels (M.S 1, M.S 2, M.S 2 and M.S 3) have 10 items in common. The second testing was administered to estimate the examinees's abilites and mean of difficulty for common items. Estimate's abilities of M.S 1 and M.S 3 and transformed to the scale of ability estimates calibrated for M.S 2.

The results of the studying show that raw scores on each level of pupils were equated by determining the ability estimate that was calculated to correspond to a particular raw score, and then determining the raw score that corresponded to the same ability estimate obtained in a separate calculation. For example and ability estimate of .54 on M.S 2 corresponds to a raw score of 25. The same ability estimate .54 obtained on M.S. 1 corresponds to a raw score of 32. Thus with this comparable basis, a raw score of 25 on M.S 2 is equated to a raw score of 14 on M.S 3.