

การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันตามเชื้อชาติของผู้สอบ
ระหว่างวิธีการของ Mantel – Haenszel วิธี Lord's χ^2 และ วิธี Closed Interval Area

ปริญญาานิพนธ์

ของ

จันทนา เปรมฤดีปริชาชาญ

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวัดผลการศึกษา

พฤษภาคม 2551

การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันตามเชื้อชาติของผู้สอบ
ระหว่างวิธีการของ Mantel – Haenszel วิธี Lord's χ^2 และ วิธี Closed Interval Area

ปริญญาานิพนธ์

ของ

จันทนา เปรมฤดีปริชาชาญ

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวัดผลการศึกษา

พฤษภาคม 2551

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันตามเชื้อชาติของผู้สอบ
ระหว่างวิธีการของ Mantel – Haenszel วิธี Lord's χ^2 และ วิธี Closed Interval Area

บทคัดย่อ
ของ
จันทนา เปรมฤดีปรีชาชาญ

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวัดผลการศึกษา

พฤษภาคม 2551

จันทนา เปรมฤดีปรีชาชาญ. (2551). การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันตามเชื้อชาติของผู้สอบระหว่างวิธีการของ Mantel – Haenszel วิธี Lord's χ^2 และ วิธี Closed Interval Area. ปริญญาานิพนธ์ กศ.ม.(การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. คณะกรรมการควบคุม : ผู้ช่วยศาสตราจารย์ ดร. อองอาจ นัยพัฒน์ , ผู้ช่วยศาสตราจารย์ ดร. นิคม ตั้งคะพิภพ.

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อเปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และศึกษาความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันตามเชื้อชาติของผู้สอบระหว่างวิธีการของ Mantel – Haenszel วิธี Lord's χ^2 และ วิธี Closed Interval Area ซึ่งมี 2 วิธีย่อย คือ วิธี Closed - Interval Signed Area และ วิธี Closed - Interval Unsigned Area กลุ่มตัวอย่างที่ใช้ในการศึกษานี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2549 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษากาญจนบุรี เขต 3 ซึ่งได้มาโดยการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) จำนวน 1,533 คน เครื่องมือที่ใช้ในการศึกษา เป็นแบบทดสอบโจทย์ปัญหากลุ่มสาระการเรียนรู้วิชาคณิตศาสตร์ ตามหลักสูตรขั้นพื้นฐาน พุทธศักราช 2544

ผลการวิจัยพบว่า

1. จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ได้จากการตรวจสอบ ระหว่าง วิธี Mantel – Haenszel วิธี Lord's χ^2 และ วิธี Closed Interval Area แตกต่างกันเมื่อขนาดของกลุ่มตัวอย่างและความยาวของแบบทดสอบต่างกัน ได้แก่ กรณีแบบทดสอบ 40 ข้อ ที่ใช้กลุ่มตัวอย่างขนาด 600 คน และ 750 คนและแบบทดสอบ 60 ข้อ ที่ใช้กลุ่มตัวอย่างขนาด 600 คน และ 750 คน
2. ความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อพิจารณาจากกลุ่มตัวอย่างขนาด 600 คน กับ 750 คน วิธีการตรวจสอบเหมือนกัน แบบทดสอบจำนวน 40 ข้อ พบว่าวิธีที่มีความสอดคล้องกันในการตรวจสอบสูงสุด คือ วิธี Mantel – Haenszel แบบทดสอบจำนวน 60 ข้อ พบว่าวิธีที่มีความสอดคล้องกันในการตรวจสอบสูงสุด คือ วิธี Closed - Interval Unsigned Area

A COMPARISON OF CONSISTENCY OF MANTEL HAENSZEL , LORD'S χ^2 AND
CLOSED INTERVAL AREA ON DETECTION DIFFERENTIAL ITEM FUNCTIONING

AN ABSTRACT

BY

JANTANA PREM RUDEEPREECHACHARN

Presented in Partial Fulfillment of Requirements for the
Master of Education Degree in Educational Measurement
at Srinakharinwirot University

May 2008

Jantana Premrudeepreechachan. (2008). *A Comparison of Consistency of Mantel – Haenszel, Lord's χ^2 and Closed Interval Area on Detection Differential Item Functioning*. Master thesis. M.Ed. (Educational Measurement).
Bangkok: Graduate School, Srinakharinwirot University. Advisor Committee:
Assist. Prof.Dr. Ong- Art Nayapat, Assist. Prof.Dr. Nikom Tangkhapipop.

The purposes of this study were to compare item number with differential item functioning and investigate the consistency of Mantel- Haenszel , Lord's χ^2 and Closed Interval Area on detection differential item functioning. The subjects consisted of 1,533 Pratom Suksa VI students of schools under Karnchanaburi Educational Area in 2006 academic year, through stratified random sampling. The instruments for collecting data included problem test

The results revealed as follows:

1. Item numbers with differential functioning through Mantel – Haenszel , Lord's χ^2 and Closed Interval area were different according to sample group and test length including 40 items for 600 and 750 persons , 60 items for 600 and 750 persons.
2. The consistence of differential item functioning considered as to 600 and 750 subjects through the same method, it was found that Mantel – Haenszel Lord's χ^2 method gained high consistence with 40 items and Closed - Interval Unsigned Area method gained high consistence with 60 items.

ประกาศคุณูปการ

ปริญญาานิพนธ์ฉบับนี้สำเร็จ ลุล่วง ได้ เป็นอย่างดี ด้วยความ กรุณา อย่างดีจาก ผู้ช่วย ศาสตราจารย์ ดร.องอาจ นัยพัฒน์ ประธานกรรมการควบคุมปริญญาานิพนธ์ ผู้ช่วยศาสตราจารย์ ดร. นิคม ตั้งคะพิภพ กรรมการควบคุมปริญญาานิพนธ์ ที่ได้เสียสละเวลาอันมีค่าในการ ให้คำปรึกษา แนะนำ การทำงานวิจัยนี้ทุกขั้นตอนจนสำเร็จสมบูรณ์ รวมทั้งทำให้ผู้วิจัยได้รับประสบการณ์ในการ ทำงานวิจัยและรู้ถึงคุณค่าของงานวิจัย ผู้วิจัยรู้สึกซาบซึ้งถึงความกรุณาดังกล่าวและขอกราบ ขอบพระคุณเป็นอย่างสูงไว้ ณ โอกาสนี้

ขอกราบขอบพระคุณ รอง ศาสตราจารย์ ชูศรี วงศ์รัตน์ และ รอง ศาสตราจารย์ นิภา ศรีไพโรจน์ ที่กรุณาเป็นกรรมการ แต่งตั้งเพิ่มเติม ในการสอบปริญญาานิพนธ์ และได้ให้ข้อเสนอแนะ เพิ่มเติมที่เป็นประโยชน์ ทำให้ปริญญาานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น

ขอกราบขอบพระคุณ ผู้ทรงคุณวุฒิทั้ง 5 ท่าน ที่กรุณา ตรวจสอบคุณภาพเครื่องมือที่ใช้ในการ วิจัย รวมไปถึงอาจารย์ทุกท่านที่ได้ประสิทธิ์ประสาทความรู้แก่ผู้วิจัยในการศึกษาตามหลักสูตรการ วัตถุประสงค์ศึกษา อันเป็นพื้นฐานสำคัญในการศึกษาวิจัยครั้งนี้

ขอขอบ คุณผู้อำนวยการ โรงเรียน ครูผู้สอน และนักเรียนกลุ่มตัวอย่าง ทุกท่าน ที่ให้ความ ร่วมมือในการเก็บข้อมูล และการวิจัยจนสำเร็จลุล่วงด้วยดี

ขอขอบ คุณพี่ ๆ น้อง ๆ เพื่อน ๆ ทุกคนที่ได้ให้คำแนะนำ ความช่วยเหลือและเป็นกำลังใจให้ ผู้วิจัยตลอดการทำปริญญาานิพนธ์ครั้งนี้จนสำเร็จลุล่วงด้วยดี

ขอกราบขอบพระคุณ นาย กมล เปรมฤดีปริชาชาญ และนางอำนาจ เปรมฤดีปริชาชาญ ซึ่ง เป็นบิดาและมารดา และครอบครัวที่ให้กำลังใจ ความห่วงใย เป็นแรงบันดาลใจ และช่วยเหลือ ทุนทรัพย์ทำให้ประสบความสำเร็จในการศึกษาในการวิจัยครั้งนี้

ท้ายสุดนี้ผู้วิจัยขอกราบขอบพระคุณทุกท่านที่ไม่ได้กล่าวนามมา ณ ที่นี้ที่สนับสนุนและให้ ความช่วยเหลือในการทำปริญญาานิพนธ์ฉบับนี้จนสำเร็จลุล่วงไปด้วยดี คุณค่าและประโยชน์ของ ปริญญาานิพนธ์ฉบับนี้ ขอมอบแต่บิดา มารดา ครูอาจารย์ ตลอดจนผู้มีพระคุณของผู้วิจัยทุกท่านที่ช่วย ให้ผู้วิจัยประสบความสำเร็จในการศึกษาเล่าเรียน

จันทนา เปรมฤดีปริชาชาญ

สารบัญ

บทที่	หน้า
1 บทนำ	
ภูมิหลัง.....	1
วัตถุประสงค์ของการวิจัย.....	4
ความสำคัญของการศึกษา.....	4
ขอบเขตของการวิจัย.....	5
สมมติฐานในการวิจัย.....	6
นิยามศัพท์เฉพาะ.....	6
กรอบการวิจัย.....	8
2 เอกสารและงานวิจัยที่เกี่ยวข้อง	
การทำหน้าที่ต่างกันของข้อสอบ.....	9
หลักการและวิธีการตรวจการทำหน้าที่ต่างกันของข้อสอบ.....	13
การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวทฤษฎีการตอบข้อสอบ.....	16
การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีแมนเทิล - แอนส์เซิล.....	23
เอกสารที่เกี่ยวข้องกับการแก้โจทย์ปัญหาคณิตศาสตร์.....	29
งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ.....	31
3 วิธีดำเนินการวิจัย	
การกำหนดประชากรและการสุ่มกลุ่มตัวอย่าง.....	51
การสร้างเครื่องมือที่ใช้ในการวิจัย.....	53
การเก็บรวบรวมข้อมูล.....	57
การวิเคราะห์ข้อมูล.....	57
4 ผลการวิเคราะห์ข้อมูล	
การนำเสนอผลการวิเคราะห์ข้อมูล.....	61
สัญลักษณ์และอักษรย่อในการวิเคราะห์ข้อมูล.....	61
ผลการวิเคราะห์ข้อมูล.....	62

สารบัญ (ต่อ)

บทที่	หน้า
5 สรุป อภิปรายผล และข้อเสนอแนะ	
วัตถุประสงค์ของการวิจัย.....	83
กลุ่มตัวอย่าง.....	83
เครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	83
วิธีดำเนินการรวบรวมข้อมูล.....	83
สรุปผลการวิจัย.....	84
อภิปรายผล.....	86
ข้อเสนอแนะในการวิจัย.....	88
บรรณานุกรม.....	90
ภาคผนวก.....	100
ประวัติย่อผู้วิจัย.....	128

บัญชีตาราง

ตาราง	หน้า
1 ผลการตอบข้อสอบข้อหนึ่งระหว่างผู้สอบกลุ่มอ้างอิงและ กลุ่มเปรียบเทียบที่มีคะแนนรวม j	24
2 สัดส่วนของผลการตอบข้อสอบข้อหนึ่งระหว่างผู้สอบกลุ่มอ้างอิงและ กลุ่มเปรียบเทียบที่มีคะแนนรวม.....	24
3 สรุปการศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบในประเทศไทย	40
4 จำนวนกลุ่มตัวอย่างจำแนกนักเรียน ในแต่ละอำเภอของ สำนักงานเขตพื้นที่การศึกษากาญจนบุรี เขต 3.....	52
5 การวิเคราะห์เนื้อหาเกี่ยวกับโจทย์ปัญหาของนักเรียนชั้นประถมศึกษาปีที่ 4-6.....	54
6 เนื้อหาแบบทดสอบโจทย์ปัญหากลุ่มสาระการเรียนรู้คณิตศาสตร์ ตามหลักสูตรขั้นพื้นฐาน พุทธศักราช 2544 ในช่วงชั้นที่ 2	56
7 ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน ค่าความเชื่อมั่นและค่าความคลาดเคลื่อน มาตรฐานของแบบทดสอบวิชาคณิตศาสตร์ จำนวน 60 ข้อ.....	62
8 ค่าไอเกนและร้อยละความแปรปรวนขององค์ประกอบ.....	63
9 ค่าพารามิเตอร์ของข้อสอบ แยกตามกลุ่มเชื้อชาติไทย – ไม่ใช่เชื้อชาติไทย	64
10 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่าง วิธี Mantel – Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA) เมื่อจำนวนข้อสอบ 40 ข้อ กลุ่มตัวอย่าง 600 คน และ 750 คน	67
11 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธี Mantel – Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA) เมื่อจำนวนข้อสอบ 60 ข้อ กลุ่มตัวอย่าง 600 คน และ 750 คน.....	69
12 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 40 ข้อ กลุ่มตัวอย่าง 600 คน ระหว่างวิธี Mantel – Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA).....	71

บัญชีตาราง (ต่อ)

ตาราง	หน้า
13 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่เมื่อจำนวนข้อสอบ 40 ข้อ กลุ่มตัวอย่าง 600 คน ระหว่างวิธี Mantel – Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA).....	72
14 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อจำนวนข้อสอบ 40 ข้อ กลุ่มตัวอย่าง 750 คน ระหว่างวิธี Mantel – Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed-Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA).....	72
15 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่เมื่อจำนวนข้อสอบ 40 ข้อ กลุ่มตัวอย่าง 750 คน ระหว่างวิธี Mantel – Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA).....	73
16 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อจำนวนข้อสอบ 60 ข้อ กลุ่มตัวอย่าง 600 คน ระหว่างวิธี Mantel – Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA).....	74
17 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่เมื่อจำนวนข้อสอบ 60 ข้อ กลุ่มตัวอย่าง 600 คน ระหว่างวิธี Mantel – Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA).....	74
18 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อจำนวนข้อสอบ 60 ข้อ กลุ่มตัวอย่าง 750 คน ระหว่างวิธี Mantel – Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA).....	75

บัญชีตาราง (ต่อ)

ตาราง	หน้า
19 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่เมื่อจำนวนข้อสอบ 60 ข้อ กลุ่มตัวอย่าง 750 คน ระหว่างวิธี Mantel – Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA).....	76
20 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อจำนวนข้อสอบ 40 ข้อ วิธีการตรวจสอบเหมือนกัน ระหว่างกลุ่มตัวอย่างขนาด 600 คน และ 750 คน.....	77
21 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อจำนวนข้อสอบ 60 ข้อ วิธีการตรวจสอบเหมือนกัน ระหว่างกลุ่มตัวอย่างขนาด 600 คน และ 750 คน.....	78
22 ร้อยละของความสอดคล้องในการตรวจสอบเมื่อกลุ่มตัวอย่างเท่ากัน วิธีตรวจสอบแตกต่างกัน.....	80
23 ร้อยละของความสอดคล้องในการตรวจสอบเมื่อกลุ่มตัวอย่างต่างกัน วิธีตรวจสอบเหมือนกัน.....	81
24 ค่าพหามิตเตอร์ของข้อสอบ แยกตามกลุ่มเชื้อชาติไทย – ไม่ใช่เชื้อชาติไทย.....	112
25 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่าง 600 คน และจำนวนข้อสอบ 40 ข้อ.....	115
26 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่าง 750 คน และจำนวนข้อสอบ 40 ข้อ.....	117
27 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่าง 600 คน และจำนวนข้อสอบ 60 ข้อ.....	119
28 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่าง 750 คน และจำนวนข้อสอบ 60 ข้อ.....	122
29 ตัวอย่างการคำนวณการทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบในแต่ละวิธีตรวจสอบ.....	125

บัญชีภาพประกอบ

ภาพประกอบ	หน้า
1 ขั้นตอนการสร้างเครื่องมือที่ใช้ในการวิจัย.....	53
2 ผลการตรวจสอบความเป็นมิติเดียวกันของข้อสอบ	63

บทที่ 1

บทนำ

ภูมิหลัง

“ ... การศึกษาเป็นเครื่องมืออันสำคัญในการพัฒนาความรู้ ความคิด ความประพฤติ ทัศนคติ ค่านิยม และคุณธรรมของบุคคล เพื่อให้เป็นพลเมืองดีมีคุณภาพและประสิทธิภาพ การพัฒนาประเทศ ก็ย่อมทำได้สะดวกราบรื่นได้ผลที่แน่นอนและรวดเร็ว ... ” (เกรียงศักดิ์ เจริญวงศ์ศักดิ์. 2543; อ้างอิงจาก แนวพระราชดำรัสด้านการศึกษและการพัฒนาคน เมื่อวันที่ 22 กรกฎาคม 2520)

การวัดผลและประเมินผลการศึกษา จึงเป็นเครื่องมืออย่างหนึ่งที่จะช่วยพัฒนาคุณภาพของการศึกษา ผลที่ได้จากการวัดผลและประเมินผลศึกษานับเป็นเครื่องมือสำคัญที่ช่วยในการปรับปรุงคุณภาพของการศึกษาเพราะผลจากการวัดผลและประเมินผลการศึกษาจะเป็นพื้นฐานสำหรับการตัดสินใจของครูและนักการศึกษา เพื่อใช้ในการปรับปรุงการเรียนการสอน การแนะแนว การประเมินผลหลักสูตร แบบเรียน การจัดระบบบริหารของโรงเรียนตลอดจนการปรับปรุงวิธีการเรียนของผู้เรียนให้มีประสิทธิภาพยิ่งขึ้นจะส่งผลให้การดำเนินการจัดการศึกษาเป็นไปตามแนวทางและบรรลุผลตามจุดมุ่งหมายที่หลักสูตรต้องการ (อนันต์ ศรีโสภณ. 2524: 1; ไพฑูรย์ เวทการ. 2536: 7)

การวัดผลและประเมินผลศึกษานั้นสามารถกระทำได้หลายวิธี แต่ที่นิยมใช้กันมากที่สุด ในโรงเรียนคือ การทดสอบ (ชวาล แพรัตกุล. 2516: 87 – 88) ซึ่งการทดสอบหรือการสอบนั้นจะจัดให้มีขึ้นในทุกระดับชั้น ตั้งแต่ชั้นอนุบาลจนถึงระดับอุดมศึกษาไม่ว่าจะเป็นโรงเรียนของภาครัฐหรือเอกชน รวมไปถึงการสอบที่กระทำทั้งในโรงเรียนและนอกโรงเรียนด้วย ดังนั้น การพัฒนาวิธีการสอบและการให้คะแนนแบบทดสอบจำเป็นต้องคำนึงถึงคุณภาพของแบบทดสอบ ซึ่งเป็นหัวใจของการวัดและการประเมินผล คุณภาพที่สำคัญที่สุดของแบบทดสอบ คือ ความเที่ยงตรง (Validity) เพราะการนำแบบทดสอบที่ขาดความเที่ยงตรงมาใช้ในการวัดเป็นการไม่ยุติธรรมสำหรับผู้สอบ ผลที่ได้จากการวัดไม่สามารถสะท้อนถึงความสามารถที่แท้จริงของผู้สอบ (Shealy ; & Stout. 1993: 197) และเมื่อนำแบบทดสอบไปสอบวัดกับกลุ่มตัวอย่างที่มีความสามารถเท่ากันแต่มีความแตกต่างกันเกี่ยวกับเพศ เชื้อชาติ ศาสนา ภูมิฐานะ และระดับสติปัญญา ฯลฯ แล้วโอกาสในการตอบข้อสอบข้อนั้นได้ถูกต้องไม่เท่ากัน เรียกว่า ข้อสอบข้อนั้นทำหน้าที่ต่างกัน (Differential Item / Test Functioning : DIF / DTF) ทำให้เกิดการได้เปรียบเสียเปรียบกันขึ้นระหว่างกลุ่มย่อยต่าง ๆ นั่นคือ เกิดการทำหน้าที่ต่างกันหรือความไม่ยุติธรรมขึ้นในการสอบ (ญาณภัทร สีหะมงคล. 2540: 1-2)

การวิเคราะห์การทำหน้าที่ต่างกันของแบบทดสอบนั้น มีวิธีหลายวิธี ดังนี้ วิธีที่ใช้หลักของทฤษฎีการวัดผลดั้งเดิม (CTT) เช่น วิธีแปลงค่าความยาก (Transformed Item Difficulty) การวิเคราะห์ความแปรปรวน (Analysis of variance) การวิเคราะห์ด้วยไคสแควร์ (Chi - Square) วิธีวิเคราะห์องค์ประกอบ (Factor Analysis) วิธีค่าอำนาจจำแนก (Item Discrimination Procedure) เป็นต้น และวิธีวิเคราะห์ที่ใช้หลักของทฤษฎีการตอบสนองข้อคำถาม (IRT) และใช้การประมาณค่าความสามารถของผู้สอบเป็นเกณฑ์การจับคู่กลุ่มผู้สอบ วิธีการตรวจในกลุ่มนี้แบ่งเป็น 3 วิธีใหญ่ ๆ คือ วิธีการวัดพื้นที่ วิธีการเปรียบเทียบค่าพารามิเตอร์ และวิธีชิปเทสท์ (Shealy ; & Stout. 1993) วิธีการวัดพื้นที่ซึ่งแบ่งออกเป็นวิธีย่อย ๆ อีกหลายวิธี เช่น วิธีการวัดพื้นที่ของราฐู (Raju. 1993) และวิธีการวัดพื้นที่ของคิมและโคเฮน (Kim and Cohen. 1993) จะพิจารณาทั้งพื้นที่แบบมีเครื่องหมาย (Closed - Interval Signed Area - CSA) และพื้นที่แบบไม่มีเครื่องหมาย (Closed - Interval Unsigned Area - CUA) เป็นต้น ส่วนวิธีการเปรียบเทียบค่าพารามิเตอร์ แบ่งออกเป็นวิธีย่อย ๆ เช่น วิธีการทดสอบไคสแควร์ของลอร์ด (Lord's chi - square test) และวิธีการทดสอบอัตราส่วนความน่าจะเป็น (Likelihood Ratio test : LR) เป็นต้น ซึ่งผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยหลักของทฤษฎีการตอบข้อคำถาม พบว่าเป็นวิธีที่ผลการวิเคราะห์ถูกต้องที่สุด แต่การใช้ทฤษฎีการตอบข้อคำถามในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบเป็นวิธีการที่ใช้การคำนวณยุ่งยากซับซ้อน ใช้กลุ่มตัวอย่างขนาดใหญ่ และสิ้นเปลืองค่าใช้จ่ายมาก (Narayanan ; & Swaminathan. 1994: 315) ทำให้เกิดความยุ่งยากในทางปฏิบัติจากเหตุผลที่กล่าวมาข้างต้น จึงมีการศึกษาวิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ โดยหลีกเลี่ยงการใช้ทฤษฎีการตอบข้อคำถาม สำหรับวิธีที่มีผู้เสนอ คือวิธีการที่ใช้สถิติแบบนอ-พาราเมตริก(Non-Parametric) ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ เช่น วิธีแมนเทล-แฮนเซล(Mantel - Haenszel:MH) วิธีชิปเทสท์ (SIBTEST:SIB) และวิธี HW 3 (HW 3 Method) เป็นวิธีการคำนวณที่ไม่ยุ่งยากซับซ้อนและสิ้นเปลืองค่าใช้จ่ายไม่มากนักเหมาะที่จะนำมาใช้ในทางปฏิบัติจึงเป็นที่น่าสนใจและนิยมใช้ในปัจจุบัน

นอกจากนี้ยังมีงานวิจัยที่ศึกษาผลกระทบของขนาดกลุ่มตัวอย่างในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ จะพบว่าวิธีการวิเคราะห์ข้อสอบที่มีการทำหน้าที่ต่างกันได้เพิ่มขึ้นเมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้น (Roger and Swaminathan.1993:105 ;Narayanan and Swaminathan.1994:315) ส่วนมาเซอร์และคณะ (Mazor ; et al. 1992) เสนอว่า การใช้กลุ่มตัวอย่างขนาด 200 คน ก็เพียงพอและไม่ควรน้อยกว่านี้ ซึ่งนารายันและสวามินาทาน (Narayanan ; & Swaminathan. 1994) เสนอว่า สำหรับวิธี MH และวิธี SIBTEST ขนาดกลุ่มตัวอย่างกลุ่มละ 300 คน ก็เพียงพอที่จะตรวจสอบอย่างมีประสิทธิภาพ ต่อมาจิตติมา วรณศรี (2539) ได้ศึกษาพบว่าเมื่อขนาดกลุ่มตัวอย่าง 200 คน และ 600 คน ทั้งวิธี MH และวิธี SIBTEST สามารถตรวจสอบได้ถูกต้อง 50 % แต่ถ้าขนาดกลุ่มตัวอย่างเพิ่มขึ้น

เป็น 1,000 คน สามารถตรวจสอบได้ถูกต้อง 100 % และการศึกษาของญาณภัทร สีหะมงคล (2540) พบว่า เมื่อกลุ่มตัวอย่างขนาด 500 คน และ 700 คน และแบบทดสอบ 40 ข้อ ควรเลือกใช้วิธี Closed Interval Area และกลุ่มตัวอย่างขนาด 1,000 คน แบบทดสอบ 20 ข้อ และแบบทดสอบ 60 ข้อ

จากการศึกษาของคณะกรรมการการศึกษาแห่งชาติ พบว่าความแตกต่างทางวัฒนธรรมของครอบครัว อันเป็นผลมาจากฐานะทางเศรษฐกิจและสังคม มีผลทำให้นักเรียนมีผลสัมฤทธิ์ทางการเรียนต่างกัน (สำนักงานคณะกรรมการการศึกษาแห่งชาติ. 2524:78) และเนื่องจากสำนักงานเขตพื้นที่การศึกษากาญจนบุรี เขต 3 มีพื้นที่รับผิดชอบ 4 อำเภอ 23 ตำบล 152 หมู่บ้าน มีอาณาเขตติดประเทศพม่า จึงมีผู้ใช้สัญชาติไทยเข้ามาอยู่ในพื้นที่ ประมาณ 1 ใน 4 ของประชากรทั้งหมด เช่น สัญชาติพม่า สัญชาติลาว นอกจากนี้ยังมีชนกลุ่มน้อยอื่น ๆ อีก เช่น กระเหรี่ยง มอญ (สำนักงานเขตพื้นที่การศึกษากาญจนบุรี เขต3. 2548: 3) ทำให้ผลการสอบของนักเรียนมีความแตกต่างกัน ซึ่งสอดคล้องกับงานวิจัยของกรีน (พิพุทธพงษ์ แนวทอง. 2544: 2; อ้างอิงจาก Intasuwan. 1979: 1; citing Gree.1972)พบว่าข้อคำถามบางข้อคำถามของแบบทดสอบวัดผลสัมฤทธิ์ของแคลิฟอร์เนีย (California Achivement Test) ซึ่งเป็นข้อคำถามที่มีคุณภาพ เมื่อไปทดสอบกับกลุ่มผู้สอบที่มีความแตกต่างกันทางเชื้อชาติ ภาษา สภาพเศรษฐกิจ และสภาพภูมิอากาศ จะให้ผลการสอบที่แตกต่างกัน ซึ่งอาจจะมีสาเหตุมาจากด้านเชื้อชาติ วัฒนธรรมและสิ่งแวดล้อมที่ต่างกันมีผลต่อความเข้าใจภาษาที่ใช้ในข้อคำถามต่างกันและงานวิจัยของผจงจิต อินทสุวรรณ (นวลอนงค์ นันทวงศ์. 2543 : 2; อ้างอิงจาก Intasuwan. 1979) ได้ศึกษาการทำหน้าที่แตกต่างของแบบทดสอบ (IEAC The International for The Evaluation of Education Achivement) กับกลุ่มนักเรียนชาวอังกฤษ และ นิวซีแลนด์ก็พบข้อที่มีการทำหน้าที่แตกต่างกันระหว่างนักเรียน 2 กลุ่ม ทั้งนี้อาจมีสาเหตุมาจากความเข้าใจในคำศัพท์ระหว่างนักเรียน 2 กลุ่ม ที่มีเชื้อชาติแตกต่างกัน สภาพแวดล้อม ความเป็นอยู่ ภาษาที่ใช้ในชีวิตประจำวันที่แตกต่างกันอาจทำให้เข้าใจภาษาที่ใช้ในแบบทดสอบแตกต่างกัน หรือภาษาที่ใช้ในแบบทดสอบอาจจะเอื้อต่อกลุ่มนักเรียนกลุ่มใดกลุ่มหนึ่งก็เป็นไปได้ ซึ่งสอดคล้องกับผลการวิจัย ชัยชัย เผ่าพงษ์ (2527) ทศนีย์ พิรมนตรี (2530) สุพัฒน์ สุขุมลันต์ (2534) นิรมล ชัยขวลิต (2537) พรรณี จินตมาศ (2540) สับโคเวียตและคณะ (Subkovaik ; et al. 1984) ที่ระบุว่า เพศ ภาษา เป็นแหล่งที่ก่อให้เกิดการทำหน้าที่ต่างกันของแบบทดสอบ ส่วน คมศักดิ์ ชื่นชม (2539) จิตสุดา ธราพร (2539) ระบุว่าศาสนาที่นักเรียนที่มีสภาพภูมิศาสตร์หรือมีพื้นที่ต่างกัน เป็นแหล่งก่อให้เกิดการทำหน้าที่แตกต่างกันของแบบทดสอบ

จากข้อความที่กล่าวมาแล้วข้างต้น จะเห็นว่ากลุ่มผู้สอบที่มีความแตกต่างกันในด้านต่างๆ ไม่ว่าจะเป็นความแตกต่างทางด้านเชื้อชาติ ภาษา ศาสนา วัฒนธรรม รวมทั้งสภาพภูมิศาสตร์ และสภาพสังคมเป็นสาเหตุที่ทำให้ข้อคำถามบางข้อเกิดการทำหน้าที่แตกต่างกันกับผู้สอบกลุ่มใดกลุ่มหนึ่ง

โดยเฉพาะ ผู้วิจัยจึงมีความสนใจที่จะศึกษาผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Mantel – Haenszel วิธี Lord's χ^2 และวิธี Closed Interval Area ระหว่างกลุ่มนักเรียนที่มีเชื้อชาติต่างกัน คือ กลุ่มนักเรียนที่มีเชื้อชาติไทยกับนักเรียนที่ไม่ใช่เชื้อชาติไทย อันจะเป็นแนวทางในการเลือกวิธีวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบมาใช้ในการปฏิบัติจริงเพื่อให้แบบทดสอบมีความยุติธรรมแก่นักเรียนทุกกลุ่มจะได้เป็นแบบทดสอบที่มีคุณภาพที่ดียิ่งขึ้นและวัดความสามารถของผู้สอบได้อย่างแท้จริง

วัตถุประสงค์ของการวิจัย

ในการศึกษาค้นคว้าครั้งนี้ ผู้วิจัยได้ตั้งวัตถุประสงค์ดังต่อไปนี้

1. เพื่อเปรียบเทียบจำนวนข้อสอบที่มีการทำหน้าที่ของผลการตรวจสอบระหว่างวิธี Mantel – Haenszel วิธี Lord's χ^2 และวิธี Closed Interval Area เมื่อขนาดของกลุ่มตัวอย่าง และความยาวของแบบทดสอบ ต่างกัน
2. เพื่อศึกษาความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่าง วิธี Mantel – Haenszel วิธี Lord's χ^2 และวิธี Closed Interval Area เมื่อขนาดของกลุ่มตัวอย่าง และความยาวของแบบทดสอบ ต่างกัน

ความสำคัญของการวิจัย

การศึกษาค้นคว้าครั้งนี้ เป็นการเปรียบเทียบผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันตามเชื้อชาติของผู้สอบตามแนวคิดทั้ง 3 วิธี ซึ่งผลที่ได้จากการศึกษามีประโยชน์และความสำคัญ ดังนี้ ด้านทางทฤษฎี ทำให้ได้ข้อสรุปที่ชัดเจนเกี่ยวกับผลของวิธีการตรวจสอบที่ทำหน้าที่ต่างกันทั้ง 3 วิธี ว่าวิธีใดบ้างที่ให้ผลสอดคล้องกันหรือขัดแย้งกันในเงื่อนไขใดของการศึกษา ซึ่งจะมีประโยชน์ต่อการเลือกใช้ให้เหมาะสมและเกิดความผิดพลาดน้อยที่สุดในการตรวจสอบ และความสำคัญในทางปฏิบัติ ระดับนโยบาย ทำให้มีแบบทดสอบที่มีมาตรฐาน มีคุณภาพและมีความยุติธรรม ทำให้การประเมินผลความรู้ ความสามารถอย่างแท้จริง เป็นประโยชน์แก่ผู้บริหารระดับเขตพื้นที่การศึกษาและโรงเรียนเพื่อจะได้มีนโยบายส่งเสริมและแก้ปัญหาทางการเรียนให้ความช่วยเหลือแก่ผู้เรียนต่อไป และระดับปฏิบัติ นั้นครูผู้สร้างข้อสอบจะได้เลือกวิธีวิเคราะห์ข้อสอบได้อย่างถูกต้อง รวดเร็ว และเหมาะสมมาพัฒนาแบบทดสอบให้มีคุณภาพและมีความยุติธรรมในการวัดและประเมินผลความรู้ความสามารถอย่างแท้จริง

ขอบเขตของการวิจัย

ประชากรที่ใช้ในการวิจัย

ประชากรที่ใช้ในการศึกษาค้นคว้าเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ภาคเรียนที่ 1 ปีการศึกษา 2549 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษากาญจนบุรี เขต 3 ที่มีนักเรียนทั้งเชื้อชาติไทยและไม่ใช่อื่นเชื้อชาติไทย จำนวนทั้งสิ้น 2,607 คน

กลุ่มตัวอย่างที่ใช้ในการวิจัย

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ ได้แก่ นักเรียนชั้นประถมศึกษาปีที่ 6 ภาคเรียนที่ 1 ปีการศึกษา 2549 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษากาญจนบุรี เขต 3 โดยนักเรียนนั้นมีความพร้อมในการเรียนรู้ในระดับใกล้เคียงกัน จำนวน 1,533 คน แบ่งเป็นนักเรียนเชื้อชาติไทยจำนวน 780 คน และนักเรียนไม่ใช่อื่นเชื้อชาติไทย จำนวน 753 คน ซึ่งได้มาโดยการสุ่มแบบแบ่งชั้น (Stratified Random Sampling)

ตัวแปรที่ศึกษา

1. ตัวแปรอิสระ ได้แก่

1.1 วิธีการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน ได้แก่

1.1.1 วิธี Mantel – Haenszel

1.1.2 วิธี Lord's χ^2

1.1.3 วิธี Closed Interval Area

1.1.3.1 Closed - Interval Signed Area (CSA)

1.1.3.2 Closed - Interval Unsigned Area (CUA)

1.2 ขนาดของกลุ่มตัวอย่าง ได้แก่ 600 และ 750 คน

1.3 ความยาวของแบบทดสอบ ได้แก่ 40 และ 60 ข้อ

2. ตัวแปรตาม ได้แก่

2.1 จำนวนข้อสอบวิชาคณิตศาสตร์ ที่ทำหน้าที่ต่างกันตามเชื้อชาติ ที่ได้จากการตรวจสอบแต่ละวิธี

2.2 ความสอดคล้องในการตรวจสอบ

สมมติฐานในการวิจัย

1. เมื่อขนาดของกลุ่มตัวอย่างเท่ากัน วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกันจากการตรวจสอบ ทั้ง 3 วิธี แตกต่างกัน
2. เมื่อขนาดของกลุ่มตัวอย่างต่างกัน วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเหมือนกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกันจากการตรวจสอบ ทั้ง 3 วิธี แตกต่างกัน

นิยามศัพท์เฉพาะ

1. การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning) หมายถึง ข้อสอบที่ผู้สอบ ณ ระดับความสามารถเดียวกันแต่อยู่ต่างกลุ่มกัน มีโอกาสของการตอบข้อสอบข้อนั้นได้ถูกต้องไม่เท่ากัน
2. การไม่ทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่ผู้สอบ ณ ระดับความสามารถเดียวกันแต่อยู่ต่างกลุ่มกัน มีโอกาสของการตอบข้อสอบข้อนั้นได้ถูกต้องเท่ากัน
3. กลุ่มอ้างอิง (Reference Group ; R) หมายถึง กลุ่มผู้สอบที่คาดว่าจะจะเป็นกลุ่มที่ได้ประโยชน์จากข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องมากกว่าผู้สอบอีกกลุ่มหนึ่งที่ได้คะแนนจากแบบทดสอบเท่ากัน สำหรับกลุ่มอ้างอิงในการวิจัยครั้งนี้ คือ กลุ่มผู้สอบที่มีเชื้อชาติไทย
4. กลุ่มเปรียบเทียบ (Focal Group : F) หมายถึง กลุ่มผู้สอบที่คาดว่าจะจะเป็นกลุ่มที่เสียประโยชน์จากข้อสอบที่ทำหน้าที่ต่างกัน คือ เป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องต่ำกว่าผู้สอบกลุ่มอ้างอิง ที่มีคะแนนจากแบบทดสอบเท่ากัน สำหรับกลุ่มเปรียบเทียบในการวิจัยครั้งนี้ คือ กลุ่มผู้สอบที่ไม่ใช่เชื้อชาติไทย
5. กลุ่มเชื้อชาติไทย คือ กลุ่มนักเรียนที่มีเชื้อชาติไทยโดยมีหลักฐานการเกิดและทะเบียนบ้านระบุว่า เป็นนักเรียนที่มีเชื้อชาติไทยและเรียนอยู่ในระบบโรงเรียน
6. กลุ่มไม่ใช่เชื้อชาติไทย คือ กลุ่มนักเรียนที่มีเชื้อชาติอื่น ๆ ที่ไม่ใช่เชื้อชาติไทย คือ พม่า , ลาว กะเหรี่ยง และมอญ โดยบิดามารดาของนักเรียนต้องเป็นคนที่ไม่มีเชื้อชาติไทยทั้ง 2 คน และนักเรียนกำลังเรียนอยู่ในระบบโรงเรียนโดยการสำรวจจากครู ผู้ใหญ่บ้าน เพื่อขอหลักฐานรับรองจากอำเภอ
7. วิธีตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน หมายถึง วิธีตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันตามแนวคิดทฤษฎีการตอบข้อสอบ ประกอบด้วย วิธี Mantel-Haenszel วิธี Lord's χ^2 และวิธี Closed Interval Area

7.1 วิธี Mantel-Haenszel หมายถึง วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่พัฒนาโดยฮอลแลนด์และเทเยอร์ (Holland ; & Thayer, 1988) ซึ่งคำนวณดัชนีการทำหน้าที่ต่างกัน

ของข้อสอบจากสัดส่วนการตอบข้อสอบระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ แล้วใช้สถิติ χ^2_{MH} ทดสอบนัยสำคัญกับดัชนี DIF ที่ระดับ .05

7.2 วิธี Lord's χ^2 หมายถึง วิธีการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันตามแนวคิดของ ลอร์ด ที่ใช้การเปรียบเทียบค่าพารามิเตอร์ของข้อสอบระหว่างกลุ่มย่อย 2 กลุ่ม

7.3 วิธี Closed Interval Area หมายถึง วิธีการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันตาม แนวคิดของคิม และโคเฮน โดยการวัดความแตกต่างพื้นที่ของฟังก์ชันการตอบข้อสอบระหว่างกลุ่ม ย่อย 2 กลุ่ม เมื่อกำหนดช่วงของความสามารถ (Ability) ผู้เข้าสอบเป็นแบบปิด (Closed Interval) แบ่งเป็น 2 วิธีย่อย คือ

- วิธี Closed - Interval Signed Area (CSA) หมายถึง การวัดขนาดพื้นที่บนช่วง ความสามารถแบบปิด โดยพิจารณาพื้นที่แบบมีเครื่องหมาย

- วิธี Closed - Interval Unsigned Area (CUA) หมายถึง การวัดขนาดพื้นที่บนช่วง ความสามารถแบบปิด โดยพิจารณาพื้นที่แบบไม่มีเครื่องหมาย

8. เกณฑ์การตัดสินข้อสอบที่ทำหน้าที่ต่างกัน หมายถึง ค่าที่แสดงว่าข้อสอบทำหน้าที่ต่างกัน จากค่าดัชนีทั้ง 3 วิธี ได้แก่

- วิธี Mantel-Haenszel พิจารณาค่า α_{MH} ไม่เท่ากับ 1 และค่า χ^2_{MH} มากกว่าค่า χ^2 ที่ df เท่ากับ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

- วิธี Lord's χ^2 ทดสอบนัยสำคัญทางสถิติที่ระดับ .05

- วิธี Closed Interval Area นำขนาดของพื้นที่ไปเทียบกับเกณฑ์ที่ตัดสินการทำหน้าที่ ต่างกันของข้อสอบ โดยใช้เกณฑ์ ดังนี้

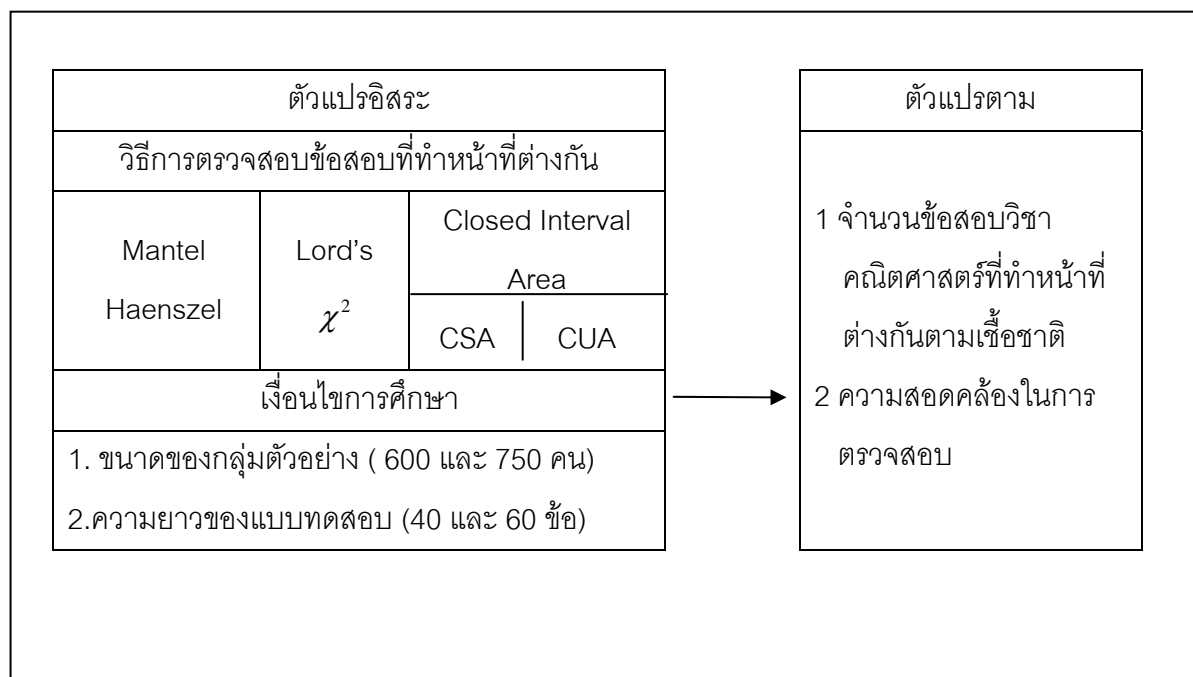
- วิธี Closed - Interval Signed Area (CSA) มีขนาดพื้นที่ตั้งแต่ .30 ขึ้นไป

- วิธี Closed - Interval Unsigned Area (CUA) มีขนาดพื้นที่ตั้งแต่ .40 ขึ้นไป

9. ความสอดคล้องในการตรวจสอบ หมายถึง สัดส่วนจำนวนข้อสอบที่ตรงกันในการตัดสิน โดยการใช่วิธีการตรวจสอบการทำหน้าที่ต่างกัน 3 วิธี ต่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดใน การเปรียบเทียบแต่ละคู่

กรอบความคิดในการวิจัย

ประเด็นในการศึกษา : เปรียบเทียบความสอดคล้องของผลการตรวจสอบที่ทำหน้าที่ต่างกัน ระหว่างวิธี Mantel – Haenszel วิธี Lord's χ^2 และวิธี Closed Interval Area จากข้อมูลแบบทดสอบวิชาคณิตศาสตร์ โดยอาศัยกรอบความคิด ดังนี้



จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวกับการตรวจสอบที่ทำหน้าที่ต่างกัน ของกาญจนา วัธนสุนทร (2538) ที่ศึกษาเพื่อพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ พบว่า วิธี Mantel – Haenszel และวิธี SIBTEST การสอดคล้องของการตัดสินภายในดัชนีเดียวกัน มีความไม่คงที่ข้ามขนาดผู้สอบแต่จะมีความสอดคล้องแนวโน้มสูงขึ้นที่ขนาดผู้สอบ ตั้งแต่ขนาด 600 คนขึ้นไป และจากการศึกษาของบากิและเฟอราธา (Baghi and Ferrara.1990) พบว่าความสอดคล้องโดยพบว่า ไคสแควร์แบบแมนเทล – แฮนส์เซล เป็นวิธีที่ใช้แทนทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ ได้ถ้าขนาดกลุ่มตัวอย่าง 750 คนขึ้นไป และจากการศึกษาของญาณภัทร สีหะมงคล (2540) พบว่า วิธี Lord's χ^2 วิธี Raju's Area Measures และวิธี Closed Interval Area มีความสอดคล้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ส่วนมากมีค่าปานกลางถึงต่ำเกือบทุกเงื่อนไขของการศึกษา ยกเว้นกรณีกลุ่มตัวอย่าง 500 คน แบบทดสอบยาว 30 , 40 และ 60 ข้อ และกลุ่มตัวอย่าง 700 คน แบบทดสอบยาว 40 ข้อ ที่มีความสอดคล้องในระดับปานกลางถึงสูงมาก และจากการศึกษาของจิตติมา วรรณศรี (2539) ที่พบว่า เมื่อใช้แบบทดสอบขนาด 60 ข้อ จะมีผลทำให้วิธี Mantel – Haenszel และวิธี SIBTEST มีประสิทธิภาพในการตรวจสอบดีที่สุด

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง และได้นำเสนอตามหัวข้อต่อไปนี้

1. การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning)
2. หลักการและวิธีการตรวจการทำหน้าที่ต่างกันของข้อสอบ
3. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวทฤษฎีการตอบข้อสอบ
4. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีแมนเทล-แฮนส์เซล (Mantel- Haenszel ; MH)
5. เอกสารที่เกี่ยวข้องกับการการแก้ไขข้อบกพร่องปัญหาคณิตศาสตร์
6. งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ

1. การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning)

1.1 ความเป็นมา

นักวัดผลการศึกษาได้ทำการศึกษาความลำเอียงของข้อสอบ (Item bias) กันอย่างกว้างขวาง ทำให้เกิดความสับสนของการใช้คำและความหมายมีประเด็นโต้แย้งกันว่าความลำเอียงของข้อสอบ เป็นผลการตัดสินว่าข้อสอบมีความยุติธรรมหรือไม่ อันส่งผลต่อการบรรลุจุดมุ่งหมายของการใช้แบบสอบหรือความลำเอียงของข้อสอบ เป็นสารสนเทศทางสถิติที่ได้จากข้อสอบเกี่ยวกับความสัมพันธ์ระหว่างคุณลักษณะที่ข้อสอบมุ่งวัด กับประสิทธิผลของผู้สอบกลุ่มต่าง ๆ ที่ทำการสอบ เมื่อกลุ่มผู้สอบต่างกลุ่มกันตอบข้อสอบเดียวกัน ความแตกต่างที่เกิดขึ้นอาจมาจากความไม่เหมาะสมของข้อคำถาม ซึ่งสามารถเกิดขึ้นได้หลายลักษณะ หรือประสิทธิผลของผู้สอบซึ่งอาจมีลักษณะพื้นฐานเดิมแตกต่างกันในหลายสถานการณ์จึงไม่เหมาะสมที่จะใช้คำว่า ข้อสอบลำเอียง (Biased item) เนื่องจากเป็นภาษาที่มีความหมายเชิงลบ ประกอบกับเกณฑ์ที่ใช้สำหรับตัดสินความลำเอียงยังมีความคลุมเครือและค่อนข้างสับสน ดังนั้น จึงควรเปลี่ยนมาใช้คำว่า การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning ; DIF) ซึ่งเป็นคำที่มีความเป็นกลางและเหมาะสมกว่า (Holland and Wainer. 1993)

การทำหน้าที่ต่างกันของข้อสอบ ได้รับความสนใจตั้งแต่ปี 1910 (Holland ; & Wainer. 1993) โดยบินेटและไซมอน (Binet and Simon) ได้ทดสอบเด็กที่มาจากฐานะทางเศรษฐกิจและสังคมต่างกัน พบว่า ข้อสอบบางข้อน่าจะวัดผลจากวัฒนธรรมการฝึก (เช่น ที่บ้าน หรือที่โรงเรียน) มากกว่าที่จะวัดความสามารถทางสมอง จากนั้นได้เอาข้อสอบเหล่านั้นออกไป เช่น ถ้าเป็นเด็กอายุ 7 ปี ข้อคำถามที่เกี่ยวกับนิ้วมือ การคัดลอกประโยค และการเรียกชื่อเหรียญถูกเอาออกจากแบบทดสอบ

โดยทั่วไปในการทดสอบจะพิจารณาเฉพาะข้อคำถามที่ไม่ขึ้นกับการฝึกฝนของที่บ้าน (Home Training) ความตั้งใจ (Attention) ภาษา (Language) นิสัยในการดูภาพ (Habit of looking at picture) และแบบฝึกหัดเชิงวิชาการ (Scholastic exercise)

ปี 1912 วิลเลียม สเตอร์น (William Stern) ได้ศึกษาความแตกต่างของชั้น (Class difference) ในประเทศเยอรมัน และพบว่ามีความคล้ายคลึงกับบินด์ทอป โดยพยายามที่จะทำความเข้าใจเกี่ยวกับแหล่งที่ทำให้เกิดความแตกต่าง เพื่อที่จะเสนอแนะว่า แบบทดสอบนั้นเข้าข้างนักเรียนชั้นหนึ่งมากกว่าชั้นอื่น ๆ อย่างชัดเจน

ต่อมาปี 1951 มีงานวิจัยที่ถือเป็นการเริ่มต้นของการศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ งานวิจัยของเอลล์ (Eells) เดวิส (Davis) ฮาวิกเฮิร์สต์ (Haighurst) เฮอริค (Herrick) และไทเลอร์ (Tyler) โดยเอลล์ และคณะสนใจความแตกต่างของการวัดบางอย่างที่พิจารณาถึงความไม่ถูกต้องของการตอบข้อสอบอันเกิดจากความสามารถ (Ability) เช่น ความแตกต่างของไอคิว โดยมุ่งไปที่ข้อคำถามแต่ละข้อในลักษณะเช่นเดียวกับ บินด์ และสเตอร์นคือดูว่าข้อคำถามใดมีการทำหน้าที่ต่างกันหรือไม่ แม้ว่า เอลล์และคณะใช้สัดส่วนการตอบถูกสำหรับกลุ่มที่แตกต่าง แต่ก็ไม่ได้รับการแนะนำให้ใช้ในเวลาไม่นานนัก ขณะเดียวกันก็เกิดคำถามเกี่ยวกับความคงที่ของความแตกต่างระหว่างกลุ่มว่าขึ้นกับเนื้อหา หรือรูปแบบคำถาม ความแตกต่างของวัฒนธรรมในการเลือกคำตอบที่ผิด มีความสัมพันธ์กับความยากของข้อสอบหรือไม่ เขาได้พยายามแก้ไขอิทธิพลของชั้นทางสังคมและภูมิหลังของเชื้อชาติที่มีผลต่อการตอบข้อสอบ โดยการเปรียบเทียบกลุ่มของเชื้อชาติที่อยู่ในชั้นทางสังคม

ผลจากการศึกษาของเอลล์ ได้นำไปสู่การศึกษาอย่างกว้างขวางในช่วง 20 ปีที่ผ่านมา เมื่อความมุ่งหวังของรัฐในการที่จะทำให้เกิดความยุติธรรมในการใช้ผลการสอบ ในช่วงปลายทศวรรษ 1960 จนถึงทศวรรษ 1970 ได้มีความวิตกกังวลเกี่ยวกับการทำหน้าที่ต่างกันของแบบทดสอบไอคิวเป็นอย่างมาก เพราะเป็นแบบทดสอบที่ถูกนำมาใช้อย่างกว้างขวางในทางด้านการศึกษา การคัดเลือกผู้สมัครงาน และ จัดบุคคลเข้าสู่ตำแหน่ง เมื่อมีการเคลื่อนไหวเรื่องสิทธิของมนุษยชน จึงได้หันกลับมาให้ความสนใจต่อแบบทดสอบต่าง ๆ ที่มีความไม่เท่าเทียมในการเข้าศึกษาและโอกาสในการทำงาน จนกระทั่งปี 1971 บริษัท Griggs v. Duke Power ซึ่งว่าผลการแบบทดสอบต่าง ๆ แล้วพบว่าแบบทดสอบเหล่านั้นให้ผลในการจ้างงานที่ไม่ได้สัดส่วนกัน เมื่อพิจารณาตามเชื้อชาติของผู้สมัครงาน

ในช่วงเวลาเดียวกันมีการโต้แย้งเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบมีทั้งบทความในวารสารต่าง ๆ เจนเซน ได้เสนอว่า ไอคิวเป็นผลมาจากการถ่ายทอดทางพันธุกรรมมากกว่ามาจากสิ่งแวดล้อม ความแตกต่างของไอคิวระหว่างคนผิวขาวกับคนผิวดำไม่ได้มาจากสิ่งแวดล้อมเพียงอย่างเดียว และได้รับการยืนยันจากนักวิทยาศาสตร์ค่อนข้างชัดเจน เขายังได้อธิบายต่อว่า พันธุกรรมอธิบายความแปรปรวนผลการปฏิบัติงานของแต่ละคนได้ประมาณ ร้อยละ 80 จากนั้นก็มีการตอบสนองงาน

ของเจเนเซน ในวารสาร Harard Education Review (1969) โดยสมาคมนักจิตวิทยาคนผิวดำ เสนอว่าการทดสอบคนผิวดำทุกชนิดควรทำให้เท่าเทียมกันมากที่สุดเท่าที่จะเป็นไปได้ (Camilli ; & Shepard. 1994 : 4 – 7)

การตรวจสอบและค้นหาการทำหน้าที่ต่างกันของข้อสอบได้มีมาอย่างต่อเนื่องควบคู่กับการพยายามคิดค้นวิธีการทางสถิติในการวิเคราะห์ การทำหน้าที่ต่างกันของข้อสอบดังปรากฏจากเอกสารการวัดผลในช่วงกลางปี 1970 และในปัจจุบันก็มีการพัฒนาเทคนิควิธีการต่าง ๆ อย่างมากมายเพื่อให้ผลการวิเคราะห์มีความถูกต้องที่สุด

1.2 ความหมายของการทำหน้าที่ต่างกันของข้อสอบ

นักวิจัยทางการวัดผลหลายท่านได้ให้ความหมายของความลำเอียงของข้อสอบและการทำหน้าที่ต่างกันของข้อสอบ ไว้ดังนี้

อลเลน และเยน (Allen ; & Yen. 1979) การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่วัดความสามารถหรือคุณลักษณะทางจิตวิทยาของผู้สอบแต่ละกลุ่มไม่ตรงกัน

ลอร์ด (Lord. 1980) แมคคอลลีย์ และเมนโดซา (McCauley ; & Mendoza. 1985) แฮมเบิลตัน และคณะ (Hambleton et. Al. 1991) ฮอลแลนด์และไวเนอร์ (Holland and Wainer. 1993) การทำหน้าที่ต่างกันของข้อสอบหมายถึง ข้อสอบที่ผู้สอบมีความสามารถเท่ากัน แต่อยู่ต่างกลุ่มกัน มีโอกาสตอบข้อสอบข้อนั้นได้ถูกต้องไม่เท่ากัน

พอร์พแฮม (Popham. 1981) การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ความชอบหรือความโอนเอียงที่ทำให้การพิจารณาตัดสินเป็นไปอย่างไม่ยุติธรรม

ฮอลแลนด์ และไวเนอร์ (Holland ; & wainer. 1993) การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่ผู้เข้าสอบ ณ ความสามารถเดียวกัน แต่อยู่ต่างกลุ่มกัน มีโอกาสในการตอบถูกไม่เท่ากัน กลุ่มที่ให้ความสนใจ หรือกลุ่มที่เสียเปรียบ เรียกว่า “Focal group” และกลุ่มที่ใช้เปรียบเทียบกับซึ่งโดยทั่วไปจะเป็นกลุ่มได้เปรียบ เรียกว่า “Reference group”

คามิลลี และเชพเพิร์ด (Camilli ; & Shepard. 1994) การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่ทำหน้าที่ต่างกันเป็นความคลาดเคลื่อนอย่างเป็นระบบ (Systematic Error) ชนิดหนึ่งของการวัดที่ทำให้เกิดการได้เปรียบเสียเปรียบต่อสมาชิกกลุ่มใดกลุ่มหนึ่งโดยเฉพาะ เช่น ข้อต่อเพศชายมากกว่าเพศหญิง ข้อต่อกลุ่มคนผิวขาวมากกว่าผิวดำ เป็นต้น

จากความหมายดังกล่าวข้างต้น สรุปได้ว่าข้อสอบหรือแบบทดสอบที่ทำหน้าที่ต่างกัน หมายถึง ข้อสอบหรือแบบทดสอบที่วัดผู้สอบที่มีความสามารถเท่ากัน วัดเรื่องเดียวกัน แต่อยู่ต่างกลุ่มกันมีโอกาสในการที่จะตอบถูกไม่เท่ากัน ซึ่งแบบทดสอบที่ดีมีมาตรฐานต้องสร้างให้มีข้อสอบที่มีการทำหน้าที่ต่างกันน้อยที่สุดไม่เป็นการเสียเปรียบแก่ฝ่ายใดฝ่ายหนึ่ง

1.3 สาเหตุที่ก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบ

จากการนำข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบมาศึกษาภายหลังจากการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยค่าสถิติแบบต่าง ๆ ในแบบทดสอบหลาย ๆ ฉบับ เพื่อศึกษาถึงสาเหตุหรือแหล่งที่ก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบในตัวคำถามเหล่านั้น พบว่า สาเหตุที่น่าจะก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบมีหลายสาเหตุด้วยกัน ซึ่งสามารถสรุปสาเหตุที่ก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบมากที่สุด ได้ดังต่อไปนี้ (Scheuneman. 1982 : 180 – 195)

1. การเดา (guessing) อาจเกิดจากข้อสอบนั้นยากเกินไปหรือเวลาไม่เพียงพอจะก่อให้เกิดความไม่เท่ากันในโอกาสการตอบข้อสอบถูกของผู้สอบแต่ละคน

2. ความเร็ว (Speed) หรือเวลาในการสอบ จำทำให้เกิดการเดาหรือกรณีในเวลาน้อยอาจจะทำข้อสอบไม่ทัน ซึ่งจะมีผลกับข้อสอบข้อหลัง ๆ โดยเฉพาะในการศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดความถนัด

3. ความกำกวมหรือความไม่ชัดเจน (Unclear) ของข้อคำถาม นั่นคือ ข้อคำถามขาดความเป็นปรนัย การใช้ภาษาถิ่นหรือใช้คำที่ไม่เป็นภาษากลางในการสื่อความหมาย ซึ่งก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบกับกลุ่มภาษาใดภาษาหนึ่งขึ้นได้

4. ลำดับชั้นของคำถาม (Series) อาจจะเป็นสิ่งที่ก่อให้เกิดความสับสน หรือชี้แนะคำตอบของข้อสอบบางข้อได้

5. สถานภาพทางสังคมหรือความเกี่ยวข้องของทางสังคม (Social implication) ก็เป็นสิ่งที่ก่อให้เกิดความแตกต่างระหว่างกลุ่มผู้สอบแต่ละกลุ่มได้

6. ประสบการณ์หรือการฝึกฝนของแต่ละกลุ่มย่อย เป็นสิ่งที่ก่อให้เกิดการได้เปรียบเสียเปรียบของแต่ละกลุ่มค่อนข้างจะชัดเจน

7. องค์ประกอบทางวัฒนธรรม ความเป็นอยู่ ขนบธรรมเนียม เชื้อชาติ ศาสนา ก็จะเอื้อให้กับบางกลุ่มย่อย จึงทำให้เกิดการได้เปรียบเสียเปรียบในบางเนื้อหาวิชาได้

นอกจากนั้นความลำเอียงของข้อสอบอาจเกิดจากสาเหตุหรือแหล่งที่สำคัญ 2 แหล่ง คือ

1. การเลือกเนื้อหา (bias in selection) คือ ผู้สร้างข้อสอบเลือกเนื้อหาเฉพาะส่วนใดส่วนหนึ่งมาสร้างข้อสอบ ทำให้ได้ข้อสอบที่มีเนื้อหาไม่ครอบคลุมและไม่ได้สัดส่วนที่สมดุลกัน

2. การสร้างข้อสอบ (bias in construction) คือ การใช้ภาษาหรือข้อความบางอย่างที่เอื้อให้เกิดประโยชน์กับผู้สอบกลุ่มใดกลุ่มหนึ่ง

2. หลักการและวิธีการตรวจการทำหน้าที่ต่างกันของข้อสอบ

การตรวจการทำหน้าที่ต่างกันของข้อสอบ (DIF detection) เป็นการเปรียบเทียบผลการตอบข้อสอบเป็นรายข้อระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่ม ที่มีความสามารถหลัก (primary ability) ที่มุ่งวัดเท่ากันแต่คาดว่าจะมีความได้เปรียบหรือเสียเปรียบกันโดยกลุ่มหนึ่งถือเป็นกลุ่มอ้างอิง (Reference group) ซึ่งคาดว่าจะได้เปรียบในการตอบข้อสอบข้อนั้น หรือมีโอกาสตอบข้อสอบได้ถูกต้องมากกว่า ส่วนอีกกลุ่มคือ กลุ่มเปรียบเทียบ (Focal group) ซึ่งเป็นกลุ่มที่สนใจศึกษาและคาดว่าจะจะเป็นกลุ่มที่เสียเปรียบ

ในการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบจำเป็นต้องจับคู่ (matching) ผู้สอบตามความสามารถ ซึ่งเป็นเงื่อนไขสำคัญของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เกณฑ์การจับคู่ (matching criteria) ที่นิยมใช้กันมี 2 วิธี ดังนี้

1) เกณฑ์ภายนอก (External Criterion)

การวิเคราะห์การทำหน้าที่ต่างกัน โดยใช้เกณฑ์ภายนอกนี้ สามารถนำไปใช้ได้ทั้งข้อสอบรายข้อและแบบสอบทั้งฉบับ โดยการใช้คะแนนจากแบบทดสอบอื่นเป็นเกณฑ์ภายนอกแล้วใช้เทคนิคการวิเคราะห์การถดถอย (Regression analysis) เพื่อทำการเปรียบเทียบเส้นกราฟความสัมพันธ์ระหว่างตัวแปรเกณฑ์ กับตัวแปรทำนายระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

หลักการนี้มีจุดมุ่งหมาย เพื่อสร้างสมการทำหน้าที่ตัวแปรเกณฑ์ ซึ่งเป็นคะแนนของแบบสอบอื่นจากตัวแปรทำนายซึ่งเป็นคะแนนรายข้อ หรือคะแนนแบบสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ จะใช้คะแนนรายข้อเป็นตัวแปรทำนาย แต่ถ้าเป็นการวิเคราะห์การทำหน้าที่ต่างกันของแบบสอบ จะใช้คะแนนรวมของแบบสอบทั้งฉบับเป็นตัวแปรทำนาย สำหรับตัวแปรเกณฑ์ที่ใช้เป็นเกณฑ์ภายนอก อาจใช้คะแนนรวมทั้งฉบับ หรือเกรดเฉลี่ย หรือผลสัมฤทธิ์ในงานที่เกี่ยวข้องของผู้สอบ (Cronbach, 1970) สมการทำนายสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแสดงได้ดังนี้

$$\text{กลุ่มอ้างอิง (R)} \quad Y_i = A_R + B_R X_i$$

$$\text{กลุ่มเปรียบเทียบ (F)} \quad Y_i = A_F + B_F X_i$$

$$\text{เมื่อ } Y_i = \text{คะแนนของตัวแปรเกณฑ์ภายนอก}$$

$$X_i = \text{คะแนนของตัวแปรทำนาย}$$

$$A = \text{ค่าคงที่หรือค่าตัดแกน (intercept)}$$

$$B = \text{ค่าความชัน (slope)}$$

จากฟังก์ชันการทำนายทั้ง 2 สมการ สามารถเปรียบเทียบค่าตัดแกน (A) และค่าความชัน (B) ของเส้นกราฟระหว่างกลุ่มอ้างอิง และกลุ่มเปรียบเทียบได้ ถ้าเส้นกราฟดังกล่าวมีค่าความชัน

หรือค่าตัดเกณฑ์ แตกต่างกันไป สำหรับข้อสอบใด หรือแบบสอบใด แสดงว่าข้อสอบหรือแบบสอบนั้นมีการทำหน้าที่ต่างกัน โดยเข้าข้างกลุ่มผู้สอบที่มีค่าตัดเกณฑ์หรือค่าความชันที่สูงกว่า

การใช้เกณฑ์ภายนอกมีข้อดี คือ เกณฑ์ที่ใช้มีความอิสระจากข้อสอบ และแบบสอบที่ต้องการตรวจสอบ แต่มีจุดอ่อนตรงที่ความเหมาะสมของเกณฑ์ที่จะนำมาใช้ ในทางปฏิบัติเป็นการยากที่จะหาตัวแปรเกณฑ์ภายนอกจากแบบสอบฉบับอื่นที่มีความเที่ยงตรงเชิงทำนาย และมีความยุติธรรมสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ถ้าตัวแปรเกณฑ์ภายนอกขาดคุณสมบัติดังกล่าวจะทำให้ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบหรือแบบสอบขาดความแม่นยำ และความสมบูรณ์

2) เกณฑ์ภายใน (Internal Criterion)

การวิเคราะห์การทำหน้าที่ต่างกัน โดยใช้เกณฑ์ภายในเป็นการนำวิธีการทางสถิติมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหรือแบบทดสอบ โดยเน้นการพิจารณาจากโครงสร้างภายในของแบบทดสอบเป็นหลัก ด้วยการวิเคราะห์ผลจากการตอบข้อสอบและความสามารถหรือคะแนนจริงของผู้สอบที่ได้จากแบบทดสอบฉบับนั้น เพื่อนำมาเปรียบเทียบระหว่างผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ที่มีความสามารถหรือคะแนนจริงเท่ากันว่าจะมีผลการตอบหรือโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันหรือไม่ เพื่อบ่งชี้ถึงการทำหน้าที่ต่างกันของข้อสอบ การวิเคราะห์ในลักษณะนี้นิยมใช้ค่าสถิติต่างๆเป็นตัวบ่งชี้ ถึงการทำหน้าที่ต่างกันของข้อสอบ และสามารถแบ่งเป็น 3 กลุ่มใหญ่ ๆ (Camilli ; & Shepard. 1994) ดังนี้

2.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของทฤษฎีการทดสอบแบบมาตรฐานดั้งเดิม (Classical Test Theory : CTT)

เป็นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ใช้กันมาตั้งแต่สมัยเริ่มต้นของการพัฒนาเทคนิควิธีการตรวจสอบ แม้ว่าปัจจุบันจะไม่ค่อยได้นำมาใช้แล้ว เนื่องจากนักวัดผลได้มีพัฒนาเทคนิควิธีการใหม่ ๆ มาใช้แทน และได้ผลดีกว่า ประกอบกับวิธีการตรวจสอบแบบนี้มีจุดอ่อนตรงไม่สามารถแยกค่าความยากของข้อสอบออกจากค่าอำนาจได้ ทำให้เกิดความคลาดเคลื่อนในการตรวจสอบค่อนข้างสูง สำหรับวิธีการในกลุ่มนี้ประกอบด้วย วิธีแปลงค่าความยาก (Transformation Item Difficulty) วิธีวิเคราะห์ความแปรปรวน (Analysis of Variance) และวิธีพอยต์ไบเซเรียล (Point Biserial) เป็นต้น

2.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของทฤษฎีการตอบข้อสอบ (Item Response Theory : IRT)

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแนวนี้ จะพิจารณาเปรียบเทียบฟังก์ชันการตอบข้อสอบ (Item Response Function) หรือเส้นโค้งลักษณะของข้อสอบ (Item Characteristic Curves) ระหว่างกลุ่ม 2 ตั้งแต่ขึ้นไป ว่ามีความแตกต่างกันหรือไม่ การตรวจสอบการทำหน้าที่ต่างกัน

ของข้อสอบตามแนวทฤษฎีการตอบข้อสอบ มีจุดเด่นกว่าวิธีตรวจสอบตามแนวคิดทฤษฎีการทดสอบแบบมาตรฐานดั้งเดิม ดังนี้

1) ทฤษฎีการตอบข้อสอบสามารถประมาณค่าพารามิเตอร์ของข้อสอบ เช่น ค่าความยาก (b) ค่าอำนาจจำแนก (C) และค่าการเดา (c) ทำให้ลดการปนกัน (Confounding) ของค่าพารามิเตอร์ลงได้

2) คุณสมบัติทางสถิติของข้อสอบ (Statistical Properties of Item) สามารถอธิบายได้ละเอียดและชัดเจนกว่า เมื่อข้อสอบทำหน้าที่ต่างกันระหว่าง 2 กลุ่ม

3) คุณสมบัติทางสถิติของข้อสอบ สามารถแสดงด้วยแผนภาพที่เป็นกราฟ ทำให้เข้าใจง่ายและนำไปใช้กว้างขวาง

2.3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้ตารางการณ์จร (Contingency Table Approach) วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแนวนี้มีลักษณะดังนี้

2.3.1 ใช้สถิติไม่พารามิเตอร์ในการทดสอบ (Non – Parametric Statistic)

2.3.2 ใช้คะแนนที่สังเกตได้ (Observe Score) ประมาณค่าความสามารถระหว่างกลุ่ม

2.3.3 มีความสอดคล้องมากกับการเปรียบเทียบระหว่างเส้นกราฟคุณลักษณะของข้อสอบตามทฤษฎีการตอบข้อสอบ

2.3.4 ไม่จำเป็นต้องใช้สถิติขั้นสูง

2.3.5 ค่อนข้างประหยัด

รวมทั้งมีมีนโทษน้อยกว่าการตรวจสอบตามแนวคิดทฤษฎีการตอบข้อสอบ ได้แก่

1. มีข้อกำหนดเบื้องต้น ในการเปรียบเทียบความแตกต่างของผลการตอบข้อสอบระหว่างผู้เข้าสอบสองกลุ่ม โดยใช้เกณฑ์ภายใน เช่น ใช้คะแนนจากการสอบเป็น 0 – 1 เหมือนกัน

2. ทั้งสองวิธีสามารถนำไปประยุกต์ใช้ในการตรวจสอบซ้ำได้ (Iterative) เช่น ตามแนวคิด Purification ของลอร์ด (Lord. 1980 : 220) เป็นต้น

3. เป็นแนวคิดการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในลักษณะการวัดแบบหลายมิติ (Multidimensionality) ชนิดหนึ่ง เช่นเดียวกับทฤษฎีการตอบข้อสอบตัวอย่างวิธีการตรวจสอบในกลุ่ม ได้แก่ ไคสแควร์ (Chi – square) แมนเทล – เฮนเซล (Mantel – Haenzel) และล็อกลิเนีย (Log – linear) เป็นต้น

อย่างไรก็ตามนักวัดผลได้พัฒนาเทคนิควิธีใหม่ ๆ ขึ้นมาอย่างต่อเนื่อง เช่น วิธีการทำให้เป็นมาตรฐาน (Standardization Method) แบบจำลองการถดถอยโลจิสติก (Logistic Regression Model) และล่าสุดวิธีซิปเทสต์ (SIBTEST) (Shealy ; & Stout. 1993)

3. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของทฤษฎีการตอบข้อสอบ

3.1 ความเป็นมาของทฤษฎีการตอบข้อสอบ

ทฤษฎีการตอบข้อสอบ (Item Response Theory) หรือ ทฤษฎีคุณลักษณะแฝง (Latent Trait Theory) ได้รับการพัฒนามานานตั้งแต่ปีเนตต์ และ ไชมอน (1916) และ ริชาร์ดสัน (1936) ได้แสดงความสัมพันธ์ระหว่างพารามิเตอร์ของแบบจำลองทฤษฎีการตอบข้อสอบและทฤษฎีแบบดั้งเดิม ต่อมาปี 1943 - 1944 ลอร์เลย์ (Hambleton ; & Swaminathan. 1985) ได้เสนอวิธีการประมาณค่าพารามิเตอร์แนวใหม่ ซึ่งมีหลักการว่าผลการสอบของผู้สอบจากแบบทดสอบใด ๆ ขึ้นอยู่กับความสามารถ (Ability or Skill) ของผู้สอบ จนกระทั่งปี 1952 ลอร์ด ได้เสนอทฤษฎีใหม่ในรูปโค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC) โดยลอร์ดได้เสนอว่าโค้งลักษณะข้อสอบมีลักษณะเป็นโค้งปกติสะสม ต่อมาเรียกว่า แบบจำลองนอร์มัลอจิว (Normal Ogive Model) ซึ่งแบบจำลองนี้จะกล่าวถึงพารามิเตอร์ 2 ตัว คือ ค่าความยาก และค่าอำนาจจำแนก แต่เนื่องจากแบบจำลองนี้มี ความยากในการคำนวณ และขาดโปรแกรมคอมพิวเตอร์ที่มีประสิทธิภาพในการใช้วิเคราะห์ข้อมูลเพื่อพิสูจน์ทฤษฎี จึงทำให้ ลอร์ด ชะงักการทำงานเกี่ยวกับทฤษฎีนี้ไประยะหนึ่ง

ปี 1960 ราสช์ (Rasch) ได้ศึกษาเกี่ยวกับเรื่องนี้โดยไม่ทราบแนวคิดของ ลอร์ด และได้เสนอแนวคิดในรูปของพารามิเตอร์ตัวเดียว คือ ค่าความยากของข้อสอบเท่านั้น

ปี 1965 ลอร์ด ได้หันมาสนใจและพัฒนาทฤษฎีการตอบข้อสอบใหม่อีกครั้ง

ปี 1968 เบิร์นบอม (Birnbaum) ได้เสนอแบบจำลองโลจิสติก (Logistic Model) ที่ใช้พารามิเตอร์ 2 ตัว คือค่าความยาก และค่าอำนาจจำแนก ซึ่งเป็นแบบจำลองที่ง่ายกว่าของลอร์ดจึงทำให้แบบจำลองนี้เป็นที่นิยมแพร่หลาย และมีการพัฒนาขึ้นเรื่อย ๆ จนกระทั่งสามารถใช้ได้กับพารามิเตอร์ตัวเดียว และพารามิเตอร์สามตัว

3.2 หลักการของทฤษฎีการตอบข้อสอบ

ทฤษฎีการตอบข้อสอบ เป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างคุณลักษณะหรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรมการตอบข้อสอบของบุคคลนั้น ๆ ในรูปของฟังก์ชันทางคณิตศาสตร์ กล่าวคือ พฤติกรรมการตอบข้อสอบของบุคคลจะถูกกำหนดโดยคุณลักษณะหรือความสามารถที่มีอยู่ในตัวบุคคล (Lord ; & Novick. 1968 : 358) ซึ่งไม่สามารถจะสังเกตได้โดยตรง เขียนในรูปฟังก์ชันทางคณิตศาสตร์ได้ ดังนี้

$$y = f(\theta)$$

เมื่อ y แทน คะแนนที่ได้จากการตอบข้อสอบ ซึ่งเป็นพฤติกรรมการตอบข้อสอบ

θ แทน คุณลักษณะหรือความสามารถภายในตัวบุคคล

f แทน ฟังก์ชัน

ซึ่ง ลอร์ด เรียกฟังก์ชันนี้ว่า ฟังก์ชันการตอบข้อสอบ (Item Response Function) หรือฟังก์ชันคุณลักษณะของข้อสอบ (Item Characteristic Function) (Hambleton ; & Swaminatan. 1985: 13)

3.3 ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบ

1) ความเป็นมิติเดียว (Unidimensionality) ถ้าความสามารถของคน (Ability or Trait) มีอยู่หลายอย่าง ซึ่งความสามารถแต่ละอย่างนี้ ต่างก็ส่งผลต่อการตอบข้อสอบ (Test Performance) ต่าง ๆ ที่รวมกันเป็นแบบทดสอบ ถ้าผลการตอบข้อสอบ หรือคะแนนของผู้สอบนั้น สามารถอธิบายได้เพียงความสามารถเดียวก็เพียงพอแล้ว เรียกข้อตกลงนี้ว่ามีความมิติเดียว (Hambleton ; & Swaminatan. 1985 : 16) สำหรับในทางปฏิบัติ ก็คือ ข้อสอบแต่ละข้อในแบบทดสอบจะต้องวัดในคุณลักษณะ(Trait)เดียว แต่อย่างไรก็ตาม แฮมเบิลตัน และสวามินาธาน (Hambleton ; & Swaminatan. 1985 : 16 -17) ได้กล่าวว่าข้อตกลงเบื้องต้นข้อนี้ไม่ได้เข้มงวดนักขอให้มีความเด่น (Dominant) ที่จะวัดในองค์ประกอบใดองค์ประกอบหนึ่งก็ใช้ได้ ซึ่งสอดคล้องกับผลการวิจัยของดราสโกว์ และพาร์สันส์ และเรคเคส (McCauley ; & Mendoza. 1985 : 390 ; citing Drasgow ; & Parsons. 1983 ; reckase. 1979) ที่พบว่า การประมาณค่าพารามิเตอร์ต่าง ๆ ยังสามารถทำได้เมื่อการวัดนั้นอาจจะมีการเบนไปในทางหลายมิติบ้างเล็กน้อย

วิธีการตรวจสอบว่าแบบทดสอบนั้นวัดในมิติเดียวหรือไม่นั้น ทำได้หลายวิธี เช่น โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) แล้วสังเกตค่าไอเกน (Eigen Value) ค่าสูงสุดว่าแตกต่างจากค่าอื่น ๆ อย่างชัดเจนหรือไม่ สำหรับ ลอร์ด และโนวิก (Lord ; & Novick. 1968 : 382 ; Lord. 1980 : 20) ยังได้แนะนำว่าการวิเคราะห์องค์ประกอบควรจะใช้ค่าสหสัมพันธ์แบบเตตราคอร์ริกไม่ควรใช้การคำนวณค่าสหสัมพันธ์แบบเพียร์สัน เพราะว่าข้อมูลที่เหมาะ (Fit) กับแบบจำลองของทฤษฎีการตอบข้อสอบไม่ได้สัมพันธ์กันในเชิงเส้นตรง (Hambleton ; & Swaminatan. 1985 : 306)

2) ความเป็นอิสระในการตอบข้อสอบ (Local Independence) หมายถึง โอกาสในการตอบข้อสอบแต่ละข้อได้ถูกต้องต้องเป็นอิสระจากกัน นั่นคือ การตอบข้อสอบข้อใดข้อหนึ่งได้ถูกหรือผิดจะไม่มีผลต่อการตอบข้ออื่น ๆ ด้วย (Lord. 1980 : 19 ; Hambleton ; & Swaminatan. 1985 : 23) หรือจะกล่าวในเชิงคณิตศาสตร์ได้ว่า ความเป็นอิสระในการตอบข้อสอบ หมายถึง ความน่าจะเป็นที่จะตอบข้อสอบถูกต้องทั้งหมด จะเท่ากับผลคูณของความน่าจะเป็นของการตอบข้อสอบถูกต้องเป็นรายข้อ (Lord. 1980 : 19 ; Hambleton ; & Swaminatan. 1985 : 23) แมกโดนัลด์ (McDonald. 1982 : 379-396) และแฮมเบิลตันกับสวามินาธาน (Hambleton ; & Swaminathan. 1985 : 22) มีความเห็นตรงกันว่าถ้ามีความมิติเดียวแล้วจะมีความเป็นอิสระในการตอบข้อสอบ ซึ่งมีความสำคัญต่อการ

ประมาณค่าพารามิเตอร์ต่าง ๆ มาก ถ้าขาดความเป็นอิสระในการตอบข้อสอบแล้ว ค่าพารามิเตอร์ต่าง ๆ ที่ประมาณค่าได้ จะมีความคลาดเคลื่อนจากความเป็นจริง

3) โค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC) โค้งแสดงลักษณะข้อสอบ เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงถึงความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบข้อนั้นได้ ถูกต้องกับระดับความสามารถที่วัดได้โดยใช้ชุดของข้อสอบ หรือแบบทดสอบนั้น (Hambleton ; & Cook. 1977 : 80) ฉะนั้นจึงทำให้ความน่าจะเป็นในการตอบข้อสอบถูกในแต่ละข้อไม่แปรเปลี่ยนด้วย

3.4 แบบจำลองในทฤษฎีการตอบข้อสอบ (Item Response Model)

แบบจำลองในทฤษฎีการตอบข้อสอบมีการพัฒนาต่อเนื่องกันมานานและมีหลายลักษณะประกอบด้วยแบบจำลองกัตต์แมนสมบูรณ์แบบ (Guttman Perfect Scale) แบบจำลองเชิงเส้น (Linear Model) แบบจำลองนอร์มินัล (Norminal Model) แบบจำลองนอร์มัลลอจิสติก (Logistic Model) แต่ในที่นี้จะกล่าวถึงบางลักษณะที่น่าสนใจ และใช้แพร่หลาย ดังนี้

แบบจำลองนอร์มัลลอจิสติก

ลอร์ด (Lord ; & Novick. 1968 : 369 ; citing Lawley. 1943,1944) เป็นผู้พัฒนาขึ้น และในปี 1952 ลอร์ด ได้พัฒนาวิธีการประมาณค่าพารามิเตอร์ และนำมาประยุกต์ใช้โดยใช้พารามิเตอร์ 2 ตัว ได้แก่ ค่าความยาก และค่าอำนาจจำแนก เขียนเป็นรูปแบบจำลองทางคณิตศาสตร์ ดังนี้

$$P_i(\theta) = \int_{-a}^{a_i(\theta - b_i)} \frac{e^{-z^2}}{\sqrt{2\pi}} dz$$

เมื่อ $P_i(\theta)$ แทน ความน่าจะเป็นที่ผู้มีความสามารถ (θ) ตอบข้อสอบข้อที่ i ได้ถูกต้อง

a_i	แทน	ค่าอำนาจจำแนกของข้อสอบข้อที่ i
b_i	แทน	ค่าความยากของข้อสอบข้อที่ i
θ	แทน	ระดับความสามารถของผู้สอบ
Z	แทน	ความเบี่ยงเบนปกติ จากการแจกแจงที่ค่าเฉลี่ย b และความเบี่ยงเบนมาตรฐาน $1/a$

แบบจำลองนอร์มัลลอจิสติก (Normal Ogive Model) เป็นแบบจำลองที่ ลอร์ด ได้ใช้เวลาอันยาวนานในการพัฒนาแบบจำลองนี้ ในระยะแรก ๆ เป็นไปอย่างเชื่องช้ามาก เนื่องจากความสลับซับซ้อนทางคณิตศาสตร์ที่เป็นพื้นฐานของทฤษฎียากแก่การปฏิบัติ หรือนำมาใช้จริง คนทั่วไปจึงไม่นิยมนำแบบจำลองนี้มาใช้ แม้แต่ ลอร์ดเอง ในระยะหลังนี้ก็ได้นำมาใช้แบบจำลองโลจิสติก

(Logistic Model) แทน เพราะเป็นแบบจำลองที่มีลักษณะคล้ายกันมากแต่วิธีการคำนวณค่าพารามิเตอร์ต่าง ๆ ได้ง่าย (Lord ; & Novick. 1968 : 366) แต่อย่างไรก็ตาม แบบจำลองนี้ก็นับว่ามีประโยชน์อย่างมาก สำหรับการพัฒนาแบบจำลองโลจิสติก (Logistic Model) ต่อมาแบบจำลองโลจิสติก (Logistic Model)

ผู้มีบทบาทสำคัญในการพัฒนาแบบจำลองนี้ คือ เบิร์นบอม (Hambleton ; & Swaminathan. 1985 : 36 ; citing Bimbaum. 1957; 1958a, 1958b 1968) โดยในครั้งแรกเบิร์น บอมได้พัฒนาแบบจำลองโลจิสติก แบบ 2 พารามิเตอร์ คือ พารามิเตอร์ค่าความยาก (b) และค่าอำนาจจำแนก (a) ซึ่งเป็นแบบจำลองที่มีโค้งแสดงลักษณะข้อสอบ (Item Characteristic Curve : ICC) คล้ายกับโค้งลักษณะข้อสอบ (ICC) ของแบบจำลองนอร์มัลออร์โจีฟ (Normal Ogive Model) แบบ 2 พารามิเตอร์ ที่ลอร์ด ได้เสนอไว้เมื่อปี ค.ศ. 1952 มาก โดยโค้งแสดงลักษณะทั้งสองมีความแตกต่างกันไม่เกิน .01 สำหรับทุก ๆ ค่าความสามารถ

เนื่องจากแบบจำลองโลจิสติก ได้ใช้แบบจำลองทางคณิตศาสตร์ที่สามารถนำไปประยุกต์ใช้ได้ง่ายแบบนอร์มัลออร์โจีฟ จึงทำให้คนนำแบบจำลองนี้ไปใช้กันมาก ต่อมา เบิร์นบอม ได้เพิ่มพารามิเตอร์ค่าการเดา (c) เข้าไปอีกตัวหนึ่ง จึงกลายเป็นแบบ 3 พารามิเตอร์ นอกจากนั้นยังได้ตัดพารามิเตอร์ค่าอำนาจจำแนก (a) ออกจากแบบ 2 พารามิเตอร์ จึงเหลือเพียงพารามิเตอร์เดียว คือ ค่าความยาก (b) ซึ่งเรียกว่า 1 พารามิเตอร์ สำหรับแบบจำลองทางคณิตศาสตร์ของแต่ละแบบจำลองย่อย มีดังนี้

1) แบบจำลองโลจิสติก แบบ 3 พารามิเตอร์ (Three Parameter Logistic Model) มีแบบจำลองทางคณิตศาสตร์ ดังนี้ (Hambleton Swaminathan ; & Rogers. 1991 : 17)

$$P_i(\theta) = c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}$$

- เมื่อ $P_i(\theta)$ คือ ความน่าจะเป็นซึ่งมีระดับความสามารถ (θ) ตอบข้อสอบข้อที่ i ได้ถูกต้อง
- a_i คือ ค่าอำนาจจำแนกของข้อสอบข้อที่ i โดยทฤษฎีแล้ว a จะมีค่าระหว่าง $-\infty$ ถึง $+\infty$ ข้อสอบที่ใช้ได้ควรมีค่าเป็นบวก โดยปกติจะมีค่าไม่เกิน 2.00
- b_i คือ ค่าความยากของข้อสอบข้อที่ i โดยทฤษฎีแล้ว b จะมีค่าระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติ b จะมีค่าอยู่ระหว่าง -2.00 ถึง +2.00 ข้อที่มีค่า b ใกล้ -2.00 เป็นข้อที่ง่ายมาก และข้อที่มีค่าใกล้ +2.00 เป็นข้อที่ยากมาก (Hambleton ; & Cook. 1977: 81)
- c_i คือ ค่าการเดาของข้อสอบ หมายถึง ค่าความน่าจะเป็นที่ผู้สอบมีความสามารถ

ต่ำสุดทำข้อสอบนี้ได้ถูกต้อง ข้อสอบที่มีค่าการเดาสูงเป็นข้อสอบที่ไม่ดี ควรมีการเดาไม่เกิน 0.30

- θ คือ ระดับความสามารถของผู้สอบ โดยทฤษฎีแล้วจะมีค่าอยู่ระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติ จะมีค่าอยู่ระหว่าง -3.00 ถึง +3.00
- D คือ ค่าคงที่(Scaling Factor) เพื่อปรับโค้งแสดงลักษณะข้อสอบ (ICC) ให้มีลักษณะใกล้เคียงกับโค้งลักษณะข้อสอบของแบบจำลองนอร์มัลออร์โงฟ (Normal Ogive Model)
- e คือ ค่าเอ็กซ์โปเนนเชียล (Exponential) เป็นค่าคงตัวหนึ่งซึ่งมีค่าเท่ากับ 2.7182818

2) แบบจำลองโลจิสติก แบบ 2 พารามิเตอร์ (Two Parameter Logistic Model) มีแบบจำลองทางคณิตศาสตร์ ดังนี้ (Hambleton Swaminathan ; & Rogers. 1991 : 15)

$$P_i(\theta) = \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}}$$

3) แบบจำลองโลจิสติก แบบ 1 พารามิเตอร์ (One Parameter Logistic Model) เป็นแบบจำลองที่ถือว่าข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากันหมด ซึ่งเท่ากับ a และมีค่าการเดา (c) เป็นศูนย์ โดยมีแบบจำลองทางคณิตศาสตร์ ดังนี้ (Hambleton Swaminathan ; & Rogers. 1991 : 12)

$$P_i(\theta) = \frac{e^{(\theta-b_i)}}{1 + e^{(\theta-b_i)}}$$

3.5 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวคิดทฤษฎีการตอบข้อสอบ

การทำหน้าที่ต่างกันของข้อสอบในแง่ของทฤษฎีการตอบข้อสอบ ก็คือ การที่ฟังก์ชันการตอบข้อสอบข้อนั้น ๆ ของประชากรย่อย 2 กลุ่ม มีความแตกต่างกันทำให้เกิดพื้นที่ระหว่างโค้งแสดงลักษณะข้อสอบ

จากนิยามดังกล่าว การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวทฤษฎีการตอบข้อสอบสามารถกระทำได้โดยเปรียบเทียบฟังก์ชันการตอบข้อสอบของผู้สอบกลุ่มย่อยที่ต้องการศึกษา ซึ่งสามารถแบ่งออกเป็น 2 แนวทางใหญ่ ๆ ดังนี้ (Hambleton Swaminathan ; & Roger. 1991 : 110)

1. การเปรียบเทียบค่าพารามิเตอร์ของข้อสอบระหว่าง 2 กลุ่ม ว่ามีความแตกต่างกันหรือไม่ มีผู้ศึกษาค้นคว้ามากพอสมควร ได้แก่ งานวิจัยของคิม และโคเฮน (Kim ; & Cohen. 1994) ลอร์ด (Lord. 1977, 1980) ทิสเซน สเติร์นเบิร์ก และวายนเนอร์ (Thissen Steinberg ; & Wainer. 1988, 1993) และไรท์ และสโตน (Wright ; & Syone. 1979) อย่างไรก็ตาม ในกลุ่มนี้ วิธีที่ได้รับความนิยมก็คือ วิธีที่อาศัยแนวคิดลอร์ด (Lord's χ^2) ในระยะหลังมีวิธีที่น่าสนใจอีกวิธี คือ วิธีทดสอบอัตราส่วนไลเคิลฮูด (Likelihood Ratio Test) ซึ่ง ทิสเซน สเติร์นเบิร์ก และวายนเนอร์ คิม และโคเฮน ก็ได้ศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีนี้ รวมทั้งเปรียบเทียบผลการตรวจสอบกับวิธีอื่นด้วย เช่น Lord's χ^2 และ Raju's Area Measures เป็นต้น

สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของลอร์ด สามารถใช้ได้กับแบบจำลองโลจิสติก ชนิด 1 พารามิเตอร์ 2 พารามิเตอร์ และ 3 พารามิเตอร์ มีขั้นตอนสรุปได้ดังนี้

1. ทดสอบข้อมูล (Test Model) กับแบบจำลองที่ศึกษา
2. ประมาณค่าพารามิเตอร์ของข้อสอบ (a, b, c)
3. แปลงหน่วยของค่าพารามิเตอร์ให้อยู่ในหน่วยเดียวกัน หรือหน่วยร่วม (Common Scale)
4. นำค่าพารามิเตอร์ของข้อสอบมาเปรียบเทียบกัน โดยใช้สถิติทดสอบ χ^2

โดยมีสมมติฐานในการทดสอบ คือ $H : b_1 = b_2, a_1 = a_2, c_1 = c_2$

สถิติทดสอบ (Hambleton Swaminathan ; & Roger. 1991 : 111)

$$\chi^2 = (a_{diff} \ b_{diff} \ c_{diff})' \Sigma^{-1} (a_{diff} \ b_{diff} \ c_{diff})$$

เมื่อ $a_{diff} = a_2 - a_1$

$$b_{diff} = b_2 - b_1$$

$$c_{diff} = c_2 - c_1$$

และ Σ แทน เมตริกซ์ Variance และ Covariance ของความแตกต่างระหว่างพารามิเตอร์ข้อสอบ หากศึกษาเป็นแบบจำลอง 1 พารามิเตอร์ สถิติทดสอบ เป็นดังนี้

$$\chi^2 = \frac{b_{diff}^2}{Var(b_1) + Var(b_2)}$$

2) การวัดพื้นที่ความแตกต่างระหว่างฟังก์ชันการตอบข้อสอบจาก 2 กลุ่ม เพื่อพิจารณาขนาดของการทำหน้าที่ต่างกันของข้อสอบ มีผู้ศึกษาค้นคว้าไว้ เช่น งานวิจัยของ คิม และโคเฮน (Kim ; & Cohen. 1991) ลินน์ เลวิน ฮาสติงส์ และวาร์ดรอป (Linn Levine Hastings ; & Wardrop. 1981) รัดเนอร์ (Rudner. 1977) และวายนเนอร์ (Wainer. 1993) เป็นต้น ในกลุ่มนี้ระยะหลังนิยมใช้แนวคิดของราฐู และแนวคิดของ คิม และโคเฮน ซึ่งราฐูเสนอให้ใช้การวัดขนาดพื้นที่บนช่วงความสามารถแบบเปิด (Open Interval) บางครั้งเรียกว่า “การวัดพื้นที่ที่สัมบูรณ์” (Exact Area Measures) ส่วนคิม และโคเฮน เสนอให้วัดขนาดพื้นที่บนช่วงความสามารถแบบปิด (Closed Interval Area) ภายใต้การวัดขนาดพื้นที่ทั้ง 2 แนวคิดนี้ ก็สามารถวัดพื้นที่ได้ 2 ลักษณะอีกด้วย คือ พื้นที่แบบมีเครื่องหมาย และไม่มีเครื่องหมาย (Signed and Unsigned Area) โดยมีสูตรในการคำนวณ ดังนี้

ถ้าเป็นแบบจำลอง 3 พารามิเตอร์

$$P(\theta) = c + (1 - c)P^*(\theta)$$

โดยที่

$$P^*(\theta) = 1 + \exp[Da(\theta - b)]^{-1}$$

a, b และ c เป็นพารามิเตอร์ที่แสดงลักษณะของข้อสอบ และ D เป็นตัวคงที่ปกติกำหนดให้ = 1.7

การคำนวณขนาดของพื้นที่ระหว่างความสามารถที่กำหนดไว้ 2 จุด บนมาตราความสามารถหาได้จาก

$$S(\theta_1 - \theta_2) = \int_{\theta_1}^{\theta_2} P(\theta)d(\theta) = c(\theta_1 - \theta_2) + (1 - c)(Da)^{-1} \ln \frac{\{1 - \exp[Da(\theta_2 - b)]\}}{\{1 + \exp[Da(\theta_2 - b)]\}}$$

ในการศึกษาการทำหน้าที่ต่างกันของข้อสอบ จะมีพารามิเตอร์ของ 2 กลุ่ม คือ กลุ่มอ้างอิง (Reference group) ได้แก่ a_R b_R และ c_R และกลุ่มเปรียบเทียบ (Focal group) ได้แก่ a_F b_F และ c_F พื้นที่ชนิดมีเครื่องหมาย (Closed - Interval Signed Area - CSA) หาได้จาก

$$CSA = \int_{\theta_1}^{\theta_2} [P_R(\theta) - P_F(\theta)]d\theta = S_R(\theta_1, \theta_2) - S_F(\theta_1, \theta_2)$$

และพื้นที่ชนิดไม่มีเครื่องหมาย (Closed-Interval Unsigned Area - CUA) หาได้จาก

$$CUA = \int_{\theta_1}^{\theta_2} |P_R(\theta) - P_F(\theta)| d\theta$$

จากการศึกษา Kim และโคเฮน (Kim and Cohen. 1991) ได้ศึกษาเปรียบเทียบความแตกต่างของพื้นที่แบบไม่มีเครื่องหมาย (Unsigned Area) และแบบมีเครื่องหมาย (Signed Area) พบว่า พื้นที่แบบมีเครื่องหมายมีความคลาดเคลื่อนต่ำกว่าแบบไม่มีเครื่องหมาย และการวัดพื้นที่ช่วงเปิดและแบบช่วงปิดจะให้ผลสอดคล้องกันค่อนข้างมาก

4. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีแมนเทล-แฮนส์เซล (Mantel-Haenszel ; MH)

วิธีแมนเทล-แฮนส์เซล พัฒนาโดย Mantel-Haenszel (Camilli ; & Shepard. 1994 ; citing Mantel-Haenszel. 1959) ซึ่งแต่เดิมเป็นสถิติสำหรับเปรียบเทียบอัตราส่วนแต่ต่อร่วม และทำการทดสอบอัตราส่วนเปรียบเทียบด้วยไคสแควร์ ต่อมาฮอลแลนด์ (Holland 1985 ; citing Holland ; & Thayer. 1988) ได้นำวิธีแมนเทล-แฮนส์เซล มาประยุกต์ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับหน่วยงานการบริการทดสอบทางการศึกษาของประเทศสหรัฐอเมริกา วิธีแมนเทล-แฮนส์เซล เป็นที่ยอมรับจากนักวิจัยอย่างกว้างขวางว่าเป็นวิธีที่ใช้ง่าย สะดวก และประหยัด โดยใช้หลักการของตารางการถัวแบบทฤษฎีการตอบข้อสอบ ไม่มีการคำนวณทวนซ้ำ จึงสามารถนำไปใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ง่ายมีขั้นตอนการคำนวณที่ไม่สลับซับซ้อน มีการทดสอบทางสถิติแบบนอนพาราเมตริก ซึ่งไม่จำเป็นต้องใช้โมเดลประมาณค่า

หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธีแมนเทล-แฮนส์เซล จะเปรียบเทียบผลการตอบข้อสอบระหว่างผู้สอบสองกลุ่ม คือกลุ่มเปรียบเทียบกับกลุ่มอ้างอิง ในการเปรียบเทียบผลการตอบข้อสอบจะเปรียบเทียบทุกระดับความสามารถของผู้สอบกลุ่มย่อยสองกลุ่มที่มีระดับความสามารถเท่ากัน ในทางปฏิบัติมักใช้คะแนนรวมของแบบทดสอบเป็นเกณฑ์การจับคู่กลุ่มผู้สอบ

ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบเป็นรายข้อ เมื่อจับคู่กลุ่มผู้สอบแล้วจะนำข้อมูลผลการตอบข้อสอบนั้น ระหว่างผู้สอบย่อยสองกลุ่มมาจัดลงในตารางการถัวแบบ 2×2 (กลุ่มผู้สอบ 2 กลุ่ม \times ผลการตอบ 2 แบบ) โดยที่ตารางการถัว 1 ตารางแทนคะแนนรวม 1 ระดับ ดังนั้นถ้ามีคะแนนรวมของกลุ่มผู้สอบทั้งสิ้น k ระดับ จะต้องสร้างตารางการถัวแบบ 2×2 ทั้งหมด k ตาราง สำหรับตารางการถัวแบบ 2×2 ของข้อสอบแต่ละข้อที่มีคะแนนรวมระดับ j แสดงดังตาราง 1

ตาราง 1 ผลการตอบข้อสอบข้อหนึ่งระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีคะแนนรวม j

กลุ่ม	คะแนนผลของการตอบข้อสอบที่ศึกษา		รวม
	ตอบถูก (1)	ตอบผิด (0)	
R	A_j	B_j	nR_j
F	C_j	D_j	nF_j
รวม	m_{1j}	m_{0j}	T_j

- เมื่อ
- A_j = จำนวนผู้สอบกลุ่มอ้างอิงที่มีคะแนนรวม j ซึ่งตอบข้อสอบถูก
 - B_j = จำนวนผู้สอบกลุ่มอ้างอิงที่มีคะแนนรวม j ซึ่งตอบข้อสอบผิด
 - C_j = จำนวนผู้สอบกลุ่มเปรียบเทียบที่มีคะแนนรวม j ซึ่งตอบข้อสอบถูก
 - D_j = จำนวนผู้สอบกลุ่มเปรียบเทียบที่มีคะแนนรวม j ซึ่งตอบข้อสอบผิด
 - m_{1j} = จำนวนผู้สอบทั้งหมดที่มีคะแนนรวม j ซึ่งตอบข้อสอบถูก
 - m_{0j} = จำนวนผู้สอบทั้งหมดที่มีคะแนนรวม j ซึ่งตอบข้อสอบผิด
 - nR_j = จำนวนผู้สอบกลุ่มอ้างอิงที่มีคะแนนรวม j
 - nF_j = จำนวนผู้สอบกลุ่มเปรียบเทียบที่มีคะแนนรวม j
 - T_j = จำนวนผู้สอบทั้งหมดที่มีคะแนนรวม j

จากผลการตอบตามตาราง 1 นำมาคำนวณสัดส่วนของผลการตอบข้อสอบถูกและผิดดัง
ตาราง 2

ตาราง 2 สัดส่วนของผลการตอบข้อสอบข้อหนึ่งระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มี
คะแนนรวม j

กลุ่ม	คะแนนผลของการตอบข้อสอบที่ศึกษา		รวม
	ตอบถูก (1)	ตอบผิด (0)	
R	pR_j	qR_j	1.00
F	pF_j	qF_j	1.00

- เมื่อ pR_j = สัดส่วนของผู้สอบกลุ่มอ้างอิงที่มีคะแนนรวม j ซึ่งตอบข้อสอบถูก

qR_j = สัดส่วนของผู้สอบกลุ่มอ้างอิงที่มีคะแนนรวม j ซึ่งตอบข้อสอบผิด

pF_j = สัดส่วนของผู้สอบกลุ่มเปรียบเทียบที่มีคะแนนรวม j ซึ่งตอบข้อสอบถูก

qF_j = สัดส่วนของผู้สอบกลุ่มเปรียบเทียบที่มีคะแนนรวม j ซึ่งตอบข้อสอบผิด

ในการทดสอบสมมติฐานของการทำหน้าที่ต่างกันของข้อสอบจะตั้งสมมติฐานศูนย์ (H_0) และสมมติฐานอื่น (H_1) ดังนี้

$$H_0: \frac{pR_j}{qR_j} = \frac{pF_j}{qF_j} \quad ; \quad j = 1, 2, 3, \dots, k$$

$$H_1: \frac{pR_j}{qR_j} = \alpha \frac{pF_j}{qF_j} \quad ; \quad j = 1, 2, 3, \dots, k \quad \text{เมื่อ } \alpha \neq 1$$

สมมติฐานศูนย์ตั้งขึ้นเป็นอิสระอย่างมีเงื่อนไขระหว่างกลุ่มผู้สอบและคะแนนผลการตอบข้อสอบที่ศึกษา ดังนั้นสัดส่วนคะแนนที่ได้จากตารางที่ 3.2 ภายใต้อสมมติฐานศูนย์ที่ตั้งไว้สามารถคำนวณเป็นค่าคาดหวัง (Expected values) ในแต่ละเซลล์ได้ดังนี้

$$E(A_j) = \frac{nR_j m_{1j}}{T_j}$$

$$E(B_j) = \frac{nR_j m_{0j}}{T_j}$$

$$E(C_j) = \frac{nF_j m_{1j}}{T_j}$$

$$E(D_j) = \frac{nF_j m_{0j}}{T_j}$$

สำหรับพารามิเตอร์ α ภายใต้อสมมติฐานอื่นเรียกว่า อัตราส่วนแอดัมต่อรวม (Common odds ratio) ซึ่งคำนวณได้จาก

$$\alpha = \frac{\frac{pR_j}{qR_j}}{\frac{pF_j}{qF_j}} = \frac{pR_j qF_j}{qR_j pF_j}$$

ถ้า $\alpha = 1$ แสดงว่าโอกาสของการตอบข้อสอบระหว่างผู้สอบทั้งสองกลุ่ม มีค่าเท่ากัน ถ้า $\alpha > 1$ แสดงว่าผู้สอบกลุ่มอ้างอิงมีโอกาสตอบข้อสอบถูกมากกว่าผู้สอบกลุ่มเปรียบเทียบและถ้า

$\alpha < 1$ แสดงว่าผู้สอบกลุ่มเปรียบเทียบมีโอกาสตอบข้อสอบถูกมากกว่าผู้สอบกลุ่มอ้างอิง แมนเทิลและแฮนเซล ได้เสนอวิธีประมาณค่า α จากตารางแบบ 2×2 จำนวน k ระดับ ดังนี้

$$\hat{\alpha}_{MH} = \frac{\sum_{j=1}^k A_j D_j / T_j}{\sum_{j=1}^k B_j C_j / T_j}$$

$\hat{\alpha}_{MH}$ เป็นค่าประมาณขนาดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบ (DIF effect size) ซึ่งมีค่าอยู่ระหว่าง 0 ถึง ∞ ฮอลแลนด์และเธเยอร์ (1985 cited in Holland ; & Thayer, 1988) ได้เสนอให้แปลงค่า $\hat{\alpha}_{MH}$ เป็นสเกลมาตรฐานเดลต้า (Delta scale ; MH_{D-DIF} หรือ MH_{DIF}) ดังนี้

$$MH_{DIF} = -2.35 \ln(\hat{\alpha}_{MH})$$

ค่า MH_{DIF} ดังกล่าวสามารถนำไปพิจารณาค่าความยากของข้อสอบ กล่าวคือ ถ้า MH_{DIF} มีค่าเป็นศูนย์ แสดงว่าข้อสอบนั้นมีความยากสำหรับแต่ละกลุ่มเท่ากัน ถ้า MH_{DIF} มีค่าเป็นลบแสดงว่าข้อสอบยากสำหรับผู้สอบกลุ่มเปรียบเทียบมากกว่ากลุ่มอ้างอิง และถ้า MH_{DIF} มีค่าเป็นบวกแสดงว่าแสดงว่าข้อสอบยากสำหรับผู้สอบกลุ่มอ้างอิงมากกว่ากลุ่มเปรียบเทียบ ส่วนการประมาณค่าความคลาดเคลื่อนมาตรฐานของ MH_{DIF} สามารถคำนวณได้จากสูตร ดังนี้

$$SE(MH_{DIF}) = 2.35 \sqrt{\text{Var}\left[\ln\left(\hat{\alpha}_{MH}\right)\right]}$$

โดยที่
$$\text{Var}\left[\ln\left(\hat{\alpha}_{MH}\right)\right] = \frac{\sum_{j=1}^k U_j V_j / T_j^2}{2 \left[\sum_{j=1}^k A_j D_j / T_j \right]^2}$$

ขณะที่
$$U_j = A_j D_j + \alpha_{MH} (B_j C_j)$$

และ
$$V_j = (A_j + D_j) + \alpha_{MH} (B_j + C_j)$$

ในการทดสอบนัยสำคัญของสมมติฐาน จะต้องนำค่า $\hat{\alpha}_{MH}$ หรือค่า MH_{DIF} ไปทดสอบกับสถิติแมนเทิล-แฮนเซลไค-สแควร์ (χ^2_{MH}) ที่ระดับชั้นความเป็นอิสระเท่ากับ 1 (df = 1) สถิติ χ^2_{MH} มีสูตรในการคำนวณดังนี้

$$\chi^2_{MH} = \frac{\left[\sum_{j=1}^k A_j - E(A_j) - 0.5 \right]^2}{\sum_{j=1}^k \text{Var}(A_j)}$$

โดยที่
$$E(A_j) = \frac{nR_j m_{1j}}{T_j}$$

$$\text{Var}(A_j) = \frac{nR_j n F_j m_j m_0}{T_j^2 (T_j - 1)}$$

เมื่อ $E(A_j)$ = ค่าคาดหวังของจำนวนผู้สอบกลุ่มอ้างอิงที่ระดับคะแนน j ซึ่งตอบข้อสอบถูก

$$\text{Var}(A_j) = \text{ค่าความแปรปรวนของจำนวนผู้สอบกลุ่มอ้างอิงที่ระดับคะแนน } j \text{ ซึ่งตอบข้อสอบถูก}$$

สำหรับเกณฑ์ในการตัดสินข้อสอบที่ทำหน้าที่ต่างกัน คือ ข้อสอบที่มีค่า $\hat{\alpha}_{MH}$ แตกต่างจาก 1 อย่างมีนัยสำคัญทางสถิติหรือค่า $\hat{\alpha}_{MH}$ แตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติ โดยมีเกณฑ์ในการพิจารณา ดังนี้

สำหรับค่า $\hat{\alpha}_{MH}$ แปลผลการทดสอบได้ดังนี้

1) ค่า $\hat{\alpha}_{MH} = 1$ หรือไม่แตกต่างจาก 1 อย่างมีนัยสำคัญ แสดงว่าข้อสอบนั้นทำหน้าที่ไม่แตกต่างกันระหว่างกลุ่ม (no DIF)

2) ค่า $\hat{\alpha}_{MH} > 1$ แสดงว่าข้อสอบนั้นทำหน้าที่แตกต่างกันระหว่างกลุ่มโดยจะเข้าข้างกลุ่มอ้างอิง

3) ค่า $\hat{\alpha}_{MH} < 1$ แสดงว่าข้อสอบนั้นทำหน้าที่แตกต่างกันระหว่างกลุ่มโดยจะเข้าข้างกลุ่มเปรียบเทียบ

ส่วนค่า MH_{DIF} แปลผลการทดสอบได้ดังนี้

1) ค่า MH_{DIF} เท่ากับ 0 หรือไม่แตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติ แสดงว่าข้อสอบนั้นทำหน้าที่ไม่แตกต่างกันระหว่างกลุ่ม (no DIF)

2) ค่า MH_{DIF} แตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติ และมีค่าเป็นบวก (positive) แสดงว่าข้อสอบนั้นทำหน้าที่แตกต่างกันระหว่างกลุ่ม โดยจะเข้าข้างกลุ่มเปรียบเทียบ (F)

3) ค่า MH_{DIF} แตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติ และมีค่าเป็นลบ (negative) แสดงว่าข้อสอบนั้นทำหน้าที่แตกต่างกันระหว่างกลุ่ม โดยจะเข้าข้างกลุ่มอ้างอิง (R)

นอกจากนี้ขนาดของ $|MH_{DIF}|$ สามารถนำไปใช้แปลผลถึงระดับของการทำหน้าที่แตกต่างกันของข้อสอบได้ ถ้า $0 < |MH_{DIF}| < 1.00$ แสดงว่าข้อสอบทำหน้าที่แตกต่างกันเล็กน้อย ถ้า $1.00 \leq |MH_{DIF}| \leq 1.50$ แสดงว่าข้อสอบทำหน้าที่แตกต่างกันปานกลาง แต่ถ้า $|MH_{DIF}| > 1.50$ แสดงว่าข้อสอบทำหน้าที่แตกต่างกันมาก (Zieky. 1993)

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยวิธี MH นี้ ใช้คะแนนรวมของแบบสอบเป็นเกณฑ์การจับคู่ ซึ่งมีจุดอ่อนในด้านความไม่เป็นอิสระของคะแนนรวม กับคะแนนรายข้อที่ทำการศึกษา

ฮอลแลนด์และเทเยอร์ (Holland ; & Thayer. 1988) ได้เสนอวิธีแก้จุดอ่อนดังกล่าว เพื่อให้เกณฑ์การจับคู่ผู้สอบระหว่างกลุ่มมีความบริสุทธิ์ยิ่งขึ้น โดยใช้วิธีการ 2 ขั้นตอน ดังนี้

ขั้นตอนแรก นำคะแนนรวมของแบบสอบทั้งฉบับเป็นเกณฑ์การจับคู่ผู้สอบระหว่างกลุ่มย่อย 2 กลุ่ม แล้ววิเคราะห์การทำหน้าที่ต่างกันของข้อสอบเป็นรายข้อ เมื่อพบว่าข้อสอบข้อใดทำหน้าที่ต่างกันให้นำคะแนนของข้อสอบข้อนั้นออกจากคะแนนรวมของผู้สอบแต่ละคน

ขั้นตอนที่สอง ใช้คะแนนรวมของแบบสอบที่นำเอาคะแนนข้อสอบที่ทำหน้าที่ต่างกัน ซึ่งตรวจพบในขั้นตอนแรกออกไป เพื่อใช้เป็นเกณฑ์การจับคู่ แล้ววิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ซ้ำอีกครั้งหนึ่ง สำหรับนำไปใช้สรุปผลการตรวจสอบ

เท่าที่ผ่านมา มีผู้นำวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิธี MH มาศึกษา เช่น คลอสเซอร์และคนอื่นๆ (Clauser and others. 1991) ใช้วิธี MH วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบอย่างสม่ำเสมอ พบว่าวิธี MH จะใช้ได้ผลดีในกรณีที่มีข้อสอบมีค่าอำนาจจำแนกสูง แต่จะไม่สามารถวิเคราะห์ข้อสอบที่มีค่าความยากมากได้ ซึ่งสอดคล้องกับข้อค้นพบของเมเซอร์และคนอื่นๆ (Mazor and other. 1991) ซุกวีคส์ และทอลแมน (Sudweeks ; & Tolman. 1990) ที่พบว่าข้อสอบที่ทำหน้าที่ต่างกันมักเป็นข้อสอบที่ยาก บากและเฟอราธา (Baghi ; & Ferrara. 1990) พบว่า เมื่อใช้กลุ่มตัวอย่างขนาด 750 คนขึ้นไป วิธี MH ใช้แทนวิธีทฤษฎีการตอบข้อสอบ 3 พารามิเตอร์ได้ สวามินาธานและโรเจอร์ (Swaminathan ; & Rogers. 1990) ได้ศึกษาเอาข้อมูลจำลองพบว่าวิธี MH วิเคราะห์ได้ดีกว่าการถดถอยแบบโลจิสต์เล็กน้อยโดยตรวจค้นได้ถูกต้องร้อยละ 75 กรณีใช้กลุ่มตัวอย่าง 250 คน และ ตรวจค้นได้ถูกต้องร้อยละ 100 กรณีกลุ่มตัวอย่าง 500 คน กรณีที่การทำหน้าที่ต่างกันของข้อสอบอย่างสม่ำเสมอ และกรณีที่ไม่สม่ำเสมอที่ติดกันปลายข้างใดข้างหนึ่ง แต่วิธี MH มีค่าใช้จ่ายน้อยกว่าวิธีโลจิสต์ประมาณ 3- 4 เท่า แฮมเบิลตันและคนอื่นๆ (Hambleton and others. 1986) พบว่า MH ให้ค่าใกล้เคียงกับทฤษฎีการตอบข้อสอบ ทั้งที่ใช้ค่าความแตกต่างของค่าเฉลี่ยกำลังสองและการตรวจสอบความแตกต่างของพื้นที่รวมได้ไค้ แต่วิธี MH มีค่าใช้จ่ายต่ำกว่าและใช้เวลาน้อยกว่าลินเคอร์ (Linacre. 1988) พบว่าใช้ได้ดีในสถานการณ์จำลองทุกสถานการณ์พอ ๆ กับโปรแกรม PROX ของราสช์ (Rasch) แต่ควรใช้เกณฑ์ในการคำนวณการทำหน้าที่ต่างกันของข้อสอบและความคลาดเคลื่อนหลาย ๆ เกณฑ์ และใช้ตัวประมาณค่าความคลาดเคลื่อนมาตรฐานมากกว่า 1 ตัว เฟอร์ลแมนและคนอื่นๆ (Perlman and others. 1988) พบว่าวิธี MH มีปัญหาด้านความเชื่อมั่น เมื่อจำนวนกลุ่มตัวอย่างน้อยกว่า 660 คน ริสเซน สไตน์เบิร์ก และวายนเนอร์ (Thissen , Steinberg ; & Wainer. 1988) พบว่าวิธี MH ให้ผลการวิเคราะห์คล้ายกับวิธีทฤษฎีการตอบข้อสอบแบบการถดถอยเชิงเส้น (IRT-LR) และอาจใช้วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบก่อนใช้วิธีทฤษฎีการตอบข้อสอบแบบการถดถอยเชิงเส้น (IRT-LR)

5. เอกสารที่เกี่ยวข้องกับการแก้โจทย์ปัญหาคณิตศาสตร์

5.1 ความหมายของโจทย์ปัญหาคณิตศาสตร์

แอนเดอร์สัน และ ฟริงกรี (Andersen ; & Pingry. 1973 : 228) ได้ให้ความหมายของโจทย์ปัญหาคณิตศาสตร์ว่าเป็นสถานการณ์หรือคำถามที่ต้องการหาข้อสรุป หรือเป็นคำตอบซึ่งผู้แก้ปัญหาจะทำได้โดยจะต้องมีกระบวนการที่เหมาะสม ซึ่งใช้ความรู้ ประสบการณ์ การวางแผน และตัดสินใจประกอบกัน

อาดัมส์ (Adams. 1977 : 176) ให้ความหมายว่า โจทย์ปัญหาคณิตศาสตร์ เป็นสถานการณ์ที่เกี่ยวกับปริมาณ และต้องมีการตัดสินใจลงมือกระทำหรือหาคำตอบ โดยปัญหานั้นจะเป็นปัญหาที่ใช้ภาษาเรื่องราวหรือคำพูดก็ได้

มนูญ อรุณไพโรจน์ (2517:17) ให้ความหมายว่า โจทย์ปัญหาคณิตศาสตร์ หมายถึง สภาพปัญหาทางคณิตศาสตร์ ซึ่งประกอบไปด้วยจำนวนและตัวเลข ตลอดจนคำห่อหุ้มที่ก่อให้เกิดปัญหา ซึ่งนักเรียนจะต้องคิดและตัดสินใจว่า จะใช้วิธีการอะไรทางคณิตศาสตร์มาแก้ปัญหานั้น

สวัสดิ จิตต์จนะ (2535:77) ให้ความหมายว่า โจทย์ปัญหาคณิตศาสตร์ เป็นข้อความที่แสดงถึง เงื่อนไข ความสัมพันธ์ของจำนวนที่กำหนดไว้ในประโยค ในลักษณะใดลักษณะหนึ่ง อันจะก่อให้เกิดจำนวนและผลลัพธ์อีกจำนวนหนึ่งที่ต้องการทราบในคำถามของโจทย์

สรุปได้ว่า โจทย์ปัญหาคณิตศาสตร์ หมายถึง สถานการณ์หรือคำถามทางคณิตศาสตร์ที่ต้องการหาข้อสรุป หรือคำตอบ โดยสถานการณ์นั้นจะเป็นปัญหาที่ใช้ภาษาแสดงถึง เงื่อนไข ความสัมพันธ์ของจำนวนที่กำหนดไว้ในแต่ละประโยค ซึ่งนักเรียนจะต้องคิดและตัดสินใจว่า จะใช้วิธีการอะไรทางคณิตศาสตร์มาแก้ปัญหา

5.2 อุปสรรคในการทำโจทย์ปัญหาคณิตศาสตร์

ในการเรียนการสอนคณิตศาสตร์ปัญหาที่ประสบมาก คือ นักเรียนไม่สามารถแก้โจทย์ปัญหาคณิตศาสตร์ได้ เนื่องจากมีอุปสรรคหลายประการ ซึ่ง บรูคเนอร์ และครอส์นิกเกิล (ศรีทอง มีทาทอง. 2534:36; อ้างอิงจาก Brueckner ; & Grossnickle. 1947 : 452-453) ได้กล่าวถึงอุปสรรคในการทำโจทย์ปัญหาไว้ดังนี้

1. นักเรียนไม่สามารถเข้าใจโจทย์ปัญหาทั้งหมด หรือบางส่วน เนื่องจากขาดประสบการณ์ และขาดความคิดรวบยอดในสภาพของโจทย์ปัญหา
2. นักเรียนมีบกพร่องในการอ่านและทำความเข้าใจ เช่น ไม่เข้าใจว่าโจทย์กำหนดอะไรให้ ไม่สามารถจดจำและจัดระบบสิ่งซึ่งเขาได้อ่านมา และไม่สามารถจะอ่านเพื่อหารายละเอียดของเนื้อหา
3. นักเรียนไม่สามารถคิดคำนวณได้ ซึ่งอาจมีสาเหตุมาจากการที่นักเรียนลืมวิธีทำหรือไม่เคยเรียนมาก่อน

4. นักเรียนขาดความเข้าใจในกระบวนการและวิธีการ เป็นผลทำให้นักเรียนหาคำตอบโดยวิธีเดาสุ่ม
 5. นักเรียนขาดความรู้ในเรื่องความสำคัญ กฎเกณฑ์ สูตร เช่น ไม่ทราบว่าหนึ่งหลายมีกี่นิ้ว หรือไม่ทราบกฎการหาเส้นรอบรูปสี่เหลี่ยมผืนผ้า เป็นต้น
 6. นักเรียนขาดความระเบียบเรียบร้อยในการเขียนคำอธิบาย
 7. นักเรียนไม่ทราบความสัมพันธ์เชิงปริมาณวิเคราะห์ ทั้งนี้มีสาเหตุมาจากการเรียนรู้ศัพท์เพียงจำกัด หรือขาดความเข้าใจในหลักเกณฑ์ต่าง ๆ เช่น ความสัมพันธ์ระหว่างราคาขาย ต้นทุน กำไร ขาดทุน เป็นต้น
 8. นักเรียนขาดความสนใจ เนื่องจากขาดความสามารถในการทำโจทย์ปัญหาซึ่งมีความยากหรือโจทย์ปัญหาไม่สนใจและไม่ได้รับประโยชน์อะไรเป็นการตอบสนอง
 9. ระดับสติปัญญาของนักเรียนต่ำเกินไปที่จะเข้าใจถึงความสัมพันธ์ต่าง ๆ ซึ่งปรากฏอยู่ในโจทย์ปัญหา
 10. นักเรียนขาดการฝึกฝนในการทำโจทย์ปัญหา
- อุปสรรคดังกล่าวนี้สอดคล้องกับที่ พันัส หันนาคินทร์ และพิทักษ์ รักษาพลเดช (2512:104) ได้กล่าวไว้เกี่ยวกับเรื่องที่ทำให้นักเรียนทำโจทย์ปัญหาไม่ได้ เนื่องจากสาเหตุต่อไปนี้
1. นักเรียนขาดทักษะในเรื่องการบวก ลบ คูณ หาร อันก่อให้เกิดความติดขัดในการทำโจทย์ต่อ ๆ ไป
 2. นักเรียนขาดความคิดหาเหตุผล มองไม่เห็นความสัมพันธ์ระหว่างข้อมูลต่าง ๆ ที่โจทย์กำหนดไว้ เพื่อที่จะบรรลุสิ่งที่โจทย์ต้องการ
 3. นักเรียนใช้วิธีการผิด ๆ ในการแก้ปัญหา เพราะไม่รู้จักนำทฤษฎีที่ได้เรียนไปแล้วใช้ในการแก้ปัญหา
 4. นักเรียนอ่านโจทย์แล้วไม่เข้าใจ ทั้งนี้อาจเป็นเพราะนักเรียนขาดความสามารถในการอ่านตลอดจนไม่รู้ความหมายของคำ (Vocabulary) ที่ใช้อย่างชัดเจน หรืออาจเป็นเพราะโจทย์นั้นซับซ้อนเกินระดับความเข้าใจของนักเรียนในเช่นนั้น ๆ
 5. นักเรียนขาดความสนใจในการทำโจทย์ปัญหา บทเรียนไม่มีลักษณะที่ช่วยความสนใจของนักเรียน
 6. นักเรียนมีความสะเพร่า ทำให้นำตัวเลขมาใช้อย่างผิด ๆ นักเรียนดีปัญหาโจทย์ผิดตลอดจนการคำนวณผิด
 7. นักเรียนชอบเดา เพราะต้องการให้เสร็จเร็ว ๆ ขาดความตั้งใจที่จะแก้ปัญหา

5.3 ปัจจัยที่ส่งผลต่อความสามารถในการแก้ปัญหา

เมื่อมีกระบวนการและยุทธวิธีการแก้ปัญหาแล้ว ก็ยังไม่ได้ประกันว่าจะสามารถแก้ปัญหาได้ดี เพราะความสามารถในการแก้ปัญหาไม่ใช่ความสามารถที่เป็นไปตามรูปแบบที่กำหนดตายตัวต้องมีการพัฒนา ปรับปรุงแบบการคิดและการแก้ปัญหาไม่ใช่ความสามารถที่เป็นไปตามรูปแบบที่กำหนดตายตัวต้องมีการพัฒนา ปรับปรุงแบบการคิดและการแก้ปัญหาใหม่ ๆ เสมอ ซึ่งต้องอาศัยทักษะและความสามารถหลายๆด้าน อัดัมส์ เอลลิส และบีสัน (Adams, Ellis ; & Beeson. 1977:174-175) ได้กล่าวถึงปัจจัยที่ส่งผลถึงความสามารถในการแก้ปัญหา 3 ด้าน คือ

1. สติปัญญา (Intelligence)

การแก้ปัญหาคำเป็นต้องใช้ในการคิดระดับสูง สติปัญญาจึงเป็นสิ่งสำคัญยิ่งประการหนึ่งในการแก้ปัญหา องค์ประกอบของสติปัญญาที่มีส่วนสัมพันธ์กับความสามารถในการแก้ปัญหา คือ องค์ประกอบทางปริมาณ (Quantitative Factors) ดังนั้น นักเรียนบางคนอาจมีความสามารถในองค์ประกอบทางด้านภาษา (Verbal Factors) แต่อาจด้อยในความสามารถที่ไม่ใช่ภาษาหรือทางด้านปริมาณ

2. การอ่าน (Reading)

การอ่านเป็นพื้นฐานที่จำเป็นสำหรับการแก้ปัญหา เพราะการแก้ปัญหาคำต้องอ่านอย่างรอบคอบ อ่านอย่างวิเคราะห์ อันจะนำไปสู่การตัดสินใจว่า ควรจะทำอะไร และอย่างไร มีนักเรียนจำนวนมากที่มีความสามารถในการอ่านแต่ไม่สามารถแก้ปัญหาได้

3. ทักษะพื้นฐาน (Basic Skills)

หลังจากวิเคราะห์สถานการณ์ปัญหาและตัดสินใจว่าทำอะไรแล้ว ก็ยังเหลือขั้นตอนการได้มาซึ่งคำตอบที่ถูกต้องเหมาะสม นั่นคือ นักเรียนจะต้องรู้การดำเนินต่าง ๆ ที่จำเป็น ซึ่งก็คือ ทักษะพื้นฐานนั่นเอง

6.งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ

งานวิจัยที่ศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบนั้น แบ่งเป็นงานวิจัยในประเทศและงานวิจัยต่างประเทศ ดังนี้

งานวิจัยในประเทศ

จากการศึกษารายงานวิจัยภายในประเทศที่ผ่านมา มีผู้ทำวิจัยเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบไว้หลายท่าน โดยศึกษาในรูปการเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันด้วยวิธีต่าง ๆ กับกลุ่มย่อยที่มีลักษณะแตกต่างกันในด้านภาษา เพศ และสภาพภูมิศาสตร์ รวมทั้งการ

เปรียบเทียบคุณภาพของแบบทดสอบหลังคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออกแล้ว นอกจากนี้ยังมี การพัฒนาเกณฑ์ที่ใช้ในการตัดสินข้อสอบที่ทำหน้าที่ต่างกันทางเพศ ซึ่งได้เสนอ ดังต่อไปนี้

ซัชชัย เผ่าพงศ์ (2527) ได้ทำการทำหน้าที่ต่างกันของข้อสอบมาตรฐานวัดความถนัดทางการ เรียนด้านคณิตศาสตร์และภาษาในระดับมัธยมศึกษาตอนต้นระหว่างกลุ่มผู้สอบเพศชายและเพศหญิง โดยการเปรียบเทียบโค้งลักษณะข้อสอบด้วยวิธีแบบจำลองโลจิสติกแบบ 3 พารามิเตอร์ กลุ่มตัวอย่าง ที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2524 เพศชายและหญิง ซึ่งประกอบด้วยทุก ภูมิภาคทั่วประเทศ 5 ภาค คือ ภาคเหนือ ภาคกลาง ภาคใต้ ภาคตะวันออก และภาค ตะวันออกเฉียงเหนือ กลุ่มซึ่งทำการวิเคราะห์ด้วยแบบทดสอบวัดความถนัดทางการเรียนด้าน คณิตศาสตร์ เป็นนักเรียนชาย 1,610 คน และนักเรียนหญิง 1,337 คน ส่วนอีกกลุ่มหนึ่งเป็นนักเรียน ชาย 1,316 คน และนักเรียนหญิง 985 คน ทำการวิเคราะห์ด้วยแบบทดสอบวัดความถนัดทางการเรียน ด้านคณิตศาสตร์ จำนวน 30 ข้อ มีข้อสอบที่ทำหน้าที่ต่างกันต่อกลุ่มใดกลุ่มหนึ่งโดยเฉพาะ 9 ข้อ คือ การทำหน้าที่ต่างกันต่อกลุ่มนักเรียนชาย 7 คน และการทำหน้าที่ต่างกันต่อกลุ่มนักเรียนหญิง 2 ข้อ ข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบต่อกลุ่มนักเรียนชายและหญิงในระดับปานกลางขึ้นไปมี จำนวน 5 ข้อ ซึ่งวัดในเรื่องร้อยละ การหาปริมาตรและการหาความยาวของเส้นในด้านของรูป สามเหลี่ยมจำนวนเรื่องละ 1 ข้อ และเรื่องเกี่ยวกับโจทย์ปัญหา จำนวน 2 ข้อ จากข้อสอบจำนวน 5 ข้อ ดังกล่าวนี้นี้เป็นข้อสอบที่มีการทำหน้าที่ต่างกันต่อกลุ่มนักเรียนหญิงในช่วงความสามารถแรกและการ ทำหน้าที่ต่างกันต่อกลุ่มนักเรียนชายในช่วงความสามารถต่อมา จำนวน 1 ข้อ ส่วนแบบทดสอบวัด ความถนัดทางการเรียนด้านภาษาจำนวน 40 ข้อ มีข้อสอบที่ทำหน้าที่ต่างกันเฉพาะกลุ่มนักเรียนชาย 3 ข้อ และเฉพาะกลุ่มนักเรียนหญิงโดยเฉพาะมี 8 ข้อ ข้อสอบที่มีการทำหน้าที่ต่างกันในระดับปาน กลางขึ้นไปมีจำนวน 9 ข้อ ซึ่งวัดความเข้าใจเกี่ยวกับการอ่านข้อความจำนวน 7 ข้อ จากแบบทดสอบ จำนวน 9 ข้อนี้เป็นข้อที่มีการทำหน้าที่ต่างกันต่อกลุ่มนักเรียนชายโดยเฉพาะจำนวน 1 ข้อ และเป็น ข้อสอบที่มีการทำหน้าที่ต่างกันต่อกลุ่มนักเรียนหญิงโดยเฉพาะจำนวน 6 ข้อ และมีข้อสอบจำนวน 2 ข้อ มีการทำหน้าที่ต่างกันทางการทดสอบมากกว่า 1 ช่วงความสามารถ

ทัศนีย์ พิรมนตรี (2530) ได้ทำการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบวิชา คณิตศาสตร์ โครงการตรวจสอบคุณภาพการศึกษาของสำนักทดสอบทางการศึกษา กรมวิชาการ กระทรวงศึกษาธิการ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2536 จำนวน 7,036 คน โดยเปรียบเทียบจำนวนข้อสอบที่มีการทำหน้าที่ต่างกันระหว่างนักเรียนในกรุงเทพมหานคร กับนักเรียนในภาคภูมิภาคทั้ง 5 ภาค คือ ภาคเหนือ ภาคกลาง ภาคตะวันออก ภาคใต้ และภาค ตะวันออกเฉียงเหนือ ด้วยวิธีตรวจสอบความลำเอียง 3 วิธี คือ วิธีกำหนดจุดเดลต้า วิธีทดสอบความ แตกต่างระหว่างกลุ่มด้วยสถิติไคสแควร์ในโมเดลลอกลิเนียร์ 2 โมเดล คือ โมเดลที่ไม่มีค่าพารามิเตอร์

ผลร่วมระหว่างระดับคะแนนกลุ่ม และโมเดลที่ไม่มีพารามิเตอร์ของผลหลักที่เกิดจากกลุ่ม และวิธีโค้ง ลักษณะข้อสอบที่ใช้พารามิเตอร์ 3 ตัว พบข้อสอบที่มีการทำหน้าที่ต่างกันจากวิธีโค้งลักษณะข้อสอบที่ใช้พารามิเตอร์ 3 ตัวมากที่สุด และในแต่ละวิธีตรวจสอบการทำงานที่ต่างกัน พบว่า มีข้อสอบที่การทำงานที่ต่างกันซ้ำกันระหว่างกลุ่มนักเรียนกรุงเทพกับนักเรียนในภาคภูมิศาสตร์ทุกภาค

สุรศักดิ์ อมรรัตนศักดิ์ (2531) ได้ทำการศึกษาเปรียบเทียบผลของวิธีวิเคราะห์การทำงานที่ต่างกันของข้อสอบที่แตกต่างกัน 4 วิธี คือ วิธีวิเคราะห์ความแปรปรวน วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 1 ตัว และวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว โดยหาสัมประสิทธิ์สหสัมพันธ์ของการทำหน้าที่ต่างกันระหว่างวิธีวิเคราะห์การทำงานที่ต่างกัน 4 วิธี และเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อนและหลังการศึกษาการทำงานที่ต่างกันของข้อสอบ ตามวิธีการคิดคะแนนรวมที่แตกต่างกัน 6 วิธี ในด้านจำนวนผู้ได้รับการคัดเลือก สัดส่วนชาย : หญิง ที่ได้รับการคัดเลือกและความเชื่อมั่นของแบบทดสอบโดยกลุ่มตัวอย่าง คือ ผู้เข้าสอบแข่งขันเพื่อบรรจุเข้ารับราชการเป็นราชการครู ในปี พ.ศ. 2529 จำนวน 2,340 คน เป็นชาย 1,170 คน หญิง 1,170 คน เครื่องมือที่ใช้ในการรวบรวมข้อมูล ได้แก่ แบบทดสอบที่ใช้ในการสอบแข่งขันเพื่อบรรจุเข้ารับราชการครูในปี พ.ศ. 2529 ซึ่งมี 4 ฉบับ คือ ฉบับที่ 1 แบบทดสอบวิชาเอก (ความรู้เกี่ยวกับการประถมศึกษา) ฉบับที่ 2 แบบทดสอบวิชาการศึกษาและกฎหมาย ฉบับที่ 3 แบบทดสอบภาษาไทย ฉบับที่ 4 แบบทดสอบความสามารถทั่วไป ซึ่งผลการศึกษา วิธีวิเคราะห์การทำงานที่ต่างกันแต่ละวิธีพบจำนวนข้อที่การทำงานที่ต่างกันแตกต่างกัน โดยวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว พบจำนวนข้อสอบที่การทำงานที่ต่างกันมากที่สุด รองลงมาได้แก่ วิธีวิเคราะห์ความแปรปรวนและวิธีแปลงค่าความยากของข้อสอบ พบจำนวนข้อสอบที่มีการทำหน้าที่ต่างกันน้อยที่สุด ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ของการทำหน้าที่ต่างกันระหว่างวิธีวิเคราะห์การทำงานที่ต่างกัน 4 วิธี พบว่า มีค่าอยู่ระหว่าง .7535 ถึง .9921 ในด้านจำนวนผู้ได้รับการคัดเลือกเมื่อเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อนและหลังการศึกษาการทำงานที่ต่างกันของข้อสอบตามวิธีคิดคะแนนดิบกับวิธีรวมคะแนนแบบอื่น ๆ อีก 5 วิธี มีจำนวนผู้ได้รับการคัดเลือกแตกต่างกันประมาณ 4 % ถึง 24 % ส่วนวิธีแปลงคะแนนดิบเป็นคะแนนมาตรฐานที่ปกติกับวิธีรวมคะแนนแบบอื่น ๆ อีก 4 วิธี มีจำนวนผู้ได้รับการคัดเลือกแตกต่างกันประมาณ 6 % ถึง 23 % ส่วนในเรื่องค่าความเชื่อมั่นพบว่า ลดลงจากเดิมเล็กน้อย

พัชรี ปิยภัณฑ (2531) ได้ทำการวิเคราะห์การทำงานที่ต่างกันของข้อสอบคณิตศาสตร์เรื่องสมการ ชั้นประถมศึกษาปีที่ 6 จำนวน 45 ข้อ ที่ผู้วิจัยสร้างขึ้นเอง ระหว่างกลุ่มผู้สอบที่มีเพศ และสภาพภูมิศาสตร์ต่างกัน ด้วยวิธีวิเคราะห์ 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีไคสแควร์ และวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ โดยทำการเปรียบเทียบจำนวนข้อสอบที่การทำงานที่ต่างกัน

การวิเคราะห์หาสัมประสิทธิ์สหสัมพันธ์ของวิธีวิเคราะห์ทั้ง 3 วิธี และศึกษาเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบก่อนและหลังคัดข้อสอบที่การทำหน้าที่ต่างกันออกแล้วกลุ่มตัวอย่างที่ใช้เป็นนักเรียนชายหญิง ชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2530 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาแห่งชาติ ในกรุงเทพมหานคร และจังหวัดสมุทรสาคร จำนวน 1,422 คน ซึ่งเลือกโดยวิธีการสุ่มแบบแบ่งชั้น ผลการวิเคราะห์พบว่า วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ พบจำนวนข้อสอบที่การทำหน้าที่ต่างกันต่อเพศมากที่สุด จำนวน 7 ข้อ รองลงมาคือวิธีโคสแควร์ และวิธีแปลงค่าความยากของข้อสอบ โดยค้นพบได้จำนวน 3 ข้อ และ 1 ข้อ ตามลำดับ สำหรับความแตกต่างในด้านสภาพภูมิศาสตร์ ได้ผลการค้นพบเช่นเดียวกัน คือ วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ พบจำนวนข้อสอบที่การทำหน้าที่ต่างกันมากที่สุดจำนวน 18 ข้อ รองลงมาคือ วิธีโคสแควร์และวิธีแปลงค่าความยากของข้อสอบ โดยค้นพบจำนวน 3 ข้อ และ 1 ข้อตามลำดับ ค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความการทำหน้าที่ต่างกันต่อเพศ ระหว่างวิธีวิเคราะห์ทั้ง 3 วิธี มีค่าระหว่าง .1713 ถึง .5618 และค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีการทำหน้าที่ต่างกันต่อกลุ่มนักเรียนกรุงเทพมหานครกับกลุ่มนักเรียนสมุทรสาคร มีค่าระหว่าง .1868 ถึง .6009 สำหรับค่าความเชื่อมั่นของแบบทดสอบก่อนและหลังการคัดข้อสอบที่มีการทำหน้าที่ต่างกันออกแล้ว พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

กาญจนา วัฒนสุทร (2538) ได้ศึกษาการพัฒนาเกณฑ์ตัดสินข้อสอบการทำหน้าที่ต่างกันทางเพศด้วยข้อมูลเชิงประจักษ์โดยใช้ดัชนีการทำหน้าที่ต่างกัน 4 ตัว คือ พื้นที่ระหว่างโค้งการตอบข้อสอบชนิดคิดเครื่องหมาย (SA) และไม่คิดเครื่องหมาย (UA) จากทฤษฎีการตอบข้อคำถามโมเดล 2 พารามิเตอร์ วิธีแมนเทิล – แฮนส์เซล (MH) และวิธี SIBTEST

ข้อมูลที่ใช้เป็นข้อมูลการตอบข้อสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของทบวงมหาวิทยาลัย ปีการศึกษา 2535 ความยาวของแบบทดสอบเป็น 20 , 30 และ 40 ข้อ สำหรับวิชาคณิตศาสตร์ และ 50 , 60 , 70 และ 80 ข้อ สำหรับวิชาภาษาอังกฤษ ขนาดกลุ่มตัวอย่างมี 6 ขนาด คือ 100 , 200 , 400 , 600 , 800 และ 1,000 คน การพัฒนาเกณฑ์กระทำโดยคำนวณค่าดัชนีทั้ง 4 ตัว จากข้อมูลการตอบข้อสอบเพศเดียวกันเพศละ 50 ค่าสำหรับแต่ละความยาวแบบทดสอบและขนาดกลุ่มผู้สอบ จากนั้นนำค่าดัชนีที่ได้ทั้งหมดมาวิเคราะห์หาค่าเฉลี่ยและกำหนดเกณฑ์จากค่าเฉลี่ย 2 ลักษณะ คือ เกณฑ์ที่กำหนดจากค่าเฉลี่ย ซึ่งรวมค่าดัชนีทุกข้อโดยไม่พิจารณาความแตกต่างในด้านความยาวของแบบทดสอบและขนาดกลุ่มผู้สอบด้วย จากนั้นนำเกณฑ์จากค่าเฉลี่ยที่กำหนดไว้ไปตัดสินค่าดัชนีที่ได้จากการวิเคราะห์ระหว่างกลุ่มผู้สอบเพศหญิงและชาย พบว่า ความสอดคล้องของการตัดสินภายในดัชนีเดียวกัน มีความไม่คงที่ข้ามขนาดผู้สอบ แต่ความสอดคล้องนี้จะมีแนวโน้มสูงขึ้นที่ขนาดผู้สอบตั้งแต่ขนาด 600 คน ขึ้นไป

นอกจากนี้ยังค้นพบว่าเกณฑ์ที่ใช้ในการตัดสินใจการทำหน้าที่ต่างกันของข้อสอบระหว่างกลุ่มผู้สอบเพศหญิงและชาย คือ $ISAI > .08$ และ $UA > .50$ สำหรับแบบทดสอบที่สั้นกว่า 50 ข้อและ $ISAI > .08$ และ $UA > 1.20$ สำหรับแบบทดสอบที่มีความยาวตั้งแต่ 50 ข้อขึ้นไป และ $.60 > MH > 1.40$ สำหรับทุกความยาวแบบทดสอบและทุกขนาดกลุ่มผู้สอบ โดย ผู้วิจัยสรุปว่า ควรใช้ขนาดกลุ่มผู้สอบอย่างน้อย 800 คน สำหรับดัชนี SA และ UA และ 600 คน สำหรับวิธีแมนเทิล – แฮนส์เชล และวิธี SIBTEST การตรวจค้นการทำหน้าที่ต่างกันของข้อสอบทางเพศมีความไม่คงที่ข้ามขนาดกลุ่มผู้สอบและความยาวของแบบทดสอบ ความสอดคล้องในการตรวจค้นข้อสอบที่มีการทำหน้าที่ต่างกันภายในวิธีเดียวกัน ข้ามขนาดกลุ่มผู้สอบค่อนข้างต่ำแต่จะสูงที่ขนาดผู้สอบตั้งแต่ 600 คนขึ้นไป และ ข้อสอบที่ทำหน้าที่ต่างกันในวิชาคณิตศาสตร์ส่วนใหญ่จะการทำหน้าที่ต่างกันเข้าสู่ผู้สอบเพศชาย แต่วิชาภาษาอังกฤษจะการทำหน้าที่ต่างกันเข้าสู่เพศหญิง

เจน จันท์แสง (2537) ได้ศึกษาดัชนีการทำหน้าที่ต่างกันของข้อสอบวัดผลสัมฤทธิ์ทางการเรียนในวิชาคณิตศาสตร์ เรื่อง จำนวนจริง ที่เกิดจากตัวแปรเพศ และที่ตั้งของโรงเรียน ด้วยวิธีวิเคราะห์ข้อสอบที่ทำหน้าที่ต่างกัน 4 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีวิเคราะห์ความแปรปรวน วิธีไคสแควร์ และวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 1 ตัว ศึกษาสัมประสิทธิ์สหสัมพันธ์ระหว่างดัชนีการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 วิธี และเปรียบเทียบความเชื่อมั่นของแบบทดสอบก่อนและหลังการคัดเลือกข้อสอบที่ทำหน้าที่ต่างกัน เครื่องมือที่ใช้เป็นแบบทดสอบชนิดเลือกตอบ จำนวน 40 ข้อ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2536 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดนครสวรรค์ จำนวน 1,000 คน พบว่า

1. วิธีวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแต่ละวิธีตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกันแตกต่างกัน โดยวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 1 ตัว ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกันได้มากที่สุด โดยพบข้อสอบที่ทำหน้าที่ต่างกันตามตัวแปรเพศ 10 ข้อ และพบข้อสอบที่ทำหน้าที่ต่างกันตามตัวแปรที่ตั้งของโรงเรียน 14 ข้อ ส่วนวิธีแปลงค่าความยากของข้อสอบเป็นวิธีที่พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด โดยพบข้อสอบที่ทำหน้าที่ต่างกันตามตัวแปรเพศ 1 ข้อ และพบข้อสอบที่ทำหน้าที่ต่างกันตามตัวแปรที่ตั้งของโรงเรียน 5 ข้อ ตามลำดับ ซึ่งแสดงว่าเพศ และที่ตั้งของโรงเรียนเป็นตัวแปรที่ทำให้ข้อสอบทำหน้าที่ต่างกัน

2. ความสัมพันธ์ระหว่างดัชนีการทำหน้าที่ต่างกันของข้อสอบของวิธีวิเคราะห์แต่ละวิธี มีค่าเป็นบวก และมีนัยสำคัญทางสถิติที่ระดับ .001 เมื่อวิเคราะห์ตามตัวแปรเพศ พบว่า สัมประสิทธิ์สหสัมพันธ์ของดัชนีแต่ละวิธีมีค่าระหว่าง .5284 ถึง .9970 เมื่อวิเคราะห์ตามตัวแปรที่ตั้งของโรงเรียน พบว่า สัมประสิทธิ์สหสัมพันธ์ของดัชนีแต่ละวิธีมีค่าระหว่าง .6008 ถึง .9475

3.แบบทดสอบหลังจากคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออกไปแล้ว มีค่าความเชื่อมั่นลดลงจากเดิม อย่างไม่มีนัยสำคัญทางสถิติ และเมื่อปรับค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออกไปแล้วให้มีจำนวนเท่าเดิมโดยใช้สูตรของ สเปียร์แมน บราวน์ พบว่า ค่าความเชื่อมั่นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

อังคณา สายยศ (2540) ได้ศึกษาผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบทดสอบความเข้าใจในการอ่านตามลักษณะการใช้ภาษาพูดในครอบครัวเป็นภาษาลาว ภาษาเขมร และภาษาส่วย โดยศึกษาเปรียบเทียบจำนวนข้อที่ทำหน้าที่ต่างกัน ความเชื่อมั่นของแบบทดสอบหลังคัดข้อที่มีการทำหน้าที่ต่างกันออกแล้ว และคะแนนที่ได้จากการสอบหลังคัดข้อที่มีการทำหน้าที่ต่างกันออกแล้ว กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ในจังหวัดสุรินทร์ ได้มาโดยการสุ่มแบบแบ่งชั้น อำเภอเป็นชั้นและโรงเรียนเป็นหน่วยการสุ่มได้จำนวนโรงเรียนทั้งสิ้น 29 โรงเรียน นักเรียน 860 คน ทำการวิเคราะห์การทำหน้าที่ต่างกันด้วยวิธีแมนเทิล - แฮนส์เชล ผลการวิจัยพบว่า จำนวนข้อสอบที่การทำหน้าที่ต่างกันของแบบทดสอบความเข้าใจในการอ่านระหว่างกลุ่มผู้สอบที่ใช้ภาษาพูดเป็นภาษาลาว ภาษาเขมร และภาษาส่วยแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ สำหรับค่าความเชื่อมั่นของแบบทดสอบหลังคัดข้อทำหน้าที่ต่างกันออกแล้วพบว่า ค่าความเชื่อมั่นระหว่างกลุ่มผู้สอบที่ใช้ภาษาพูดเป็นภาษาลาว ภาษาเขมร และภาษาส่วยแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ส่วนคะแนนที่ได้จากการสอบด้วยแบบทดสอบที่คัดข้อการทำหน้าที่ต่างกันออกแล้วมีคะแนนระหว่างกลุ่มผู้สอบที่พูดภาษาเขมรมีคะแนนแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

เรวดี อินทสระระ (2539) ได้ศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่วิเคราะห์การทำหน้าที่ต่างกันต่อเพศ ด้วยวิธีทฤษฎีการตอบข้อคำถาม (IRT) วิธีแมนเทิล - แฮนส์เชล (MH) และวิธี SIBTEST การตัดสินผลการสอบที่คิดคะแนนมาตรฐานที่ - ปกติ และ คะแนนน้ำหนักความสามารถและสาเหตุการทำหน้าที่ต่างกันของข้อสอบ โดยศึกษาการทำหน้าที่ต่างกันของข้อสอบคัดเลือกเข้าศึกษาในชั้นปีที่ 1 ประเภทรับตรง ปีการศึกษา 2538 ของมหาวิทยาลัยสงขลานครินทร์ ในวิชาภาษาไทย ก วิชาสังคมศึกษา ก วิชาภาษาอังกฤษ กข วิชาละ 8,127 คน (ชาย 2,722 คน หญิง 3,961 คน) ความเที่ยงตรงเชิงพยากรณ์ศึกษาจากคะแนนสอบคัดเลือกกับระดับผลการเรียนภาคเรียนที่ 1 ปีการศึกษา 2538 ของนักศึกษาที่ได้รับการคัดเลือกประเภทรับตรง สายวิทยาศาสตร์ 763 คน และสายศิลปศาสตร์ 281 คน และสาเหตุการทำหน้าที่ต่างกันของข้อสอบจากระบุสาเหตุของนักวัดผลการศึกษาหรืออาจารย์ผู้สอนจำนวน 50 คน และนิสิตนักศึกษาที่เรียนในสาขานั้น ๆ วิชาละ 50 คน ผลการศึกษาพบว่า วิธีตรวจสอบการทำหน้าที่ต่างกันทั้ง 3 วิธี ตัดสินจำนวนข้อสอบที่การทำหน้าที่ต่างกันแตกต่างกันวิชาภาษาไทย ก ฉบับที่ 2 และวิชาสังคมศึกษา ก ฉบับที่ 1 ที่ระดับนัยสำคัญทางสถิติที่ .05 นอกนั้นต่างกันที่ระดับนัยสำคัญทางสถิติที่ .01 โดยใช้ทฤษฎีการตอบข้อคำถามตัดสินจำนวน

ข้อสอบที่การทำหน้าที่ต่างกันได้มากที่สุด ความสัมพันธ์ของลำดับที่ของการสอบไม่ว่าจะคิดคะแนนมาตรฐานที่ - ปกติ หรือ คิดคะแนนน้ำหนักความสามารถและใช้ข้อสอบจำนวนทั้งหมด หรือใช้เฉพาะข้อสอบที่ปราศจากการทำหน้าที่ต่างกันต่างมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการเรียนในสายวิทยาศาสตร์ การคิดคะแนนน้ำหนักความสามารถนั้นใช้ข้อสอบจำนวนทั้งหมดและใช้เฉพาะข้อสอบที่ปราศจากการทำหน้าที่ต่างกันมีประสิทธิภาพในการทำนายถูกสูงกว่การคิดคะแนนมาตรฐานที่ - ปกติ ส่วนสายศิลปศาสตร์ การคิดคะแนนมาตรฐานที่ - ปกติ ใช้ข้อสอบทั้งหมด และใช้เฉพาะข้อสอบที่ปราศจากการทำหน้าที่ต่างกัน มีประสิทธิภาพในการทำนายสูงกว่การคิดคะแนนน้ำหนักความสามารถ และสาเหตุของการทำหน้าที่ต่างกันของข้อสอบต่อเพศทั้งเพศชายและหญิง เกิดจากข้อสอบเป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะจะมีโอกาสตอบถูกมากกว่า เป็นเรื่องราวที่กลุ่มนั้น ๆ สนใจ และเป็นข้อสอบที่ถามความจำ

ญาณภัทร สีหะมงคล (2540) ได้ศึกษาเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord's χ^2 วิธี Raju's Area Measures และวิธี Closed Interval Area เมื่อขนาดกลุ่มตัวอย่าง ความยาวแบบทดสอบ และสัดส่วนจำนวนข้อสอบที่ทำหน้าที่ต่างกันแบบทดสอบแตกต่างกัน ข้อมูลที่ใช้ในการศึกษาเป็นผลการสอบประเมินคุณภาพและความก้าวหน้าทางการศึกษาวิชาคณิตศาสตร์ ของนักเรียนชั้นประถมศึกษาปีที่ 4 ปีการศึกษา 2536 สังกัดสำนักงานการประถมศึกษาแห่งชาติ จำนวน 11,404 คน เครื่องมือที่ใช้ในการศึกษาเป็นแบบทดสอบแบบเลือกตอบ จำนวน 80 ข้อ แบ่งเป็น 3 ตอน คือ การคิดคำนวณ ความรู้ความเข้าใจ และโจทย์ปัญหา การศึกษาแบ่งออกเป็น 2 ส่วน ได้แก่ ส่วนที่ 1 เป็นการตรวจสอบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ความสอดคล้องในการตรวจสอบ ความสัมพันธ์ระหว่างวิธีตรวจสอบ และส่วนที่ 2 เป็นการตรวจสอบอัตราความไม่สอดคล้องในการตรวจสอบ และอัตราความสอดคล้องในการตรวจสอบ ผลการศึกษาพบว่า

1. จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ได้จากการตรวจสอบ ระหว่าง Lord's χ^2 วิธี Raju's Area Measures และวิธี Closed Interval Area แตกต่างกันเมื่อขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบต่างกัน ยกเว้นกรณีกลุ่มตัวอย่าง 501 คน แบบทดสอบความยาว 20 , 30 และ 40 ข้อ และกรณีกลุ่มตัวอย่าง 700 คน แบบทดสอบยาว 40 ข้อ

2. ความสอดคล้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ส่วนมากมีค่าปานกลางถึงต่ำเกือบทุกเงื่อนไขของการศึกษา ยกเว้นกรณีกลุ่มตัวอย่าง 500 คน แบบทดสอบความยาว 30 , 40 และ 60 ข้อ และกลุ่มตัวอย่าง 700 คน แบบทดสอบยาว 40 ข้อ ที่มีความสอดคล้องกันในระดับปานกลางถึงสูงมาก

3. ความสัมพันธ์ระหว่างวิธีการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันทั้ง 3 วิธี โดยภาพรวมมีสัมประสิทธิ์สหสัมพันธ์ค่อนข้างสูง และมีนัยสำคัญทางสถิติเกือบทุกเงื่อนไขของการศึกษา

4. อัตราความไม่สอดคล้องในการตรวจสอบ เมื่อแบบทดสอบยาวมากขึ้นและมีสัดส่วนจำนวนข้อที่ต่างกันมากขึ้น อัตราความไม่สอดคล้องชนิดที่ 1 ส่วนใหญ่เพิ่มขึ้น แต่อัตราความไม่สอดคล้องชนิดที่ 2 ส่วนใหญ่ลดลง โดยที่อัตราความไม่สอดคล้องชนิดที่ 1 และอัตราความไม่สอดคล้องชนิดที่ 2 ส่วนมากมีค่าใกล้เคียงกัน

เมื่อกลุ่มตัวอย่างขนาดใหญ่ Lord's χ^2 วิธี Raju's Area Measures มีอัตราความไม่สอดคล้องสูงกว่าวิธี Closed Interval Area และเมื่อกลุ่มตัวอย่างขนาดเล็ก วิธี Lord's χ^2 วิธี Raju's Area Measures มีอัตราความไม่สอดคล้องต่ำกว่าวิธี Closed Interval Area

5. อัตราความสอดคล้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อแบบทดสอบยาวมากขึ้น และสัดส่วนจำนวนข้อสอบที่ทำหน้าที่ต่างกันมากขึ้น อัตราความสอดคล้องชนิดที่ 1 ส่วนใหญ่ลดลงแต่อัตราความสอดคล้องชนิดที่ 2 ส่วนใหญ่เพิ่มขึ้น

พรรณี จินตมาศ (2540) ได้ศึกษาเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันต่อเพศของแบบทดสอบวิชาคณิตศาสตร์ โดยวิธีวิเคราะห์ต่างกัน 3 วิธี คือ วิธีแมนเทล-แฮนส์เซล, วิธี SIBTEST และวิธีแปลงค่าความยากของข้อสอบ โดยใช้ขนาดกลุ่มตัวอย่างต่าง ๆ กัน ผลการศึกษาพบว่าขนาดกลุ่มตัวอย่าง 500 คน วิธีแปลงค่าความยากของข้อสอบ พบจำนวนข้อสอบที่การทำหน้าที่ต่างกันน้อยที่สุด วิธีแปลงค่าความยากของข้อสอบจะตรวจพบจำนวนข้อสอบที่มีความการทำหน้าที่ต่างกันซ้ำกับวิธีของแมนเทล - แฮนส์เซล และวิธี SIBTEST ทุกข้อ วิธี SIBTEST พบจำนวนข้อสอบที่มีการทำหน้าที่ต่างกันมากที่สุดและเมื่อวิเคราะห์จากขนาดกลุ่มตัวอย่าง 1,000 คน วิธีแมนเทล - แฮนส์เซล พบจำนวนข้อสอบที่มีการทำหน้าที่ต่างกันมากที่สุด วิธีแปลงค่าความยากของข้อสอบไม่พบข้อสอบที่การทำหน้าที่ต่างกัน โดยจำนวนข้อสอบที่วิเคราะห์ด้วยวิธีแปลงค่าความยากกับวิธีแมนเทล - แฮนส์เซล และวิธีแปลงค่าความยากกับวิธี SIBTEST แตกต่างกันอย่างมีนัยสำคัญ ที่ระดับ .05 นอกนั้น มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ สำหรับจำนวนข้อสอบที่วิเคราะห์ด้วยวิธีแปลงค่าความยากระหว่างกลุ่มผู้สอบขนาด 500 คน และขนาด 1,000 คน จะมีจำนวนข้อสอบที่มีการทำหน้าที่ต่างกันแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้น มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ค่าความเชื่อมั่นของแบบทดสอบหลังการคัดเลือกข้อสอบที่มีการทำหน้าที่ต่างกันออกแล้ว เมื่อวิเคราะห์จากกลุ่มตัวอย่างขนาด 500 คน จากการวิเคราะห์การทำหน้าที่ต่างกันด้วยวิธีแปลงค่าความยากและวิธี SIBTEST แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้น มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และเมื่อวิเคราะห์จากกลุ่มตัวอย่างขนาด 1,000 คน ค่าความเชื่อมั่นของแบบทดสอบหลังการคัดเลือกข้อสอบที่มีการทำหน้าที่ต่างกันออกแล้วจากการตรวจสอบการทำหน้าที่

ต่างกัน 3 วิธีนั้น มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ สำหรับค่าความเชื่อมั่นของแบบทดสอบ หลังการคัดเลือกข้อสอบที่มีการทำหน้าที่ต่างกันออกแล้วจากการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธี SIBTEST ระหว่างกลุ่มผู้สอบขนาด 500 คน และ 1,000 คน มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับ .05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

ทองอยู่ สาระ (2543) ได้ศึกษาเปรียบเทียบอำนาจการตรวจสอบและจำแนกผิดพลาดในการ ตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบสม่ำเสมอและแบบไม่สม่ำเสมอ ระหว่าง วิธีแมนเทล – แชนส์ เซล และวิธีถดถอยโลจิสติก โดยใช้ความยาวแบบทดสอบและขนาดกลุ่มตัวอย่างต่างกัน พบว่า

1. อำนาจการตรวจสอบและการจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันทั้ง แบบสม่ำเสมอและแบบไม่สม่ำเสมอ ระหว่างวิธีแมนเทล – แชนส์เซล และวิธีถดถอยโลจิสติก ภายใต ความยาวแบบทดสอบและขนาดกลุ่มตัวอย่างที่ศึกษา เกือบทุกเงื่อนไขไม่แตกต่างกัน

2. ความยาวของแบบทดสอบไม่มีผลต่ออำนาจการตรวจสอบและการจำแนกผิดพลาดในการ ตรวจสอบด้วยวิธีแมนเทล – แชนส์เซล และวิธีถดถอยโลจิสติก

3. ขนาดกลุ่มตัวอย่างมีผลต่ออำนาจการตรวจสอบด้วย วิธีแมนเทล – แชนส์เซล และวิธี ถดถอยโลจิสติก เกือบทุกเงื่อนไขของการศึกษา กล่าวคือ เมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้น อำนาจการ ตรวจสอบจะมีค่าเพิ่มขึ้น แต่พบว่าขนาดกลุ่มตัวอย่างจะไม่มีผลต่อการจำแนกผิดพลาดในเกือบทุก เงื่อนไขของการศึกษา

นวลอนงค์ นันทวงศ์ (2543) ได้ศึกษาเปรียบเทียบการทำหน้าที่ต่างกันของข้อสอบของ แบบทดสอบความเข้าใจในการอ่าน ตามลักษณะการใช้ภาษาพูดในครอบครัว โดยกลุ่มอ้างอิงเป็น กลุ่มใช้ภาษาพูดเป็นภาษาไทยและกลุ่มเปรียบเทียบเป็นกลุ่มที่ใช้ภาษาพูดเป็นภาษาลาวพวนกับ ภาษาลาว ซึ่งทำการวิเคราะห์การทำหน้าที่ต่างกันด้วยวิธี 3 วิธี คือ วิธีไคสแควร์ วิธีแมนเทล – แชนส์ เซล และวิธีวิเคราะห์ความแปรปรวน โดยการเปรียบเทียบจำนวนข้อสอบที่มีการทำหน้าที่ต่างกันและ เปรียบเทียบค่าความเชื่อมั่นที่คำนวณด้วยสูตรแบบแบ่งครึ่งฉบับของแบบทดสอบหลังการคัดเลือก ข้อสอบที่มีการทำหน้าที่ต่างกันออกแล้ว กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักเรียนชายและนักเรียน หญิง ในระดับชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 2 ปีการศึกษา 2542 ของโรงเรียนสังกัดกรมสามัญ ศึกษา จังหวัดลพบุรีจำนวน 900 คน ซึ่งได้มาโดยการสุ่มแบบสองขั้นตอน ผลการวิจัยพบว่า ในกลุ่ม ผู้สอบที่ใช้ภาษาพูดเป็นภาษาลาวพวน จำนวนข้อสอบที่มีการทำหน้าที่ต่างกันจากการตรวจสอบการ ทำหน้าที่ต่างกันระหว่างวิธีไคสแควร์กับวิธีแมนเทล – แชนส์เซล และวิธีไคสแควร์กับวิธีวิเคราะห์ความ แปร ปรวน แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีแมนเทล – แชนส์เซล กับวิธี วิเคราะห์ความแปรปรวนรวมแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ส่วนในกลุ่มผู้สอบที่ใช้ภาษาพูด เป็นภาษาลาว พบว่า จำนวนข้อสอบที่มีการทำหน้าที่ต่างกันระหว่างวิธีไคสแควร์กับวิธีวิเคราะห์ความ

แปรปรวน แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีไคสแควร์กับวิธีแมนเทิล – แฮนส์เซล และวิธีแมนเทิล – แฮนส์เซล กับวิธีวิเคราะห์ความแปรปรวนรวม แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ จำนวนข้อสอบที่มีการทำหน้าที่ต่างกัน เมื่อมีวิธีวิเคราะห์ทั้งวิธีไคสแควร์ วิธีแมนเทิล – แฮนส์เซล และวิธีวิเคราะห์ความแปรปรวน ระหว่างกลุ่มผู้สอบที่ใช้ภาษาพูดเป็นภาษาลาวและภาษาลาวพวน พบว่า จำนวนข้อสอบที่มีการทำหน้าที่ต่างกันแตกต่างอย่างไม่มีนัยสำคัญทางสถิติ ค่าความเชื่อมั่นของแบบทดสอบความเข้าใจในการอ่านหลังจากคัดเลือกข้อสอบที่มีการทำหน้าที่ต่างกันออกแล้ว เมื่อกลุ่มผู้สอบที่ใช้ภาษาพูดเป็นภาษาลาวและภาษาลาวพวน พบว่า ต่างก็มีค่าความเชื่อมั่นระหว่างวิธีตรวจสอบการทำหน้าที่ต่างกันทั้งวิธีไคสแควร์ วิธีแมนเทิล – แฮนส์เซล และวิธีวิเคราะห์ความแปรปรวน แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ส่วนค่าความเชื่อมั่นของแบบทดสอบความเข้าใจในการอ่านหลังจากคัดเลือกข้อสอบที่มีการทำหน้าที่ต่างกันออกแล้ว เมื่อมีวิธีตรวจสอบการทำหน้าที่ต่างกันวิธีเดียวกัน เมื่อกลุ่มผู้สอบที่ใช้ภาษาพูดเป็นภาษาลาวและภาษาลาวพวน มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

สำหรับการศึกษาค้นคว้าเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบเพื่อทำวิทยานิพนธ์ระดับบัณฑิตศึกษาในประเทศไทยสรุปโดยสังเขปได้ดังนี้

ตาราง 3 สรุปการศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบในประเทศไทย

ลำดับที่	ผู้วิจัย	ปีการศึกษา	วิธีการตรวจสอบที่ใช้	ตัวแปรแบ่งกลุ่มประชากร
1	ชัชชัย เผ่าพงศ์	2527	ทฤษฎีการตอบข้อสอบชนิด 3 พารามิเตอร์	เพศ
2	ทัศนีย์ พิรมนตรี	2530	1. กำหนดจุดเดลด้า 2. ทดสอบความแตกต่างระหว่างกลุ่มด้วย χ^2 ในแบบจำลองลอกลิเนียร์ 3. ทฤษฎีการตอบข้อสอบชนิด 3 พารามิเตอร์	ภาคภูมิศาสตร์

ตาราง 3 (ต่อ)

ลำดับที่	ผู้วิจัย	ปีการศึกษา	วิธีการตรวจสอบที่ใช้	ตัวแปรแบ่งกลุ่มประชากร
3	พัชรี ปิยภัณฑ์	2531	1. แปลงค่าความยาก 2. χ^2 3. ทฤษฎีการตอบข้อสอบชนิด 1 พารามิเตอร์	1. เพศ 2. สภาพพื้นที่ต่างกัน
4	สุรศักดิ์ อมรรัตนศักดิ์	2531	1. การวิเคราะห์ความแปรปรวน 2. แปลงค่าความยากของข้อสอบ 3. ทฤษฎีการตอบข้อสอบชนิด 1 และ 3 พารามิเตอร์	เพศ
5	สุพัฒน์ สุขกุลสันต์	2534	1. กำหนดจุดเดลด้า 2. $\chi^2 - 3$ 3. ทฤษฎีการตอบข้อสอบชนิด 3 พารามิเตอร์	1. เพศ 2. ภาควิชาศาสตร์
6	นิรมล ชัยชวลิต	2537	1. แปลงค่าความยาก 2. χ^2 3. การวิเคราะห์ความแปรปรวน	เพศ
7	เจน จันท์แสง	2537	1. แปลงค่าความยาก 2. การวิเคราะห์ความแปรปรวน 3. χ^2 4. ทฤษฎีการตอบข้อสอบชนิด 1 พารามิเตอร์	1. เพศ 2. ที่ตั้งของโรงเรียน
8	กาญจนา วัฒนสุนทร	2538	1. การตอบข้อสอบชนิดคิดพื้นที่ (SA) 2. การตอบข้อสอบชนิดไม่คิดพื้นที่ (UA) 3. MH 4. SIBTEST	เพศ
9	คมศักดิ์ ชื่นชม	2539	1. วิธีค่าอำนาจจำแนก 2. วิเคราะห์องค์ประกอบ 3. การถดถอยโลจิสติก	1. ภาษาที่ใช้ในชีวิตประจำวัน ในครอบครัว 2. ศาสนาที่นับถือ
10	จิตติมา วรรณศรี	2539	MH, SIBTEST	เพศ

ตาราง 3 (ต่อ)

ลำดับที่	ผู้วิจัย	ปีการศึกษา	วิธีการตรวจสอบที่ใช้	ตัวแปรแบ่งกลุ่มประชากร
11	เสรี ชัดเข้ม	2540	1. MH แบบปกติ 2. MH แบบแบ่งกลุ่ม ความสามารถและความยาก ของข้อสอบ	เพศ
12	จิตสุดา ธราพร	2539	1. แปลงค่าความยาก 2. ทฤษฎีการตอบข้อสอบชนิด 3 พารามิเตอร์ 3. MH	ศาสนาที่นับถือ
13	ญาณภัทร สีหะมงคล	2540	1. Lord's χ^2 2. Raju's Area Measures 3. วิธี Closed Interval Area	เพศ
14	เววดี อินทะสระ	2539	1. IRT 2. MH 3. SIBTEST	เพศ
15	พรรณี จิตมาศ	2540	1. แปลงค่าความยาก 2. MH 3. SIBTEST	เพศ
16	ดารณี แดดภู	2542	1. วิธีราส์ชโมเดล 2. MH 3. SIBTEST	1. เพศ 2. ภูมิลำเนาของนักเรียน
17	ทองอยู่ สาระ	2543	1. MH 2. การถดถอยโลจิสติก	เพศ
18	นวลอนงค์ นันทวงศ์	2543	1. χ^2 2. MH 3. การวิเคราะห์ความแปรปรวน	ลักษณะการใช้ภาษาพูดใน บ้าน
19	อารี วัชรโสติกุล	2543	1. SIBTEST 2. การถดถอยโลจิสติก	รูปแบบการตรวจสอบการทำ หน้าที่ต่างกัน
20	วลีมาศ แซ่อึ้ง	2544	1. MH 2. SIBTEST และ Modified SIBTEST 3. การถดถอยโลจิสติก	

ตาราง 3 (ต่อ)

ลำดับที่	ผู้วิจัย	ปีการศึกษา	วิธีการตรวจสอบที่ใช้	ตัวแปรแบ่งกลุ่มประชากร
21	พิพุดพงษ์ แนวทอง	2544	1. HW 3 3. ไบซีเรียล 4. Item-total Correlation	เพศ
22	วิภา จำมัน	2544	1. MH 2. SIBTEST	เพศ
23	นุชริน ไบโพธิ์	2544	1. Lord's χ^2 2. MH 3. SIBTEST	เพศ
24	ชุติมา แสงดารารัตน์	2546	1. MH 2. การวิเคราะห์ความแปรปรวน	เพศ

งานวิจัยในต่างประเทศ

สำหรับงานวิจัยในต่างประเทศที่เกี่ยวกับการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบนั้นมีผู้วิจัยทำวิจัยไว้หลายท่าน ที่สำคัญสรุปได้ดังนี้

ผจงจิต อินทสุวรรณ (เหวดี อินทสระ. 2539 : 30 ; อ้างอิงจาก ผจงจิต อินทสุวรรณ. 1979) ได้ศึกษาวิธีตรวจสอบการทำหน้าที่ต่างกันที่ต่างกัน 3 วิธี คือ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ วิธีราล์ช และวิธีไคสแควร์ ของแบบทดสอบ IEA กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักเรียนอายุ 14 ปี ชาวอังกฤษ อเมริกา และนิวซีแลนด์จำนวน 8,601 คน ผลการวิจัยพบว่า สหสัมพันธ์ระหว่างวิธีทั้ง 3 มีค่าอยู่ระหว่าง .51 - .98 และวิธีราล์ช กับไคสแควร์มีค่าสหสัมพันธ์สูงถึง .98

สงบ ลักษณะ (เหวดี อินทสระ. 2539 : 30 ; อ้างอิงจาก สงบ ลักษณะ. 1979 : citing Burrill. 1982 : 168) ได้ศึกษาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบ ITBS ฉบับภาษาด้านคำศัพท์และฉบับคณิตศาสตร์ ด้วยวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และวิธีวิเคราะห์ความแปรปรวน ผลการวิจัย พบว่า วิธีวิเคราะห์ความแปรปรวนเป็นวิธีที่ใช้ได้ดีในทางปฏิบัติ แต่เป็นวิธีที่ไม่ครอบคลุม ส่วนวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ เป็นวิธีที่ครอบคลุมกว่าแต่ต้องใช้กลุ่มตัวอย่างขนาดใหญ่และการคำนวณยุ่งยากซับซ้อน

เชปเพิร์ด คามิลลี และเอเวอร์ริลล์ (จิตสุดา ธราพร. 2539 : 73 ; อ้างอิงจาก Berk. 1982 : 169 ; citing Shepard Camilli ; & Averill. 1980) ได้เปรียบเทียบวิธีวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบจากแบบทดสอบ Lorge – Thorndike แบ่งเป็นฉบับภาษากับไม่ใช่ภาษา ใช้กลุ่มตัวอย่างเป็น

นักเรียนผิวขาว ผิวดำ และชิวาโน (Chicano) เชื้อชาติละ 500 คน วิเคราะห์การทำหน้าที่ต่างกันด้วยวิธีไคสแควร์ วิธีกำหนดจุดเดลตา หรือวิธีแปลงค่าความยากพอยท์ไบซีเรียล ไค้งลักษณะข้อสอบ 3 พารามิเตอร์ และไค้งลักษณะข้อสอบ 1 พารามิเตอร์ (Rasch Model) การวิเคราะห์การทำหน้าที่ต่างกันใช้ทั้งเกณฑ์ภายนอกและเกณฑ์ภายใน โดยที่เกณฑ์ภายนอกคือ คะแนนจาก Raven' s Coloured Progressive Matrices เกณฑ์ภายในคือคะแนนรวมจากแบบทดสอบ ผลการวิจัยพบว่าวิธีไค้งลักษณะข้อสอบ 3 พารามิเตอร์ เป็นวิธีที่เหมาะสมในทางทฤษฎีถูกต้องและตรงที่สุด แต่มีข้อจำกัดคือต้องใช้กลุ่มตัวอย่างขนาดใหญ่ ส่วนวิธีอื่น ๆ ก็สามารถใช้ได้

เพิร์ลแมน และคณะ (ญาณภัทร สีหะมงคล. 2540 : 27 ; อ้างอิงจาก Periman; et al. 1988) ได้ศึกษาความคงที่ของวิธีประมาณค่าการทำหน้าที่ต่างกันของข้อสอบ 4 วิธี ได้แก่ วิธีแปลงค่าความยาก วิธีแปลงค่าความยากที่ดัดแปลงโดย Shepard วิธีวิเคราะห์ค่าที่เหลือแบบ 1 พารามิเตอร์ของราสช์ และวิธีแมนเทล – แฮนส์เซล กลุ่มตัวอย่างที่ใช้มี 30 กลุ่ม ตั้งแต่ 600 คนจนถึง 2,000 คน ซึ่งสุ่มมาจากนักเรียนระดับ 9 จำนวน 54,896 คน ประกอบด้วยคนผิวขาว ผิวดำ และชนชาติที่พูดภาษาสเปนให้ทำแบบทดสอบทักษะความสามารถขั้นต่ำของรัฐชิคาโก ซึ่งเป็นแบบทดสอบแบบเลือกตอบ 46 ข้อ การศึกษาพบว่า แต่ละวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ไม่คงที่เมื่อขนาดกลุ่มตัวอย่างต่างกัน และความเชื่อมั่นของตัวบ่งชี้การทำหน้าที่ต่างกันของข้อสอบจะมีปัญหา ถ้าใช้กลุ่มตัวอย่างต่ำกว่า 666 คน

แฮมเบิลตัน และโรเจอร์ส (Hambleton ; & Rogers.1989) ได้เปรียบเทียบวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี คือ วิธีทฤษฎีการตอบข้อสอบ และวิธีแมนเทล แฮนเซล (MH) ข้อมูลที่ใช้เป็นผลการสอบแบบทดสอบ New Mexico High School Proficiency Exam (NMHSPE) จำนวน 150 ข้อ ของนักเรียนผิวขาว 8,000 คน และนักเรียนพื้นเมือง 2,600 คน

NMHSPE เป็นแบบทดสอบชนิดเลือกตอบ 4 ตัวเลือก วัดทักษะการดำเนินชีวิต 5 ด้าน คือ ความรู้เกี่ยวกับแหล่งทรัพยากรของชุมชน ความรู้เกี่ยวกับผู้บริโภค ความรู้เกี่ยวกับรัฐบาลและกฎหมาย ความรู้เกี่ยวกับสุขภาพทางกายและจิต และความรู้เกี่ยวกับอาชีพ

กลุ่มตัวอย่างที่ศึกษา สุ่มมากลุ่มละ 2,000 คน และแบ่งเป็นกลุ่มย่อย ๆ ละ 1,000 คน การศึกษาการทำหน้าที่ต่างกัน โดยพิจารณาร้อยละของข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันหรือไม่ต่างกันซ้ำกันในการตรวจสอบ 2 ครั้ง

ผลการศึกษา พบว่า วิธีทฤษฎีการตอบข้อสอบมีความสอดคล้องของผลการตรวจสอบร้อยละ 73 ส่วนวิธีแมนเทล แฮนเซล มีความสอดคล้องของผลการตรวจสอบร้อยละ 80 และพบว่าร้อยละ 47 ของข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันในการตรวจสอบคู่ที่ 1 เป็นข้อสอบทำหน้าที่ต่างกันในการตรวจสอบคู่ที่ 2 ด้วย สำหรับวิธีแมนเทล แฮนเซล และร้อยละ 61 สำหรับวิธีทฤษฎีการตอบข้อสอบ

ในทางกลับกันพบว่า ร้อยละ 64 ของข้อสอบที่ทำหน้าที่ต่างกันในการตรวจสอบคู่ที่ 2 เป็นข้อสอบที่ทำหน้าที่ต่างกันในการตรวจสอบคู่ที่ 1 สำหรับวิธีแมนเทิล แชนเซล และร้อยละ 56 สำหรับวิธีทฤษฎีการตอบข้อสอบ จะเห็นได้ว่าความคงที่ของการตรวจสอบทั้ง 2 วิธี อยู่ในระดับปานกลาง

บากิ และเฟอราธา (Baghi ; & Ferrara. 1990) ได้เปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยทฤษฎีการตอบข้อสอบชนิด 3 พารามิเตอร์ และสถิติไคสแควร์แบบแมนเทิล แชนส์เซล โดยใช้ข้อมูลผลการสอบจากแบบทดสอบทักษะความเป็นพลเมืองดีของรัฐแมริแลนด์ กลุ่มตัวอย่างเป็นนักเรียนระดับ 9 จำนวน 50,000 คน ขนาดกลุ่มตัวอย่างที่ใช้ในการตรวจสอบด้วย วิธีแมนเทิล – แชนส์เซล มี 4 ขนาด ได้แก่ 200 , 500 , 750 และ 1,000 คน ในแต่ละกลุ่มเปรียบเทียบ

ผลการศึกษา พบว่า วิธีทฤษฎีการตอบข้อสอบตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างกลุ่มหญิงและชาย พบข้อสอบที่ทำหน้าที่ต่างกัน 4 ข้อ และพบข้อสอบที่ทำหน้าที่ต่างกัน 3 ข้อ ในการเปรียบเทียบระหว่างผู้สอบผิวขาวและผิวดำ และในขณะที่ผลจากการหาค่าสหสัมพันธ์ระหว่างดัชนีทั้งสองตัว แสดงความไม่สอดคล้องกันในบ่งชี้ข้อสอบที่ทำหน้าที่ต่างกันนั้น ความสอดคล้องโดยรวมพบว่า ไคสแควร์ แบบแมนเทิล – แชนส์เซล เป็นวิธีที่ใช้แทนทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์ ได้ถ้าขนาดกลุ่มตัวอย่าง 750 คน ขึ้นไป

คิม และโคเฮน (Kim and Cohen. 1991) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบที่สร้างขึ้นอย่างจงใจให้ทำหน้าที่ต่างกัน โดยการเปรียบเทียบผลการคำนวณพื้นที่ความแตกต่างระหว่างโค้งลักษณะข้อสอบของกลุ่มตัวอย่างผิวขาว และผิวดำ 2 วิธี ประกอบด้วย การคำนวณพื้นที่ภายใต้ความสามารถที่มีขอบเขตจำกัด (ปิด) และที่ไม่มีขอบเขตจำกัด (เปิด)

ขนาดของกลุ่มตัวอย่าง 1,000 คน สุ่มมาจากคนผิวขาว 1,021 คน และผิวดำ 1,008 คน เครื่องมือที่ใช้ในการศึกษาเป็นแบบทดสอบคำศัพท์แบบเลือกตอบชนิด 4 ตัวเลือก จำนวน 50 ข้อ เป็นคำศัพท์ภาษาอังกฤษมาตรฐาน 40 ข้อ อีก 10 ข้อ เป็นคำศัพท์ที่มีความง่ายสำหรับกลุ่มผิวดำ มากกว่าผิวขาว

การคำนวณพื้นที่ระหว่างฟังก์ชันการตอบ ทั้งภายใต้ช่วงความสามารถแบบปิด และแบบเปิด จะพิจารณา 2 ลักษณะ คือ พื้นที่ชนิดมีเครื่องหมาย (Signed Area) และพื้นที่ชนิดไม่มีเครื่องหมาย (Unsigned Area) แบบจำลองที่ใช้ในการศึกษาเป็นแบบจำลอง 3 พารามิเตอร์ และ 3 พารามิเตอร์ ที่กำหนดให้ c คงที่เท่ากับ .23

ผลการศึกษา พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างการคำนวณพื้นที่แบบช่วงเปิด และแบบช่วงปิด มีความไวต่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมากกว่าแบบมีเครื่องหมาย นอกจากนั้นยังพบอีกว่าพื้นที่แบบมีเครื่องหมาย และไม่มีเครื่องหมายของการคำนวณแบบช่วงเปิด มีความสัมพันธ์กันมากกว่าการคำนวณพื้นที่แบบช่วงปิด

สำหรับความคลาดเคลื่อนในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน พบว่าเมื่อใช้แบบจำลอง 3 พารามิเตอร์ พื้นที่แบบมีเครื่องหมาย มีอัตราความคลาดเคลื่อนต่ำกว่าวิธีอื่น ๆ และมีความคลาดเคลื่อนในการยอมรับผิดพลาด จำนวน 2 ข้อ ที่ระดับ .05 ส่วนพื้นที่แบบไม่มีเครื่องหมายมีอัตราความคลาดเคลื่อนสูงกว่าวิธีอื่น ๆ เล็กน้อย

โคเฮน และคิม (Cohen ; & Kim. 1993) ได้ศึกษาเปรียบเทียบการใช้สถิติทดสอบ χ^2 ของ ลอร์ด และสถิติทดสอบพื้นที่ความแตกต่างระหว่างโค้งลักษณะข้อสอบชนิดมีเครื่องหมาย และไม่มีเครื่องหมาย และสถิติ Z ของราชู โดยจำลองข้อมูลตามทฤษฎีการตอบข้อสอบชนิด 2 พารามิเตอร์ กำหนดให้แบบทดสอบยาว 20 และ 60 ข้อ จำนวนผู้สอบ 100 และ 500 และมีจำนวนข้อสอบที่ทำหน้าที่ต่างกันแบบทดสอบจำนวนร้อยละ 0, 10 และ 20 พบว่า

1.แบบทดสอบที่มีความยาวมากขึ้นและมีจำนวนข้อสอบที่ทำหน้าที่ต่างกันแบบทดสอบน้อย ความคลาดเคลื่อนในการยอมรับผิดพลาด (False Positive : FP) จะมากขึ้น และการกำหนดระดับนัยสำคัญต่างกันก็มีผลต่อความคลาดเคลื่อนประเภทนี้ (ที่ระดับ .05 จะเกิด FP มากกว่าที่ระดับ .01) สำหรับสถิติทดสอบ Z จะเกิดขึ้นทุกเงื่อนไขของการศึกษา ส่วนสถิติ χ^2 จะเกิดขึ้นบางเงื่อนไข

2.แบบทดสอบที่มีความยาวมากขึ้นและมีจำนวนข้อสอบที่ทำหน้าที่ต่างกันแบบทดสอบมาก ความคลาดเคลื่อนในการปฏิเสธผิดพลาด (False Negative : FN) จะมากขึ้น โดยสถิติทดสอบ Z จะเกิดขึ้นมากกว่าสถิติทดสอบ χ^2

3.ความแตกต่างของการเกิดความคลาดเคลื่อนในการปฏิเสธผิดพลาดระหว่างสถิติทดสอบ Z และ χ^2 เมื่อจำนวนผู้เข้าสอบ 100 คน จะน้อยกว่าเมื่อจำนวนผู้เข้าสอบ 500 คน และสถิติทดสอบ χ^2 เกิดความคลาดเคลื่อนในการปฏิเสธผิดพลาดน้อยกว่า สถิติทดสอบ Z โดยเฉพาะกรณีที่มีจำนวนข้อสอบทำหน้าที่ต่างกันแบบทดสอบมากที่สุด คือร้อยละ 20

ราชู และเจฟฟรี (Raju ; & Jeffrey. 1993 : 301 – 319) ได้เปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของ ข้อสอบระหว่าง วิธีลอร์ด วิธีราชู ละวิธี Mantel – Haenszel เครื่องมือที่ใช้ในการศึกษาเป็นแบบทดสอบวัดคำศัพท์ จำนวน 45 ข้อ ชนิด 5 ตัวเลือก กลุ่มตัวอย่างเป็นนักเรียนระดับ 10 และ 12 จำนวน 839 คน แยกเป็นชาย 399 คน หญิง 440 คน ผิวขาว 436 คน และผิวดำ 245 คน (อีก 158 คน เป็นเชื้อชาติอื่น)

ผลการศึกษาพบว่า การตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างนักเรียนผิวดำและผิวขาวมีข้อสอบที่ทำหน้าที่ต่างกัน 1 ข้อ สำหรับการทดสอบด้วยวิธีลอร์ด และวิธีราชู ทั้งพื้นที่ชนิดที่มีเครื่องหมายและไม่มีเครื่องหมาย ได้แก่ ข้อ 41 ทั้ง 3 วิธี ส่วนวิธี Mantel – Haenszel พบ 5 ข้อ ซ้ำกับ 3 วิธีแรก 4 ข้อ (ร้อยละ 80) ได้แก่ ข้อ 2, 18, 23, 33 และข้อที่ไม่ซ้ำกับ 3 วิธีแรกคือ ข้อ 41 ทุกข้อมีค่าความยากสูงกว่าข้อที่ทำหน้าที่ไม่ต่างกัน

เป็นที่น่าสังเกตว่า วิธี Mantel – Haenszel ตรวจพบข้อสอบที่ทำหน้าที่ต่างกันได้มากกว่าวิธี ทฤษฎีการตอบข้อสอบ ซึ่งขัดแย้งกับการศึกษาของ สวามินาธาน และโรเจอร์ (1990) ที่พบว่า วิธี ทฤษฎีการตอบข้อสอบตรวจพบข้อสอบที่ทำหน้าที่ต่างกันได้มากกว่าวิธี Mantel – Haenszel

นารายานาน และสวามินาธาน (Narayanan ; & Swaminathan. 1994 ; 315- 328) ได้ศึกษา เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันแบบยูนิฟอร์มด้วยวิธี MH และวิธี SIBTEST โดยใช้ ข้อมูลที่จำลองภายใต้เงื่อนไข 1,296 เงื่อนไข ($9 \times 3 \times 2 \times 4 \times 6$) แล้วทำการคำนวณซ้ำ 100 ครั้ง ใน แต่ละเงื่อนไข เงื่อนไขที่ทำการศึกษาประกอบด้วย

1. ขนาดกลุ่มตัวอย่างที่ต่างกัน 3 ขนาดคือ กลุ่มเปรียบเทียบใช้ขนาด 100 , 200 และ 300 คน กลุ่มอ้างอิงใช้ขนาด 300 , 500 และ 1,000 คน วางไขว้กันได้กลุ่มตัวอย่างทั้งหมด 9 กลุ่ม
2. ความแตกต่างในการแจกแจงความสามารถ 3 ระดับ ระดับที่ 1 ให้ทั้ง 2 กลุ่มมีค่าเฉลี่ยการ แจกแจงความสามารถอยู่ที่ 0.0 ระดับที่ 2 ให้ค่าเฉลี่ยของความสามารถในกลุ่มอ้างอิง และกลุ่ม เปรียบเทียบเป็น 0.0 และ 0.5 ตามลำดับ และระดับที่ 3 ค่าเฉลี่ยความสามารถในกลุ่มอ้างอิงและกลุ่ม เปรียบเทียบเป็น 0.0 และ -1.0 ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานของทั้ง 3 ระดับเป็น 1.0
3. ร้อยละของข้อสอบที่การทำหน้าที่ต่างกันมี 2 ขนาด คือ ร้อยละ 10 และ ร้อยละ 20 จาก แบบทดสอบจำนวน 40 ข้อ
4. ขนาดของการทำหน้าที่ต่างกัน 4 ระดับ คือ 0.4 , 0.6 , 0.8 และ 1.0
5. ชนิดของแบบทดสอบ 6 ชนิด คือ ชนิดของแบบทดสอบที่ประกอบด้วยค่าความยากและค่า อำนาจจำแนก 3 ระดับ คือ สูง ปานกลาง ต่ำ ทั้ง 2 ค่าออกจำนวน 3 ฉบับ

ผลการวิจัยพบว่า MH และวิธี SIBTEST สามารถตรวจค้นการทำหน้าที่ต่างกันได้ใกล้เคียงกัน โดยจะตรวจค้นได้ดีในขนาดกลุ่มตัวอย่าง 600 คน (กลุ่มอ้างอิงและกลุ่มเปรียบเทียบกลุ่มละ 300 คน) ขนาดของการทำหน้าที่ต่างกันทั้ง 4 ขนาดมีผลต่อการตรวจค้นการทำหน้าที่ต่างกันทั้ง 2 วิธีอย่างมี นัยสำคัญทางสถิติ ความแตกต่างในการแจกแจงความสามารถไม่มีผลต่อวิธีการวิเคราะห์การทำหน้าที่ ต่างกันแบบ SIBTEST แต่มีผลต่อวิธี MH ที่เป็นเช่นนี้อาจเป็นเพราะวิธี SIBTEST ได้มีการปรับแก้การ ถดถอย ขนาดของแบบทดสอบและร้อยละของข้อสอบที่การทำหน้าที่ต่างกันไม่มีผลต่อวิธีทั้งสอง สำหรับการเกิดความคลาดเคลื่อนแบบที่ 1 วิธี MH จะมีอัตราการเกิดอยู่ในระดับปกติ แต่วิธี SIBTEST จะเพิ่มขึ้นเล็กน้อยในกลุ่มที่มีความสามารถเท่ากัน แต่ในกลุ่มที่มีความสามารถไม่เท่ากันการเกิดความ คลาดเคลื่อนแบบที่ 1 จะมีอัตราการเกิดสูงขึ้นทั้ง 2 วิธี ซึ่งการเพิ่มความยาวของแบบทดสอบจะทำให้ เกิดความเชื่อมั่นสูงขึ้น ซึ่งมีผลให้ความคลาดเคลื่อนแบบที่ 1 นำจะมีอัตราการเกิดลดลง

รูสโซ และสเตาท์ (Roussos ; & Stout. 1996 : 215 – 230) ได้ศึกษาผลการใช้จำนวนกลุ่ม ตัวอย่างขนาดเล็ก ความแตกต่างของค่าพารามิเตอร์ข้อสอบ รวมทั้งความแตกต่างในค่าเฉลี่ยของ

ความสามารถของกลุ่มอ้างอิง และกลุ่มเปรียบเทียบเพื่อศึกษาผลที่มีต่อความคลาดเคลื่อน แบบที่ 1 (Type I error) คือการยอมรับว่าข้อสอบการทำหน้าที่ต่างกันทั้ง ๆ ที่ไม่การทำหน้าที่ต่างกัน โดยวิธี MH และวิธี SIBTEST

การศึกษาที่ 1 ใช้กลุ่มตัวอย่างเป็นกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ กลุ่มละ 100 , 200 , 500 และ 1,000 คน เลือกข้อสอบ 1 ข้อ จากแบบทดสอบ ASVAB เลือกข้อที่มีพารามิเตอร์ของข้อสอบ คือ a , b และ c ใกล้เคียงกับค่าเฉลี่ยของแบบทดสอบมากที่สุดคือ 1.32 , 0.03 และ 0.25 (ค่าเฉลี่ย 1.22 0.09 และ 0.02) และเป็นข้อสอบที่ไม่ทำหน้าที่ต่างกันของความสามารถ (θ) ของกลุ่มตัวอย่างทั้ง 2 กลุ่มถูกจำลองให้มีการกระจายเป็นปกติมีค่าความแปรปรวนเป็น 1.00 ความแตกต่างระหว่างค่าเฉลี่ยของการแจกแจงของ 2 กลุ่ม ถูกจำลองให้มีค่า 0.0 , 0.5 และ 1.0 แล้วคำนวณค่าความสามารถเฉลี่ยของแต่ละกลุ่มให้จุดกลางของค่าเฉลี่ยอยู่บนระดับความยากเฉลี่ยของแบบทดสอบ จำลองการตอบ 400 ครั้ง และวิเคราะห์ 400 ครั้งในแต่ละสถานการณ์

ผลการใช้ SIBTEST และ MH ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เลือกขึ้นมา พบว่า นอกจากกลุ่มตัวอย่างขนาด 100 คน และค่าความแตกต่างของความสามารถเป็น 0.0 แล้วอัตราการเกิดความคลาดเคลื่อนแบบที่ 1 ของวิธีทั้ง 2 แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ แต่ในขนาดกลุ่มตัวอย่าง 100 คน และความแตกต่างของความสามารถเป็น 0.0 จะมีอัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 2 วิธี แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 โดยที่วิธี SIBTEST จะมีอัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 มากกว่าวิธี MH และในขนาดกลุ่มตัวอย่างไม่เกิน 1,000 คน จะมีแนวโน้มการเกิดความคลาดเคลื่อนสูงทั้ง 2 วิธี

การศึกษาที่ 2 ใช้แบบทดสอบชุดเดิม แต่กำหนดค่าความยากและค่าอำนาจจำแนกที่แตกต่างกัน ดังนี้ ค่าอำนาจจำแนกเป็น 0.4 , 1.0 และ 2.5 ค่าความยากเป็น -1.5 , -0.5 , 0.0 , 0.5 และ 1.5 ค่าการเดาเป็น 0.20 ทำการจำลองการแจกแจงความสามารถให้เป็น 2 แบบ คือ 0.0 และ 1.0 นอกนั้นจำลองทุกอย่างเช่นเดียวกับการศึกษาที่ 1 แต่ใช้จำนวนผู้สอบขนาดใหญ่คือ 500 , 1,000 และ 3,000 คน ในแต่ละสถานการณ์คำนวณซ้ำ 100 ครั้ง และตรวจสอบการทำหน้าที่ต่างกันโดยใช้ MH และ SIBTEST ผลการวิจัยพบว่า เมื่อค่าอำนาจจำแนกสูงขึ้นและค่าความยากต่ำ อัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 จะสูงทั้ง 2 วิธีที่ระดับนัยสำคัญ 0.5 และวิธี SIBTEST เมื่อการแจกแจงความสามารถเป็น 1.0 ค่าการเดาเป็น 0.20 วิธี SIBTEST มีอัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธี MH แต่เมื่อใช้การแจกแจงความสามารถที่ 0.0 ไม่มีการเดาแล้วอัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 2 วิธีไม่แตกต่างกัน

สรุปได้ว่า เมื่อใช้กลุ่มตัวอย่างขนาดเล็กไม่เกิน 1,000 คน วิธีทั้ง 2 ให้ผลด้านความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกัน แต่เมื่อใช้ขนาดกลุ่มตัวอย่างที่มากกว่า 1,000 คน และค่าเฉลี่ยการแจกแจง

ความสามารถของทั้ง 2 กลุ่มแตกต่างกัน รวมทั้ง ข้อสอบมีค่าความยากต่ำ ค่าอำนาจจำแนกสูง วิธี SIBTEST มีอัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธี MH

ซวิก,ไทเยอร์ และมาสซิโอ (Zwick,Thyer ; & Mazzeo. 1997:321 – 344) ได้ทำการศึกษาเกี่ยวกับการค้นหาการทำหน้าที่ต่างกันจากแบบทดสอบที่มีการตรวจให้คะแนนแบบหลายค่า โดยใช้วิธีวิเคราะห์ 5 วิธีด้วยกัน คือ SMD-H , SMD-M , วิธี MH และวิธีSIBTEST 2 วิธี ได้แก่วิธี Standrad SIBTEST และ Modifed SIBTEST โดยใช้เงื่อนไขแบบทดสอบที่มีการตรวจให้คะแนนแบบ 0-1 จำนวน 50 ข้อ และแบบทดสอบที่มีการตรวจให้คะแนนแบบหลายค่า จำนวน 18 ข้อ ใช้กับกลุ่มตัวอย่าง 1,000 คน แบ่งเป็นกลุ่มอ้างอิง 500 คน และกลุ่มเปรียบเทียบ 500 คน โดยมีการจับคู่ระหว่างข้อสอบแบบให้คะแนน 2 ค่า กับแบบให้คะแนนหลายค่า โดยใช้โมเดลลอจิติสส์ 3 พารามิเตอร์ จากข้อมูลจำนวน 75 ข้อ ที่มีค่าอำนาจจำแนกในช่วง .74 ถึง 1.00 ค่าความยากอยู่ระหว่าง – 1.96 ถึง 1.95 ค่าการเดาอยู่ที่ .15 และสถานการณ์ที่จำลองขึ้นมากับกลุ่มตัวอย่างทั้ง 2 กลุ่มที่มีการแจกแจงดัชนีการปฏิบัติที่ดีดูจาก SMD เมื่อความแตกต่างของค่าเฉลี่ยของกลุ่มที่มีความคลาดเคลื่อนมาตรฐานเป็น 1 วิธีการ Modifed SIBTEST เป็นการวัดผลกระทบของขนาดกลุ่มตัวอย่างค่อนข้างดี จาก 5 วิธี ในทางปฏิบัติ จะไม่สามารถมองเห็นความแตกต่างกันและข้อสอบที่ศึกษามีค่าอำนาจจำแนกสูง วิธีการ SIBTEST จะแสดงให้เห็นดีกว่าวิธีอื่นในรูปแบบคลาดเคลื่อนประเภทที่ 1 มากกว่าวิธี SMD และวิธี MH จากแบบทดสอบตอบสั้น ๆ ค่าอำนาจจำแนกจาก 5 วิธี จะไม่เหมือนกันเป็นความแตกต่างในการวิเคราะห์การทำหน้าที่ต่างกันและองค์ประกอบอื่น ๆ วิธีการที่ใช้กันบ่อย ๆ ในการประยุกต์ใช้การทำหน้าที่ต่างกันจากแบบทดสอบแบบPolytomous เมื่อมีการสอบตัวแปรคือการจับคู่กันระหว่างการสอบกับแบบฝึกหัดอัตราความคลาดเคลื่อนประเภทที่ 1 จากวิธี MH และวิธี SMD จะเกี่ยวกับภายใต้เงื่อนไขของเกณฑ์ วิธีการปัจจุบันของ SIBTEST ไม่สามารถใช้กับการจับคู่ที่ง่าย ๆ ได้ นั่นคือ ใช้กับแบบทดสอบแบบถูก – ผิดไม่ได้ งานวิจัยในปัจจุบันจึงไม่นิยมกัน

จากการศึกษาผลการวิจัยข้างต้นจะเห็นได้ว่าการศึกษาเปรียบเทียบวิธีวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบมีหลายวิธี บางวิธีเป็นวิธีที่ยืดหยุ่นวิธีการวัดผลแบบดั้งเดิม เช่น วิธีการแปลงค่าความยาก วิธีวิเคราะห์ความแปรปรวน วิธีโคสแควร์ วิธีค่าอำนาจจำแนก เป็นต้น บางวิธียืดหยุ่นวิธีการตอบข้อคำถาม เช่น วิธีโค้งลักษณะข้อสอบ เป็นต้น และบางวิธีก็ใช้สถิติแบบนอนพาราเมตริก เช่น วิธีแมนเทล – แฮนส์เซล , วิธีชิปเทสท์ , วิธี Modifed SIBTEST และ Polytomous STND เป็นต้น จากวิธีต่าง ๆ ดังกล่าวบางวิธีวิธีเป็นวิธีที่ให้ผลการวิเคราะห์ได้ถูกต้อง แม่นยำ แต่ไม่เหมาะที่จะนำมาใช้ในทางปฏิบัติ เพราะต้องใช้กลุ่มตัวอย่างขนาดใหญ่ การคำนวณและการแปลผลซับซ้อน ยุ่งยาก ต้องเสียค่าใช้จ่ายสูง ในขณะที่บางวิธีเป็นวิธีที่ใช้กลุ่มตัวอย่างไม่มาก การคำนวณและการแปลผลง่าย ไม่ซับซ้อน แต่ให้ผลการวิเคราะห์ที่มีความถูกต้องน้อยกว่าวิธีอื่น ๆ และบางวิธีใช้กลุ่มตัวอย่างไม่มาก การ

คำนวณและการแปลผลง่าย เหมาะที่จะนำมาใช้ในทางปฏิบัติได้และให้ผลการวิเคราะห์ที่มีความถูกต้อง ซึ่งวิธีที่มีผู้แนะนำมาใช้ เช่น วิธีแมนเทล – แฮนส์เซล , วิธีชิปเทสท์ และวิธีแปลงค่าความยากสำหรับตัวแปรที่ทำหน้าที่ต่างกันที่นิยมมาศึกษาคือ ภาษา เพศ วัฒนธรรม ศาสนา สภาพภูมิศาสตร์ เป็นต้น นอกจากนี้จากการศึกษางานวิจัยก็พบว่า มี ตัวแปรบางตัวที่ส่งผลต่อการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีต่าง ๆ เช่น ภาษาที่ใช้พูดในครอบครัว เชื้อชาติ ศาสนา และตัวแปรจากแบบทดสอบ เช่น ความยาวแบบทดสอบ การแจกแจงความสามารถของผู้สอบในแต่ละกลุ่ม เป็นต้น ซึ่งผลการวิจัยที่ผ่านมายังไม่มีผู้ศึกษาเปรียบเทียบการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้านเชื้อชาติที่ไม่ใช่ไทย แต่ศึกษาอยู่ในประเทศไทย ด้วยวิธีวิเคราะห์วิธี Mantel – Haenszel วิธี Lord's χ^2 และ วิธี Closed Interval Area

บทที่ 3

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการตามขั้นตอนดังนี้

1. การกำหนดประชากรและการสุ่มกลุ่มตัวอย่าง
2. การสร้างเครื่องมือที่ใช้ในการวิจัย
3. การเก็บรวบรวมข้อมูล

การกำหนดประชากรและการสุ่มกลุ่มตัวอย่าง

ประชากร

ประชากรที่ใช้ในการศึกษาค้นคว้าเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ภาคเรียนที่ 1 ปีการศึกษา 2549 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษากาญจนบุรี เขต 3 ที่มีนักเรียนทั้งเชื้อชาติไทยและไม่ใช่อื่นเชื้อชาติไทย จำนวนทั้งสิ้น 2,607 คน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ภาคเรียนที่ 1 ปีการศึกษา 2549 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษากาญจนบุรี เขต 3 ซึ่งเลือกมาโดยการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) (สุชาติ กิระนนท์. 2542) ผู้วิจัยดำเนินการสุ่มตัวอย่าง ดังนี้

1. สืบหาข้อมูลหน่วยสมาชิกของประชากร จากแหล่งข้อมูลทุติยภูมิ คือ ข้อมูลสารสนเทศจำนวนนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2548 สำนักงานเขตพื้นที่การศึกษากาญจนบุรี เขต 3 จัดทำกรอบของการสุ่ม (Sampling Frame) โดยอาศัยการแบ่งตามอำเภอที่สำนักงานเขตพื้นที่การศึกษากาญจนบุรี เขต 3 รับผิดชอบ

2. การกำหนดขนาดของกลุ่มตัวอย่างแบบแบ่งชั้น โดยมีอำเภอเป็นตัวแบ่งชั้นจากจำนวนประชากรทั้งหมด ดำเนินการสุ่มโดยอาศัยหลักของการสุ่มที่อาศัยการกำหนดขนาดของความคลาดเคลื่อน (limit of error) เท่ากับ 0.5 คะแนน จากแบบทดสอบใจหายปัญหากลุ่มสาระการเรียนรู้ วิชาคณิตศาสตร์ ระดับความเชื่อมั่น 95 % ($\alpha = .05$) ในการประมาณค่าเฉลี่ยประชากร และค่าความแปรปรวนเท่ากับ 49 ซึ่งผู้วิจัยถือว่าเป็นขนาดที่เพียงพอที่จะนำผลการวิจัยไปใช้ในการตัดสินใจในกรณีต่าง ๆ ที่เกี่ยวข้องได้ เมื่อคำนวณจากสูตรการกำหนดขนาดกลุ่มตัวอย่างของการสุ่ม พบว่าต้องใช้

กลุ่มตัวอย่างอย่างน้อย 361 คน แต่การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Mantel – Haenzel ควรใช้กลุ่มผู้สอบระหว่าง 200 ถึง 1,000 คน (ทงอยู่ สาระ. 2543 : 38 ; อ้างอิงจาก Hambleton and others. 1989) ดังนั้นในการวิจัยครั้งนี้ต้องการกลุ่มตัวอย่าง ที่มีเชื้อชาติไทยอย่างน้อย 750 คน และกลุ่มตัวอย่างที่ไม่มีเชื้อชาติไทยอย่างน้อย 750 คน ใช้เป็นประชากรเทียบในการสุ่มกลุ่มตัวอย่างขนาด ต่าง ๆ

3. สุ่มนักเรียนจากแต่ละอำเภอ โดยวิธีสุ่มอย่างง่าย (Simple Random Sampling) ตามสัดส่วนของจำนวนนักเรียน เพื่อให้ได้นักเรียนกลุ่มตัวอย่างจำนวนที่ต้องการ ดังแสดงในตาราง 1

ตาราง 4 จำนวนกลุ่มตัวอย่างจำแนกนักเรียน ในแต่ละอำเภอของสำนักงานเขตพื้นที่การศึกษา กาญจนบุรี เขต 3

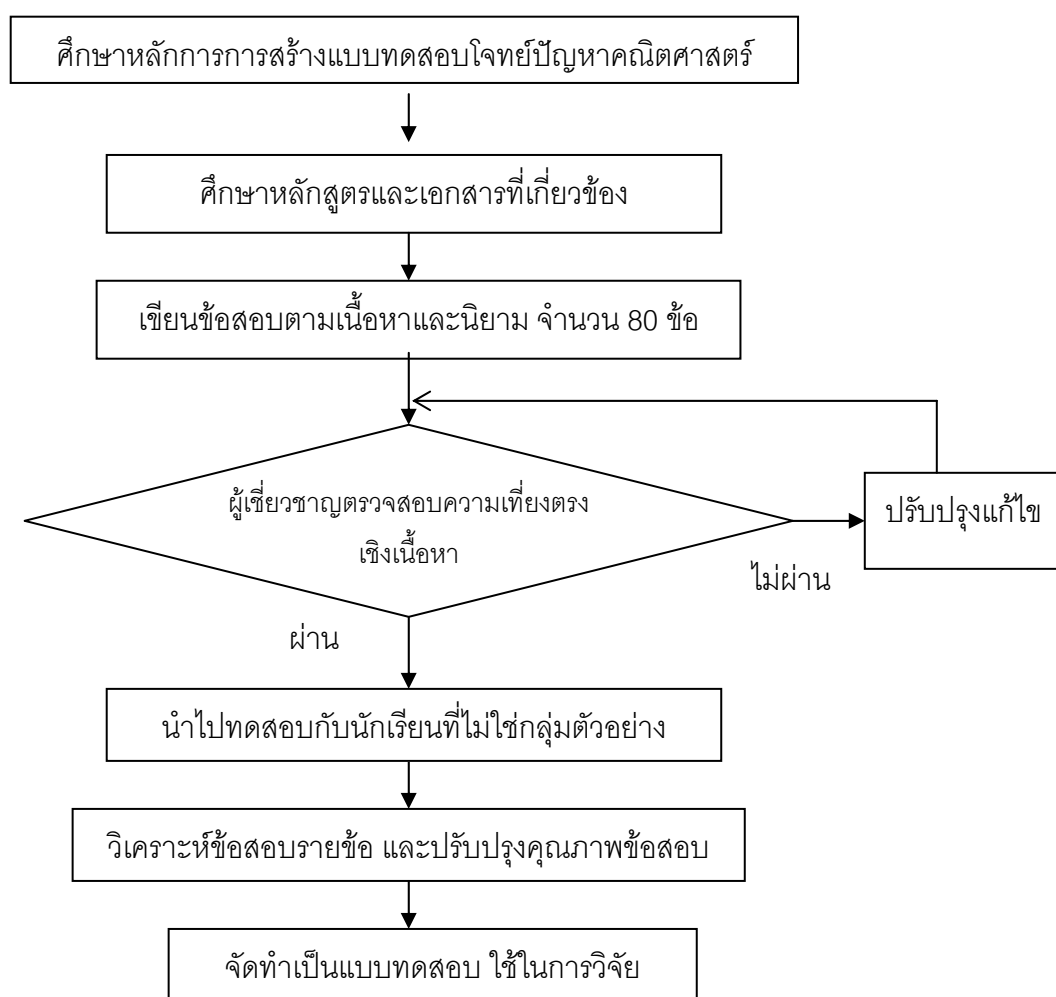
อำเภอ	นักเรียน (คน)		รวม (คน)
	มีเชื้อชาติไทย	ไม่ใช่เชื้อชาติไทย	
ทองผาภูมิ	229	330	559
ไทรโยค	283	131	414
ศรีสวัสดิ์	146	62	208
สังขละบุรี	122	230	352
รวม (คน)	780	753	1,533

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นแบบทดสอบเจตคติปัญหาในกลุ่มสาระการเรียนรู้วิชา คณิตศาสตร์ ตามหลักสูตรขั้นพื้นฐาน พุทธศักราช 2544 ในช่วงชั้นที่ 2 (ประถมศึกษาปีที่ 4 – 6) เป็นแบบเลือกตอบชนิด 4 ตัวเลือก ที่ผู้วิจัยสร้างขึ้นเอง

การสร้างเครื่องมือที่ใช้ในการศึกษา

วิธีดำเนินการสร้างแบบทดสอบคณิตศาสตร์ ช่วงชั้นที่ 2 ผู้วิจัยดำเนินการตามขั้นตอน ดังแสดงในภาพประกอบ 1



ภาพประกอบ 1 ขั้นตอนการสร้างเครื่องมือที่ใช้ในการวิจัย

จากแผนภูมิการดำเนินการสร้างเครื่องมือที่ใช้ในการวิจัย มีรายละเอียดของการดำเนินการสร้าง ดังนี้

1. กำหนดจุดมุ่งหมายในการสร้างแบบทดสอบคณิตศาสตร์ในการแก้ปัญหาของนักเรียนชั้นประถมศึกษาปีที่ 4-6 เพื่อศึกษาเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันตามเชื้อชาติของผู้สอบโดยวิธีวิเคราะห์ต่างกัน

2. ศึกษาทฤษฎีเกี่ยวกับการสร้างแบบทดสอบเลือกตอบที่ให้คะแนน 2 ค่า และการวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ จากเอกสารและงานวิจัยที่เกี่ยวข้อง แล้วทำความเข้าใจเพื่อเป็นแนวทางในการสร้างข้อสอบ

3. ศึกษาหลักสูตรขั้นพื้นฐาน พุทธศักราช 2544 คู่มือครูและแบบเรียนวิชาคณิตศาสตร์ เพื่อวิเคราะห์เนื้อหา ในการแก้ปัญหาของนักเรียนชั้นประถมศึกษาปีที่ 4-6

4. สร้างตารางวิเคราะห์เนื้อหาเกี่ยวกับโจทย์ปัญหาของนักเรียนชั้นประถมศึกษาปีที่ 4-6

ตาราง 5 การวิเคราะห์เนื้อหาเกี่ยวกับโจทย์ปัญหาของนักเรียนชั้นประถมศึกษาปีที่ 4-6

เนื้อหา	จำนวน(ข้อ)
การบวก การลบ การคูณ และการหารจำนวน	7
การบวก การลบ การคูณ และการหารระคนจำนวน	7
การบวกเศษส่วน การลบเศษส่วน การคูณเศษส่วน และการหารเศษส่วน	9
การบวก การลบ การคูณ และการหารเศษส่วนระคน	9
การบวก การลบ การคูณ และการหารทศนิยมไม่เกินสามตำแหน่ง	9
การบวก การลบ การคูณ และการหารทศนิยมระคนที่มีผลลัพธ์ไม่เกินสามตำแหน่ง	9
โจทย์ปัญหาร้อยละที่มีผลลัพธ์เป็นจำนวนนับ	6
พื้นที่และความยาวรอบรูปของรูปสี่เหลี่ยม	6
โจทย์ปัญหาเกี่ยวกับเงิน	6
โจทย์ปัญหาเกี่ยวกับเวลา	6
เมื่อกำหนดสมการเกี่ยวกับการบวก การลบ การคูณ หรือการหารที่มีตัวไม่ทราบค่าหนึ่งตัวให้ สามารถหาคำตอบและแสดงวิธีแก้สมการได้	6
รวม	80

5.เขียนข้อสอบ ตามเนื้อหาที่กำหนดไว้ ได้แบบทดสอบโจทย์ปัญหาในกลุ่มสาระการเรียนรู้วิชาคณิตศาสตร์ ตามหลักสูตรขั้นพื้นฐาน พุทธศักราช 2544 ในช่วงชั้นที่ 2 ชนิดเลือกตอบ 4 ตัวเลือก จำนวน 80 ข้อ จัดพิมพ์ข้อสอบที่สร้างแล้วนำไปให้ผู้เชี่ยวชาญซึ่งมีความรู้ในเนื้อหาและการเขียนข้อสอบวิชาคณิตศาสตร์และมีความรู้ทางการวัดผลการศึกษา จำนวน 5 ท่าน เป็นผู้ตรวจสอบความเที่ยงตรงเชิงเนื้อหาของแบบทดสอบรายข้อ

6.ปรับปรุงแก้ไขข้อสอบตามข้อเสนอแนะของผู้เชี่ยวชาญ และวิเคราะห์ดัชนีความสอดคล้องของความคิดเห็นของผู้เชี่ยวชาญเกี่ยวกับความสอดคล้องระหว่างเนื้อหาข้อสอบแต่ละข้อกับระดับพฤติกรรมการวัดที่กำหนดไว้ในตารางวิเคราะห์เนื้อหา ซึ่งมีสูตรดังนี้

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

เมื่อ IOC แทน ดัชนีความสอดคล้องของความคิดเห็นของผู้เชี่ยวชาญ
 $\sum R$ แทน ผลรวมของการตัดสินของผู้เชี่ยวชาญ
 N แทน จำนวนผู้เชี่ยวชาญ

โดยผู้เชี่ยวชาญพิจารณาว่า ข้อสอบวัดได้ตรงกับระดับพฤติกรรมที่กำหนด ตามเกณฑ์ ดังนี้

1 แทน แน่ใจ ข้อสอบวัดได้ ตรง กับระดับพฤติกรรมที่กำหนด
 0 แทน ไม่แน่ใจ ข้อสอบวัดได้ ตรง กับระดับพฤติกรรมที่กำหนด
 -1 แทน แน่ใจว่าข้อสอบวัดได้ ไม่ตรง กับระดับพฤติกรรมที่กำหนด

การแปลความหมายดัชนีความสอดคล้องที่คำนวณได้โดยถือว่า ถ้าค่า IOC มีค่าตั้งแต่ 0.5 ขึ้นไป แสดงว่า ข้อสอบวัดได้ตรงกับระดับพฤติกรรมที่กำหนด

7.คัดเลือกข้อสอบ จำนวน 70 ข้อ ที่มีค่าดัชนีความสอดคล้องของความคิดเห็นของผู้เชี่ยวชาญ (IOC) ตั้งแต่ 0.5 ขึ้นไป และสอดคล้องกับตารางกำหนดข้อสอบ

8. จัดพิมพ์และนำไปทดลองสอบกับนักเรียนชั้นประถมศึกษาปีที่ 6 ภาคเรียนที่ 2 ปีการศึกษา 2549 ประกอบด้วยโรงเรียนสมาคมป่าไม้แห่งประเทศไทยอุทิศ จำนวน 30 คน และโรงเรียนอนุบาลทองผาภูมิ จำนวน 70 คน ระหว่างวันที่ 7 – 8 ธันวาคม 2549 แล้วนำผลการสอบมาวิเคราะห์ข้อสอบหาคุณภาพของข้อสอบด้านความยากง่าย และค่าอำนาจจำแนก จากการคำนวณ ค่าความยากมีค่าตั้งแต่ 0.17 ถึง 0.84 และค่าอำนาจจำแนก มีค่าตั้งแต่ -0.02 ถึง 0.76 แล้วคัดเลือกข้อสอบที่มีค่าความยากง่ายระหว่าง .20 - .80 และค่าอำนาจจำแนกตั้งแต่ .20 ขึ้นไป คัดเลือกมาจำนวน 60 ข้อ ซึ่งมีค่าความยากง่ายอยู่ระหว่าง 0.21 ถึง 0.80 และค่าอำนาจจำแนกอยู่ระหว่าง 0.20 ถึง 0.76

9.นำข้อสอบที่ผ่านการวิเคราะห์และการปรับปรุงแล้วจากข้อ 8 จำนวน 60 ข้อ จัดพิมพ์เป็นแบบทดสอบ เพื่อนำไปใช้เป็นเครื่องมือในการเก็บรวบรวมข้อมูล โดยมีรายละเอียดของเนื้อหา ดังตาราง 6

ตาราง 6 เนื้อหาแบบทดสอบโจทย์ปัญหากลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตร
ขั้นพื้นฐาน พุทธศักราช 2544 ในช่วงชั้นที่ 2 (ประถมศึกษาปีที่ 4 – 6)

เนื้อหา	จำนวน(ข้อ)
การบวก การลบ การคูณ และการหารจำนวน	5
การบวก การลบ การคูณ และการหารระคนจำนวน	5
การบวกเศษส่วน การลบเศษส่วน การคูณเศษส่วน และการหารเศษส่วน	7
การบวก การลบ การคูณ และการหารเศษส่วนระคน	8
การบวก การลบ การคูณ และการหารทศนิยมไม่เกินสามตำแหน่ง	7
การบวก การลบ การคูณ และการหารทศนิยมระคนที่มีผลลัพธ์ไม่เกินสามตำแหน่ง	8
โจทย์ปัญหาร้อยละที่มีผลลัพธ์เป็นจำนวนนับ	4
พื้นที่และความยาวรอบรูปของรูปสี่เหลี่ยม	4
โจทย์ปัญหาเกี่ยวกับเงิน	4
โจทย์ปัญหาเกี่ยวกับเวลา	4
เมื่อกำหนดสมการเกี่ยวกับการบวก การลบ การคูณ หรือการหารที่มีตัวไม่ทราบค่า หนึ่งตัวให้ สามารถหาคำตอบและแสดงวิธีแก้สมการได้	4
รวม	60

ตัวอย่างข้อสอบคณิตศาสตร์

(0) มีน้ำหวาน $\frac{8}{9}$ ของขวด แบ่งให้น้อง $\frac{2}{9}$ ของขวด เหลือน้ำหวานในขวดเท่าไร ?

ก. $\frac{11}{9}$ ขวด

ข. $\frac{6}{9}$ ขวด

ค. $\frac{2}{3}$ ขวด

ง. $\frac{9}{6}$ ขวด

(คำตอบที่ถูกต้อง คือ ข้อ ข)

(00) คุณพ่อมีเงิน 5,000.75 บาท แบ่งให้คุณแม่ 2,000.25 บาท ที่เหลือแบ่งให้ลูก 2 คน
คนละเท่า ๆ กัน ลูกจะได้รับคนละเท่าไร ?

ก. 1,400.25 บาท

ข. 1,500.00 บาท

ค. 1,500.25 บาท

ง. 3,000.50 บาท

(คำตอบที่ถูกต้อง คือ ข้อ ค)

การตรวจให้คะแนน

การตรวจให้คะแนนของแบบทดสอบเลือกตอบวิชาคณิตศาสตร์ ชนิด 4 ตัวเลือก ใช้เกณฑ์การตรวจให้คะแนนดังนี้ คือ ถ้าตอบถูกให้ 1 คะแนน ตอบผิดหรือไม่ตอบ หรือตอบมากกว่า 1 แห่ง ในข้อนั้น ๆ ให้ 0 คะแนน

การเก็บรวบรวมข้อมูล

ผู้วิจัยได้ดำเนินการ ดังนี้

1. ขอหนังสือจากบัณฑิตวิทยาลัยมหาวิทยาลัยศรีนครินทรวิโรฒเพื่อขอความอนุเคราะห์ในการเก็บรวบรวมข้อมูล โดยทำหนังสือถึงผู้อำนวยการโรงเรียนที่เป็นกลุ่มตัวอย่าง แล้วนำไปติดต่อขอความร่วมมือกับโรงเรียนที่เป็นกลุ่มตัวอย่าง เพื่อนัดหมายและกำหนดวัน เวลา ในการนำแบบทดสอบ ไปทดสอบกับนักเรียนที่เป็นกลุ่มตัวอย่าง

2. จัดเตรียมแบบทดสอบให้เพียงพอกับจำนวนนักเรียนที่จะทำการสอบในแต่ละโรงเรียน

3. ดำเนินการสอบ โดยนำแบบทดสอบ ไปทำการทดสอบกับนักเรียนที่เป็นกลุ่มตัวอย่าง ในช่วง 5 กุมภาพันธ์ ถึง 16 กุมภาพันธ์ 2550

4. ชี้แจงให้นักเรียนที่เป็นกลุ่มตัวอย่างทราบวัตถุประสงค์ของการสอบ และขอความร่วมมือในการทำแบบทดสอบ เพื่อให้ได้ผลการสอบตรงตามความเป็นจริงมากที่สุด

5. นำแบบทดสอบไปทดสอบกับกลุ่มตัวอย่าง จำนวน 1,533 คน นำผลการตอบข้อสอบในแบบทดสอบ มาตรวจดูความเรียบร้อยแล้วลงรหัสผลการตอบข้อสอบทุกข้อ (ตอบถูกได้ 1 และตอบผิดได้ 0) แล้วคัดเลือกกระดาษคำตอบที่มีความสมบูรณ์ มาจำนวน 1,500 คน

6. นำข้อมูลที่ได้มาเตรียมข้อมูลในการศึกษา วิเคราะห์ทางสถิติและทดสอบสมมติฐานที่ตั้งไว้

การวิเคราะห์ข้อมูลและสถิติที่ใช้ในการวิเคราะห์ข้อมูล

1. การจัดเตรียมข้อมูลให้สอดคล้องกับวัตถุประสงค์ของการศึกษา เพื่อการตรวจสอบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และความสอดคล้องในการตรวจสอบทั้ง 3 วิธี มีขั้นตอนดังนี้

1.1 สุ่มข้อมูลจากกลุ่มตัวอย่างให้มีขนาด 600 และ 750 คน แยกเชื้อชาติไทยและไม่ใช่เชื้อชาติไทย

1.2 สุ่มข้อสอบ เพื่อจัดชุดแบบทดสอบใหม่ ขนาด 40 และ 60 ตามลำดับ โดยใช้การสุ่มอย่างง่าย (จับฉลากหมายเลขข้อที่เขียนไว้)

ได้กรอบหรือแบบแผนของการศึกษาในส่วนที่ดังนี้

I	N	วิธีตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน			
		MH	LC	CSA	CUA
40	600				
	750				
60	600				
	750				

I แทน จำนวนข้อสอบ

N แทน ขนาดกลุ่มตัวอย่าง

ได้แบบทดสอบที่ใช้ในการศึกษา 2 ชุด ดังนี้

แบบทดสอบชุดที่ 1 จำนวน 40 ข้อ เพื่อนำไปตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มผู้สอบ 600 คน และ 750 คน

แบบทดสอบชุดที่ 2 จำนวน 60 ข้อ เพื่อนำไปตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มผู้สอบ 600 คน และ 750 คน

2. ตรวจสอบความเป็นมิติเดียวของแบบทดสอบ ใช้การวิเคราะห์องค์ประกอบ (Factor Analysis) ด้วยการวิเคราะห์ประกอบหลัก (Principal Component Analysis) หมุนแกนด้วยวิธีแวนิแมกซ์ (Varimax) ตามข้อเสนอของลอร์ด สำหรับการตรวจสอบความเป็นมิติเดียวครั้งนี้ใช้โปรแกรมสำเร็จรูป SPSS FOR WINDOWS Version 11.5

3. การประมาณค่าพารามิเตอร์ของข้อสอบ ใช้โปรแกรม BILOG 3 (Mislevy and Bock, 1994) ซึ่งเป็นโปรแกรมที่สามารถใช้กับไมโครคอมพิวเตอร์ เพื่อประมาณค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าอำนาจจำแนก (a) และค่าความยาก (b) ตามแนวคิดทฤษฎีการตอบข้อสอบ (Item Response Theory) วิธีประมาณค่ามีทั้งมาร์จินัล แมกซิมัม ไคลลิสูด (Marginal Maximum Likelihood Method) และวิธีเบย์ (Bayesian Method) ข้อมูลที่ใช้เป็นแบบ 0 – 1 (Binary) แบ่งเป็นโปรแกรมย่อยที่สำคัญ ดังนี้

INPUT เป็นโปรแกรมย่อย ที่ใช้ในขั้นตอนการกำหนดลักษณะของข้อมูล การแปลงข้อมูลดิบและแบบจำลองที่ใช้ในการวิเคราะห์

CALIBRATE เป็นโปรแกรมย่อยที่ใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบสำหรับผลการวิเคราะห์ ประกอบด้วย 3 ส่วน (Phase) คือ

ส่วนที่ 1 เป็นผลการวิเคราะห์ข้อมูลที่เกี่ยวข้องกับค่าอำนาจจำแนก (r) และค่าความยาก (p) ตามแนวคิดทฤษฎีดั้งเดิม (Classical Test Theory)

ส่วนที่ 2 เป็นการประมาณค่าพารามิเตอร์ของข้อสอบ ตามแนวคิดทฤษฎีการตอบข้อสอบ ได้แก่ ค่าอำนาจจำแนก (a) และค่าความยาก (b) เมตริกซ์ของความแปรปรวน และความแปรปรวนร่วม รวมทั้งจะมีผลการทดสอบความเหมาะสมระหว่างแบบจำลองกับข้อมูลในการศึกษาด้วย ซึ่งการคำนวณเพื่อการเปรียบเทียบค่าพารามิเตอร์ของข้อสอบ จะต้องใช้ไฟล์พารามิเตอร์ของข้อสอบ และไฟล์เมตริกซ์ของความแปรปรวนและความแปรปรวนร่วม ในโปรแกรม EQUATE และโปรแกรม IRTDIF ต่อไป

ส่วนที่ 3 เป็นผลการวิเคราะห์เกี่ยวกับฟังก์ชันสารสนเทศของข้อสอบหรือแบบทดสอบ (Test Information Function)

4. การแปลงหน่วยของพารามิเตอร์ข้อสอบ (Matric Transformation) ก่อนที่จะนำมาเปรียบเทียบกัน นั่นคือ ต้องแปลงให้อยู่ในหน่วยเดียวกันหรือสเกลร่วม (Common Scale) จึงจะเปรียบเทียบกันได้ โดยใช้โปรแกรม EQUATE 2.1 (Baker. 1991) ซึ่งโปรแกรมนี้ได้ใช้แนวคิดของสต็อกกิง และลอร์ด (Stocking and Lord. 1983 : 201 – 210) ในการแสดงความสัมพันธ์ระหว่างเมตริกซ์ของค่าพารามิเตอร์ที่ประมาณค่าได้ระหว่างสองกลุ่ม เป็นเชิงเส้นตรง (Linear) ค่าที่ได้จากโปรแกรม EQUATE จะให้ค่า A และ K เพื่อนำไปใช้ในการแปลงหน่วยของค่าพารามิเตอร์ของข้อสอบในโปรแกรม IRTDIF ต่อไป

5. การตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน โดยวิธี Mantel – Haenzel ใช้โปรแกรม SIBTEST (Stout และ Rousos. 1992)

6. การตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน ใช้โปรแกรม IRTDIF version 1.0 พัฒนาโดยคิม และโคเฮน (Kim and Cohen. 1991) ซึ่งสามารถตรวจวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบได้ทั้ง 2 วิธี ได้แก่ วิธีตามแนวคิดของลอร์ด (Lord's χ^2) และวิธีตามแนวคิดของคิม และโคเฮน (Closed Interval Area)

7. การหาความสอดคล้องในการตรวจสอบระหว่างวิธีตรวจสอบ เป็นการวิเคราะห์ความสอดคล้องของผลการตัดสินการทำหน้าที่ต่างกันของข้อสอบ คำนวณได้จาก

สูตร

$$\text{ความสอดคล้อง} = \frac{R}{T} \times 100$$

เมื่อ

R แทน จำนวนข้อสอบที่ซ้ำกันในการตัดสิน

T แทน จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในการเปรียบเทียบ

แต่ละคู่

8. การทดสอบความแตกต่างของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี
ตรวจสอบ โดยใช้สถิติทดสอบ คือ Cochran Q Test

$$Q = \frac{(K - 1) \left[\left(K \sum_{j=1}^k G_j^2 \right) - \left(\sum_{j=1}^k G_j \right)^2 \right]}{K \sum_{i=1}^N L_i - \sum_{i=1}^N L_i^2} \quad (\text{Siegel. 1956 : 162})$$

เมื่อ	Q	แทน	การทดสอบแบบคิว
	K	แทน	จำนวนกลุ่มตัวอย่าง (จำนวนวิธีการตรวจสอบ)
	G_j	แทน	ผลรวมในแนวตั้ง (จำนวนข้อสอบที่ทำหน้าที่ต่างกันแต่ละวิธี)
	L_i	แทน	ผลรวมในแนวนอน

บทที่ 4


ผลการวิเคราะห์ข้อมูล

การนำเสนอผลการวิเคราะห์ข้อมูล ได้เรียงตามลำดับขั้นตอนดังนี้

- ตอนที่ 1 ค่าสถิติพื้นฐานของข้อมูลการสอบ คุณภาพของแบบทดสอบ การตรวจสอบความเป็นมิติเดียวของแบบทดสอบ และค่าพารามิเตอร์ของข้อสอบ
- ตอนที่ 2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่าง วิธี Mantel – Haenszel วิธี Lord's χ^2 และวิธี Closed Interval Area
- ตอนที่ 3 ผลการทดสอบความแตกต่างของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี Mantel – Haenszel วิธี Lord's χ^2 และวิธี Closed Interval Area แบ่งเป็น
- 3.1 ผลการทดสอบความแตกต่างของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเมื่อกลุ่มตัวอย่างเท่ากัน วิธีการตรวจสอบต่างกัน
- 3.2 ผลการทดสอบความแตกต่างของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเมื่อกลุ่มตัวอย่างต่างกัน วิธีการตรวจสอบเหมือนกัน
- ตอนที่ 4 ความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี Mantel – Haenszel วิธี Lord's χ^2 และวิธี Closed Interval Area

เพื่อให้การนำเสนอข้อมูลเข้าใจตรงกันและง่ายขึ้น จึงกำหนดสัญลักษณ์แทนดังนี้

N	แทน	จำนวนคน
I	แทน	จำนวนข้อสอบ
\bar{x}	แทน	ค่าเฉลี่ยเลขคณิต
S.D.	แทน	ค่าส่วนเบี่ยงเบนมาตรฐาน
KR.20	แทน	ความเชื่อมั่นของแบบทดสอบ
SE	แทน	ค่าความคลาดเคลื่อนมาตรฐาน
a_t	แทน	ค่าอำนาจจำแนกของข้อสอบรวม
b_t	แทน	ค่าความยากของข้อสอบรวม
a_1	แทน	ค่าอำนาจจำแนกของข้อสอบ กลุ่มเชื้อชาติไทย
b_1	แทน	ค่าความยากของข้อสอบ กลุ่มเชื้อชาติไทย

a_2	แทน	ค่าอำนาจจำแนกของข้อสอบ กลุ่มไม่ใช่เชื้อชาติไทย
b_2	แทน	ค่าความยากของข้อสอบ กลุ่มไม่ใช่เชื้อชาติไทย
MH	แทน	การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธีMantel-Haenszel
LC	แทน	วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของLord's χ^2
CSA	แทน	วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของคิม และโคเฮน โดยคิดเครื่องหมาย เรียกว่า “วิธีClosed - Interval Signed Area”
CUA	แทน	วิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของคิมและโคเฮน โดยไม่คิดเครื่องหมาย เรียกว่า “วิธีClosed - Interval Unsigned Area”
*	แทน	$p < .05$
**	แทน	$p < .01$
***	แทน	$p < .001$
	แทน	ข้อสอบที่ตรวจสอบพบว่าทำหน้าที่ต่างกันตรงกัน

ตอนที่ 1 ค่าสถิติพื้นฐานของข้อมูลผลการสอบ คุณภาพของแบบทดสอบ การตรวจสอบความเป็นมิติเดียวของแบบทดสอบ และค่าพารามิเตอร์ของข้อสอบ

1) ค่าสถิติพื้นฐานของข้อมูลผลการตอบข้อสอบ และผลการวิเคราะห์คุณภาพของแบบทดสอบ ได้แก่ ค่าเฉลี่ย (\bar{x}) ค่าส่วนเบี่ยงเบนมาตรฐาน (S.D) ค่าความเชื่อมั่นของแบบทดสอบ (KR-20) และค่าความคลาดเคลื่อนมาตรฐาน (SE) ดังตาราง 7

ตาราง 7 ค่าเฉลี่ย ค่าส่วนเบี่ยงเบนมาตรฐาน ค่าความเชื่อมั่นและค่าความคลาดเคลื่อนมาตรฐานของแบบทดสอบวิชาคณิตศาสตร์ จำนวน 60 ข้อ

กลุ่ม	N (คน)	\bar{x}	S.D.	KR-20	SE
เชื้อชาติไทย	750	48.61	6.15	.812	2.67
ไม่ใช่เชื้อชาติไทย	750	35.56	8.02	.808	3.51
รวม	1,500	42.08	9.58	.894	3.12

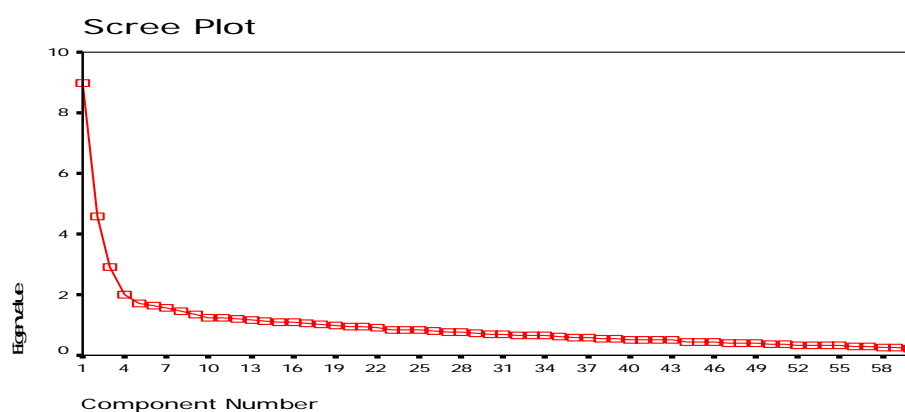
จากตาราง 7 พบว่า กลุ่มเชื้อชาติไทย 750 คน กลุ่มไม่ใช่เชื้อชาติไทย 750 คน ค่าเฉลี่ยของคะแนนสอบกลุ่มเชื้อชาติไทยเท่ากับ 48.61 คะแนน กลุ่มไม่ใช่เชื้อชาติไทยเท่ากับ 35.56 คะแนน ค่าความเชื่อมั่นของแบบทดสอบ โดยใช้สูตรคูเดอริชาร์ดสัน (KR-20) เท่ากับ .894 และค่าความคลาดเคลื่อนมาตรฐาน(SE) เท่ากับ 3.12

2) การตรวจสอบความเป็นมิติเดียวของแบบทดสอบ

ผู้วิจัยได้นำคะแนนจากการสอบ มาตรวจสอบความเป็นมิติเดียวกันของแบบทดสอบโดยใช้การวิเคราะห์องค์ประกอบ (Factor Analysis) ด้วยการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis) หมุนแกนด้วยวิธีแวนิแมกซ์ (Varimax) โดยใช้โปรแกรมสำเร็จรูป SPSS For Windows Version 11.5 และหาค่าไอเกน แล้วนำค่าไอเกนมาลงจุดกราฟ แสดงผลดังตาราง 8 และภาพประกอบ 2

ตาราง 8 ค่าไอเกนและร้อยละความแปรปรวนขององค์ประกอบ

องค์ประกอบ	ค่าไอเกน	ร้อยละความแปรปรวน
1	8.973	14.954
2	4.583	7.638
3	2.912	4.854
4	1.997	3.328
5	1.691	2.818
6	1.625	2.708



ภาพประกอบ 2 ผลการตรวจสอบความเป็นมิติเดียวกันของข้อสอบ

ผลการวิเคราะห์ตาราง 8 และภาพประกอบ 2 พบว่า ค่าไอเกนขององค์ประกอบที่ 1 มีค่าสูงกว่าไอเกนขององค์ประกอบที่ 2 ประมาณ 1.96 เท่า ส่วนเส้นกราฟจากองค์ประกอบที่ 1 มา 2 จะลาดชัน และตั้งแต่องค์ประกอบที่ 2 เป็นต้นไปจะใกล้เคียงกัน นับว่าองค์ประกอบที่ 1 มีค่าเด่นกว่าองค์ประกอบอื่น แม้ว่าค่าร้อยละของความแปรปรวนจะไม่สูงนัก (ร้อยละ 14.954) หากพิจารณาตามเกณฑ์ที่เรคเคส (Reckase. อ้างถึงใน Raju. 1993 : 304) เสนอค่าร้อยละของความแปรปรวนควรมีค่าไม่น้อยกว่าร้อยละ 20 จึงจะถือว่าแบบทดสอบมีความมิติเดียว ซึ่งในทางปฏิบัติหาพบได้ยากมาก เนื่องจากแบบทดสอบประกอบด้วยเนื้อหาหลายอย่างและวัดหลายสมรรถภาพ

3) ค่าพารามิเตอร์ของข้อสอบ (a,b) แยกตามกลุ่มเชื้อชาติไทย – ไม่ใช่เชื้อชาติไทย

ผู้วิจัยได้นำคะแนนจากการสอบ มาประมาณค่าพารามิเตอร์ ใช้โปรแกรม BILOG 3 (Mislevy and Bock. 1994) เพื่อประมาณค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าอำนาจจำแนก (a) และค่าความยาก (b) ตามแนวคิดทฤษฎีการตอบข้อสอบ (Item Response Theory) แสดงผลดังตาราง 9

ตาราง 9 ค่าพารามิเตอร์ของข้อสอบ แยกตามกลุ่มเชื้อชาติไทย – ไม่ใช่เชื้อชาติไทย

ข้อที่	พารามิเตอร์ข้อสอบ					
	a_1	b_1	a_2	b_2	a_t	b_t
1	0.804	3.652	0.378	5.474	0.651	3.789
2	0.277	-5.240	0.272	-5.188	0.125	-5.373
3	1.359	-1.980	0.251	8.401	0.586	4.448
4	1.081	2.787	0.264	5.988	0.756	2.872
5	0.747	3.931	0.330	5.192	0.904	2.726
6	1.434	-1.474	1.434	-0.414	-0.64	0.545
7	0.615	5.074	0.373	5.187	0.570	4.183
8	0.792	3.937	0.204	6.596	0.502	3.724
9	0.434	5.139	0.243	4.738	0.585	2.843
10	2.575	-1.274	0.127	-5.000	-0.77	0.592
11	0.228	6.283	0.484	0.647	0.441	1.755
12	0.072	-4.932	0.478	5.087	0.203	5.101

ตาราง 9 (ต่อ)

ข้อที่	พารามิเตอร์ข้อสอบ					
	a_1	b_1	a_2	b_2	a_t	b_t
13	0.400	-1.923	0.086	-0.848	0.079	-4.598
14	0.894	2.499	0.261	3.797	0.691	2.200
15	0.153	-5.999	0.202	-5.386	0.093	-3.614
16	0.608	3.656	0.806	1.551	0.636	2.401
17	0.940	2.951	0.263	-4.578	0.089	-6.625
18	0.360	5.304	0.193	7.406	0.568	3.356
19	0.399	-1.354	0.399	-0.174	-0.23	0.392
20	0.581	1.389	0.366	4.006	0.334	4.905
21	0.467	2.682	0.196	4.751	0.213	3.772
22	0.669	1.950	0.221	4.523	0.267	4.066
23	0.593	2.972	0.419	2.103	0.520	2.036
24	0.460	2.952	0.197	5.753	0.354	3.873
25	0.732	2.226	0.734	1.632	0.392	2.955
26	0.572	2.700	0.186	3.490	0.455	2.241
27	0.307	5.697	0.217	5.840	0.353	3.870
28	0.587	2.829	0.211	4.065	0.390	3.175
29	1.185	1.736	0.227	5.814	0.403	3.634
30	0.466	4.894	0.154	4.479	0.572	1.947
31	0.423	2.981	0.388	4.816	0.398	5.101
32	0.121	6.544	0.298	3.487	0.380	3.022
33	0.442	3.656	0.228	2.263	0.109	5.827
34	0.660	2.805	0.353	3.217	0.369	3.605
35	0.187	7.487	0.465	2.582	0.523	2.774
36	0.252	4.476	0.336	3.657	0.202	6.366
37	0.444	2.227	0.201	7.001	0.145	8.516
38	0.610	3.167	0.157	7.661	0.166	6.282

ตาราง 9 (ต่อ)

ข้อที่	พารามิเตอร์ข้อสอบ					
	a_1	b_1	a_2	b_2	a_t	b_t
39	0.395	4.307	0.318	5.111	0.417	4.156
40	0.313	5.270	0.341	2.345	0.450	2.616
41	0.619	1.858	0.268	3.667	0.239	5.151
42	0.969	2.096	0.194	6.299	0.261	4.240
43	0.155	6.180	0.324	3.919	0.533	2.827
44	0.203	4.828	0.243	4.859	0.120	8.731
45	0.192	8.173	0.325	3.298	0.192	5.236
46	0.195	6.223	0.556	3.278	0.184	8.685
47	0.258	5.118	1.109	2.052	0.156	8.720
48	0.503	5.845	0.214	4.513	0.247	4.572
49	0.754	2.993	0.263	6.759	0.483	4.590
50	0.122	-1.154	0.215	4.840	0.708	2.284
51	0.988	3.023	0.322	-2.780	0.074	-4.361
52	0.766	3.650	0.226	5.340	0.820	2.366
53	1.885	-1.746	1.885	-0.703	0.483	4.711
54	0.371	3.985	0.259	5.905	0.536	3.704
55	0.132	-8.923	0.350	2.536	0.495	2.417
56	0.860	3.483	0.180	-1.849	0.061	-1.020
57	0.854	3.379	0.317	2.004	0.746	2.580
58	1.140	2.916	0.171	3.699	0.531	3.960
59	1.203	3.239	0.238	6.696	0.603	3.506
60	1.660	-1.739	0.175	7.062	0.579	3.199
\bar{x}	0.542	0.536	0.305	0.645	0.402	0.214
S.D	0.299	0.287	0.171	0.324	0.404	0.541

จากตาราง 9 ค่าความยากของข้อสอบเฉลี่ยเท่ากับ 0.214 และค่าอำนาจจำแนกเฉลี่ยเท่ากับ 0.402 ซึ่งค่าความยากของข้อสอบ อยู่ในระดับปานกลาง และค่าอำนาจจำแนกอยู่ในเกณฑ์ดี และค่าความยากของกลุ่มไม่ใช่เชื้อชาติไทยสูงกว่ากลุ่มเชื้อชาติไทย

ตอนที่ 2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธี Mantel – Haenszel วิธี Lord's χ^2 และวิธี Closed Interval Area

การวิเคราะห์ตอนที่ 2 ผู้วิจัยได้นำผลที่ได้จากการสอบมาสุ่มกลุ่มตัวอย่างเป็นสองขนาดคือ ขนาด 600 คน และ 750 คน แยกตามกลุ่มเชื้อชาติไทย - ไม่ใช่เชื้อชาติไทย แล้วสุ่มข้อสอบเพื่อจัดชุดแบบทดสอบใหม่ โดยใช้การสุ่มอย่างง่าย (จับฉลากหมายเลขข้อที่เขียนไว้) ได้แบบทดสอบสองชุด คือ แบบทดสอบที่มีจำนวน 40 ข้อ และ 60 ข้อ หลังจากการสุ่มจะได้กลุ่มที่จะทำการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 4 กลุ่ม ได้แก่ กลุ่มที่หนึ่งแบบทดสอบ 40 ข้อ กลุ่มตัวอย่าง 600 คน กลุ่มที่สองแบบทดสอบ 40 ข้อ กลุ่มตัวอย่าง 750 คน กลุ่มที่สามแบบทดสอบ 60 ข้อ กลุ่มตัวอย่าง 600 คน กลุ่มที่สี่แบบทดสอบ 60 ข้อ กลุ่มตัวอย่าง 750 คน จากนั้นทำการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยวิธี Mantel – Haenszel ใช้โปรแกรม SIBTEST ส่วนวิธี Lord's χ^2 และวิธี Closed Interval Area ใช้โปรแกรม IRTDIF version 1.0 แล้วนำเสนอผลการวิเคราะห์ข้อมูลเฉพาะข้อที่ตรวจสอบพบว่าทำหน้าที่ต่างกันเท่านั้น แสดงผลดังตาราง 10 - 11

ตาราง 10 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธี Mantel – Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA) เมื่อจำนวนข้อสอบ 40 ข้อ กลุ่มตัวอย่าง 600 คน และ 750 คน

ข้อที่	600 คน				750 คน			
	MH	LC	CSA	CUA	MH	LC	CSA	CUA
3	1.65 [*]				1.66 [*]			
4	1.73 [*]				1.83 [*]			
5	2.70 ^{***}	19.86 ^{***}	- 0.517	0.624	2.93 ^{***}	33.52 ^{***}	0.425	0.446
12	3.45 ^{***}	19.57 ^{***}	- 0.519	0.524	1.83 ^{**}	12.82 ^{**}	0.385	0.410
13	0.47 [*]	9.99 ^{**}			3.18 ^{***}	15.86 ^{**}	0.421	0.429
14						6.08 [*]		

ตาราง 10 (ต่อ)

ข้อที่	600 คน				750 คน			
	MH	LC	CSA	CUA	MH	LC	CSA	CUA
15								0.538
32					0.68*	6.83*		
33	0.36**	10.11**	-0.335		0.59**	6.02*		
34	1.73*	8.22*	0.317	0.428	1.83**	12.35**	0.579	0.585
35	1.66*	6.88*	0.407	0.483	1.75**	9.42**	0.375	0.476
36	1.83**	7.48*		0.463	1.66*	7.25*		0.503
38	0.53**	12.98**	-0.448	0.526	2.15***	12.51**	-0.375	0.456
39	0.52**	9.73**	-0.549	0.549	2.93***	16.04***	0.362	0.506
40	1.59*	8.72*	-0.545	0.595	2.69**	10.51**	0.437	0.528
รวม	12	10	7	9	13	12	8	10

ผลการวิเคราะห์ตามตาราง 10 พบว่า เมื่อกลุ่มตัวอย่าง 600 คน วิธี Mantel-Haenszel (MH) ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 12 ข้อ วิธี Lord's χ^2 (LC) ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 10 ข้อ วิธี Closed - Interval Signed Area (CSA) ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 7 ข้อ และวิธี Closed - Interval Unsigned Area (CUA) ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 9 ข้อ โดยทั้งสี่วิธีตรวจพบข้อที่ทำหน้าที่ต่างกัน 7 ข้อ ได้แก่ ข้อ 5 , 12 , 34 , 35 , 38 , 39 , 40 ตามลำดับ ข้อที่ 13 กับ 36 ตรวจพบด้วยวิธี Mantel-Haenszel วิธี Lord's χ^2 (LC) และวิธี Closed - Interval Unsigned Area (CUA) ตรงกัน ข้อที่ 33 ตรวจพบด้วยวิธี Mantel-Haenszel (MH) วิธี Lord's (LC) และวิธี Closed - Interval Signed Area (CSA) ตรงกัน ส่วนข้อที่เหลือคือ ข้อ 3 กับ 4 วิธี Mantel-Haenszel (MH) ตรวจพบ

เมื่อกลุ่มตัวอย่าง 750 คน วิธี Mantel-Haenszel (MH) ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 13 ข้อ วิธี Lord's χ^2 (LC) ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 12 ข้อ วิธี Closed - Interval Signed Area (CSA) ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 8 ข้อ และวิธี Closed - Interval Unsigned Area (CUA) ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 10 ข้อ โดยทั้งสี่วิธีตรวจพบข้อที่ทำหน้าที่ต่างกัน 8 ข้อ ได้แก่ ข้อ 5 , 12 , 13 , 34 , 35 , 38 , 39 , 40 ตามลำดับ ข้อที่ 36 ตรวจพบด้วยวิธี Mantel-Haenszel (MH) วิธี Lord's χ^2 (LC) และวิธี Closed - Interval Unsigned Area (CUA) ตรงกัน ข้อที่ 32 , 33 ตรวจพบด้วยวิธี Mantel-Haenszel (MH) และ วิธี

Lord's χ^2 (LC) ตรงกัน ข้อที่ 3 กับ 4 ตรวจพบด้วยวิธี Mantel-Haenszel (MH) ข้อที่ 14 ตรวจพบด้วยวิธี Lord's χ^2 (LC) และข้อที่ 15 ตรวจพบด้วยวิธี Closed - Interval Unsigned Area (CUA)

ตาราง 11 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธี Mantel-Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA) เมื่อจำนวนข้อสอบ 60 ข้อ กลุ่มตัวอย่าง 600 คน และ 750 คน

ข้อที่	600 คน				750 คน			
	MH	LC	CSA	CUA	MH	LC	CSA	CUA
4	0.63 [*]				2.93 ^{***}	10.22 ^{**}		0.562
5	0.34 [*]				0.68 [*]	6.29 [*]		
6	3.18 ^{**}	14.86 ^{***}	-0.340	0.667	1.83 ^{**}	10.54 ^{**}	-0.312	0.483
8	1.75 ^{**}	10.40 ^{**}	-0.359	0.411	1.83 ^{**}	10.19 ^{**}	-0.406	0.445
9					0.40 [*]	6.69 [*]		
10					0.47 [*]	7.99 [*]		
12	4.32 ^{***}	14.39 ^{***}	0.401	0.401	0.59 ^{**}	11.99 ^{**}	-0.418	0.457
17	2.69 ^{**}	16.12 ^{***}	-0.355	0.420	0.38 ^{***}	13.94 ^{***}	-0.437	0.526
18	0.47 [*]	7.62 [*]		0.576	1.75 ^{**}	11.19 ^{**}	-0.36	0.412
44		6.03 [*]	0.421	0.429		9.29 ^{**}	0.375	0.456
49	2.62 ^{***}	17.79 ^{***}	-0.586	0.724	2.65 ^{***}	14.30 ^{**}	-0.346	0.401
50	2.93 ^{***}				2.15 ^{***}	8.23 ^{**}		
51	1.83 ^{**}				0.57 ^{**}	6.08 [*]		
52	0.53 ^{**}	10.51 ^{**}	-0.345	0.407	0.38 ^{***}	15.43 ^{**}	-0.424	0.424
53	0.57 ^{**}	10.37 ^{**}	-0.419	0.507	0.36 ^{**}	10.48 ^{**}	-0.348	0.411
54					2.69 ^{**}	9.04 [*]		0.434
55	1.75 ^{**}	6.18 [*]						
57	2.69 ^{**}	9.07 [*]		0.465	1.83 ^{**}	11.19 ^{**}	-0.330	0.403
58	2.62 ^{***}	16.12 ^{***}	0.579	0.585	2.65 ^{**}	13.94 ^{**}	-0.402	0.451
59	0.31 ^{***}	16.81 ^{***}	0.385	0.410	0.45 ^{***}	26.11 ^{***}	-0.359	-0.420
60	3.38 ^{***}	17.59 ^{***}	-0.406	0.445	2.15 ^{***}	23.17 ^{***}	-0.419	0.507
รวม	17	14	11	13	19	20	13	15

ผลการวิเคราะห์ตามตาราง 11 พบว่า เมื่อกลุ่มตัวอย่าง 600 คน วิธี Mantel-Haenszel (MH) ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 17 ข้อ วิธี Lord's χ^2 (LC) ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 14 ข้อ วิธี Closed - Interval Signed Area (CSA) ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 11 ข้อ และวิธี Closed - Interval Unsigned Area (CUA) ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 13 ข้อ โดยทั้งสี่วิธีตรวจพบข้อที่ทำหน้าที่ต่างกัน 10 ข้อ ได้แก่ ข้อ 6 , 8 , 12 , 17 , 49 , 52 , 53 , 58 , 59 , 60 ตามลำดับ ส่วนข้อที่ 18 กับ 57 ตรวจพบด้วยวิธี Mantel-Haenszel (MH) วิธี Lord's χ^2 (LC) และวิธี Closed - Interval Unsigned Area (CUA) ตรงกัน ข้อที่ 44 ตรวจพบด้วยวิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA) ตรงกัน ข้อที่ 55 ตรวจพบด้วยวิธี Mantel-Haenszel (MH) วิธี Lord's χ^2 (LC) ตรงกัน ส่วนข้อที่เหลือ คือ ข้อที่ 4 , 5 , 50 , 51 Mantel-Haenszel (MH) ตรวจพบ

เมื่อกลุ่มตัวอย่าง 750 คน วิธี Mantel-Haenszel (MH) ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 19 ข้อ วิธี Lord's χ^2 (LC) ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 20 ข้อ วิธี Closed - Interval Signed Area (CSA) ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 13 ข้อ และวิธี Closed - Interval Unsigned Area (CUA) ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 15 ข้อ โดยทั้งสี่วิธีตรวจพบข้อที่ทำหน้าที่ต่างกัน 12 ข้อ ได้แก่ ข้อ 6 , 8 , 12 , 17 18 , 49 , 52 , 53 , 57 , 58 , 59 , 60 ตามลำดับ ข้อที่ 4 , 54 ตรวจพบด้วยวิธี Mantel-Haenszel (MH) วิธี Lord's χ^2 (LC) และวิธี Closed - Interval Unsigned Area (CUA) ตรงกัน ข้อที่ 44 ตรวจพบด้วย วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA) ตรงกัน ส่วนข้อที่เหลือ คือข้อที่ 5 , 9 , 10 , 50 , 51 , 56 ตรวจพบด้วยวิธี Mantel - Haenszel (MH) วิธี Lord's χ^2 (LC) ตรงกัน

ตอนที่ 3 ผลการทดสอบความแตกต่างของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ระหว่างวิธี Mantel - Haenszel วิธี Lord's χ^2 และวิธี Closed Interval Area

การวิเคราะห์ตอนที่ 3 ผู้วิจัยได้นำผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี Mantel - Haenszel วิธี Lord's χ^2 และวิธี Closed Interval Area ที่ตรวจพบในตอนที่ 2 มาทดสอบความแตกต่าง โดยใช้สถิติทดสอบ คือ Cochran Q - Test (Siegel. 1956 : 162) วิเคราะห์โดยใช้โปรแกรมสำเร็จรูป Microsoft Excel 2003 โดยแบ่งการทดสอบเป็นสองกรณีตามสมมติฐานการวิจัยตามลำดับดังนี้

3.1 ผลการทดสอบความแตกต่างของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อกลุ่มตัวอย่างเท่ากัน วิธีการตรวจสอบต่างกัน แสดงผลดังตาราง 12 – 19

ตาราง 12 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 40 ข้อ กลุ่มตัวอย่าง 600 คน ระหว่างวิธี Mantel – Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed – Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA)

ข้อสอบที่ถูกตรวจพบว่า (ข้อ)	MH	LC	CSA	CUA	Q – Test
ทำหน้าที่ต่างกัน	12	10	7	9	7.00*
ไม่ทำหน้าที่ต่างกัน	28	30	33	31	

ผลการวิเคราะห์ตามตาราง 12 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Mantel – Haenszel (MH) ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 12 ข้อ ไม่ทำหน้าที่ต่างกัน 28 ข้อ วิธี Lord's χ^2 (LC) ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 10 ข้อ ไม่ทำหน้าที่ต่างกัน 30 ข้อ วิธี Closed - Interval Signed Area (CSA) ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 7 ข้อ ไม่ทำหน้าที่ต่างกัน 33 ข้อ และวิธี Closed - Interval Unsigned Area (CUA) ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 9 ข้อ ไม่ทำหน้าที่ต่างกัน 31 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วย Q – Test พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบแตกต่างกัน และได้ทำการเปรียบเทียบความแตกต่างของจำนวนข้อสอบที่ตรวจพบเป็นรายคู่ แสดงผลดังตาราง 13

ตาราง 13 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่ เมื่อจำนวนข้อสอบ 40 ข้อ กลุ่มตัวอย่าง 600 คน ระหว่างวิธี Mantel – Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA)

วิธีตรวจสอบ	การทดสอบแบบคิวเป็นรายคู่ (Q – Test)			
	MH	LC	CSA	CUA
MH	-	2.00	4.00*	3.00
LC		-	3.00	1.00
CSA			-	2.00
CUA				-

การวิเคราะห์ตามตาราง 13 พบว่า เมื่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ จากข้อสอบ 40 ข้อ กลุ่มตัวอย่าง 600 คน มาทดสอบความแตกต่างเป็นรายคู่ (Q – Test) พบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบโดยวิธี Mantel-Haenszel (MH) กับ วิธี Closed - Interval Signed Area (CSA) แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนค่าคู่ที่เหลือคือ วิธี Mantel-Haenszel (MH) กับ วิธี Lord's χ^2 (LC) ,วิธี Mantel-Haenszel (MH) กับ วิธี Closed - Interval Unsigned Area (CUA) , วิธี Lord's χ^2 (LC) กับ วิธี Closed - Interval Signed Area(CSA) , วิธี Lord's χ^2 (LC) กับ วิธี Closed - Interval Unsigned Area (CUA) และวิธี Closed - Interval Signed Area(CSA) กับ วิธี Closed - Interval Unsigned Area (CUA) แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

ตาราง 14 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 40 ข้อ กลุ่มตัวอย่าง 750 คน ระหว่างวิธี Mantel-Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed – Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA)

ข้อสอบที่ถูกตรวจพบว่า (ข้อ)	MH	LC	CSA	CUA	Q – Test
ทำหน้าที่ต่างกัน	13	12	8	10	8.00*
ไม่ทำหน้าที่ต่างกัน	27	28	32	30	

ผลการวิเคราะห์ตามตาราง 14 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Mantel-Haenszel (MH) ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 13 ข้อ ไม่ทำหน้าที่ต่างกัน 27 ข้อ วิธี Lord's χ^2 (LC) ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 12 ข้อ ไม่ทำหน้าที่ต่างกัน 28 ข้อ วิธี Closed - Interval Signed Area (CSA) ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 8 ข้อ ไม่ทำหน้าที่ต่างกัน 32 ข้อ และวิธี Closed - Interval Unsigned Area (CUA) ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 10 ข้อ ไม่ทำหน้าที่ต่างกัน 30 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วย Q - Test พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบแตกต่างกัน และได้ทำการเปรียบเทียบความแตกต่างของจำนวนข้อสอบที่ตรวจพบเป็นรายคู่ แสดงผลดังตาราง 15

ตาราง 15 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่ เมื่อจำนวนข้อสอบ 40 ข้อ กลุ่มตัวอย่าง 750 คน ระหว่างวิธี Mantel-Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA)

วิธีตรวจสอบ	การทดสอบแบบคิดเป็นรายคู่ (Q - Test)			
	MH	LC	CSA	CUA
MH	-	0.33	5.00*	1.80
LC		-	4.00*	1.00
CSA			-	2.00
CUA				-

การวิเคราะห์ตามตาราง 15 พบว่า เมื่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ จากข้อสอบ 40 ข้อ กลุ่มตัวอย่าง 750 คน มาทดสอบความแตกต่างเป็นรายคู่ (Q - Test) พบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบโดยวิธี Mantel-Haenszel (MH) กับ วิธี Closed - Interval Signed Area(CSA) และ วิธี Lord's χ^2 (LC) กับ วิธี Closed - Interval Signed Area(CSA) แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนคู่ที่เหลือคือ วิธี Mantel-Haenszel (MH) กับ วิธี Lord's χ^2 (LC) ,วิธี Mantel-Haenszel (MH) กับ วิธี Closed - Interval Unsigned Area (CUA) , วิธี Lord's χ^2 (LC) กับ วิธี Closed - Interval Unsigned Area (CUA) และวิธี Closed - Interval Signed Area(CSA) กับ วิธี Closed - Interval Unsigned Area (CUA) แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

ตาราง 16 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 60 ข้อ กลุ่มตัวอย่าง 600 คน ระหว่างวิธี Mantel-Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA)

ข้อสอบที่ถูกตรวจพบว่า (ข้อ)	MH	LC	CSA	CUA	Q - Test
ทำหน้าที่ต่างกัน	17	14	11	13	9.00*
ไม่ทำหน้าที่ต่างกัน	43	46	49	47	

ผลการวิเคราะห์ตามตาราง 16 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Mantel-Haenszel (MH) ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 17 ข้อ ไม่ทำหน้าที่ต่างกัน 43 ข้อ วิธี Lord's χ^2 (LC) ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 14 ข้อ ไม่ทำหน้าที่ต่างกัน 46 ข้อ วิธี Closed - Interval Signed Area (CSA) ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 11 ข้อ ไม่ทำหน้าที่ต่างกัน 49 ข้อ และวิธี Closed - Interval Unsigned Area (CUA) ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 13 ข้อ ไม่ทำหน้าที่ต่างกัน 47 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วย Q - Test พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบแตกต่างกัน และได้ทำการเปรียบเทียบความแตกต่างของจำนวนข้อสอบที่ตรวจพบเป็นรายคู่ แสดงผลดังตาราง 17

ตาราง 17 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่ เมื่อจำนวนข้อสอบ 60 ข้อ กลุ่มตัวอย่าง 600 คน ระหว่างวิธี Mantel-Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA)

วิธีตรวจสอบ	การทดสอบแบบคิเป็นรายคู่ (Q - Test)			
	MH	LC	CSA	CUA
MH	-	1.80	9.31*	7.36*
LC		-	6.00*	4.00*
CSA			-	2.00
CUA				-

การวิเคราะห์ตามตาราง 17 พบว่า เมื่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ จากข้อสอบ 60 ข้อ กลุ่มตัวอย่าง 600 คน มาทดสอบความแตกต่างเป็นรายคู่ (Q – Test) พบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบโดยวิธี Mantel-Haenszel (MH) กับ วิธีClosed - Interval Signed Area(CSA) , วิธี Mantel-Haenszel (MH) กับ วิธีClosed - Interval Unsigned Area (CUA) , วิธี Lord's χ^2 (LC) กับ วิธีClosed - Interval Signed Area (CSA) , วิธี Lord's χ^2 (LC) กับ วิธีClosed - Interval Unsigned Area (CUA) แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนสองคู่ที่เหลือคือ วิธี Mantel-Haenszel (MH) กับ วิธี Lord's χ^2 (LC) และวิธีClosed - Interval Signed Area (CSA) กับ วิธีClosed - Interval Unsigned Area (CUA) แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

ตาราง 18 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 60 ข้อ กลุ่มตัวอย่าง 750 คน ระหว่างวิธี Mantel-Haenszel (MH) วิธี Lord's χ^2 (LC) วิธีClosed - Interval Signed Area (CSA) และวิธีClosed - Interval Unsigned Area (CUA)

ข้อสอบที่ถูกตรวจพบว่า (ข้อ)	MH	LC	CSA	CUA	Q – Test
ทำหน้าที่ต่างกัน	19	20	13	15	13.55 [*]
ไม่ทำหน้าที่ต่างกัน	41	40	47	45	

ผลการวิเคราะห์ตามตาราง 18 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Mantel-Haenszel (MH) ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 19 ข้อ ไม่ทำหน้าที่ต่างกัน 41 ข้อ วิธี Lord's χ^2 (LC) ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 20 ข้อ ไม่ทำหน้าที่ต่างกัน 40 ข้อ วิธีClosed - Interval Signed Area (CSA) ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 13 ข้อ ไม่ทำหน้าที่ต่างกัน 47 ข้อ และวิธีClosed - Interval Unsigned Area (CUA) ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 15 ข้อ ไม่ทำหน้าที่ต่างกัน 45 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วย Q – Test พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบแตกต่างกัน และได้ทำการเปรียบเทียบความแตกต่างของจำนวนข้อสอบที่ตรวจพบเป็นรายคู่ แสดงผลดังตาราง 19

ตาราง 19 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่ เมื่อจำนวนข้อสอบ 60 ข้อ กลุ่มตัวอย่าง 750 คน ระหว่างวิธี Mantel-Haenszel (MH) วิธี Lord's χ^2 (LC) วิธี Closed - Interval Signed Area (CSA) และวิธี Closed - Interval Unsigned Area (CUA)

วิธีตรวจสอบ	การทดสอบแบบคิเป็นรายคู่ (Q – Test)			
	MH	LC	CSA	CUA
MH	-	1.00	9.00*	7.14*
LC		-	14.00*	12.00*
CSA			-	2.00
CUA				-

การวิเคราะห์ตามตาราง 19 พบว่า เมื่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ จากข้อสอบ 60 ข้อ กลุ่มตัวอย่าง 750 คน มาทดสอบความแตกต่างเป็นรายคู่ (Q – Test) พบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบโดยวิธี Mantel-Haenszel (MH) กับ วิธี Closed - Interval Signed Area (CSA), วิธี Mantel-Haenszel (MH) กับ วิธี Closed - Interval Unsigned Area (CUA) , วิธี Lord's χ^2 (LC) กับ วิธี Closed - Interval Signed Area (CSA) , วิธี Lord's χ^2 (LC) กับ วิธี Closed - Interval Unsigned Area (CUA) แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนสองคู่ที่เหลือคือ วิธี Mantel-Haenszel (MH) กับ วิธี Lord's χ^2 (LC) และวิธี Closed - Interval Signed Area (CSA) กับ วิธี Closed - Interval Unsigned Area (CUA) แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

3.2 ผลการทดสอบความแตกต่างของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อกลุ่มตัวอย่างต่างกัน วิธีการตรวจสอบเหมือนกัน แสดงผลดังตาราง 20 – 21

ตาราง 20 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 40 ข้อ
วิธีการตรวจสอบเหมือนกัน ระหว่างกลุ่มตัวอย่างขนาด 600 คน และ 750 คน

วิธีตรวจสอบ	การทำหน้าที่ของข้อสอบ	600	750	Q – Test
MH	ทำหน้าที่ต่างกัน	15	18	3.00
	ไม่ทำหน้าที่ต่างกัน	25	22	
LC	ทำหน้าที่ต่างกัน	10	19	6.23*
	ไม่ทำหน้าที่ต่างกัน	30	21	
CSA	ทำหน้าที่ต่างกัน	7	8	1.00
	ไม่ทำหน้าที่ต่างกัน	33	32	
CUA	ทำหน้าที่ต่างกัน	8	10	2.00
	ไม่ทำหน้าที่ต่างกัน	32	30	

ผลการวิเคราะห์ตามตาราง 20 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Mantel-Haenszel (MH) จากกลุ่มตัวอย่างขนาด 600 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 15 ข้อ ไม่ทำหน้าที่ต่างกัน 25 ข้อ กลุ่มตัวอย่างขนาด 750 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 18 ข้อ ไม่ทำหน้าที่ต่างกัน 22 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วย Q – Test พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 2 ขนาด ไม่แตกต่างกัน

จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Lord's χ^2 (LC) จากกลุ่มตัวอย่างขนาด 600 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 10 ข้อ ไม่ทำหน้าที่ต่างกัน 30 ข้อ กลุ่มตัวอย่างขนาด 750 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 19 ข้อ ไม่ทำหน้าที่ต่างกัน 21 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วย Q – Test พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 2 ขนาดแตกต่างกัน

จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Closed - Interval Signed Area (CSA) จากกลุ่มตัวอย่างขนาด 600 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 7 ข้อ ไม่ทำหน้าที่ต่างกัน 33 ข้อ กลุ่มตัวอย่างขนาด 750 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 8 ข้อ ไม่ทำหน้าที่ต่างกัน 32 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วย Q – Test พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 2 ขนาดไม่แตกต่างกัน

และจำนวนข้อสอบที่ตรวจสอบด้วยวิธี วิธี Closed - Interval Unsigned Area (CUA) จากกลุ่มตัวอย่างขนาด 600 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 8 ข้อ ไม่ทำหน้าที่ต่างกัน 32 ข้อ กลุ่มตัวอย่างขนาด 750 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 10 ข้อ ไม่ทำหน้าที่ต่างกัน 30 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วย Q – Test พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 2 ขนาดไม่แตกต่างกัน

ตาราง 21 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 60 ข้อ
วิธีการตรวจสอบเหมือนกัน ระหว่างกลุ่มตัวอย่างขนาด 600 คน และ 750 คน

วิธีตรวจสอบ	การทำหน้าที่ของข้อสอบ	600	750	Q – Test
MH	ทำหน้าที่ต่างกัน	22	25	1.29
	ไม่ทำหน้าที่ต่างกัน	38	35	
LC	ทำหน้าที่ต่างกัน	17	27	5.56 [*]
	ไม่ทำหน้าที่ต่างกัน	43	33	
CSA	ทำหน้าที่ต่างกัน	11	13	2.00
	ไม่ทำหน้าที่ต่างกัน	49	47	
CUA	ทำหน้าที่ต่างกัน	13	15	2.00
	ไม่ทำหน้าที่ต่างกัน	47	45	

ผลการวิเคราะห์ตามตาราง 21 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Mantel-Haenszel (MH) จากกลุ่มตัวอย่างขนาด 600 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 22 ข้อ ไม่ทำหน้าที่ต่างกัน 38 ข้อ กลุ่มตัวอย่างขนาด 750 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 25 ข้อ ไม่ทำหน้าที่ต่างกัน 35 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วย Q – Test พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 2 ขนาดไม่แตกต่างกัน

จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Lord's χ^2 (LC) จากกลุ่มตัวอย่างขนาด 600 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 17 ข้อ ไม่ทำหน้าที่ต่างกัน 43 ข้อ กลุ่มตัวอย่างขนาด 750 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 27 ข้อ ไม่ทำหน้าที่ต่างกัน 33 ข้อ เมื่อทดสอบความแตกต่างของจำนวน

ข้อสอบที่ทำหน้าที่ต่างกันด้วย Q – Test พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 2 ขนาดแตกต่างกัน

จำนวนข้อสอบที่ตรวจพบด้วยวิธี วิธี Closed - Interval Signed Area (CSA) จากกลุ่มตัวอย่างขนาด 600 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 11 ข้อ ไม่ทำหน้าที่ต่างกัน 49 ข้อ กลุ่มตัวอย่างขนาด 750 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 13 ข้อ ไม่ทำหน้าที่ต่างกัน 47 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วย Q – Test พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 2 ขนาดไม่แตกต่างกัน

จำนวนข้อสอบที่ตรวจพบด้วยวิธี วิธี Closed - Interval Unsigned Area (CUA) จากกลุ่มตัวอย่างขนาด 600 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 13 ข้อ ไม่ทำหน้าที่ต่างกัน 47 ข้อ กลุ่มตัวอย่างขนาด 750 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 15 ข้อ ไม่ทำหน้าที่ต่างกัน 45 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วย Q – Test พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 2 ขนาดไม่แตกต่างกัน

ตอนที่ 4 ความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ระหว่างวิธี Mantel – Haenszel วิธี Lord's χ^2 และวิธี Closed Interval Area

การวิเคราะห์ตอนที่ 4 ผู้วิจัยได้นำผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี Mantel – Haenszel วิธี Lord's χ^2 และวิธี Closed Interval Area จากตอนที่ 2 มาหาค่าความสอดคล้องในการตรวจสอบระหว่างวิธีตรวจสอบ แสดงผลดังตาราง 22 – 23

ตาราง 22 ร้อยละของความสอดคล้องในการตรวจสอบเมื่อกลุ่มตัวอย่างเท่ากัน วิธีตรวจสอบแตกต่างกัน

I	N	N	600				750			
			วิธี	MH	LC	CSA	CUA	MH	LC	CSA
40	600	MH	-	83.3	66.7	75				
		LC		-	80	90				
		CSA			-	70				
		CUA				-				
	750	MH					-	78.6	61.5	64.3
		LC						-	66.67	69.23
		CSA							-	80
		CUA								-
60	600	MH	-	72.2	55.6	70.6				
		LC		-	78.57	92.86				
		CSA			-	84.62				
		CUA				-				
	750	MH					-	95	60	73.7
		LC						-	65	75
		CSA							-	86.67
		CUA								-

ผลการวิเคราะห์ตามตาราง 22 พบว่า ความสอดคล้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ อยู่ระหว่างร้อยละ 55.6 ถึง ร้อยละ 95

เมื่อจำนวนข้อสอบ 40 ข้อ กลุ่มตัวอย่าง 600 คน วิธีที่สอดคล้องกันสูงสุดได้แก่ วิธี Lord's χ^2 (LC) กับ วิธี Closed - Interval Unsigned Area (CUA) (ร้อยละ 90) ส่วนวิธีที่สอดคล้องกันต่ำสุดได้แก่ วิธี Mantel-Haenszel (MH) กับ วิธี Closed - Interval Signed Area (CSA) (ร้อยละ 66.7)

ผลการวิเคราะห์ตามตาราง 23 พบว่า ความสอดคล้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ มีค่าอยู่ระหว่างร้อยละ 65 ถึง ร้อยละ 92.31

เมื่อจำนวนข้อสอบ 40 ข้อ ตรวจสอบด้วยวิธี Mantel-Haenszel (MH) และขนาดกลุ่มตัวอย่าง 600 คน กับ 750 คน ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่สอดคล้องกันสูงสุด (ร้อยละ 92.31) ส่วนที่สอดคล้องกันต่ำสุด ได้แก่ ตรวจสอบด้วยวิธี Closed - Interval Signed Area (CSA) เมื่อมีกลุ่มตัวอย่าง 600 คน กับ 750 คน (ร้อยละ 77.78)

เมื่อจำนวนข้อสอบ 60 ข้อ ตรวจสอบด้วยวิธี วิธี Closed - Interval Unsigned Area (CUA) และขนาดกลุ่มตัวอย่าง 600 คน กับ 750 คน ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่สอดคล้องกันสูงสุด (ร้อยละ 86.67) ส่วนที่สอดคล้องกันต่ำสุด ได้แก่ ตรวจสอบด้วยวิธี Lord's χ^2 (LC) เมื่อมีกลุ่มตัวอย่าง 600 คน กับ 750 คน (ร้อยละ 65)

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

วัตถุประสงค์ของการวิจัย

ในการศึกษาค้นคว้าครั้งนี้ ผู้วิจัยได้ตั้งวัตถุประสงค์ดังต่อไปนี้

1. เพื่อเปรียบเทียบจำนวนข้อสอบที่มีการทำหน้าที่ของผลการตรวจสอบระหว่างวิธี Mantel – Haenszel วิธี Lord's χ^2 และวิธี Closed Interval Area เมื่อขนาดของกลุ่มตัวอย่าง และความยาวของแบบทดสอบ ต่างกัน
2. เพื่อศึกษาความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่าง วิธี Mantel – Haenszel วิธี Lord's χ^2 และวิธี Closed Interval Area เมื่อขนาดของกลุ่มตัวอย่าง และความยาวของแบบทดสอบ ต่างกัน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ ได้แก่ นักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2549 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษากาญจนบุรี เขต 3 โดยนักเรียนนั้นมีความพร้อมในการเรียนรู้ ในระดับใกล้เคียงกัน ซึ่งได้มาโดยการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) แบ่งเป็นนักเรียนเชื้อชาติไทยจำนวน 780 คน และนักเรียนไม่ใช่เชื้อชาติไทย จำนวน 753 คน

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

ข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลผลการสอบวิชาคณิตศาสตร์ ตามหลักสูตรขั้นพื้นฐาน พุทธศักราช 2544 ในช่วงชั้นที่ 2 (ประถมศึกษาปีที่ 4 – 6) เป็นแบบทดสอบใจหายปัญหาจำนวน 60 ข้อ เป็นแบบเลือกตอบชนิด 4 ตัวเลือก ที่ผู้วิจัยสร้างขึ้นเอง หากคุณภาพโดยใช้สูตรคูเดอริชาร์ดสัน (KR-20) ได้ความเชื่อมั่นเท่ากับ .894 และวิเคราะห์องค์ประกอบ (Factor Analysis) เพื่อตรวจสอบความเป็นมิติเดียวของแบบทดสอบ พบว่าค่าไอเกนขององค์ประกอบที่ 1 มีค่าสูงกว่าไอเกนขององค์ประกอบที่ 2 ประมาณ 1.96 เท่า

วิธีดำเนินการรวบรวมข้อมูล

1. ขอหนังสือจากบัณฑิตวิทยาลัยมหาวิทยาลัยศรีนครินทรวิโรฒเพื่อขอความอนุเคราะห์ในการเก็บรวบรวมข้อมูล โดยทำหนังสือถึงผู้อำนวยการโรงเรียนที่เป็นกลุ่มตัวอย่าง แล้วนำไปติดต่อขอความ

ร่วมมือกับโรงเรียนที่เป็นกลุ่มตัวอย่าง เพื่อนัดหมายและกำหนดวัน เวลา ในการนำแบบทดสอบ ไปทดสอบกับนักเรียนที่เป็นกลุ่มตัวอย่าง

2. จัดเตรียมแบบทดสอบให้เพียงพอกับจำนวนนักเรียนที่จะทำการสอบในแต่ละโรงเรียน

3. ดำเนินการสอบ โดยนำแบบทดสอบ ไปทำการทดสอบกับนักเรียนที่เป็นกลุ่มตัวอย่าง ในช่วง 5 กุมภาพันธ์ ถึง 16 กุมภาพันธ์ 2550

4. ชี้แจงให้นักเรียนที่เป็นกลุ่มตัวอย่างทราบวัตถุประสงค์ของการสอบ และขอความร่วมมือในการทำแบบทดสอบ เพื่อให้ได้ผลการสอบตรงตามความเป็นจริงมากที่สุด

5. นำผลการตอบข้อสอบในแบบทดสอบ มาตรวจดูความเรียบร้อยแล้วลงรหัสผลการตอบข้อสอบทุกข้อ (ตอบถูกได้ 1 และตอบผิดได้ 0)

6. นำข้อมูลที่ได้มาเตรียมข้อมูลในการศึกษา วิเคราะห์ทางสถิติและทดสอบสมมติฐานที่ตั้งไว้

สรุปผลการวิจัย

ผลการวิจัยสรุปได้ดังต่อไปนี้

1. จำนวนข้อสอบที่ทำหน้าที่ต่างกันตรวจสอบพบ จากแบบทดสอบ 2 ชุด คือ แบบทดสอบที่มีจำนวน 40 ข้อ และ 60 ข้อ กับกลุ่มตัวอย่าง 2 ขนาด คือ กลุ่มตัวอย่าง 600 คน และ 750 คน ตรวจสอบด้วยวิธี Mantel-Haenszel , วิธี Lord's χ^2 , วิธี Closed - Interval Signed Area และ วิธี Closed - Interval Unsigned Area สรุปได้ดังนี้

1.1 แบบทดสอบจำนวน 40 ข้อ กลุ่มตัวอย่าง 600 คน พบว่า วิธี Mantel-Haenszel พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด คือ จำนวน 12 ข้อ รองลงมาคือ วิธี Lord's χ^2 พบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน คือ จำนวน 10 ข้อ วิธี Closed - Interval Unsigned Area พบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน คือ จำนวน 9 ข้อ และวิธี Closed - Interval Signed Area พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด คือ จำนวน 7 ข้อ เมื่อพิจารณาจำนวนข้อสอบที่ตรวจสอบพบกันทั้ง 4 วิธี จำนวน 7 ข้อ

1.2 แบบทดสอบจำนวน 40 ข้อ กลุ่มตัวอย่าง 750 คน พบว่า วิธี Mantel-Haenszel พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด คือ จำนวน 13 ข้อ รองลงมาคือ วิธี Lord's χ^2 พบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน คือ จำนวน 12 ข้อ วิธี Closed - Interval Unsigned Area พบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน คือ จำนวน 10 ข้อ และวิธี Closed - Interval Signed Area พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด คือ จำนวน 8 ข้อ เมื่อพิจารณาจำนวนข้อสอบที่ตรวจสอบพบกันทั้ง 4 วิธี จำนวน 8 ข้อ

1.3 แบบทดสอบจำนวน 60 ข้อ กลุ่มตัวอย่าง 600 คน พบว่า วิธี Mantel-Haenszel พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด คือ จำนวน 17 ข้อ รองลงมาคือ วิธี Lord's χ^2 พบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน คือ จำนวน 14 ข้อ วิธี Closed - Interval Unsigned Area พบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน คือ จำนวน 13 ข้อ และวิธี Closed - Interval Signed Area พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด คือ จำนวน 11 ข้อ เมื่อพิจารณาจำนวนข้อสอบที่ตรวจสอบพบกันทั้ง 4 วิธี จำนวน 10 ข้อ

1.4 แบบทดสอบจำนวน 60 ข้อ กลุ่มตัวอย่าง 750 คน พบว่า วิธี Lord's พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด คือ จำนวน 20 ข้อ รองลงมาคือ วิธี Mantel-Haenszel พบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน คือ จำนวน 19 ข้อ วิธี Closed - Interval Unsigned Area พบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน คือ จำนวน 15 ข้อ และวิธี Closed - Interval Signed Area พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด คือ จำนวน 13 ข้อ เมื่อพิจารณาจำนวนข้อสอบที่ตรวจสอบพบกันทั้ง 4 วิธี จำนวน 12 ข้อ

2. เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อขนาดกลุ่มตัวอย่างเท่ากัน แต่วิธีการตรวจสอบต่างกัน จากจำนวนข้อสอบ 40 ข้อ และ 60 ข้อ ตามลำดับ ตรวจสอบกับกลุ่มตัวอย่างขนาดเดียวกัน ในแต่ละวิธีตรวจสอบ คือ วิธี Mantel-Haenszel , วิธี Lord's χ^2 , วิธี Closed - Interval Signed Area และวิธี Closed - Interval Unsigned Area โดยใช้ Q - test ผลการทดสอบ สรุปว่า เมื่อจำนวนข้อสอบ 40 ข้อ กับกลุ่มตัวอย่างขนาด 600 คน และ 750 คน และจากจำนวนข้อสอบ 60 ข้อ กับกลุ่มตัวอย่างขนาด 600 คน และ 750 คน จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบโดยวิธี Mantel-Haenszel , วิธี Lord's χ^2 , วิธี Closed - Interval Signed Area และวิธี Closed - Interval Unsigned Area แตกต่างกัน

3. เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อขนาดกลุ่มตัวอย่างต่างกัน แต่วิธีการตรวจสอบเหมือนกัน จากจำนวนข้อสอบ 40 ข้อ และ 60 ข้อ ตามลำดับ จากวิธีการตรวจสอบเดียวกัน ในแต่ละขนาดของกลุ่มตัวอย่าง คือ 600 คน และ 750 คน โดยใช้ Q - test ผลการทดสอบ สรุปว่า เมื่อจำนวนข้อสอบ 40 ข้อ ตรวจสอบโดยวิธี Lord's χ^2 และ จำนวนข้อสอบ 60 ข้อ ตรวจสอบโดยวิธี Lord's χ^2 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบจากกลุ่มตัวอย่างสองขนาดแตกต่างกัน ส่วนจากจำนวนข้อสอบ 40 ข้อ ตรวจสอบโดยวิธี Mantel-Haenszel , วิธี Closed - Interval Signed Area และวิธี Closed - Interval Unsigned Area จากจำนวนข้อสอบ 60 ข้อ ตรวจสอบโดยวิธี Mantel-Haenszel , วิธี Closed - Interval Signed Area และ วิธี Closed - Interval Unsigned Area จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบจากกลุ่มตัวอย่างสองขนาดไม่แตกต่างกัน

4. ความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อกลุ่มตัวอย่างเท่ากันแต่วิธีตรวจสอบต่างกัน สรุปได้ดังนี้

4.1 แบบทดสอบจำนวน 40 ข้อ กลุ่มตัวอย่าง 600 คน พบว่าวิธีที่มีความสอดคล้องกันในการตรวจสอบสูงสุด ได้แก่ วิธีLord's กับ วิธี Closed - Interval Unsigned Area และวิธีที่มีความสอดคล้องกันในการตรวจสอบต่ำสุด ได้แก่ วิธีMantel-Haenszel กับ วิธี Closed - Interval Signed Area

4.2 แบบทดสอบจำนวน 40 ข้อ กลุ่มตัวอย่าง 750 คน พบว่าวิธีที่มีความสอดคล้องกันในการตรวจสอบสูงสุด ได้แก่ วิธีClosed - Interval Signed Area กับ วิธี Closed - Interval Unsigned Area และวิธีที่มีความสอดคล้องกันในการตรวจสอบต่ำสุด ได้แก่ วิธีMantel-Haenszel กับ วิธี Closed - Interval Signed Area

4.3 แบบทดสอบจำนวน 60 ข้อ กลุ่มตัวอย่าง 600 คน พบว่าวิธีที่มีความสอดคล้องกันในการตรวจสอบสูงสุด ได้แก่ วิธีLord's กับ วิธี Closed - Interval Unsigned Area และวิธีที่มีความสอดคล้องกันในการตรวจสอบต่ำสุด ได้แก่ วิธีMantel-Haenszel กับ วิธี Closed - Interval Signed Area

4.4 แบบทดสอบจำนวน 60 ข้อ กลุ่มตัวอย่าง 750 คน พบว่าวิธีที่มีความสอดคล้องกันในการตรวจสอบสูงสุด ได้แก่ วิธีMantel-Haenszel กับ วิธี Lord's χ^2 และวิธีที่มีความสอดคล้องกันในการตรวจสอบต่ำสุด ได้แก่ วิธีMantel-Haenszel กับ วิธี Closed - Interval Signed Area

5. ความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อขนาดกลุ่มตัวอย่างต่างกัน คือ 600 คน กับ 750 คน วิธีการตรวจสอบเหมือนกัน สรุปได้ดังนี้

5.1 แบบทดสอบจำนวน 40 ข้อ พบว่าวิธีที่มีความสอดคล้องกันในการตรวจสอบสูงสุดคือ วิธีMantel-Haenszel และวิธีที่มีความสอดคล้องกันในการตรวจสอบต่ำสุดคือ วิธี Closed - Interval Signed Area

5.2 แบบทดสอบจำนวน 60 ข้อ พบว่าวิธีที่มีความสอดคล้องกันในการตรวจสอบสูงสุดคือ วิธี Closed - Interval Unsigned Area และวิธีที่มีความสอดคล้องกันในการตรวจสอบต่ำสุดคือ วิธีLord's χ^2

อภิปรายผล

จากการวิจัยสามารถอภิปรายผลได้ดังนี้

1. จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจสอบเมื่อขนาดกลุ่มตัวอย่างเท่ากัน วิธีการตรวจสอบต่างกัน พบว่า เมื่อจำนวนข้อสอบ 40 ข้อ กลุ่มตัวอย่าง 600 คน จำนวนข้อสอบที่ทำหน้าที่

ต่างกันที่ตรวจพบโดยวิธี Mantel-Haenszel กับ วิธี Closed - Interval Signed Area แตกต่างกันอย่างมีนัยสำคัญทางสถิติ เมื่อจำนวนข้อสอบ 40 ข้อ กลุ่มตัวอย่าง 750 คน จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบโดยวิธี Mantel-Haenszel กับ วิธี Closed - Interval Signed Area และ วิธี Lord's กับ วิธี Closed - Interval Signed Area แตกต่างกันอย่างมีนัยสำคัญทางสถิติ เมื่อจำนวนข้อสอบ 60 ข้อ กลุ่มตัวอย่าง 600 คน จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบโดยวิธี Mantel-Haenszel กับ วิธี Closed - Interval Signed Area, วิธี Mantel-Haenszel กับ วิธี Closed - Interval Unsigned Area, วิธี Lord's χ^2 กับ วิธี Closed - Interval Signed Area, วิธี Lord's χ^2 กับ วิธี Closed - Interval Unsigned Area แตกต่างกันอย่างมีนัยสำคัญทางสถิติ เมื่อจำนวนข้อสอบ 60 ข้อ กลุ่มตัวอย่าง 750 คน จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบโดยวิธี Mantel-Haenszel กับ วิธี Closed - Interval Signed Area, วิธี Mantel-Haenszel กับ วิธี Closed - Interval Unsigned Area, วิธี Lord's χ^2 กับ วิธี Closed - Interval Signed Area, วิธี Lord's χ^2 กับ วิธี Closed - Interval Unsigned Area แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ซึ่งสอดคล้องกับสมมติฐานการวิจัยที่ตั้งไว้ และยังสอดคล้องกับงานวิจัยของญาณภัทร สีหะมงคล (2540 : 92) ที่ศึกษาเปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบด้วยวิธี Lord's χ^2 และวิธี Closed Interval Area เมื่อกลุ่มตัวอย่างมีขนาดเท่ากันพบจำนวนข้อสอบที่ทำหน้าที่ต่างกันแตกต่างกันอย่างมีนัยสำคัญ

นอกจากนี้จากการศึกษายังพบว่าทั้งวิธี Lord's χ^2 และวิธี Closed Interval Area ตั้งอยู่บนพื้นฐานทฤษฎีการตอบข้อสอบ (Item Response Theory) เหมือนกัน แต่ผลจำนวนข้อสอบที่พบว่าทำหน้าที่ต่างกัน ทั้งนี้ขึ้นอยู่กับขนาดของกลุ่มตัวอย่าง ความยาวของแบบทดสอบ เกณฑ์ที่ใช้ในการตัดสินใจการทำหน้าที่ต่างกันของข้อสอบ และความไวต่อการตัดสินใจข้อสอบที่ทำหน้าที่ต่างกัน ของสถิติทดสอบต่างกัน ซึ่งสอดคล้องกับงานวิจัยของสุรศักดิ์ อมรรัตนศักดิ์ (2531:200-207) และงานวิจัยของกาญจนา วัฒนสุนทร (2537:267) ที่พบว่าวิธีการใช้ทฤษฎีการตอบข้อสอบ ค้นพบข้อสอบที่ทำหน้าที่ต่างกัน กับวิธี Mantel-Haenszel แตกต่างกัน ทั้งนี้เนื่องจากวิธีการใช้ทฤษฎีการตอบข้อสอบมีความไวในการตรวจสอบการทำหน้าที่ต่างกัน เพราะเป็นวิธีการที่ใช้ความแตกต่างของพื้นที่ระหว่างโค้งลักษณะข้อสอบ (Kim and Cohon.1991) และขึ้นอยู่กับเกณฑ์ที่กำหนดขนาดของดัชนีการทำหน้าที่ต่างกันถือว่าการทำหน้าที่ต่างกันด้วย

2. จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจสอบเมื่อขนาดกลุ่มตัวอย่างต่างกัน วิธีการตรวจสอบเหมือนกัน พบว่า เมื่อจำนวนข้อสอบ 40 ข้อ ตรวจสอบด้วยวิธี Lord's χ^2 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบระหว่างกลุ่มตัวอย่างขนาด 600 คน กับ 750 คน แตกต่างกันอย่างมีนัยสำคัญทางสถิติ และจำนวนข้อสอบ 60 ข้อ ตรวจสอบด้วยวิธี Lord's χ^2 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบระหว่างกลุ่มตัวอย่างขนาด 600 คน กับ 750 คน แตกต่างกันอย่างมี

นัยสำคัญทางสถิติ ซึ่งสอดคล้องกับสมมติฐานการวิจัยที่ตั้งไว้ ส่วนกรณีจำนวนข้อสอบ 40 ข้อ ตรวจสอบด้วยวิธี Mantel-Haenszel , วิธี Closed - Interval Signed Area และวิธี Closed - Interval Unsigned Area กลุ่มตัวอย่างขนาด 600 คน กับ 750 คน แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และจำนวนข้อสอบ 60 ข้อ ตรวจสอบด้วยวิธี Mantel-Haenszel , วิธี Closed - Interval Signed Area และวิธี Closed - Interval Unsigned Area กลุ่มตัวอย่างขนาด 600 คน กับ 750 คน แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ซึ่งไม่สอดคล้องกับสมมติฐานการวิจัยที่ตั้งไว้ ทั้งนี้อาจเนื่องมาจากจำนวนข้อสอบที่ทำหน้าที่ต่างกันตรวจพบเมื่อขนาดกลุ่มตัวอย่าง 600 คน กับ 750 คน ทั้ง 4 วิธีสามารถตรวจสอบได้ใกล้เคียงกัน ดังนั้นเมื่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันมาทดสอบ นัยสำคัญทางสถิติจึง แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และเมื่อพิจารณาเป็นรายข้อแล้ว พบว่า วิธี Mantel-Haenszel , วิธี Closed - Interval Signed Area และวิธี Closed - Interval Unsigned Area เมื่อขนาดกลุ่มตัวอย่าง 600 คน กับ 750 คน ส่วนใหญ่จะตรวจพบข้อสอบที่ทำหน้าที่ต่างกันซ้ำกัน (นุชริน ไบโพธิ์.2544 : 86)

นอกจากนี้จากการศึกษายังพบว่า เมื่อขนาดกลุ่มตัวอย่างใหญ่ขึ้นสำหรับวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของ Lord's χ^2 โดยใช้การเปรียบเทียบค่าพารามิเตอร์ของข้อสอบแล้วทดสอบนัยสำคัญทางสถิติด้วยสถิติทดสอบไคกำลังสอง (χ^2) สามารถตรวจพบข้อสอบที่ทำหน้าที่ต่างกันได้มากขึ้น ซึ่งสอดคล้องกับข้อเสนอแนะของ มิลแชป แลอีเวอร์สัน (Millsap and Everson. 1993 : 314) และแฮมเบิลตัน สวามินาทาน และโรเจอร์ (Hambleton, Swaminathan and Rogers. 1991 : 113) ว่าวิธีนี้มีความไวต่อการตัดสินข้อสอบที่ทำหน้าที่ต่างกันมาก

3. ความสอดคล้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี Mantel - Haenszel วิธี Lord's χ^2 และวิธี Closed Interval Area พบว่าส่วนใหญ่มีค่าสอดคล้องปานกลางถึงสูง และมีความไม่คงที่ในเกือบทุกเงื่อนไขของการศึกษา ซึ่งสอดคล้องกับกาญจนา วัฒนสุนทร (2537) ที่พบว่ามีความไม่คงที่ในแบบทดสอบที่มีจำนวนข้อสอบต่างกัน และ ขนาดของกลุ่มตัวอย่างต่างกันเนื่องจากเกณฑ์ที่ใช้ตัดสินข้อสอบที่ทำหน้าที่ต่างกันยังไม่สามารถกำหนดแน่นอนและความไวต่อการตัดสินของสถิติทดสอบที่ใช้ในการศึกษาต่างกัน

ข้อเสนอแนะในการวิจัย

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1งานวิจัยครั้งนี้ได้ข้อค้นพบว่า เมื่อใช้จำนวนข้อสอบ 40 ข้อ และ 60 ข้อ ผลการตรวจสอบระหว่างวิธี Mantel - Haenszel วิธี Lord's χ^2 ไม่แตกต่างกัน ดังนั้นควรเลือกใช้ วิธี Mantel - Haenszel เพราะเมื่อขนาดกลุ่มตัวอย่างต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ

ระหว่างกลุ่มตัวอย่างขนาด 600 คน กับ 750 คน ไม่แตกต่างกันและคำนวณง่ายกว่า แต่วิธี Lord's χ^2 เมื่อขนาดกลุ่มตัวอย่างต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบระหว่างกลุ่มตัวอย่างขนาด 600 คน กับ 750 คน แตกต่างกัน

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

2.1 ควรศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้ข้อมูลจำลอง (Simulated data) เพื่อยืนยันผลการศึกษาว่าได้ข้อสรุปสอดคล้องกันหรือไม่

2.2 ควรศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ กับตัวแปรอื่น ๆ เช่น แบบจำลองในการศึกษา วิธีประมาณค่าพารามิเตอร์ เป็นต้น จะทำให้ได้ข้อสรุปในแง่มุมต่าง ๆ ที่สมบูรณ์ยิ่งขึ้น

2.3 ควรศึกษาความคลาดเคลื่อนของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ทั้งความคลาดเคลื่อนชนิดที่ 1 และความคลาดเคลื่อนชนิดที่ 2

2.4 ควรศึกษาถึงความสัมพันธ์ระหว่างการทำหน้าที่ต่างกันของข้อสอบ กับความเป็นหลายมิติของข้อสอบ

บรรณานุกรม

บรรณานุกรม

- กรมการศึกษานอกโรงเรียน. (2535). *ชาวเขาเรียนเลข*. ฝ่ายส่งเสริมและพัฒนาเทคโนโลยีทางการศึกษา ศูนย์เทคโนโลยีทางการศึกษา. กรุงเทพฯ.
- กัลยา วานิชย์บัญชา. (2544). *การใช้ SPSS for Windows ในการวิเคราะห์ข้อมูล*. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย .
- กาญจนา วัชรสุนทร. (2538) *การพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ*. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย . อัดสำเนา.
- เกรียงศักดิ์ เจริญวงศ์ศักดิ์. (2543). *จอมปราชญ์นักการศึกษา : สังเคราะห์ วิเคราะห์ และประยุกต์แนวพระราชดำริด้านการศึกษาและการพัฒนาคน*. กรุงเทพฯ : ชัคเชส มีเดีย.
- คมศักดิ์ ชื่นชม. (2539) *การศึกษาผลการวิเคราะห์ความลำเอียงที่ใช้วิธีต่างกันของแบบทดสอบวัดจริยธรรมด้านความซื่อสัตย์ ตามทฤษฎีคลาสสิกอล*. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.
- เจน จันทรแสง. (2537) *การวิเคราะห์ความลำเอียงของข้อสอบของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3* . วิทยานิพนธ์ กศ.ม.(การวัดผลการศึกษา) พิษณุโลก : บัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร. อัดสำเนา.
- จิตติมา วรรณศรี. (2539) *การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทิล-แฮนส์กับวิธีชิปเทสต์ เมื่อขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบต่างกัน*. วิทยานิพนธ์ ค.ม.(การวัดและประเมินผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย . อัดสำเนา.
- จิตสุดา ธารพร. (2539) *การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา*. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.
- ชัชชัย เผ่าพงศ์. (2527) *การวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์และภาษา ระดับชั้นมัธยมศึกษาตอนต้น*. วิทยานิพนธ์ คศ.ม.(การวัดและประเมินผลการศึกษา) ขอนแก่น : บัณฑิตวิทยาลัย มหาวิทยาลัยขอนแก่น . อัดสำเนา.
- ชวาล แพร์ตกุล. (2516). *เทคนิคการวัดผล*. กรุงเทพฯ : โรงเรียนแพร์ตันอนุสรณ์.

- ชุติมา แสงคารารัตน์. (2546) การเปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบวัดความถนัดทางการเรียนตามความรู้สึกคุ้นเคย ความรู้สึกสนใจและความรู้สึกพอใจในข้อสอบด้วยวิธีการตรวจสอบต่างกัน. ปรินญาณิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.
- ญาณภัทร สีหะมงคล. (2540). การเปรียบเทียบผลการตรวจสอบที่ทำหน้าที่ต่างกัน ระหว่างวิธี Lord's χ^2 วิธี Raju's Area Measurement และวิธี Closed Interval Area . ปรินญาณิพนธ์ กศ.ค. (การทดสอบและการวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.
- ดารณี แดคงู. (2542). การศึกษาผลการวิเคราะห์ความลำเอียงของข้อสอบวัดความเข้าใจในการอ่านภาษาไทยที่ใช้วิธีวิเคราะห์ความลำเอียงต่างกัน. ปรินญาณิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.
- ทองอยู่ สาระ. (2543). การเปรียบเทียบอำนาจการตรวจสอบและการจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบสม่าเสมอ ระหว่างวิธีแมนเทล-แฮนส์กับวิธีถดถอยโลจิสติก โดยใช้ความยาวแบบทดสอบและขนาดกลุ่มตัวอย่างต่างกัน . ปรินญาณิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.
- ทัศนีย์ พิรมนตรี. (2530) การวิเคราะห์ความลำเอียงของแบบทดสอบวิชาคณิตศาสตร์โครงการตรวจสอบคุณภาพการศึกษาชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526. วิทยานิพนธ์ ค.ม.(การวัดและประเมินผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย . อัดสำเนา.
- นิรมล ชัยเซาวลิต. (2537) การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของแบบทดสอบความเข้าใจในการอ่านภาษาไทยตามทฤษฎีคลาสสิกคอลที่ใช้วิธีวิเคราะห์ต่างกัน. ปรินญาณิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.
- นุชริน ไบโพธิ์. (2544) การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord's χ^2 วิธี Mantel – Haenszel และวิธี SIBTEST. ปรินญาณิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.
- นวลอนงค์ นันทวงศ์. (2543) การเปรียบเทียบความลำเอียงของแบบทดสอบความเข้าใจการอ่านตามลักษณะการใช้ภาษาพูดในครอบครัว โดยวิธีวิเคราะห์ต่างกัน. ปรินญาณิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.

- ผจงจิต อินทสุวรรณ. (2530, พฤษภาคม – สิงหาคม). แนวทางการศึกษาความลำเอียงของข้อสอบ. *การวัดผลการศึกษา*. 9(1) : 54-62.
- พัชรีย์ ปิยะภักดิ์. (2531) *การวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน*. ปรินญาณิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.
- พรรณี จินตมาศ. (2540) *การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบ โดยใช้ขนาดกลุ่มผู้สอบและวิธีวิเคราะห์ต่างกัน*. ปรินญาณิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.
- พิพุรพงษ์ แนวทอง. (2544) *การเปรียบเทียบผลของความลำเอียงของแบบทดสอบวัดความเข้าใจทางภาษาโดยใช้ขนาดกลุ่มผู้สอบและวิธีวิเคราะห์ต่างกัน*. ปรินญาณิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.
- ไพฑูรย์ เวทการ. (2536). *การประเมินผลและการสร้างแบบทดสอบ*. ลำปาง : ภาควิชาทดสอบและวิจัยทางการศึกษา วิทยาลัยลำปาง.
- เรวดี อินทสระ. (2539) *ผลการตรวจสอบความลำเอียงของข้อสอบต่อการศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่คิดคะแนนต่างกัน*. ปรินญาณิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.
- วิภา จำมัน. (2544) *การเปรียบเทียบผลการตรวจสอบข้อสอบที่ลำเอียงของแบบทดสอบวัดความสามารถด้านภาษา ตามระดับความง่ายของข้อสอบและวิธีที่ต่างกัน*. ปรินญาณิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.
- วลีมาศ แซ่อึ้ง. (2544, พฤษภาคม-สิงหาคม). การเปรียบเทียบอำนาจจำแนกการทดสอบและอัตราความคลาดเคลื่อนประเภทที่ 1 ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบอนุกรมระหว่างวิธีชิปเทสต์ปรับใหม่ วิธีชิปเทสต์ วิธีแมนเทล – แอนส์เซล. *วารสารวิธีวิทยาการวิจัย*. 14(2) : 216-228.
- สิริรัตน์ วิภาสศิลป์. (2545). *การเปรียบเทียบวิธีชิปเทสต์และดีเอฟไอทีในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบ หมวดข้อสอบ และแบบทดสอบ จากข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ*. ปรินญาณิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.
- เสรี ชัดเข้ม. (2540) *การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันแบบไม่สม่ำเสมอของข้อสอบระหว่างวิธีแมนเทล แอนเซล แบบแบ่งกลุ่มตามความสามารถและความยาวของ*

- ข้อสอบ . วิทยานิพนธ์ ค.ค.(การวัดและประเมินผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย . อัดสำเนา.
- สุชาดา กิระนันท์. (2542). *ทฤษฎีและวิธีการสำรวจตัวอย่าง*. กรุงเทพฯ : ศูนย์หนังสือจุฬาลงกรณ์มหาวิทยาลัย .
- สุพัฒน์ สุกมลสันต์. (2534). *การวิเคราะห์ความลำเอียงของข้อสอบภาษาอังกฤษเข้ามหาวิทยาลัยปี 2531 – 2533*. กรุงเทพฯ : สถาบันภาษาจุฬาลงกรณ์มหาวิทยาลัย . อัดสำเนา.
- สุรศักดิ์ อมรรัตนศักดิ์. (2531). *การศึกษาเปรียบเทียบผลของวิธีความลำเอียงของข้อสอบที่ต่างกัน 4 วิธี*. วิทยานิพนธ์ ค.ค. (การวัดและประเมินผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย . อัดสำเนา.
- สำนักงานคณะกรรมการการศึกษาแห่งชาติ. สำนักงานกฤษฎมนตรี. (2524). *รายงานวิจัยองค์ประกอบบางประการที่มีอิทธิพลต่อสัมฤทธิ์ผลทางการเรียนของนักเรียนชั้นมัธยมศึกษา*. กรุงเทพฯ : ม.ป.พ.
- สำนักงานเขตพื้นที่การศึกษากาญจนบุรี เขต 3. (2548). *แผนปฏิบัติการ ปีงบประมาณ 2548*.
กาญจนบุรี : สำนักงานเขตพื้นที่การศึกษากาญจนบุรี เขต 3.
- อนันต์ ศรีโสภณ. (2524). *การวัดและการประเมินผลการศึกษา*. กรุงเทพฯ : ไทยวัฒนาพานิช.
- อังคณา สายยศ. (2540). ผลการวิเคราะห์ความลำเอียงของแบบทดสอบความเข้าใจในการอ่านตามลักษณะการใช้ภาษาในครอบครัว. *การวัดผลการศึกษา*. 19(55) : 41-45.
- อารี วัชรโสทธิกุล. (2543). *การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้รูปแบบและวิธีการต่างกัน*. ปรินญาณิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.
- อุทัยวรรณ สายพัฒนา. (2547). *การเปรียบเทียบประสิทธิภาพของผลการตรวจการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าระหว่าง GMH และวิธี Polytomous SIBTEST*. ปรินญาณิพนธ์ กศ.ค. (การทดสอบและวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อัดสำเนา.
- Allen, Mary J ; & Wendy M. Yen. (1979) *Introduction to Measurment Theory*. California : Brooks/Cole Publishing Company.
- Baghi, Heibatollah ; & Steven F. Ferrara.(1990, February). Detecting Differential item Functioning Using IRT and Mantel-Haenszel Techniques : Implementing Procedures

- and Comparing Results. *Paper Presented at the Annual Meeting of the Eastern Educational Research Association*. Clearwater , Florida. 10(1) : pp14-17.
- Baker, F.B.. A Al – Karni ; & I.M.Al-Dosary. (1991) EQUATE : A Computer Program for the test Characteristic method of IRT equating. *Applied Psychological Measurement*. 15(3) : 78.
- Camilli, Gregory ; & Lorrie A. Shepard. (1994) *Methods for Identifying Biased Test Items*. California : Sage Publications.
- Clauser, Brian E. and Others. (1991, April). Examination of Various Influence on the Mantel-Haenszel Statistic. *Paper Presented at the Annual Meeting of American Educational Research Association*. Chicago, Illinois. 11(4) : pp3-7.
- Cohen, A.S., Kim, S.H. ; & Subkoviak, M.J. (1991) Influence of Distributions on Detection of DIF, *Journal of Educational Measurement*. 28(1) : pp49-59.
- Conbach, Lee J. (1970). *Essentials of Psychological Testing*. 3 rd ed. New York : Harper and Row Pubrishers.
- Doolittle, Allen E. ; & T. Anne. Cleary. (1987, Summer) Gender- Based Differential Item Performance In Mathematics Achievement Items, *Journal of Educational Measurement*. 24(2) : pp157-166.
- Hambleton, R.K. ; & L.L. Cook. (1977, Summer) Latent Trait Model and their Use in the Analysis of Educational Test Data. *Journal of Educational Measurement*. 14(2) : pp75-96.
- Hambleton, R.K. ; & H. Jane. Rogers. (1989, April) Detecting Potential Biased Test Items. Comparison of IRT Area and Mantel-Haenszel Methods. *Applied Measurement in Education*. 2(4) : pp313-334.
- Hambleton, R.K. ; & Swaminathan. H. (1985) *Item Response Theory Principles and Applications*. Boston, Kluwer-Nilhoff Publishing.
- Hambleton, R.K. ; & Swaminathan. H. & Rogers. H.J. (1991). *Fundamentals of Item Response Theory*. Newbury Park, California : Sage Publications, Inc.
- Hambleton, R.K. and Others. (1986, April) Identifying Potentially Biased Test Items. Comparison of the Mantel-Haenszel Statistic and Several Item Response

- Theory Methods. *Paper presented at the Annual Meeting of the American Educational Research Association*. San Francisco, California, 12(4) : pp16-20.
- Holland, P.W. ; & Thayer, D.T.. (1988) Differential Item Performance and the Mantel-Haenszel Procedure. *Test Validity*. In Wainer, Howard and Brgun, Henty ,Hillsdale NJ : Erlbaum.
- Holland, P.W. ; & H. Wainer. (1993) *Differential Item Functioning*. Hillsdale NJ : Erlbaum.
- Kim, Seock-Ho and Allen S. Cohen. (1991,September) A Comparison of Two Area Measures for Detecting Differential Item Functioning, *Applied Psychological Measurement*. 15(3) : pp269-278.
- Kim, Seock-Ho ; & Allen S.Cohen. (1993,March) A Comparison of Lord's χ^2 and Raju's Area Measures In Detection of DIF , *Applied Psychological Measurement*. 17(1) : pp39-52.
- Kim, Seock-Ho ; & Allen S.Cohen. (1994,April) An Inveatigation of Lord's χ^2 Procedure for the Detection of Differential Item Functioning, *Applied Psychological Measurement*. 18(3) : pp217-228.
- Linacre, John Michael. (1988). *The Practical of the Standara Error of the Mantel-Haenszel Statistics*. Paper present at the Annual Meeting of the American Educational Research Association. New Orleans. Los angelrs. pp 5-9.
- Linn, Robert L. ; & Levine, Michael V. , Hastings, Nicholas C. and Wardrop James L. (1981). Item Bias in a Test of Reading Comprenension, *Applied Psychological Measurement*. 5(2) : pp159-173.
- Lord, F.M.(1980). *Application of Item Response Theory to Practical Testing Problems*. Hillsdale, N.J. : Lawrence Eribaum Associates.
- Lord, F.M. ; & M.R. Novick. (1968). *Statistical Theories of Mental Test Score*. Massachusetts : Addison Wesley Publishing Company.
- Mazor, K.M. ; & Clauser, B.E. and Hambleton, R.K. (1992). The Effect of Sample Size on the Functioning of the Mantel-Haenszel Statistic, *Educational and Psychological Measurement*. 52(3) : pp443-451.

- Mazor, K.M. ; Clauser, B.E. ; & Hambleton, R.K. (1995). Using Logistic Regression and Mantel-Haenszel with Multiple Ability Estimates to Detect Differential Item Functioning, *Journal of Educational Measurement*. 32(3) : pp131-144.
- McCauley, Cynthia D. ; & Jorge. Mendoza. (1985,December). A Simulation Study of Item Bias Using a Two-Parameter Item Response Model, *Applied Psychological Measurement*. 9(4) : pp389-400.
- Mislevy, R.J. ; & R.D. Bock. (1994,March). BILOG 3 : Item Analysis and Test Scoring with Binary Logistic Model (Computer Program). Mooresville in : Scientific Software.
- Mislevy, Robert J. and Stocking, MARTHA I. Consumer's Guide to LOGIST and BILOG, *Applied Psychological Measurement*. 13(1) : pp57-75.
- Narayanan, Pankaja ; & H. Swaminathan. (1994,December) Performance of the Mantel-Haenszel and Simultaneous Item Bias Procedures for Detecting Differential Item Function, *Applied Psychological Measurement*. 18(4) : pp315-328.
- Perlman, C.L; et al. (1988). *Investigating the Stability of Four Methods for Estimating Item Bias*. Paper Presented at the Annual Meeting of the National Council for Measurement in Education. New Orleans, LA. pp5-9.
- Popham, W.J. (1981). *Modern educational Measurement*. Englewood Cliffs, New Jersey : Prentice-Hall.
- Raju, Drasgow Fritz ; & Jeffrey A. Slinde. (1993,March) An Empirical Comparison of the Area Methods, Lord's Chi-Square Test and the Mantel-Haenszel for Assessing Differential Item Functioning, *Educational and Psychological Measurement*. 53(2) : pp301-319.
- Rogers, H.J. and Swaminathan H. (1993). A Comparison of Logistic Regression and Mantel-Haenszel Procedures for Detection of Differential Item Functioning, *Applied Psychological Measurement*. 17(2) : pp105-116.
- Roussos, Louis A. ; & Stout, William F. (1996, Summer) Simulation Studies of Effects of Small Sample Size and Studied Item Parameters on SIBTEST and Mantel-Haenszel Type I Error Performance. *Journal of Educational Measurement*. 33(2) : pp215-230.
- Rudner, L.M. (1977). *An Evaluation of Select Approaches for Biased Item Identification*. Unpublished Doctoral Dissertation Catholic University of America.

- Seigel, Sidney. (1956). *Nonparametric Statistics for the Behavioral Sciences*. McGraw-Hill Kogakusha Tokyo.
- Shealy, R. ; & W. Stout. (1993). A Model-Based Standardization Approach that Separates True Bias/DIF from Group Ability Differences and Detects test Bias / DTF as well as Bias/DIF, *Psychometrika*. (58) : pp159-194.
- Shepard , L. ; Camilli, G. ; & Williams , D. M. (1995). Validity of approximation techniques for detecting item bias. *Journal of Educational Measurement*. 22 : pp77-105.
- Stocking, M.L. ; & Lord, F.M. (1983,December). Developing a common metric in item response theory, *Applied Psychological Measurement*. 7(3) : pp201-210.
- Subkoviak, M.J.; et al. (1984). Empirical Comparison of Selected Item Bias Detection Procedures with Bias Manipulation, *Journal of Educational Measurement*. 21(1) : pp49-58.
- Sudweeks, Richard, R. ; & Tolman, Richard R. (1990, April). The Use of Empirical versus Subjective Procedures for Identifying Science Test Items which Function Differentially for Females and Males, *Paper presented at the Annual Meeting of the National Association of Research in Science Teaching*. Atlanta, Georgia. 10(1) : pp8-11.
- Swaminathan H. ; & Rogers, H.J. (1990). Detection of Differential Item Functioning Using Logistic Regression Procedures , *Journal of Educational Measurement*. 27(4) : pp361-370.
- Thissen, David ; Lynne Steinberg ; & Howard. Wainer, (1988). *Use of Item Response Theory In The Study of Group Differences in Trace Test Validity*. pp147-170.
- _____. (1993). Detection of Differential Item Functioning using the Parameters of Item Response Models. In P.W. Holland and H. Wainer *Differential Item Functioning*. Hillsdale NJ : Erlbaum. pp67-113.
- Wilson, Jame W. (1971). An Evaluation of Learning in Secondary School Mathematics. *Handbook on Formative Evaluation of Student Learning*. USA : McGraw-Hill.
- Wright, B.D. ; & M.H. Stone. (1979). *Best Test Design*. Chicago : Mesa Press.

Zieky, Micheal. (1993). Practical Questions in the use of DIF Statistics in Test Development. *Differential Item Functioning*. Hillsdale, New Jersey : Lawrence Erlbaum Associates. Pp337-347.

Zwick, R.; Thayer, D.T. ; & Mazzeo. J. (1997). Descriptive and Inferential Procedures for Assessing Differential Item Functioning in Polytomous Items, *Applied Measurement in Education*. 10(4) : pp321-344.

ภาคผนวก

ภาคผนวก ก

รายนามผู้เชี่ยวชาญในการตรวจสอบคุณภาพเครื่องมือ

รายชื่อผู้เชี่ยวชาญในการตรวจสอบคุณภาพเครื่องมือ

1. ดร. สุวพร เข้มแข็ง ภาควิชาการวัดผลและวิจัยทางการศึกษา
คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
2. ผอ. อุไรวรรณ วิเศษสรรค์ ผู้อำนวยการโรงเรียนบ้านปากลำปีลึก
สำนักงานเขตพื้นที่การศึกษากาญจนบุรี เขต 3
3. รศ. สัตยา รัตนวราภักษ์ กลุ่มสาระการเรียนรู้คณิตศาสตร์
โรงเรียนสาธิตมหาวิทยาลัยศรีนครินทรวิโรฒ ปทุมวัน
4. ผศ. นภาลักษณ์ รุ่งสุวรรณ กลุ่มสาระการเรียนรู้คณิตศาสตร์
โรงเรียนสาธิตมหาวิทยาลัยศรีนครินทรวิโรฒ ปทุมวัน
5. นายไมตรี สมบูรณ์ หัวหน้ากลุ่มสาระการเรียนรู้คณิตศาสตร์
โรงเรียนจุฬาภรณราชวิทยาลัย ชลบุรี

ภาคผนวก ข

แบบทดสอบโจทย์ปัญหาในกลุ่มสาระการเรียนรู้วิชาคณิตศาสตร์

แบบทดสอบโจทย์ปัญหากลุ่มสาระการเรียนรู้วิชาคณิตศาสตร์
ตามหลักสูตรขั้นพื้นฐาน พุทธศักราช 2544 ในช่วงชั้นที่ 2

- คำชี้แจง** 1. แบบทดสอบมีทั้งหมด 1 ฉบับ จำนวน 60 ข้อ ใช้เวลา 150 นาที
2. คำถามทุกข้อเป็นแบบปรนัยชนิดเลือกตอบ 4 ตัวเลือก ให้นักเรียนเลือกคำตอบที่ถูกต้องที่สุดเพียงข้อเดียว จากข้อ ก ข ค หรือ ง โดยกาเครื่องหมาย X ลงในช่องสี่เหลี่ยมใต้ข้อที่นักเรียนเลือกในกระดาษคำตอบ ดังตัวอย่างต่อไปนี้

(0) ซื้อเข็มกานจวนภิเชกอันละ 50 บาท ขายไปอันละ 70 บาท ถ้าซื้อมา 40 อัน จะได้กำไรกี่บาท?

ก. 1,500 บาท

ข. 800 บาท

ค. 700 บาท

ง. 150 บาท

จะเห็นว่าคำตอบที่ถูกต้องคือ ข้อ ข ให้นักเรียนกาเครื่องหมาย X ลงในช่องสี่เหลี่ยมให้ตรงกับข้อ

(0) ได้ตัวเลือก ข ในกระดาษคำตอบดังนี้

ข้อ	ก	ข	ค	ง
0		X		

3. ให้นักเรียนทดเลขในกระดาษที่แจกให้ ห้ามนักเรียนทำเครื่องหมายใด ๆ ลงในแบบทดสอบนี้

แบบทดสอบโจทย์ปัญหากลุ่มสาระการเรียนรู้วิชาคณิตศาสตร์
ตามหลักสูตรขั้นพื้นฐาน พุทธศักราช 2544 ในช่วงชั้นที่ 2

1. ขายปลาทุ 34 เข่ง ได้เงินทั้งหมด 408 บาท ขายปลาทุเข่งละเท่าไร ?
 ก. 11 บาท ข. 12 บาท ค. 13 บาท ง. 14 บาท
2. เดือนมกราคมสมชายจ่ายค่าอาหาร 6,980 บาท จ่ายค่าน้ำมันรถ 1,295 บาท จ่ายค่าเบ็ดเตล็ด 3,086 บาท เดือนนี้สมชายจ่ายเงินทั้งหมดเท่าไร ?
 ก. 11,361 บาท ข. 11,631 บาท ค. 13,161 บาท ง. 16,131 บาท
3. ปัจจุบัน พ.ศ. 2550 อรดีมีอายุ 21 ปี อรดีเกิดเมื่อพ.ศ. ไດ ?
 ก. พ.ศ. 2522 ข. พ.ศ. 2525 ค. พ.ศ. 2529 ง. พ.ศ. 2550
4. มีข้าวโพดอ่อน 630 ฟัก บรccoli 14 ฟัก จะต้องใช้ถุงกี่ถุง ?
 ก. 43 ถุง ข. 44 ถุง ค. 45 ถุง ง. 46 ถุง
5. มีดอกกุหลาบอยู่ 132 กำ แต่ละกำมีดอกกุหลาบ 25 ดอก มีดอกกุหลาบทั้งหมดกี่ดอก ?
 ก. 3,000 ดอก ข. 3,300 ดอก ค. 3,600 ดอก ง. 3,900 ดอก
6. กำหนดเงินรางวัลให้แก่ผู้แข่งขัน รางวัลที่ 1 เป็นเงิน 25,000 บาท รางวัลที่ 2 เป็นเงิน 10,080 บาท รางวัลที่ 3 เป็นเงิน 7,750 บาท แต่ปรากฏว่ามีผู้ได้รับรางวัลที่ 1 หนึ่งคน รางวัลที่ 2 ถึง 2 ราย ส่วนรางวัลที่ 3 ไม่มีผู้ได้รับ ดังนั้นต้องหาเงินรางวัลเพิ่มขึ้นอีกกี่บาท ?
 ก. 2,330 บาท ข. 19,780 บาท ค. 27,95 บาท ง. 30,025 บาท
7. อุดรปลูกมะเขือได้ 255 แถว อุดรปลูกได้ 209 แถว ถ้าแต่ละแถวมีมะเขือ 22 ต้น อุดรและ อุดรปลูกมะเขือรวมกันทั้งหมดกี่ต้น ?
 ก. 10,800 ต้น ข. 10,108 ต้น ค. 10,280 ต้น ง. 10,208 ต้น
8. สุรัตน์เก็บมะขามได้ 2,800 ฟัก ขายไป 640 ฟัก ที่เหลือแบ่งใส่ถุง ๆ ละ 10 ฟัก จะได้ทั้งหมดกี่ถุง ?
 ก. 214 ถุง ข. 215 ถุง ค. 216 ถุง ง. 217 ถุง
9. ขายมะละกอ 314 ผล ราคาผลละ 7 บาท นำเงินไปซื้อนม 8 โหล ราคาโหลละ 84 บาท เหลือเงินกี่บาท ?
 ก. 1,346 บาท ข. 1,436 บาท ค. 1,526 บาท ง. 1,636 บาท
10. ซื้อเนื้อปู 3 กิโลกรัม กิโลกรัมละ 110 บาท ซื้อเนื้อหมู 2 กิโลกรัม กิโลกรัมละ 65 บาท ให้ธนบัตร 1,000 บาท จะได้รับเงินทอนเท่าไร ?
 ก. 540 บาท ข. 530 บาท ค. 520 บาท ง. 460 บาท

11. รถยนต์คันหนึ่งเมื่อนำมาขับจะสิ้นเปลืองน้ำมันในอัตรา $10\frac{1}{2}$ กิโลเมตรต่อลิตร ถ้ามีน้ำมันอยู่ $20\frac{4}{5}$ ลิตร จะขับไปได้ระยะทางไกลเท่าไร ?
 ก. $200\frac{4}{10}$ กิโลเมตร ข. $215\frac{1}{2}$ กิโลเมตร ค. $218\frac{2}{5}$ กิโลเมตร ง. $230\frac{1}{5}$ กิโลเมตร
12. มีน้ำหวาน $\frac{8}{9}$ ของขวด แบ่งให้น้อง $\frac{2}{9}$ ของขวด เหลือน้ำหวานในขวดเท่าไร ?
 ก. $\frac{11}{9}$ ขวด ข. $\frac{6}{9}$ ขวด ค. $\frac{2}{3}$ ขวด ง. $\frac{9}{6}$ ขวด
13. กิตติมีเงิน 500 บาท นำเงินซื้ออุปกรณ์การเรียน $\frac{3}{4}$ ของเงินที่มี กิตติเหลือเงินเท่าไร ?
 ก. 125 บาท ข. 250 บาท ค. 375 บาท ง. 425 บาท
14. มีผักหวาน 99 กิโลกรัม แบ่งใส่ถุง ถุงละ $2\frac{3}{4}$ กิโลกรัม จะได้ทั้งหมดกี่ถุง ?
 ก. 35 ถุง ข. 36 ถุง ค. 37 ถุง ง. 38 ถุง
15. วีระและอำพลช่วยกันปลูกผักได้ $\frac{7}{10}$ ของแปลง แต่วีระปลูกได้ $\frac{4}{10}$ ของแปลง ดังนั้นอำพลปลูกผักได้เป็นเศษส่วนเท่าไรของแปลง ?
 ก. $\frac{3}{10}$ ของแปลง ข. $\frac{4}{10}$ ของแปลง ค. $\frac{5}{10}$ ของแปลง ง. $\frac{7}{10}$ ของแปลง
16. ซีนมมา 3 โหล ตีหมดไปแล้ว $\frac{3}{4}$ ของโหล ตีมนมไปที่กล่อง ?
 ก. 22 กล่อง ข. 24 กล่อง ค. 26 กล่อง ง. 27 กล่อง
17. ขนมปังกรอบถุงหนึ่งหนัก $\frac{3}{10}$ กิโลกรัม ขนมปังกรอบ 5 ถุง จะหนักเท่าไร ?
 ก. $\frac{15}{10}$ กิโลกรัม ข. $\frac{16}{10}$ กิโลกรัม ค. $\frac{18}{10}$ กิโลกรัม ง. $\frac{19}{10}$ กิโลกรัม
18. เลี้ยงเป็ด 150 ตัว ออกไข่ทุกวัน วันละ $\frac{4}{5}$ ของเป็ดทั้งหมด เวลา 1 เดือน จะได้ไข่กี่ฟอง ?
 ก. 3,400 ฟอง ข. 3,500 ฟอง ค. 3,600 ฟอง ง. 3,700 ฟอง
19. แม่ค้าซื้อพุทรา 25 กิโลกรัม ราคา กิโลกรัมละ $3\frac{1}{4}$ บาท ขายไป กิโลกรัมละ $5\frac{1}{2}$ บาท จะได้กำไรเท่าไร ?
 ก. $56\frac{1}{4}$ บาท ข. $81\frac{1}{4}$ บาท ค. $137\frac{1}{2}$ บาท ง. $218\frac{3}{4}$ บาท

20. พ่อค้าซื้อน้ำตาลมะพร้าวมา 2 ปีบ ปีบแรกหนัก $52\frac{3}{5}$ กิโลกรัม ปีบที่สองหนัก 60 กิโลกรัม
ราคากิโลกรัมละ 10 บาท ต้องจ่ายเงินเท่าไร ?
ก. 1,011 บาท ข. 1,020 บาท ค. 1,106 บาท ง. 1,126 บาท
21. บัญชาเก็บไข่มา 280 ฟอง ทำไข่ดาวไป $\frac{1}{7}$ ของที่เก็บมา ใช้ทำขนมไปอีก $\frac{1}{2}$ ของที่เหลือ บัญชา
ใช้ไข่ทำขนมไปอีกกี่ฟอง ?
ก. 40 ฟอง ข. 120 ฟอง ค. 180 ฟอง ง. 240 ฟอง
22. วันแรกราณีอ่านหนังสือได้ $10\frac{1}{2}$ หน้า วันที่สองอ่านได้ $12\frac{2}{5}$ หน้า เขาอ่านหนังสือเฉลี่ยวันละ
กี่หน้า ?
ก. $10\frac{9}{20}$ หน้า ข. $11\frac{9}{20}$ หน้า ค. $12\frac{9}{20}$ หน้า ง. $13\frac{9}{20}$ หน้า
23. ในบึงแห่งหนึ่งมีงูคิดเป็น $\frac{1}{20}$ ของจำนวนกบในบึง ถ้าจับกบไป 10 ตัวแล้ว ยังเหลือกบอีก 250
ตัว ในบึงมีงูกี่ตัว ?
ก. 11 ตัว ข. 13 ตัว ค. 15 ตัว ง. 17 ตัว
24. นักเรียนชั้น ป.6 มี 50 คน เป็นนักเรียนหญิง $\frac{2}{5}$ ของนักเรียนทั้งหมด ในวันหนึ่งมีนักเรียนชายไม่
มาเรียน 4 คน อยากทราบว่าวันนั้นมีนักเรียนชายมาเรียนกี่คน ?
ก. 20 คน ข. 26 คน ค. 30 คน ง. 34 คน
25. แม่ค้ามีผลไม้ 70 กิโลกรัม ขายครั้งแรก $5\frac{1}{2}$ กิโลกรัม ขายครั้งที่สอง $28\frac{4}{5}$ กิโลกรัม แม่ค้าเหลือ
ผลไม้อยู่เท่าไร ?
ก. $34\frac{3}{10}$ กิโลกรัม ข. $35\frac{7}{10}$ กิโลกรัม ค. $36\frac{3}{10}$ กิโลกรัม ง. $36\frac{7}{10}$ กิโลกรัม
26. เส้นผมของวิกานดายาว 1.39 เซนติเมตรต่อเดือน ในเวลา 1 ปี ถ้าวิกานดาไม่ตัดผม เส้นผม
ของวิกานดาจะยาวเท่าไร ?
ก. 13.39 เซนติเมตร ข. 16.68 เซนติเมตร ค. 31.39 เซนติเมตร ง. 41.70 เซนติเมตร
27. มีน้ำมันเบนซิน 68 ขวด ถ้า 1 ขวด เท่ากับ 1.5 ลิตร มีน้ำมันอยู่ที่ลิตร ?
ก. 103 ลิตร ข. 102 ลิตร ค. 101 ลิตร ง. 99 ลิตร
28. กัลยามีที่ดิน 10.8 ไร่ แบ่งที่ดินแปลงละ 1.2 ไร่ จะได้ทั้งหมดกี่แปลง ?
ก. 12 แปลง ข. 11 แปลง ค. 10 แปลง ง. 9 แปลง

29. ข้าวสารหนักถุงละ 4.5 กิโลกรัม ข้าวสาร 250 ถุง จะหนักเท่าไร ?
 ก. 1,025 กิโลกรัม ข. 1,052 กิโลกรัม ค. 1,125 กิโลกรัม ง. 1,215 กิโลกรัม
30. ละครุดราคา กิโลกรัมละ 25 บาท ซื้อละครุด 5.6 กิโลกรัม ต้องจ่ายเงินเท่าไร ?
 ก. 130 บาท ข. 140 บาท ค. 150 บาท ง. 160 บาท
31. กัลยามีเงิน 30.25 บาท กาญจนามีเงินมากกว่ากัลยา 7.50 บาท กาญจนามีเงินเท่าไร ?
 ก. 33.25 บาท ข. 35.50 บาท ค. 37.75 บาท ง. 38.25 บาท
32. มีขนมปัง 2.45 กิโลกรัม ต้องการแบ่งใส่ 7 ถุง ถุงละเท่า ๆ กัน แต่ละถุงมีขนมปังหนักกี่ กิโลกรัม ?
 ก. 0.15 กิโลกรัม ข. 0.25 กิโลกรัม ค. 0.35 กิโลกรัม ง. 0.45 กิโลกรัม
33. ปัญญามีเงิน 224.75 บาท อรดีมีเงิน 4.2 เท่าของปัญญา อรดีและปัญญามีเงินรวมกันกี่บาท ?
 ก. 943.95 บาท ข. 1,168.70 บาท ค. 1,240.50 บาท ง. 1,384.25 บาท
34. ไข่ไก่ราคาฟองละ 1.50 บาท ไข่เป็ดราคาฟองละ 1.75 บาท ซื้อไข่ไก่ 13 ฟอง ไข่เป็ด 14 ฟอง จะต้องจ่ายเงินเท่าไร ?
 ก. 19.50 บาท ข. 24.50 บาท ค. 44.00 บาท ง. 44.50 บาท
35. ปรีชาขายพุทรา 28 กิโลกรัม ราคา กิโลกรัม 13.75 บาท นำเงินไปซื้อไก่ 49.25 บาท ปรีชา เหลือเงินเท่าไร ?
 ก. 385.00 บาท ข. 358.00 บาท ค. 335.75 บาท ง. 323.25 บาท
36. ประกอบซื้อหนังสือ 3 เล่ม ราคาเล่มละ 11.25 บาท และซื้อปากกา 2 ด้าม ราคาด้ามละ 9.25 บาท ให้เงิน 100 บาท จะได้รับเงินทอนเท่าไร ?
 ก. 54.50 บาท ข. 53.75 บาท ค. 52.25 บาท ง. 47.75 บาท
37. ก๋วยเตี๋ยวเล็ก 4 ก๋วยเตี๋ยว และก๋วยเตี๋ยวใหญ่ 1 ก๋วยเตี๋ยว มีน้ำหนักรวมกัน 52.5 กิโลกรัม ถ้าก๋วยเตี๋ยวใหญ่หนัก 20.3 กิโลกรัม ก๋วยเตี๋ยวเล็กแต่ละก๋วยเตี๋ยวจะหนักเท่าไร ?
 ก. 8.00 กิโลกรัม ข. 8.05 กิโลกรัม ค. 16.10 กิโลกรัม ง. 32.20 กิโลกรัม
38. นมสดราคาถุงละ 3.75 บาท ซื้อ 9 ถุง ให้ธนบัตร ใบละ 50 บาท จะได้รับเงินทอนเท่าไร ?
 ก. 15.25 บาท ข. 16.25 บาท ค. 16.75 บาท ง. 33.75 บาท
39. คุณพ่อมีเงิน 5,000.75 บาท แบ่งให้คุณแม่ 2,000.25 บาท ที่เหลือแบ่งให้ลูก 2 คน คนละเท่า ๆ กัน ลูกจะได้รับคนละเท่าไร ?
 ก. 1,400.25 บาท ข. 1,500.00 บาท ค. 1,500.25 บาท ง. 3,000.50 บาท

40. หนังสือพิมพ์เก่า 15 ฉบับ หน้า 1 กิโลกรัม ถ้ามีหนังสือพิมพ์เก่า 456 ฉบับ จะขายได้เงินเท่าไร ถ้าขายได้ในราคา กิโลกรัมละ 2.50 บาท ?
- ก. 76.50 บาท ข. 76.25 บาท ค. 76.00 บาท ง. 30.50 บาท
41. เสื้อตัวหนึ่งราคา 500 บาท ลดให้ผู้ซื้อ 20 % ผู้ซื้อต้องจ่ายเงินเท่าใด ?
- ก. 100 บาท ข. 200 บาท ค. 400 บาท ง. 420 บาท
42. ใหม่บอกขายเครื่องคอมพิวเตอร์ในราคา 25,000 บาท แต่เสียค่านายหน้า 2% ใหม่เสียค่านายหน้ากี่บาท ?
- ก. 100 บาท ข. 200 บาท ค. 400 บาท ง. 500 บาท
43. ฝากเงินไว้ 11,000 บาท ธนาคารให้ดอกเบี้ยร้อยละ 4.5 ต่อปี สิ้นปีจะได้ดอกเบี้ยเท่าไร ?
- ก. 505 บาท ข. 495 บาท ค. 465 บาท ง. 475 บาท
44. ขายเสื้อยืดราคา ตัวละ 84 บาท ได้กำไร 5 % ถ้าต้องการขายให้ได้กำไร 20 % ต้องขายเสื้อยืดตัวละเท่าไร ?
- ก. 90 บาท ข. 94 บาท ค. 96 บาท ง. 98 บาท
45. มีเงิน 160,800 บาท ซื้อที่ดินได้ 96 ตารางวา ที่ดินราคาตารางวาละเท่าไร ?
- ก. 1,345 บาท ข. 1,675 บาท ค. 1,765 บาท ง. 1,795 บาท
46. หน้าต่างบานหนึ่งกว้าง 22 เซนติเมตร ยาว 38 เซนติเมตร ความยาวรอบขอบหน้าต่างเป็นเท่าไร ?
- ก. 50 เซนติเมตร ข. 60 เซนติเมตร ค. 100 เซนติเมตร ง. 120 เซนติเมตร
47. สมหวังวิ่งรอบสนามที่มีด้านกว้าง 45 เมตร ด้านยาว 60 เมตร ถ้าวิ่ง 3 รอบ จะวิ่งได้ระยะทางเท่าไร ?
- ก. 210 เมตร ข. 315 เมตร ค. 360 เมตร ง. 630 เมตร
48. ผ้าเช็ดตัวผืนหนึ่งกว้าง 80 เซนติเมตร ยาว 150 เซนติเมตร ถ้าตัดแบ่งออกเป็น 2 ผืนตามเส้นทแยงมุมแล้วแต่ละผืนจะมีพื้นที่เท่าไร ?
- ก. 460 ตารางเซนติเมตร ข. 920 ตารางเซนติเมตร
ค. 6,000 ตารางเซนติเมตร ง. 12,000 ตารางเซนติเมตร
49. มีธนบัตรฉบับละหนึ่งร้อยบาท 1 ฉบับ แลกเหรียญห้าบาทได้กี่เหรียญ ?
- ก. 10 เหรียญ ข. 15 เหรียญ ค. 20 เหรียญ ง. 25 เหรียญ
50. วีระมีเหรียญ 5 บาท 13 เหรียญ ธนบัตรใบละ 20 บาท 7 ใบ วีระมีเงินกี่บาท ?
- ก. 185 บาท ข. 190 บาท ค. 200 บาท ง. 205 บาท

51. ปราณี่มีเงิน 700 บาท ต้องการซื้อกระเป๋าราคา 250 บาท รองเท้าราคา 580 บาท จะต้องหาเงินมาเพิ่มอีกกี่บาทจึงจะพอซื้อ ?
 ก. 130 บาท ข. 150 บาท ค. 180 บาท ง. 200 บาท
52. อุไรมีเงิน 375 บาท แบ่งใช้ 5 วัน วันละเท่า ๆ กัน อุไรจะต้องใช้เงินวันละกี่บาท ?
 ก. 60 บาท ข. 65 บาท ค. 70 บาท ง. 75 บาท
53. ขณะนี้เวลา 10.05 นาฬิกา อีก 35 นาที ต่อมาจะเป็นเวลา เท่าไร ?
 ก. 10.30 นาฬิกา ข. 10.35 นาฬิกา ค. 10.40 นาฬิกา ง. 10.45 นาฬิกา
54. ขณะนี้เวลา 19.55 นาฬิกา อีกกี่นาฬิกาจะเป็นเวลา 20.15 นาฬิกา ?
 ก. 5 นาที ข. 10 นาที ค. 15 นาที ง. 20 นาที
55. นิยมซื้อรถยนต์คันหนึ่งโดยผ่อนส่งเป็นเวลา 55 เดือน อยากทราบว่านิยมจะต้องผ่อนรถเป็นเวลานานกี่ปี กี่เดือน ?
 ก. 4 ปี 5 เดือน ข. 4 ปี 7 เดือน ค. 5 ปี 5 เดือน ง. 5 ปี 7 เดือน
56. กานต์นอนเวลา 21.10 นาฬิกา ตื่นนอนเวลา 6.30 นาฬิกา กานต์ใช้เวลาอนกี่ชั่วโมง ?
 ก. 9 ชั่วโมง 20 นาที ข. 9 ชั่วโมง 25 นาที ค. 8 ชั่วโมง 15 นาที ง. 8 ชั่วโมง 20 นาที
57. พ่อค้ามีขนมอยู่จำนวนหนึ่ง ซื้อมาอีก 315 ชิ้น ทำให้พ่อค้ามีขนม 682 ชิ้น เดิมพ่อค้ามีขนมกี่ชิ้น ?
 ก. 997 ชิ้น ข. 377 ชิ้น ค. 373 ชิ้น ง. 367 ชิ้น
58. แพรมีลูกหินอยู่ 512 ลูก แบ่งให้เพื่อนไปจำนวนหนึ่ง ทำให้เหลือลูกหิน 207 ลูก แพรแบ่งลูกหินให้เพื่อนกี่ลูก ?
 ก. 215 ลูก ข. 305 ลูก ค. 315 ลูก ง. 719 ลูก
59. พ่อมีที่ดินจำนวนหนึ่ง แบ่งให้ลูก 4 คนได้รับคนละ 245 ตารางวา พ่อมีที่ดินกี่ตารางวา ?
 ก. 950 ตารางวา ข. 980 ตารางวา ค. 990 ตารางวา ง. 995 ตารางวา
60. สมศักดิ์ขับรถยนต์ได้ระยะทางวันละ A กิโลเมตร ขับรถยนต์นาน 1 สัปดาห์ ได้ระยะทางทั้งหมด 476 กิโลเมตร สมศักดิ์ขับรถยนต์ได้วันละกี่กิโลเมตร ?
 ก. 63 กิโลเมตร ข. 65 กิโลเมตร ค. 68 กิโลเมตร ง. 74 กิโลเมตร

ภาคผนวก ค

ค่าพารามิเตอร์ของข้อสอบและผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน

ตารางที่ 24 ค่าพารามิเตอร์ของข้อสอบ แยกตามกลุ่มเชื้อชาติไทย – ไม่ใช่เชื้อชาติไทย

ข้อที่	พารามิเตอร์ข้อสอบ					
	a_1	b_1	a_2	b_2	a_t	b_t
1	0.804	3.652	0.378	5.474	0.651	3.789
2	0.277	-5.240	0.272	-5.188	0.125	-5.373
3	1.359	-1.980	0.251	8.401	0.586	4.448
4	1.081	2.787	0.264	5.988	0.756	2.872
5	0.747	3.931	0.330	5.192	0.904	2.726
6	1.434	-1.474	1.434	-0.414	-0.64	0.545
7	0.615	5.074	0.373	5.187	0.570	4.183
8	0.792	3.937	0.204	6.596	0.502	3.724
9	0.434	5.139	0.243	4.738	0.585	2.843
10	2.575	-1.274	0.127	-5.000	-0.77	0.592
11	0.228	6.283	0.484	0.647	0.441	1.755
12	0.072	-4.932	0.478	5.087	0.203	5.101
13	0.400	-1.923	0.086	-0.848	0.079	-4.598
14	0.894	2.499	0.261	3.797	0.691	2.200
15	0.153	-5.999	0.202	-5.386	0.093	-3.614
16	0.608	3.656	0.806	1.551	0.636	2.401
17	0.940	2.951	0.263	-4.578	0.089	-6.625
18	0.360	5.304	0.193	7.406	0.568	3.356
19	0.399	-1.354	0.399	-0.174	-0.23	0.392
20	0.581	1.389	0.366	4.006	0.334	4.905
21	0.467	2.682	0.196	4.751	0.213	3.772
22	0.669	1.950	0.221	4.523	0.267	4.066
23	0.593	2.972	0.419	2.103	0.520	2.036
24	0.460	2.952	0.197	5.753	0.354	3.873
25	0.732	2.226	0.734	1.632	0.392	2.955

ตาราง 24 (ต่อ)

ข้อที่	พารามิเตอร์ข้อสอบ					
	a_1	b_1	a_2	b_2	a_t	b_t
26	0.572	2.700	0.186	3.490	0.455	2.241
27	0.307	5.697	0.217	5.840	0.353	3.870
28	0.587	2.829	0.211	4.065	0.390	3.175
29	1.185	1.736	0.227	5.814	0.403	3.634
30	0.466	4.894	0.154	4.479	0.572	1.947
31	0.423	2.981	0.388	4.816	0.398	5.101
32	0.121	6.544	0.298	3.487	0.380	3.022
33	0.442	3.656	0.228	2.263	0.109	5.827
34	0.660	2.805	0.353	3.217	0.369	3.605
35	0.187	7.487	0.465	2.582	0.523	2.774
36	0.252	4.476	0.336	3.657	0.202	6.366
37	0.444	2.227	0.201	7.001	0.145	8.516
38	0.610	3.167	0.157	7.661	0.166	6.282
39	0.395	4.307	0.318	5.111	0.417	4.156
40	0.313	5.270	0.341	2.345	0.450	2.616
41	0.619	1.858	0.268	3.667	0.239	5.151
42	0.969	2.096	0.194	6.299	0.261	4.240
43	0.155	6.180	0.324	3.919	0.533	2.827
44	0.203	4.828	0.243	4.859	0.120	8.731
45	0.192	8.173	0.325	3.298	0.192	5.236
46	0.195	6.223	0.556	3.278	0.184	8.685
47	0.258	5.118	1.109	2.052	0.156	8.720
48	0.503	5.845	0.214	4.513	0.247	4.572
49	0.754	2.993	0.263	6.759	0.483	4.590
50	0.122	-1.154	0.215	4.840	0.708	2.284

ตาราง 24 (ต่อ)

ข้อที่	พารามิเตอร์ข้อสอบ					
	a_1	b_1	a_2	b_2	a_t	b_t
51	0.988	3.023	0.322	-2.780	0.074	-4.361
52	0.766	3.650	0.226	5.340	0.820	2.366
53	1.885	-1.746	1.885	-0.703	0.483	4.711
54	0.371	3.985	0.259	5.905	0.536	3.704
55	0.132	-8.923	0.350	2.536	0.495	2.417
56	0.860	3.483	0.180	-1.849	0.061	-1.020
57	0.854	3.379	0.317	2.004	0.746	2.580
58	1.140	2.916	0.171	3.699	0.531	3.960
59	1.203	3.239	0.238	6.696	0.603	3.506
60	1.660	-1.739	0.175	7.062	0.579	3.199
\bar{x}	0.542	0.536	0.305	0.645	0.402	0.214
S.D	0.299	0.287	0.171	0.324	0.404	0.541

ตาราง 25 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่าง 600 คน และจำนวน
ข้อสอบ 40 ข้อ

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ			
	a ₁	b ₁	a ₂	b ₂	MH	LC	CSA	CUA
1	0.391	3.652	0.378	5.474	0.97	2.54	-0.281	0.297
2	1.846	-8.24	0.272	-5.188	1.22	2.13	-0.254	0.226
3	0.658	2.787	0.251	8.401	1.65*	2.45	0.190	0.150
4	0.804	3.931	0.264	5.988	1.73*	2.42	-0.242	0.214
5	0.277	5.074	0.33	5.192	2.70***	19.86***	<u>-0.517</u>	<u>0.624</u>
6	1.081	3.937	0.373	5.187	0.97	1.45	0.920	0.817
7	0.434	5.139	0.204	6.596	0.70	1.26	-0.113	0.214
8	0.228	6.283	0.243	4.738	0.87	2.87	-0.238	0.233
9	0.400	-1.923	0.127	-5.000	1.25	1.59	-0.258	0.243
10	0.894	2.499	0.484	0.647	1.06	1.22	-0.286	0.298
11	0.153	-9.999	0.478	5.087	0.78	2.13	-0.192	0.192
12	0.608	3.656	0.086	-0.848	3.45***	19.57***	<u>-0.519</u>	<u>0.524</u>
13	0.94	2.951	0.261	3.797	0.47*	9.99**	0.263	0.172
14	0.36	5.304	0.202	-5.386	0.72	1.52	-0.163	0.154
15	0.581	1.389	0.806	1.551	1.25	1.98	0.186	0.188
16	0.467	2.682	0.263	-4.578	1.22	0.89	-0.265	0.243
17	0.669	1.950	0.193	7.406	1.13	1.77	-0.270	0.197
18	0.593	2.972	0.366	4.006	0.45	1.25	0.014	0.021
19	0.460	2.952	0.196	4.751	1.18	0.29	0.028	0.023
20	0.732	2.226	0.221	4.523	1.06	3.21	0.113	0.123
21	0.572	2.700	0.419	2.103	0.87	3.59	-0.238	0.241
22	0.307	5.697	0.197	5.753	1.02	1.45	0.204	0.264
23	0.587	2.829	0.734	1.632	0.72	1.26	-0.286	0.251
24	1.185	1.736	0.186	3.490	0.97	2.87	-0.150	0.178

ตาราง 25 (ต่อ)

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ			
	a ₁	b ₁	a ₂	b ₂	MH	LC	CSA	CUA
25	0.466	4.894	0.217	5.840	0.72	3.59	0.153	0.194
26	0.423	2.981	0.211	4.065	1.18	2.12	-0.224	0.294
27	0.121	6.544	0.227	5.814	1.01	0.72	0.001	0.012
28	0.442	3.656	0.154	4.479	1.24	1.25	0.091	0.224
29	0.660	2.805	0.388	4.816	1.13	3.27	-0.225	0.225
30	0.187	7.487	0.298	3.487	1.02	0.98	-0.178	0.199
31	0.252	4.476	0.228	2.263	1.22	3.36	-0.016	0.009
32	0.444	2.227	0.353	3.217	0.72	0.28	-0.179	0.188
33	0.610	3.167	0.465	2.582	0.36**	10.11**	<u>-0.335</u>	0.273
34	0.395	4.307	0.336	3.657	1.73*	8.22*	<u>0.317</u>	<u>0.428</u>
35	0.313	5.27	0.201	7.001	1.66*	6.88*	<u>0.407</u>	<u>0.483</u>
36	0.619	1.858	0.157	7.661	1.83**	7.48*	-0.201	<u>0.463</u>
37	0.969	2.096	0.318	5.111	0.78	0.98	0.001	0.002
38	0.155	6.18	0.341	2.345	0.53**	12.98**	<u>-0.448</u>	<u>0.526</u>
39	0.203	4.828	0.268	3.667	0.52**	9.73**	<u>-0.549</u>	<u>0.549</u>
40	0.192	8.173	0.194	6.299	1.59*	8.72*	<u>-0.545</u>	<u>0.595</u>
รวม					12	10	7	9

ตาราง 26 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่าง 750 คน และจำนวน
ข้อสอบ 40 ข้อ

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ			
	a ₁	b ₁	a ₂	b ₂	MH	LC	CSA	CUA
1	0.373	5.187	0.353	3.217	0.87	2.88	-0.245	0.326
2	0.204	6.596	0.465	2.582	1.22	2.15	-0.154	0.126
3	0.243	4.738	0.336	3.657	1.66*	2.45	0.190	0.150
4	0.127	-5.000	0.201	7.001	1.83*	2.42	-0.242	0.214
5	0.484	0.647	0.157	7.661	2.93***	33.52***	<u>0.425</u>	<u>0.446</u>
6	0.478	5.087	0.318	5.111	0.97	1.45	0.920	0.817
7	0.086	-0.848	0.341	2.345	0.70	1.26	-0.113	0.214
8	0.228	6.283	0.121	6.544	0.87	2.87	-0.238	0.233
9	0.400	-1.923	0.442	3.656	1.25	1.59	-0.258	0.243
10	0.894	2.499	0.660	2.805	1.06	1.22	-0.286	0.298
11	0.153	-9.999	0.187	7.487	0.78	2.13	-0.192	0.192
12	0.608	3.656	0.252	4.476	1.83**	12.82**	<u>0.385</u>	<u>0.410</u>
13	0.940	2.951	0.261	3.797	3.18***	15.86**	<u>0.421</u>	<u>0.429</u>
14	0.307	5.697	0.202	-5.386	1.25	6.08*	-0.221	0.221
15	0.587	2.829	0.806	1.551	0.97	1.98	0.186	<u>0.538</u>
16	1.185	1.736	0.263	-4.578	1.22	0.89	-0.265	0.243
17	0.593	2.972	0.193	7.406	1.13	1.77	-0.270	0.197
18	0.460	2.952	0.366	4.006	0.45	1.25	0.014	0.021
19	0.732	2.226	0.196	4.751	1.18	0.29	0.028	0.023
20	0.388	4.816	0.221	4.523	1.06	3.21	0.113	0.123
21	0.298	3.487	0.419	2.103	0.87	3.59	-0.238	0.241
22	0.228	2.263	0.197	5.753	1.02	1.45	0.204	0.264
23	0.378	5.474	0.734	1.632	0.72	1.26	-0.286	0.251
24	0.608	3.656	0.186	3.490	0.97	2.87	-0.150	0.178

ตาราง 26 (ต่อ)

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ			
	a ₁	b ₁	a ₂	b ₂	MH	LC	CSA	CUA
25	0.360	5.304	0.819	0.220	1.10	2.01	-0.143	0.154
26	0.581	1.389	0.913	0.045	1.18	2.12	-0.224	0.294
27	0.467	2.682	0.890	0.8484	1.01	0.72	0.001	0.012
28	0.669	1.950	0.615	0.531	1.24	1.25	0.091	0.224
29	0.593	2.972	0.716	1.403	1.13	3.27	-0.225	0.225
30	0.460	2.952	0.549	1.519	1.02	0.98	-0.178	0.199
31	0.732	2.226	0.661	1.422	1.22	3.36	-0.016	0.009
32	0.572	2.700	0.378	5.474	0.68*	6.83*	-0.150	0.192
33	0.610	3.167	0.272	-5.188	0.59**	6.02*	0.001	0.062
34	0.395	4.307	0.251	8.401	1.83**	12.35**	<u>0.579</u>	<u>0.585</u>
35	0.313	5.27	0.264	5.988	1.75**	9.42**	<u>0.375</u>	<u>0.476</u>
36	0.619	1.858	0.33	5.192	1.66*	7.25*	-0.192	<u>0.503</u>
37	0.969	2.096	0.466	4.894	1.11	0.76	0.024	0.081
38	0.155	6.18	0.423	2.981	2.15***	12.51**	<u>-0.375</u>	<u>0.456</u>
39	0.203	4.828	0.268	3.667	2.93***	16.04***	<u>0.362</u>	<u>0.506</u>
40	0.672	0.168	2.782	2.176	2.69**	10.51**	<u>0.437</u>	<u>0.528</u>
รวม					13	12	8	10

ตาราง 27 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่าง 600 คน และจำนวน
ข้อสอบ 60 ข้อ

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ			
	a_1	b_1	a_2	b_2	MH	LC	CSA	CUA
1	0.557	0.165	0.523	-0.241	1.22	0.12	-0.224	0.294
2	0.155	6.180	0.298	3.487	1.28	0.81	0.001	0.012
3	0.203	4.828	0.228	2.263	1.35	0.59	0.091	0.224
4	0.192	8.173	0.353	3.217	0.63*	3.78	-0.225	0.225
5	0.195	6.223	0.465	2.582	0.34*	3.09	-0.178	0.199
6	0.258	5.118	0.336	3.657	3.18**	14.86***	<u>-0.340</u>	<u>0.667</u>
7	0.503	5.845	0.201	7.001	1.03	0.44	0.092	0.144
8	0.754	2.993	0.157	7.661	1.75**	10.40**	<u>-0.359</u>	<u>0.411</u>
9	0.122	-1.154	0.318	5.111	1.91	2.18	-0.296	0.363
10	0.988	3.023	0.341	2.345	0.79	0.32	-0.168	0.214
11	0.153	-5.999	0.317	2.004	1.11	5.52	-0.183	0.233
12	0.608	3.656	0.171	3.699	4.32***	14.39***	<u>0.401</u>	<u>0.401</u>
13	0.940	2.951	0.238	6.696	0.47	0.96	-0.243	0.342
14	0.360	5.304	1.093	-0.173	0.64	1.23	0.014	0.052
15	0.399	-1.354	0.272	-5.188	0.74	3.28	-0.265	0.298
16	1.359	-1.980	0.251	8.401	1.36	1.35	-0.270	0.341
17	1.081	2.787	0.264	5.988	2.69**	16.12***	<u>-0.355</u>	<u>0.420</u>
18	0.747	3.931	0.330	5.192	0.47*	7.62*	0.298	<u>0.576</u>
19	1.434	-1.474	1.434	-0.414	0.89	2.14	-0.201	0.243
20	0.615	5.074	0.373	5.187	1.09	2.29	-0.175	0.197
21	0.792	3.937	0.204	6.596	1.28	5.21	-0.158	0.180
22	0.395	4.307	0.734	1.632	0.72	3.27	0.001	0.003
23	0.313	5.270	0.186	3.490	0.53	4.29	-0.189	0.215
24	0.619	1.858	0.217	5.840	1.04	2.89	-0.132	0.132

ตาราง 27 (ต่อ)

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ			
	a ₁	b ₁	a ₂	b ₂	MH	LC	CSA	CUA
25	0.593	2.972	0.806	1.551	1.90	2.13	-0.121	0.152
26	0.460	2.952	0.263	-4.578	1.27	2.45	-0.153	0.194
27	0.732	2.226	0.193	7.406	1.35	2.42	-0.251	0.294
28	0.572	2.700	0.399	-0.174	1.03	1.45	0.012	0.012
29	0.307	5.697	0.366	4.006	1.91	1.26	-0.224	0.224
30	0.587	2.829	0.196	4.751	0.79	2.87	0.001	0.002
31	0.228	6.283	0.221	4.523	1.11	1.46	-0.187	0.230
32	0.434	5.139	0.419	2.103	0.64	1.59	0.091	0.143
33	2.575	-1.274	0.197	5.753	1.22	1.22	-0.225	0.225
34	0.277	-5.240	0.268	3.667	1.25	2.13	-0.178	0.224
35	0.072	-4.932	0.194	6.299	0.72	1.52	-0.016	0.154
36	0.400	-1.923	0.324	3.919	1.06	1.98	-0.227	0.347
37	0.894	2.499	0.243	4.859	0.97	2.15	-0.179	0.214
38	1.185	1.736	0.215	4.840	0.70	1.04	-0.136	0.172
39	0.466	4.894	0.322	-2.780	1.18	2.19	-0.149	0.188
40	0.423	2.981	0.226	5.340	0.87	1.96	0.001	0.001
41	0.121	6.544	1.885	-0.703	1.01	0.89	-0.201	0.243
42	0.442	3.656	0.259	5.905	1.24	1.77	-0.175	0.197
43	0.660	2.805	0.350	2.536	0.78	2.01	-0.158	0.180
44	0.187	7.487	0.180	-1.849	1.02	6.03*	<u>0.421</u>	<u>0.429</u>
45	0.252	4.476	0.243	4.738	1.22	1.25	0.001	0.003
46	0.444	2.227	0.127	-5.000	1.25	0.29	-0.189	0.215
47	0.610	3.167	0.484	0.647	0.72	2.54	-0.132	0.132
48	0.969	2.096	0.478	5.087	0.97	1.02	-0.230	0.267
49	0.132	-8.923	0.086	-0.848	2.62***	17.79***	<u>-0.586</u>	<u>0.724</u>

ตาราง 27 (ต่อ)

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ			
	a ₁	b ₁	a ₂	b ₂	MH	LC	CSA	CUA
50	0.860	3.483	0.325	3.298	2.93 ^{***}	0.54	-0.283	0.254
51	0.854	3.379	0.556	3.278	1.83 ^{**}	1.53	-0.245	0.297
52	0.766	3.650	1.109	2.052	0.53 ^{**}	10.51 ^{**}	<u>-0.345</u>	<u>0.407</u>
53	1.885	-1.746	0.214	4.513	0.57 ^{**}	10.37 ^{**}	<u>-0.419</u>	<u>0.507</u>
54	0.371	3.985	0.263	6.759	0.72	1.87	-0.293	0.246
55	1.140	2.916	0.211	4.065	1.75 ^{**}	6.18 [*]	0.092	0.144
56	1.203	3.239	0.227	5.814	0.97	4.98	0.028	0.048
57	1.685	-0.548	0.154	4.479	2.69 ^{**}	9.07 [*]	-0.238	<u>0.465</u>
58	0.581	1.389	0.388	4.816	2.62 ^{***}	16.12 ^{***}	<u>0.579</u>	<u>0.585</u>
59	0.467	2.682	0.261	3.797	0.31 ^{***}	16.81 ^{***}	<u>0.385</u>	<u>0.410</u>
60	0.669	1.950	0.202	-5.386	3.38 ^{***}	17.59 ^{***}	<u>-0.406</u>	<u>0.445</u>
รวม					17	14	11	13

ตาราง 28 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่าง 750 คน และจำนวน
ข้อสอบ 60 ข้อ

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ			
	a_1	b_1	a_2	b_2	MH	LC	CSA	CUA
1	0.804	3.652	0.378	5.474	1.18	2.19	-0.149	0.188
2	0.277	-5.240	0.272	-5.188	0.87	1.96	0.001	0.001
3	1.359	-1.980	0.251	8.401	1.01	0.89	-0.201	0.243
4	1.081	2.787	0.264	5.988	2.93 ^{***}	10.22 ^{**}	-0.297	<u>0.562</u>
5	0.747	3.931	0.330	5.192	0.68 [*]	6.29 [*]	0.191	0.234
6	1.434	-1.474	1.434	-0.414	1.83 ^{**}	10.54 ^{**}	<u>-0.312</u>	<u>0.483</u>
7	0.615	5.074	0.373	5.187	0.72	1.87	-0.293	0.246
8	0.792	3.937	0.204	6.596	1.83 ^{**}	10.19 ^{**}	<u>-0.406</u>	<u>0.445</u>
9	0.434	5.139	0.243	4.738	0.40 [*]	6.69 [*]	-0.187	0.281
10	2.575	-1.274	0.127	-5.000	0.47 [*]	7.99 [*]	-0.094	0.094
11	0.228	6.283	0.484	0.647	0.87	2.54	-0.040	0.241
12	0.072	-4.932	0.478	5.087	0.59 ^{**}	11.99 ^{**}	<u>-0.418</u>	<u>0.457</u>
13	0.400	-1.923	0.086	-0.848	0.12	5.84	-0.187	0.281
14	0.894	2.499	0.261	3.797	0.74	1.43	-0.094	0.094
15	0.153	-5.999	0.202	-5.386	0.69	1.34	-0.078	0.131
16	0.608	3.656	0.806	1.551	1.23	2.42	-0.040	0.241
17	0.940	2.951	0.263	-4.578	0.38 ^{***}	13.94 ^{***}	<u>-0.437</u>	<u>0.526</u>
18	0.360	5.304	0.193	7.406	1.75 ^{**}	11.19 ^{**}	<u>-0.36</u>	<u>0.412</u>
19	0.399	-1.354	0.399	-0.174	0.72	1.52	-0.016	0.154
20	0.581	1.389	0.366	4.006	1.06	1.98	-0.227	0.347
21	0.467	2.682	0.196	4.751	0.97	2.15	-0.179	0.214
22	0.669	1.950	0.221	4.523	0.70	1.04	-0.136	0.172
23	0.593	2.972	0.419	2.103	1.18	2.19	-0.149	0.188
24	0.460	2.952	0.197	5.753	0.87	1.96	0.001	0.001

ตาราง 28 (ต่อ)

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ			
	a ₁	b ₁	a ₂	b ₂	MH	LC	CSA	CUA
25	0.732	2.226	0.734	1.632	1.90	2.13	-0.121	0.152
26	0.572	2.700	0.186	3.490	1.27	2.45	-0.153	0.194
27	0.307	5.697	0.217	5.840	1.35	2.42	-0.251	0.294
28	0.587	2.829	0.211	4.065	1.03	1.45	0.012	0.012
29	1.185	1.736	0.227	5.814	1.91	1.26	-0.224	0.224
30	0.466	4.894	0.154	4.479	0.79	2.87	0.001	0.002
31	0.423	2.981	0.388	4.816	1.11	1.46	-0.187	0.230
32	0.121	6.544	0.298	3.487	0.64	1.59	0.091	0.143
33	0.442	3.656	0.228	2.263	1.22	1.22	-0.225	0.225
34	0.660	2.805	0.353	3.217	1.25	2.13	-0.178	0.224
35	0.187	7.487	0.465	2.582	0.72	1.52	-0.016	0.154
36	0.252	4.476	0.336	3.657	1.06	1.98	-0.227	0.347
37	0.444	2.227	0.201	7.001	0.97	2.15	-0.179	0.214
38	0.610	3.167	0.157	7.661	0.70	1.04	-0.136	0.172
39	0.395	4.307	0.318	5.111	1.18	2.19	-0.149	0.188
40	0.313	5.270	0.341	2.345	0.87	1.96	0.001	0.001
41	0.619	1.858	0.268	3.667	1.01	0.89	-0.201	0.243
42	0.969	2.096	0.194	6.299	1.24	1.77	-0.175	0.197
43	0.155	6.180	0.324	3.919	0.78	2.01	-0.158	0.180
44	0.203	4.828	0.243	4.859	1.23	9.29**	<u>0.375</u>	<u>0.456</u>
45	0.192	8.173	0.325	3.298	1.22	1.22	-0.225	0.225
46	0.195	6.223	0.556	3.278	1.25	2.13	-0.178	0.224
47	0.258	5.118	1.109	2.052	0.72	1.52	-0.016	0.154
48	0.503	5.845	0.214	4.513	1.06	1.98	-0.227	0.347
49	0.754	2.993	0.263	6.759	2.65***	14.30**	<u>-0.346</u>	<u>0.401</u>

ตาราง 28 (ต่อ)

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ			
	a ₁	b ₁	a ₂	b ₂	MH	LC	CSA	CUA
50	0.122	-1.154	0.215	4.840	2.15 ^{***}	8.23 ^{**}	-0.178	0.224
51	0.988	3.023	0.322	-2.780	0.57 ^{**}	6.08 [*]	-0.016	0.154
52	0.766	3.650	0.226	5.340	0.01	15.43 ^{**}	<u>-0.424</u>	<u>0.424</u>
53	1.885	-1.746	1.885	-0.703	0.36 ^{**}	10.48 ^{**}	<u>-0.348</u>	<u>0.411</u>
54	0.371	3.985	0.259	5.905	2.69 ^{**}	9.04 [*]	-0.287	<u>0.434</u>
55	0.132	-8.923	0.350	2.536	1.27	2.45	-0.153	0.194
56	0.860	3.483	0.180	-1.849	1.35	2.42	-0.251	0.294
57	0.854	3.379	0.317	2.004	1.83 ^{**}	11.19 ^{**}	<u>-0.330</u>	<u>0.403</u>
58	1.140	2.916	0.171	3.699	2.65 ^{**}	13.94 ^{**}	<u>-0.402</u>	<u>0.451</u>
59	1.203	3.239	0.238	6.696	0.45 ^{***}	26.11 ^{***}	<u>-0.359</u>	<u>-0.420</u>
60	1.660	-1.739	0.175	7.062	2.15 ^{***}	23.17 ^{***}	<u>-0.419</u>	<u>0.507</u>
รวม					19	20	13	15

ตาราง 29 ตัวอย่างการคำนวณการทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่
ตรวจพบในแต่ละวิธีตรวจสอบ

ข้อที่	วิธีตรวจสอบ				L_i	L_i^2
	MH	LC	CSA	CUA		
4	0.63*				1	1
5	0.34*				1	1
6	3.18**	14.86***	-0.340	0.667	4	16
8	1.75**	10.40**	-0.359	0.411	4	16
12	4.32***	14.39***	0.401	0.401	4	16
17	2.69**	16.12***	-0.355	0.420	4	16
18	0.47*	7.62*		0.576	3	9
44		6.03*	0.421	0.429	3	9
49	2.62***	17.79***	-0.586	0.724	4	16
50	2.93***				1	1
51	1.83**				1	1
52	0.53**	10.51**	-0.345	0.407	4	16
53	0.57**	10.37**	-0.419	0.507	4	16
55	1.75**	6.18*			2	4
57	2.69**	9.07*		0.465	3	9
58	2.62***	16.12***	0.579	0.585	4	16
59	0.31***	16.81***	0.385	0.410	4	16
60	3.38***	17.59***	-0.406	0.445	4	16
G_j	17	14	11	13	ΣL_i	ΣL_i^2

สูตร

$$Q = \frac{(K - 1) \left[\left(K \sum_{j=1}^k G_j^2 \right) - \left(\sum_{j=1}^k G_j \right)^2 \right]}{K \sum_{i=1}^N L_i - \sum_{i=1}^N L_i^2}$$

เมื่อ K แทน จำนวนกลุ่มตัวอย่าง (จำนวนวิธีการตรวจสอบ)
 G_j แทน ผลรวมในแนวตั้ง (จำนวนข้อสอบที่ทำหน้าที่ต่างกันแต่ละวิธี)
 L_i แทน ผลรวมในแนวนอน

แทนค่า

$$\begin{aligned} Q &= \frac{(4 - 1) \left[4(17^2 + 14^2 + 11^2 + 13^2) - (17 + 14 + 11 + 13)^2 \right]}{4(55) - 195} \\ &= \frac{(3) [3100 - 3025]}{220 - 195} \\ &= \frac{225}{25} \\ &= 9^{**} \end{aligned}$$

$$\chi^2_{.05,3} = 7.815 \quad , \quad \chi^2_{.01,3} = 6.251$$

ดังนั้น จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบในแต่ละวิธีแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

การทดสอบรายคู่

ให้ Q_1 แทน คู่ระหว่าง MH กับ LC

$$Q_1 = \frac{(2-1)[(2(17^2 + 14^2) - (17+14)^2)]}{2(31) - 57}$$

$$= \frac{970 - 961}{62 - 57}$$

$$= \frac{9}{5}$$

$$= 1.80$$

$$\chi^2_{.05,1} = 3.841, \chi^2_{.01,1} = 2.706$$

ดังนั้น จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบระหว่าง MH กับ LC ไม่แตกต่างกัน

ประวัติย่อผู้วิจัย

ประวัติย่อผู้วิจัย

ชื่อ ชื่อสกุล	นางสาวจันทนา เปรมฤดีปรีชาชาณู
วันเดือนปีเกิด	12 มิถุนายน 2522
สถานที่เกิด	จังหวัดชลบุรี
สถานที่อยู่ปัจจุบัน	241 ม.2 ต.หนองอิรุณ อ.บ้านบึง จ.ชลบุรี 20220
ตำแหน่งหน้าที่การงานปัจจุบัน	ครู อันดับ คศ. 1
สถานที่ทำงานปัจจุบัน	โรงเรียนจุฬาภรณราชวิทยาลัย ชลบุรี สำนักงานเขตพื้นที่การศึกษาชลบุรี เขต 1

ประวัติการศึกษา

พ.ศ. 2533	ประถมศึกษาปีที่ 6 จากโรงเรียนวัดเขาไผ่ “เขียววิทยาคุณ”
พ.ศ. 2537	มัธยมศึกษาปีที่ 3 จากโรงเรียนบ้านบึงมัญญวิทยาคาร
พ.ศ. 2540	มัธยมศึกษาปีที่ 6 จากโรงเรียนชลราษฎรอำรุง
พ.ศ. 2544	กศ.บ. วิชาเอกคณิตศาสตร์ จากมหาวิทยาลัยศรีนครินทรวิโรฒ
พ.ศ. 2551	กศ.ม. สาขาวิชาการวัดผลการศึกษา จากมหาวิทยาลัยศรีนครินทรวิโรฒ