

การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน
ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

ปริญญานิพนธ์

ของ

นุชริน ไบโพธิ์

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา

พฤษภาคม 2544

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน
ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

บทคัดย่อ
ของ
นุชริน ไบโพธิ์

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา
พฤษภาคม 2544

นุชริน โบโพธิ์. (2544). การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน

ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST ปริญญาณิพนธ์ กศ.ม.

(การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

คณะกรรมการควบคุม : รองศาสตราจารย์ ดร. บุญเชิด ภิญโญนนตพงษ์, ดร.สุภาพร เข้มเฮง

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อ เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และศึกษาความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST กลุ่มตัวอย่างที่ใช้เป็น นักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2543 ในเขตกรุงเทพมหานคร จำนวน 790 คน ได้จากการสุ่มแบบสองขั้นตอน เครื่องมือที่ใช้ในการศึกษา เป็นแบบทดสอบวัดความถนัดทางคณิตศาสตร์ จำนวน 60 ข้อ ผลการวิจัยสรุปได้ดังนี้

1. ข้อสอบที่ทำหน้าที่ต่างกันที่ได้จากการตรวจสอบระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST มีจำนวนข้อที่ตรวจพบแตกต่างกัน เมื่อขนาดกลุ่มตัวอย่างและจำนวนข้อสอบต่างกัน ได้แก่ กรณีแบบทดสอบ 30 ข้อ ที่ใช้กลุ่มตัวอย่างขนาด 300 คน และ 500 คน และแบบทดสอบ 40 ข้อ ที่ใช้กลุ่มตัวอย่างขนาด 500 คนและ 700 คน

2. ความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อพิจารณาจากข้อสอบจำนวน 20 ข้อและ 30 ข้อ ที่ตรวจสอบโดยวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST พบว่ากลุ่มตัวอย่างขนาด 500 คนและ 700 คน มีจำนวนข้อสอบที่ตรวจพบสอดคล้องกันสูงสุด ส่วนข้อสอบจำนวน 40 ข้อ ซึ่งตรวจสอบด้วยวิธี Lord 's χ^2 พบว่ากลุ่มตัวอย่างขนาด 300 คน และ 500 คน มีจำนวนข้อสอบที่ตรวจพบสอดคล้องกันสูงสุด แต่เมื่อตรวจสอบโดยวิธี Mantel - Haenszel และวิธี SIBTEST พบว่ากลุ่มตัวอย่างขนาด 500 คนและ 700 คน มีจำนวนข้อสอบที่ตรวจพบสอดคล้องกันสูงสุด

A COMPARISON OF CONSISTENCY OF LORD 'S χ^2 ,MANTEL – HAENSZEL AND
SIBTEST IN DETECTION OF DIFFERENTIAL ITEM FUNCTIONING

AN ABSTRACT
BY
NOOTCHARIN BAIPO

Presented in partial fulfillment of the requirements
for the Master of Education degree in Educational Measurement
at Srinakharinwirot University
May 2001

Nootcharin Baipo. (2001). *A Comparison of Consistency of Lord 's χ^2 , Mantel – Haenszel and SIBTEST in Detection of Differential Item Functioning*. Master thesis, M.Ed (Educational Measurement). Bangkok : Graduate School, Srinakharinwirot University.
Advisor Committee : Assoc. Prof. Dr. Boonchird Pinyoanuntapong,
Dr. Suwaporn Semheng .

The purpose of this study was to compare the consistency of three different methods of differential item function detecting ; Lord 's χ^2 , Mantel - Haenszel and SIBTEST. The comparisons were investigated under various situations focusing on sample sizes and test lengths. The sample of the study was 790 Mathayom Suksa 1 students in the academic year 2000 from secondary school under Department Of General Education, Bangkok, selected by two - stage random sampling. The instrument used in this study was a mathematics aptitude test consisted of 60 test items. The first part of the study was to investigate the amount of DIF-items, the second part of the study investigate the degree of consistency among the methods. The findings were as follows :

1. The number of DIF-item were statistically significant different among Lord 's χ^2 , Mantel - Haenszel and SIBTEST with respect to the difference between sample sizes and test lengths. The differences were found at the 30 items for 300 and 500 persons, 40 items for 500 and 700 persons.

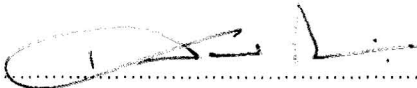
2. The level of agreement of detecting DIF among Lord 's χ^2 , Mantel - Haenszel and SIBTEST were moderate to high at the test length of 20 and 30 items for 500 and 700 persons but at the test length of 40 items for 300 and 500 persons detecting DIF by Lord 's χ^2 and at the test length 40 items for 500 and 700 persons detecting DIF by Mantel - Haenszel and SIBTEST were moderate to high .

ปริญญานิพนธ์
เรื่อง

การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน
ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

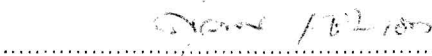
ของ
นางสาวนุชริน ไบโพธิ์

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา
ของมหาวิทยาลัยศรีนครินทรวิโรฒ

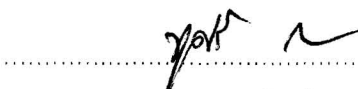

.....คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ ดร.เสริมศักดิ์ วิศาลาภรณ์)
วันที่ เดือน พฤษภาคม พ.ศ. 2544

คณะกรรมการสอบปริญญานิพนธ์


.....ประธาน
(รองศาสตราจารย์ ดร.บุญเชิด ปิญโญนนตพงษ์)


.....กรรมการ
(อาจารย์ ดร.สุพร เข้มเฮง)


.....กรรมการที่แต่งตั้งเพิ่มเติม
(รองศาสตราจารย์นิภา ศรีไพโรจน์)


.....กรรมการที่แต่งตั้งเพิ่มเติม
(รองศาสตราจารย์ชูศรี วงศ์รัตนะ)

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ โดยได้รับความช่วยเหลือจากบุคคลที่เกี่ยวข้องหลายท่าน ผู้วิจัยขอกราบขอบพระคุณมา ณ ที่นี้ ดังนี้ รองศาสตราจารย์ ดร. บุญเชิด ภิญโญนนตพงษ์ ประธานที่ปรึกษาปริญญานิพนธ์ ดร.สุพร เข้มเฮง กรรมการในการทำปริญญานิพนธ์ ที่ได้ช่วยแนะแนวในการแก้ปัญหาและให้กำลังใจแก่ผู้วิจัย รวมทั้ง รองศาสตราจารย์นิภา ศรีไพโรจน์ และรองศาสตราจารย์ชูศรี วงศ์รัตน์ ที่ร่วมเป็นกรรมการสอบปริญญานิพนธ์และให้คำแนะนำในการปรับปรุงแก้ไขปริญญานิพนธ์เป็นอย่างดี

ในการตรวจสอบและพิจารณาเครื่องมือที่ใช้ในการวิจัย ได้รับความเอื้อเฟื้อจากผู้ช่วยศาสตราจารย์ ดร.สมทรง อัสวกุล ผู้ช่วยศาสตราจารย์ ดร.สมบูรณ์ ตันยะ ผู้ช่วยศาสตราจารย์จิรนาถสมพงษ์ ผู้ช่วยศาสตราจารย์วันิดา เดชตานนท์ และผู้ช่วยศาสตราจารย์อาภรณ์ บางเจริญพรพงศ์ ซึ่งผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง

การเก็บรวบรวมข้อมูลได้รับความร่วมมือจากผู้บริหาร คณะครูและนักเรียนของโรงเรียนในสังกัด กรมสามัญศึกษา กรุงเทพมหานคร ที่เป็นกลุ่มตัวอย่าง ซึ่งผู้วิจัยรู้สึกซาบซึ้งในความกรุณาที่ทุกท่านให้ความช่วยเหลือเป็นอย่างดี

สุดท้ายนี้ผู้วิจัยขอกราบขอบพระคุณทุกท่านที่ไม่ได้กล่าวนามมา ณ ที่นี้ที่สนับสนุนและให้ความช่วยเหลือในการทำปริญญานิพนธ์ฉบับนี้จนสำเร็จลุล่วงไปด้วยดี คุณค่าและประโยชน์ของปริญญานิพนธ์ฉบับนี้ ขอมอบเป็นเครื่องบูชาแด่ บิดา-มารดา ครู-อาจารย์ ตลอดจนผู้มีพระคุณของผู้วิจัยทุกท่านที่ช่วยให้ผู้วิจัยประสบความสำเร็จในการศึกษา

นุชริน ไบโพธิ์

สารบัญ

บทที่	หน้า
1 บทนำ.....	1
ภูมิหลัง.....	1
จุดมุ่งหมายของการศึกษาค้นคว้า.....	4
ความสำคัญของการศึกษาค้นคว้า.....	4
ขอบเขตของการศึกษาค้นคว้า.....	4
นิยามศัพท์เฉพาะ.....	5
สมมติฐานของการศึกษาค้นคว้า.....	6
2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	7
การทำหน้าที่ต่างกันของข้อสอบ.....	7
การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ.....	9
การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี Lord's χ^2	11
การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี Mantel - Haenszel.....	16
การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี SIBTEST.....	20
ทฤษฎีสมรรถภาพสมองและความถนัด.....	23
งานวิจัยที่เกี่ยวข้อง.....	26
3 วิธีดำเนินการศึกษาค้นคว้า.....	47
ประชากร.....	47
กลุ่มตัวอย่าง.....	47
เครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	48
วิธีดำเนินการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	49
วิธีดำเนินการเก็บรวบรวมข้อมูล.....	51
การวิเคราะห์ข้อมูลและสถิติที่ใช้วิเคราะห์ข้อมูล.....	52
4 ผลการวิเคราะห์ข้อมูล.....	55
การเสนอผลการวิเคราะห์ข้อมูล.....	55
สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล.....	55
ผลการวิเคราะห์ข้อมูล.....	56

สารบัญ (ต่อ)

บทที่	หน้า
5 สรุป อภิปรายผล และข้อเสนอแนะ.....	82
จุดมุ่งหมายของการศึกษาค้นคว้า.....	82
กลุ่มตัวอย่าง.....	82
เครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	82
วิธีดำเนินการเก็บรวบรวมข้อมูล.....	82
สรุปผลการวิเคราะห์ข้อมูล.....	83
อภิปรายผล.....	85
ข้อเสนอแนะในการวิจัย.....	87
บรรณานุกรม.....	88
ภาคผนวก.....	95
ประวัติย่อผู้วิจัย.....	119

บัญชีตาราง

ตาราง		หน้า
1	ข้อมูลผลการสอบของกลุ่มอ้างอิง (R) และกลุ่มเปรียบเทียบ (F) ที่มีคะแนนรวม อยู่ในช่วงคะแนน j.....	16
2	สัดส่วนการตอบข้อสอบของประชากร 2 กลุ่ม.....	17
3	จำนวนกลุ่มตัวอย่างแยกตามขนาดโรงเรียน	48
4	ค่าไอเกนและร้อยละของความแปรปรวนขององค์ประกอบ	56
5	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST เมื่อจำนวนข้อสอบ 20 ข้อ กลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน.....	58
6	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST เมื่อจำนวนข้อสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน.....	59
7	ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST เมื่อจำนวนข้อสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน.....	61
8	เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 20 ข้อ กลุ่มตัวอย่าง 300 คน ระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST.....	63
9	เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 20 ข้อ กลุ่มตัวอย่าง 500 คน ระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST.....	64
10	เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 20 ข้อ กลุ่มตัวอย่าง 700 คน ระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST.....	64
11	เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 30 ข้อ กลุ่มตัวอย่าง 300 คน ระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST.....	65
12	เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่ เมื่อจำนวน ข้อสอบ 30 ข้อ กลุ่มตัวอย่าง 300 คน ระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST.....	65
13	เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 30 ข้อ กลุ่มตัวอย่าง 500 คน ระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST.....	66

บัญชีตาราง (ต่อ)

[illegible]

บัญชีตาราง (ต่อ)

[illegible]

บัญชีตาราง (ต่อ)

ตาราง		หน้า
37	จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ ระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST จากจำนวนข้อสอบ 3 ชุด และกลุ่มตัวอย่าง 3 ขนาด.....	83
38	สรุปผลการศึกษาความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกัน ของข้อสอบ เมื่อขนาดกลุ่มตัวอย่างเท่ากัน วิธีการตรวจสอบต่างกัน.....	84
39	สรุปผลการศึกษาความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อ สอบ เมื่อขนาดกลุ่มตัวอย่างต่างกัน วิธีการตรวจสอบเท่ากัน.....	85
40	ค่าพารามิเตอร์ของข้อสอบแยกตามเพศชาย – หญิง	99
41	จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อกลุ่มตัวอย่างขนาด 300 คน และจำนวนข้อสอบ 20 ข้อ	101
42	จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อกลุ่มตัวอย่างขนาด 500 คน และจำนวนข้อสอบ 20 ข้อ	102
43	จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อกลุ่มตัวอย่างขนาด 700 คน และจำนวนข้อสอบ 20 ข้อ	103
44	จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อกลุ่มตัวอย่างขนาด 300 คน และจำนวนข้อสอบ 30 ข้อ	104
45	จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อกลุ่มตัวอย่างขนาด 500 คน และจำนวนข้อสอบ 30 ข้อ	105
46	จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อกลุ่มตัวอย่างขนาด 700 คน และจำนวนข้อสอบ 30 ข้อ	106
47	จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อกลุ่มตัวอย่างขนาด 300 คน และจำนวนข้อสอบ 40 ข้อ	107
48	จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อกลุ่มตัวอย่างขนาด 500 คน และจำนวนข้อสอบ 40 ข้อ	108
49	จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อกลุ่มตัวอย่างขนาด 700 คน และจำนวนข้อสอบ 40 ข้อ	109

บัญชีภาพประกอบ

ภาพประกอบ

หน้า

1	การทำหน้าที่ต่างกันของข้อสอบเมื่อใช้วิธีภายนอกตรวจสอบ.....	9
2	โค้งแสดงลักษณะข้อสอบ 3 พารามิเตอร์	14
3	แสดงลำดับขั้นในการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	49
4	ผลการตรวจสอบความเป็นมิติเดียวกันของแบบทดสอบ.....	56

บทที่ 1

บทนำ

ภูมิหลัง

เครื่องมือวัดผลการศึกษาที่นิยมใช้กันอย่างแพร่หลายคือ แบบทดสอบแบบเลือกตอบ เนื่องจากสามารถวัดสมรรถภาพสมองด้านความรู้ความจำ ความเข้าใจ การนำไปใช้ การวิเคราะห์ การสังเคราะห์ และการประเมินค่า ได้อย่างเป็นปรนัยสูงและวัดได้ครอบคลุมเนื้อหา การจะนำแบบทดสอบมาใช้ให้เหมาะสมและตรงกับจุดมุ่งหมายที่ต้องการวัดจะต้องศึกษาถึงวิธีการสร้าง ลักษณะเด่น และคุณภาพของแบบทดสอบด้วย แบบทดสอบที่ดีจะต้องมี ความเชื่อมั่น (Reliability) งามลึก (Searching) ยั่วให้ห้อยากตอบ (Exemplary) มีความเป็นปรนัย(Objectivity) สูง มีความยุติธรรม (Fairness) มีความจำเพาะเจาะจง (Definite) มีประสิทธิภาพ (Efficiency) มีความยาก (Difficulty) พอเหมาะ มีอำนาจจำแนก (Discrimination) และสำคัญที่สุดคือ ความเที่ยงตรง (Validity) เพราะความเที่ยงตรงของเครื่องมือวัดจะเป็นตัวชี้ให้เห็นว่าเครื่องมือวัดนั้นๆ วัดได้ตรงตามเป้าหมายของคุณลักษณะ (Trait) ที่ต้องการจะวัด ถ้าหากแบบทดสอบขาดความเที่ยงตรงแล้ว ก็จะทำให้เกิดความคลาดเคลื่อนในการแปลผลคะแนนที่ได้จากการสอบวัด (ญาณภัทร สีหะมงคล .2540 : 1)

ข้อสอบที่สร้างขึ้นเพื่อวัดลักษณะแฝง (Latent Trait) หลังจากนำไปใช้แล้วพบว่า ผู้สอบนอกจากต้องใช้ความรู้ในเรื่องที่ต้องการวัดแล้ว ยังต้องใช้ความรู้ในเรื่องอื่น เข้ามาช่วยในการทำข้อสอบนั้น เรียกว่า ข้อสอบนั้นทำหน้าที่เบี่ยงเบน (Differential Item Functioning) และเมื่อใดก็ตามที่นำข้อสอบเบี่ยงเบนนั้น ไปวัดประชากรที่ประกอบด้วยประชากรกลุ่มย่อยตั้งแต่สองกลุ่มขึ้นไป โดยที่ประชากรย่อยกลุ่มหนึ่งมีความรู้ในเรื่องอื่นสูงกว่าประชากรย่อยอีกกลุ่มหนึ่ง ทำให้มีการได้เปรียบในการตอบข้อสอบได้ถูกต้อง ทั้ง ๆ ที่มีระดับความรู้ในเรื่องที่ต้องการวัดเท่ากัน ถ้าแปลความหมายของผลการวัดโดยอ้างอิงเฉพาะในเรื่องที่ต้องการวัดเพียงอย่างเดียว จะทำให้ข้อสอบเกิดการทำหน้าที่ต่างกัน (Differential Item Functioning) ทั้งนี้เพราะที่จริงแล้วผลการสอบที่แตกต่างกันระหว่างกลุ่มประชากรย่อย มิใช่เพราะความแตกต่างของความรู้ในเรื่องที่ต้องการวัดเพียงอย่างเดียว แต่มีความแตกต่างของความรู้ในเรื่องอื่นที่มีได้ตั้งใจวัดเข้ามาด้วย ทำให้เกิดการได้เปรียบเสียเปรียบกันระหว่างกลุ่มย่อยต่างๆ นั่นคือ เกิดความลำเอียงหรือความไม่ยุติธรรมขึ้นในการสอบ (กาญจนา วัฒนสุนทร. 2538 : 1) ซึ่งได้รับการวิจารณ์และเป็นประเด็นสำคัญอย่างหนึ่งที่นักวัดผลหรือผู้ที่เกี่ยวข้องกับกระบวนการทดสอบ จะต้องคำนึงถึงและให้ความสนใจ ทั้งนี้เพราะหากข้อสอบที่มีความลำเอียงเข้าข้างกลุ่มใดกลุ่มหนึ่ง หรือเอื้อประโยชน์ต่อคนบางกลุ่มแล้ว จะทำให้ผลการสอบขาดความน่าเชื่อถือ และเกิดความไม่เสมอภาคในการตัดสิน

จากการวิจัยของกรีน (Intasuwan. 1979 : 1 ; citing Green. 1972) พบว่า ข้อคำถามบางข้อของแบบทดสอบวัดผลสัมฤทธิ์แคลิฟอร์เนีย (California Achievement Test) ซึ่งเป็นข้อคำถามที่มีคุณภาพเมื่อนำไปทดสอบกับกลุ่มผู้สอบที่มีความแตกต่างกันทางเชื้อชาติ ภาษา สภาพเศรษฐกิจ และสภาพภูมิอากาศ จะให้ผลสอบที่แตกต่างกัน ซึ่งอาจมีสาเหตุมาจากวัฒนธรรม และสิ่งแวดล้อมที่ต่างกันมีผลต่อความเข้าใจภาษาที่ใช้ในข้อคำถามต่างกัน นอกจากนี้นงานวิจัยของ ดูลิทเทิลและเคลียร์ (Doolittle and Cleary. 1987) ยังพบอีกว่า เพศชายมีความสามารถในการทำแบบทดสอบวิชาเรขาคณิตได้ดีกว่าเพศหญิง เพราะเนื้อหาเป็นแผนภาพ (Diagram)ซึ่งง่ายต่อเพศชาย เนื่องจากเพศชายมีทักษะเชิงอวกาศดีกว่าเพศหญิง นอกจากนั้นยังพบว่า เพศชายมีทักษะทางการใช้เหตุผลทางคณิตศาสตร์มาก ในขณะที่เพศหญิงมีทักษะทางการคำนวณ

มาก และฮาร์นิช (Harnisch. 1991) ได้ศึกษาการที่ทำหน้าที่ต่างกันต่อเพศของข้อสอบ Second International Mathematics Study (SIMS) โดยใช้กลุ่มตัวอย่างชาวอเมริกัน และชาวญี่ปุ่น พบลักษณะการทำหน้าที่ต่างกันของข้อสอบที่คล้ายคลึงกัน โดยที่ข้อสอบที่ใช้ทักษะคำนวณ เพศชายจะมีความถนัดมากกว่าเพศหญิง ทั้งในกลุ่มชาวอเมริกันและชาวญี่ปุ่น ในขณะเดียวกัน ชัชชัย เผ่าพงษ์ (2527) ดูลิทเทิลและเคลียร์ (Doolittle and Cleary. 1987) ยังพบว่า เพศชายจะมีความถนัดด้านคณิตศาสตร์มากกว่าเพศหญิง ในขณะที่เพศหญิงจะมีความถนัดด้านภาษามากกว่าเพศชาย

สำหรับกระบวนการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสามารถทำได้สองวิธี (อุทุมพร จามรมา. 2537 : 129) ดังนี้ วิธีที่หนึ่ง ได้แก่ การตรวจสอบในขั้นตอนการสร้างแบบทดสอบ วิธีที่สอง ได้แก่ การตรวจสอบโดยใช้สถิติ หลังจากที้นำแบบทดสอบไปใช้แล้ว ซึ่งสามารถทำได้สองแนวทางดังนี้ แนวทางที่หนึ่ง ได้แก่ ใช้วิธีเกณฑ์ภายนอก (External Criterion) โดยใช้แบบทดสอบที่สร้างขึ้นหาความสัมพันธ์กับเกณฑ์ที่เป็นมาตรฐาน เช่น เกรตเฉลี่ย เป็นต้น ส่วนแนวทางที่สอง ได้แก่ ใช้วิธีเกณฑ์ภายใน (Internal Criterion) สามารถศึกษาจากการตรวจสอบโครงสร้างภายในของข้อสอบ การตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันตามแนวนี้สามารถแบ่งออกเป็นสามกลุ่มใหญ่ ๆ (Camilli and Shepard. 1994 : 22 – 131) ดังนี้ กลุ่มที่หนึ่ง ได้แก่ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ตามทฤษฎีการทดสอบมาตรฐานดั้งเดิม (Classical Test Theory) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามวิธีนี้ ได้แก่ วิธีการแปลงค่าความยาก (Transformation Item Difficulty) วิธี Golden Rule วิธีวิเคราะห์ความแปรปรวน (Analysis of Variance) และวิธีพอยต์ไบเซรียล (Point Biserial) กลุ่มที่สอง ได้แก่ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้ตารางการณัจร (Contingency Table) การตรวจสอบในแนวนนี้ ได้แก่ วิธีแมนเทล – แฮนส์เซล (Mantel - Haenszel หรือ MH) และวิธีกำลังสอง (Chi-square) เป็นต้น วิธีนี้ค่อนข้างสะดวก เพราะสามารถคำนวณด้วยเครื่องคิดเลขได้ไม่ต้องอาศัยเครื่องคอมพิวเตอร์ กลุ่มที่สาม ได้แก่ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของทฤษฎีการตอบข้อสอบ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบกลุ่มนี้ สามารถจำแนกเป็นสองแนวทาง (Hambleton, Swaminathan and Rogers. 1991 : 110) ดังนี้ แนวทางที่หนึ่ง ได้แก่ การเปรียบเทียบค่าพารามิเตอร์ของข้อสอบระหว่างสองกลุ่มย่อยว่ามีความแตกต่างกันหรือไม่ ในกลุ่มนี้นิยมใช้วิธี Lord's χ^2 และล่าสุดใช้วิธีทดสอบอัตราส่วนน่าจะเป็นสูงสุด (Likelihood Ratio Test) ส่วนแนวทางที่สอง ได้แก่ การวัดพื้นที่ความแตกต่างของฟังก์ชันการตอบข้อสอบระหว่างสองกลุ่มย่อย ในระยะหลังนี้นิยมใช้วิธี Raju's Area Measures และวิธี Closed Interval Area

ผลจากการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยใช้ข้อมูลจำลองขึ้น ซึ่งมีทั้งข้อสอบที่ทำหน้าที่ต่างกันและไม่ทำหน้าที่ต่างกันในลักษณะต่างกัน ส่วนมากพบว่าวิธีตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน ได้ถูกต้องและสอดคล้องกับข้อมูลจำลองมากที่สุดได้แก่ วิธีทฤษฎีการตอบข้อสอบสองและสามพารามิเตอร์ และวิธีที่ถูกต้องรองลงมาซึ่งมีค่าใกล้เคียงกัน ได้แก่ การทดสอบค่า χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST ซึ่งมีความสอดคล้องสูงกับวิธี Mantel - Haenszel ในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน เช่น การศึกษาของ ซิลลี่และสเตาท์ (Shealy and Stout. 1993)* แคมเบิลตันและคณะ (Hambleton and Others. 1986) ทำให้วิธี IRT วิธี χ^2 วิธี Mantel - Haenszel และ SIBTEST มีความน่าสนใจในทางปฏิบัติ

นอกจากนี้ยังมีงานวิจัยที่ศึกษาผลกระทบของขนาดกลุ่มตัวอย่างและจำนวนข้อสอบ ต่อการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ พบว่า ขนาดกลุ่มผู้สอบและจำนวนข้อสอบมีผลต่อความคงที่ของ

การวิเคราะห์ข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องในการใช้ข้อมูลจำลอง เช่น สวามินาธาน และโรเจอร์ (Swaminathan and Rogers. 1990) พบว่า วิธี Mantel - Haenszel วิเคราะห์ได้ถูกต้องร้อยละ 75 ในกรณีใช้ผู้สอบขนาด 250 คนและร้อยละ 100 กรณีใช้ผู้สอบ 500 คน และงานวิจัยของโรเจอร์และสวามินาธาน (Rogers and Swaminathan. 1993 : 105) ที่ศึกษาเปรียบเทียบการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ วิธี Mantel - Haenszel กับวิธีถดถอยโลจิสติก (Logistic Regression) โดยศึกษาผลกระทบของขนาดกลุ่มตัวอย่างต่อการวิเคราะห์ทั้งสองวิธี กลุ่มตัวอย่างที่ใช้ศึกษาขนาด 250 คน และ 500 คน พบว่า วิธี Mantel - Haenszel จะวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบได้เพิ่มขึ้นเมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้น คิมและโคเฮน (Kim and Cohen. 1994) และโคเฮน คิม และวอลแลค (Cohen Kim and Wollack. 1995) ได้ศึกษาอัตราการเกิดความคลาดเคลื่อนชนิดที่ 1 (Type I error rates) ของวิธี Lord's χ^2 และวิธีทดสอบอัตราส่วนความน่าจะเป็นสูงสุด (Likelihood Ratio Test) จำลองข้อมูลโดยการกำหนดขนาดกลุ่มตัวอย่างเป็น 250 และ 1,000 คน จำนวนข้อสอบ 50 ข้อ และแบบจำลองที่ศึกษาเป็นแบบจำลอง 2 กับ 3 พารามิเตอร์ และ 3 พารามิเตอร์ ที่กำหนดให้ค่าการเดาคงที่ พบว่า โดยภาพรวมอัตราการเกิดความคลาดเคลื่อนชนิดที่ 1 ของวิธีการทดสอบอัตราส่วนความน่าจะเป็นสูงสุด จะมีค่าใกล้เคียงค่าคาดหวัง ในทางทฤษฎีมากกว่าวิธีของลอร์ด ณ ระดับนัยสำคัญ .01 และ .001 แต่ ณ ระดับนัยสำคัญ .05 ทั้งสองวิธีให้ค่าใกล้เคียงกัน เมื่อใช้แบบจำลอง 2 กับ 3 พารามิเตอร์ ที่กำหนดให้ค่าการเดาคงที่ อย่างไรก็ตาม สำหรับวิธีของลอร์ดเมื่อใช้แบบจำลอง 3 พารามิเตอร์ จะเกิดความคลาดเคลื่อนชนิดที่ 1 สูงกว่าเมื่อใช้แบบจำลอง 2 กับ 3 พารามิเตอร์ ที่กำหนดให้ค่าการเดาคงที่ และงานวิจัยของนารายานานและสวามินาธาน (Narayanan and Swaminathan. 1994 : 315) ที่ศึกษาวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยวิธี Mantel - Haenszel และวิธี SIBTEST พบว่า วิธีการทั้งสองใช้วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบเข้าสู่กลุ่มใดกลุ่มหนึ่งเพียงกลุ่มเดียว (Uniform DIF) เมื่อกลุ่มตัวอย่างมีการแจกแจงความสามารถเท่าเทียมกัน แต่วิธี SIBTEST จะให้ผลการวิเคราะห์ถูกต้องกว่าวิธี Mantel - Haenszel เมื่อกลุ่มตัวอย่างมีการแจกแจงความสามารถไม่เท่าเทียมกัน และทั้งสองวิธีจะใช้ได้ดีและมีประสิทธิภาพเมื่อกลุ่มตัวอย่างแต่ละกลุ่มมีขนาด 300 คน

จากที่กล่าวมาข้างต้นและจากการศึกษารายงานการวิจัยที่ผ่านมา จะเห็นว่ามีการศึกษาถึงวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบหลายวิธี แต่ยังไม่มีการวิจัยที่ศึกษาความสอดคล้องของผลการตรวจสอบและจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และ วิธี SIBTEST ซึ่งเป็นวิธีที่ใช้ได้ดีในทางปฏิบัติ ย่อยต่อการคำนวณและแปลผล

ดังนั้น ผู้วิจัยจึงสนใจจะศึกษา จำนวนข้อสอบและความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน ระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และ วิธี SIBTEST โดยใช้แบบทดสอบความถนัดทางคณิตศาสตร์ จำนวน 20 ข้อ 30 ข้อ และ 40 ข้อ และมีกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน อันจะเป็นแนวทางในการเลือกวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมาใช้ในการปฏิบัติ เพื่อพัฒนาแบบทดสอบให้มีความยุติธรรมต่อไป

จุดมุ่งหมายของการศึกษาค้นคว้า

การศึกษาค้นคว้าครั้งนี้มีจุดมุ่งหมายสำคัญเพื่อศึกษาการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST ว่าจะให้ผลเป็นอย่างไร โดยมีจุดมุ่งหมายเฉพาะดังนี้

1. เพื่อเปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันระหว่าง วิธี Lord's χ^2 วิธี Mantel - Haenszel และ วิธี SIBTEST เมื่อขนาดกลุ่มตัวอย่าง และจำนวนข้อสอบในแบบทดสอบ ต่างกัน
2. เพื่อศึกษาความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน ระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST เมื่อขนาดกลุ่มตัวอย่าง และจำนวนข้อสอบในแบบทดสอบ ต่างกัน

ความสำคัญของการศึกษาค้นคว้า

การศึกษาค้นคว้านี้ เป็นการเปรียบเทียบความสอดคล้องของผลการตรวจสอบที่ทำหน้าที่ต่างกัน 3 วิธี ซึ่งผลที่ได้จากการศึกษามีประโยชน์และความสำคัญ คือ ทำให้ได้ข้อสรุปที่ชัดเจนเกี่ยวกับผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน 3 วิธี ว่าให้ผลสอดคล้องกันหรือขัดแย้งกันในเงื่อนไขของการศึกษา ซึ่งมีประโยชน์ต่อการเลือกใช้ให้เหมาะสม และเกิดความผิดพลาดในการตรวจสอบน้อยที่สุด แล้วยังเป็นการพัฒนาทฤษฎีการทดสอบในด้านการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน ซึ่งทำให้เกิดแง่มุมใหม่ ๆ ในการศึกษาครั้งต่อไป

ขอบเขตของการศึกษาค้นคว้า

1. ประชากร

ประชากรที่ใช้ในการศึกษาค้นคว้า ได้แก่ นักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2543 สังกัดกรมสามัญศึกษา จังหวัดกรุงเทพมหานคร จำนวน 49,708 คน เป็นนักเรียนชาย 25,872 คนและนักเรียนหญิง 23,836 คน

2. กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้า ได้แก่ นักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2543 สังกัดกรมสามัญศึกษา จังหวัดกรุงเทพมหานคร จำนวน 10 โรงเรียน จำนวน 790 คน แบ่งเป็นนักเรียนชาย 378 คน และนักเรียนหญิง 412 คน ซึ่งได้มาโดยการสุ่มแบบสองขั้นตอน (Two Stage Random Sampling)

3. ตัวแปรที่ศึกษา

3.1 ตัวแปรอิสระ ประกอบด้วย

3.1.1 วิธีการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน ได้แก่

3.1.1.1 วิธี Lord's χ^2

3.1.1.2 วิธี Mantel - Haenszel

3.1.1.3 วิธี SIBTEST

3.1.2 ขนาดกลุ่มตัวอย่าง แบ่งเป็น

3.1.2.1 300 คน

3.1.2.2 500 คน

3.1.2.3 700 คน

3.1.3 จำนวนข้อสอบในแบบทดสอบ กำหนดดังนี้

3.1.3.1 20 ข้อ

3.1.3.2 30 ข้อ

3.1.3.3 40 ข้อ

3.2 ตัวแปรตาม ได้แก่

3.2.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

3.2.2 ความสอดคล้องในการตรวจสอบ

นิยามศัพท์เฉพาะ

1. ข้อสอบที่ทำหน้าที่ต่างกัน หมายถึง ข้อสอบที่ผู้สอบซึ่งมีความสามารถเท่ากันแต่อยู่ต่างกลุ่มกัน มีโอกาสตอบข้อสอบข้อนั้นถูกต่างกัน ในการศึกษาครั้งนี้กลุ่มย่อย (Subgroup) ที่ใช้ในการศึกษา คือ เพศ แบ่งเป็น กลุ่มอ้างอิง (Reference Group) ในที่นี้คือ เพศชาย และกลุ่มเปรียบเทียบ (Focal Group) ในที่นี้คือ เพศหญิง

2. กลุ่มอ้างอิง (Reference Group) หมายถึง กลุ่มที่ได้เปรียบในการทำข้อสอบ ในการศึกษาครั้งนี้คือ เพศชาย

3. กลุ่มเปรียบเทียบ (Focal Group) หมายถึง กลุ่มที่เสียเปรียบในการทำข้อสอบ ในการศึกษาครั้งนี้คือ เพศหญิง

4. การตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน หมายถึง การนำผลการสอบมาวิเคราะห์เพื่อเปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันและความสอดคล้องในการตรวจสอบ ระหว่างกลุ่มอ้างอิง และกลุ่มเปรียบเทียบ ในการศึกษาครั้งนี้ใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธี ได้แก่ วิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

4.1 วิธี Lord's χ^2 หมายถึง วิธีการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันตามแนวคิดของลอร์ด ที่ใช้การเปรียบเทียบค่าพารามิเตอร์ของข้อสอบระหว่างกลุ่มย่อยสองกลุ่ม

4.2 วิธี Mantel - Haenszel หมายถึง วิธีการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน ตามแนวคิดของแมนเทลและแฮนส์เซล โดยพิจารณาเปรียบเทียบสัดส่วนของการตอบของกลุ่มผู้สอบสองกลุ่ม ข้อสอบที่ทำหน้าที่ต่างกันจะมีค่า α_{MH} ไม่เท่ากับ 1 และค่า χ^2_{MH} มากกว่าค่า χ^2 ที่ df เท่ากับ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

4.3 วิธี SIBTEST หมายถึง วิธีการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน โดยพิจารณาค่า β ที่ได้จากการทดสอบความแตกต่างของการตอบถูกของกลุ่มผู้สอบแต่ละกลุ่ม ข้อสอบที่ทำหน้าที่ต่างกันจะมีค่า B_U หรือค่า Z_{SIB} แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

5. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ หมายถึง จำนวนข้อสอบที่ตรวจพบว่า ทำหน้าที่ต่างกัน

6. ดัชนีการทำหน้าที่ต่างกันของข้อสอบ หมายถึง ค่าสถิติที่ใช้ในการแสดงความแตกต่างของการตอบข้อสอบระหว่างกลุ่มย่อยสองกลุ่ม ซึ่งขึ้นอยู่กับวิธีที่ใช้ในการตรวจสอบ ซึ่งการศึกษาค้นคว้าครั้งนี้มี 3 วิธี คือ

6.1 ดัชนีการทำหน้าที่ต่างกันของข้อสอบจาก วิธี Lord 's χ^2 หมายถึง ค่าสัดส่วนความแตกต่างของการตอบถูกหรือผิดระหว่างกลุ่มย่อยสองกลุ่ม คือค่า χ^2

6.2 ดัชนีการทำหน้าที่ต่างกันของข้อสอบจากวิธี Mantel – Haenszel หมายถึง ค่าสัดส่วนความแตกต่างของการตอบระหว่างกลุ่มย่อยสองกลุ่ม คือค่า α_{MH}

6.3 ดัชนีการทำหน้าที่ต่างกันของข้อสอบจากวิธี SIBTEST หมายถึง ค่าที่ได้จากการทดสอบความแตกต่างของการตอบระหว่างกลุ่มย่อยสองกลุ่ม คือค่า β

7. เกณฑ์การตัดสินข้อสอบที่ทำหน้าที่ต่างกัน หมายถึง ค่าที่แสดงว่าข้อสอบทำหน้าที่ต่างกัน จากค่าดัชนีทั้ง 3 วิธี ได้แก่

7.1 วิธี Lord 's χ^2 ทดสอบนัยสำคัญทางสถิติที่ระดับ .05

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

7.3 วิธี SIBTEST พิจารณาค่า B_U หรือค่า Z_{SIB} แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

8. ความสอดคล้องในการตรวจสอบ หมายถึง สัดส่วนจำนวนข้อสอบที่ตรงกันในการตัดสินโดยใช้วิธีการตรวจสอบการทำหน้าที่ต่างกัน 3 วิธี ต่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในการเปรียบเทียบแต่ละคู่คำนวณได้ดังนี้

$$\text{ความสอดคล้องในการตรวจสอบ} = \frac{\text{จำนวนข้อสอบที่ตรงกันในการตัดสิน}}{\text{จำนวนข้อสอบที่ทำหน้าที่ต่างกันในการเปรียบเทียบแต่ละคู่}}$$

สมมติฐานในการศึกษาค้นคว้า

1. เมื่อขนาดกลุ่มตัวอย่างต่างกัน วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเหมือนกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกันจากการตรวจสอบทั้ง 3 วิธีแตกต่างกัน

2. เมื่อขนาดกลุ่มตัวอย่างเท่ากัน วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบต่างกัน จำนวนข้อสอบที่ทำหน้าที่ต่างกันจากการตรวจสอบทั้ง 3 วิธีแตกต่างกัน

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในการศึกษาครั้งนี้ มีเอกสารและงานวิจัยที่เกี่ยวข้อง ดังนี้

1. การทำหน้าที่ต่างกันของข้อสอบ
2. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
3. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี Lord 's χ^2
4. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี Mantel - Haenszel
5. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี SIBTEST
6. ทฤษฎีสมรรถภาพสมองและความถนัด
7. งานวิจัยที่เกี่ยวข้อง

1. การทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning)

การทำหน้าที่ต่างกันของข้อสอบ เดิมใช้คำว่าความลำเอียงของข้อสอบ (Item Bias) ซึ่งเริ่มได้รับความสนใจตั้งแต่ปี 1910 (Holland and Wainer, 1993) โดยบินเน็ต และไซมอน (Binet and Simon) ได้ทดสอบเด็กที่มีฐานะทางเศรษฐกิจและสังคมต่างกัน พบว่า ข้อสอบบางข้อน่าจะวัดผลจากวัฒนธรรมการฝึก (เช่น ที่บ้าน หรือ ที่โรงเรียน) มากกว่าที่จะวัดความสามารถทางสมอง ปี 1912 วิลเลียม สเตอร์น (William Stern) ได้ศึกษาความแตกต่างของชั้นเรียน (Class Difference) ในประเทศเยอรมัน พบว่ามีความคล้ายคลึงกับที่บินเน็ตพบโดยพยายามที่จะทำความเข้าใจเกี่ยวกับแหล่งที่ทำให้เกิดความแตกต่าง เพื่อที่จะเสนอแนะว่า แบบทดสอบนั้นเข้าข้างนักเรียนชั้นหนึ่งมากกว่าชั้นอื่น ๆ อย่างชัดเจน

ปี 1951 มีการวิจัยที่ถือเป็นการเริ่มต้นของการศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบ ซึ่งเรียกว่า "ข้อสอบลำเอียง" ได้แก่ งานวิจัยของเอลส์ (Eells) เดวิส (Davis) ฮาวิกเกิร์ต (Havighurst) เฮร์ริค (Herrick) และไทเลอร์ (Tyler) โดยเอลส์และคณะ สนใจความแตกต่างของการวัดบางอย่างที่พิจารณาถึงความไม่ถูกต้องของการตอบข้อสอบอันเกิดจากความสามารถ (Ability) เช่น ความแตกต่างของไอคิวเด็กนักเรียน จุดมุ่งหมายหลักเพื่อที่จะอธิบายความแตกต่างระหว่างกลุ่มในแบบทดสอบไอคิว โดยมุ่งไปที่ข้อคำถามแต่ละข้อ ในลักษณะเช่นเดียวกับบินเน็ตและสเตอร์น คือดูว่าข้อคำถามใดมีการทำหน้าที่ต่างกันหรือไม่ แม้ว่าเอลส์และคณะ จะใช้สัดส่วนการตอบถูกสำหรับกลุ่มที่แตกต่างกัน แต่ก็ไม่ได้รับการแนะนำให้ใช้ในเวลาไม่นานนัก ขณะเดียวกันก็เกิดคำถามเกี่ยวกับความคงที่ของความแตกต่างระหว่างกลุ่มว่าขึ้นอยู่กับเนื้อหาหรือรูปแบบคำถาม ความแตกต่างของวัฒนธรรมในการเลือกคำตอบที่มีความสัมพันธ์กับความยากของข้อสอบหรือไม่ และเขาได้พยายามแก้ไขอิทธิพลของชั้นทางสังคมและภูมิหลังของเชื้อชาติที่มีผลต่อการตอบข้อสอบ โดยการเปรียบเทียบกลุ่มของเชื้อชาติที่อยู่ในชั้นทางสังคม

ผลจากการศึกษาของเอลส์ ได้นำไปสู่การศึกษาอย่างกว้างขวางในช่วง 20 ปีต่อมา เพื่อความมุ่งหวังของรัฐในการที่จะทำให้เกิดความยุติธรรมในการใช้ผลการสอบ ในปี 1960 และ 1970 การพัฒนาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบได้มุ่งไปที่แบบทดสอบไอคิว เนื่องจากแบบทดสอบไอคิวได้ถูกนำไปใช้ในการคัดเลือก (Selection) และ จัดบุคคลเข้าสู่ตำแหน่ง (Placement) อย่างแพร่หลาย จนกระทั่งปี 1971 บริษัท Griggs v. Duke Power ซึ่งการใช้ผลการทดสอบไอคิวในการคัดเลือกอาจไม่เหมาะสม ยกเว้น

กรณีที่ผลของการปฏิบัติงานมีความสอดคล้องกับผลการทดสอบ ในช่วงเวลาเดียวกันมีการโต้แย้งเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบมีทั้งบทความในวารสารต่าง ๆ เจนเซนได้เสนอว่า ไอคิวเป็นผลมาจากการถ่ายทอดทางพันธุกรรมมากกว่าสิ่งแวดล้อม ความแตกต่างของไอคิวระหว่างคนผิวขาวกับคนผิวดำไม่ได้มาจากสิ่งแวดล้อมเพียงอย่างเดียว และได้รับการยืนยันจากนักวิทยาศาสตร์อย่างค่อนข้างชัดเจน เขายังได้อธิบายต่อว่า พันธุกรรมอธิบายความแปรปรวนของผลการปฏิบัติงานของแต่ละคนได้ประมาณร้อยละ 80 จากนั้นก็มีการตอบสนองงานของเจนเซน ในวารสาร Harvard Education Review (1960) โดยสมาคมนักจิตวิทยาคนผิวดำเสนอว่า การทดสอบคนผิวดำทุกชนิดควรทำให้เท่าเทียมกันมากเท่าที่จะเป็นไปได้ (Camilli and Shepard. 1994 : 4-7)

การตรวจสอบและค้นหาการทำหน้าที่ต่างกันของข้อสอบได้มีมาอย่างต่อเนื่องควบคู่กับการพยายามคิดค้นวิธีการทางสถิติในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ดังปรากฏจากเอกสารการวัดผลในช่วงกลางปี 1970 และในปัจจุบันก็มีการพัฒนาเทคนิควิธีการต่าง ๆ อย่างมากมาย เพื่อให้ผลการวิเคราะห์มีความถูกต้องที่สุด

ความหมายของการทำหน้าที่ต่างกันของข้อสอบ

สำหรับความหมายของการทำหน้าที่ต่างกันของข้อสอบมีผู้ให้ความเห็นไว้ต่าง ๆ ดังนี้ (ญาณภัทร สีหะมงคล. 2540 : 15)

อลเลนและเยน (Allen and Yen. 1979 : 129) ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบว่า เป็นข้อสอบที่วัดความสามารถหรือคุณลักษณะทางจิตวิทยาของผู้สอบแต่ละกลุ่มไม่ตรงกัน

ลอร์ด (Lord. 1980 : 212) แมคคอลลีย์และเมนโดซา (McCauley and Mendoza. 1985 : 389) แฮมเบิลตันและคณะ (Hambleton, et. al. 1991 : 110) ฮอลล์แลนด์และไวเนอร์ (Holland and Wainer. 1993 : 4) ได้ให้ความหมายของข้อสอบที่ทำหน้าที่ต่างกันว่า เป็นข้อสอบที่ผู้สอบมีความสามารถเท่ากัน แต่อยู่ต่างกลุ่มกันมีโอกาสตอบข้อสอบข้อนั้นได้ถูกต้องไม่เท่ากัน

ปอบแฮม (Popham. 1990 : 178) ได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบว่า คือ ความชอบหรือความไม่เอียงที่ทำให้เกิดการตัดสินอย่างไม่ยุติธรรม

มิลล์แซป และอีเวอร์สัน (Millsap and Everson. 1993 : 297) ได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบว่าเป็นความไม่แม่นยำ (Inaccuracy) อย่างเป็นระบบของการวัด

คามิลลีและเชพเพิร์ด (Camilli and Shepard. 1994 :8) ได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบว่า เป็นความคลาดเคลื่อนอย่างเป็นระบบ (Systematic Error) หรือความไม่เที่ยงตรง (Invalidity) ของการวัดชนิดหนึ่ง ทำให้สมาชิกบางกลุ่มที่ได้ประโยชน์หรือเสียประโยชน์ เช่น ข้อสอบที่เข้าข้างกลุ่มนักเรียนชายมากกว่านักเรียนหญิง หรือคนผิวขาวกับคนผิวดำ หรือระหว่างคนที่อยู่ต่างภูมิภาคกัน เป็นต้น

ฮอลล์แลนด์และไวเนอร์ (Holland and Wainer. 1993) ได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบว่า เป็นข้อสอบที่ผู้เข้าสอบ ณ ความสามารถเดียวกันแต่อยู่ต่างกลุ่มกัน มีโอกาสในการตอบถูกไม่เท่ากัน กลุ่มที่เสียเปรียบจะเรียกว่า "Focal Group" และกลุ่มได้เปรียบ เรียกว่า "Reference Group"

โดยสรุป ข้อสอบที่ทำหน้าที่ต่างกัน หมายถึง ข้อสอบที่ทำให้ผู้สอบที่มีความสามารถเท่ากัน ภายใต้ประชากรเดียวกันแต่อยู่ต่างกลุ่มกันมีโอกาสในการที่จะตอบถูกไม่เท่ากัน โดยจะเอื้อประโยชน์ต่อผู้เข้าสอบบางกลุ่ม

2. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

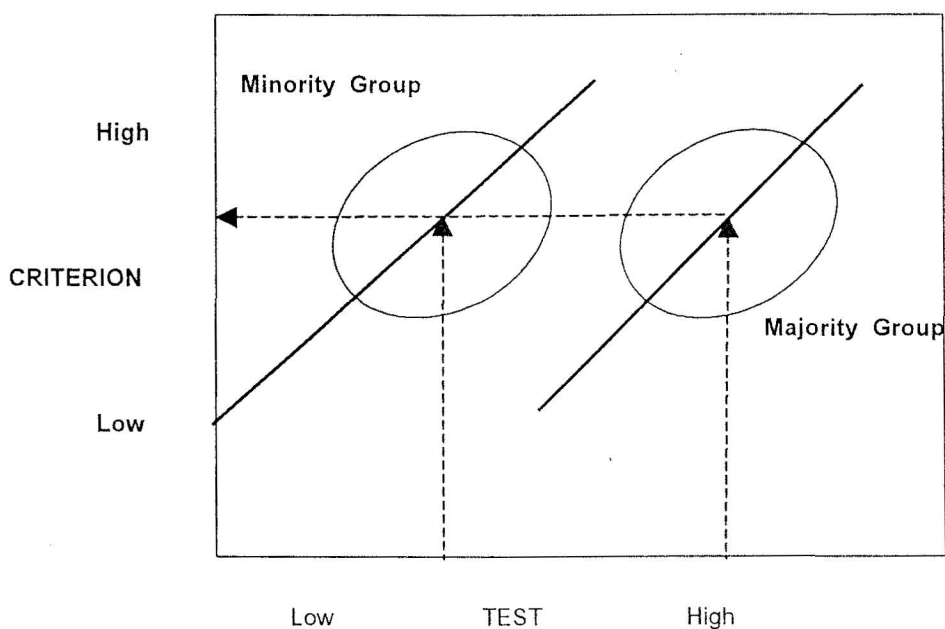
การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสามารถทำได้ 2 ขั้นตอน (ญานภักดิ์ สี่หมะมงคล. 2540 : 15) ได้แก่

ขั้นแรก เป็นขั้นของการสร้างแบบทดสอบ โดยให้ผู้เชี่ยวชาญตรวจสอบว่า ข้อสอบหรือแบบทดสอบทำหน้าที่ต่างกันหรือไม่ นั่นคือ เป็นการตรวจสอบความเที่ยงตรง (Validity) ของแบบทดสอบ

ขั้นที่สอง เป็นการใช้เทคนิคทางสถิติเข้ามาช่วยหลังจากนำแบบทดสอบไปใช้แล้ว นำผลมาวิเคราะห์ ดูลักษณะการตอบข้อสอบของกลุ่มคนที่อยู่ต่างกลุ่มกัน เช่น เพศ เชื้อชาติ ศาสนา ท้องถิ่น และระดับสติปัญญา ว่าข้อสอบหรือแบบทดสอบยังคงทำหน้าที่ไม่ต่างกัน หรือเอื้อประโยชน์ต่อคนบางกลุ่มหรือไม่ ซึ่งการศึกษาการทำหน้าที่ต่างกันของข้อสอบที่มีพื้นฐานจากการใช้หลักทางสถิติสามารถทำได้ 2 แนวทาง (Camilli and Shepard. 1994 :9) ได้แก่

การใช้วิธีภายนอก (External Method) หรือการใช้เกณฑ์ภายนอก(External Criterion Method)

เป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยการหาความเที่ยงตรงเชิงพยากรณ์ (Predictive Validity) นั่นเอง ทำได้โดยการเขียนกราฟแสดงความสัมพันธ์ระหว่างตัวทำนาย (ข้อสอบ) กับเกณฑ์ (เกรดเฉลี่ยหรือผลการปฏิบัติงาน เป็นต้น) แล้วพิจารณาค่าความลาดชันของกราฟ (Slope) และค่าจุดตัดแกนตั้ง (Intercept) ระหว่างสองกลุ่ม ถ้าต่างกันแสดงว่าข้อสอบทำหน้าที่ต่างกันเข้าข้างกลุ่มใดกลุ่มหนึ่ง (ดังภาพประกอบ 1)



ภาพประกอบ 1 การทำหน้าที่ต่างกันของข้อสอบ เมื่อใช้วิธีภายนอกตรวจสอบ

สำหรับวิธีนี้จะมีปัญหามากตรงความเป็นมาตรฐานของเกณฑ์ ดังนั้นจึงไม่ได้รับความนิยมในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การใช้วิธีภายใน (Internal Method) หรือเกณฑ์ภายใน (Internal Criterion Method)

วิธีการนี้ได้รับความนิยมอย่างมาก เนื่องจากไม่ต้องกังวลเรื่องเกณฑ์ สามารถศึกษาจากการตรวจสอบโครงสร้างภายในของข้อสอบ ซึ่งพิจารณาคะแนนที่ได้จากผลการตอบของผู้เข้าสอบต่างกลุ่มว่าวัดในคุณลักษณะที่ต้องการวัด ตามโครงสร้างเช่นเดียวกันหรือไม่ ถือเป็นกระบวนการของการสนับสนุนการตรวจสอบความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแนวนี้ สามารถแบ่งเป็น 3 กลุ่มใหญ่ ๆ (Camilli and Shepard, 1994) ดังนี้

2.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของทฤษฎีการทดสอบแบบมาตรฐานดั้งเดิม (Classical Test Theory : CTT)

เป็นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ใช้กันมาตั้งแต่สมัยเริ่มต้นของการพัฒนาเทคนิควิธีการตรวจสอบ แม้ว่าปัจจุบันจะไม่ค่อยได้นำมาใช้แล้ว เนื่องจากนักวัดผลได้มีการพัฒนาเทคนิควิธีการใหม่ ๆ มาใช้แทน และได้ผลดีกว่า ประกอบกับวิธีการตรวจสอบแบบนี้มีจุดอ่อนตรงไม่สามารถแยกค่าความยากของข้อสอบออกจากค่าอำนาจจำแนกได้ ทำให้เกิดความคลาดเคลื่อนในการตรวจสอบค่อนข้างสูง สำหรับวิธีการในกลุ่มนี้ประกอบด้วย วิธีแปลงค่าความยาก (Transformation Item Difficulty) วิธี Golden Rule วิธีวิเคราะห์ความแปรปรวน (Analysis of Variance) และวิธีพอยต์ไบเซเรียล (Point Biserial) เป็นต้น

2.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของทฤษฎีการตอบข้อสอบ (Item Response Theory ; IRT)

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแนวนี้ จะพิจารณาเปรียบเทียบฟังก์ชันการตอบข้อสอบ (Item Response Function) หรือเส้นโค้งคุณลักษณะของข้อสอบ (Item Characteristic Curves) ระหว่างกลุ่มตั้งแต่ 2 กลุ่มขึ้นไป ว่ามีความแตกต่างกันหรือไม่ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวคิดทฤษฎีการตอบข้อสอบ มีจุดเด่นกว่าวิธีตรวจสอบตามแนวคิดทฤษฎีมาตรฐานดั้งเดิม ดังนี้

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

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

3. คุณสมบัติทางสถิติของข้อสอบ สามารถแสดงด้วยแผนภาพที่เป็นกราฟ ทำให้เข้าใจง่ายและนำไปใช้กว้างขวาง

2.3 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้ตารางการถ้อย (Contingency Table Approach) วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแนวนี้มีลักษณะดังนี้

2.3.1 ใช้สถิติไม่พารามิเตอร์ในการทดสอบ (Non - Parametric Statistics)

2.3.2 ใช้คะแนนที่สังเกตได้ (Observe Score) ประมาณค่าความสามารถระหว่างกลุ่ม

2.3.3 มีความสอดคล้องมากกับการเปรียบเทียบระหว่างเส้นกราฟคุณลักษณะของ

ข้อสอบตามทฤษฎีการตอบข้อสอบ

2.3.4 ไม่จำเป็นต้องใช้สถิติขั้นสูง

2.3.5 ค่อนข้างประหยัด

รวมทั้งมีมีโนทัศน์คล้ายกับการตรวจสอบตามแนวคิดทฤษฎีการตอบข้อสอบ ได้แก่

1. มีข้อกำหนดเบื้องต้น ในการเปรียบเทียบความแตกต่างของผลการตอบข้อสอบระหว่างผู้เข้าสอบสองกลุ่ม โดยใช้เกณฑ์ภายใน เช่น ใช้คะแนนจากการสอบเป็น 0 – 1 เหมือนกัน
2. ทั้งสองวิธีสามารถนำไปประยุกต์ในการตรวจสอบซ้ำได้ (Iterative) เช่น ตามแนวคิด Purification ของลอร์ด (Lord, 1980 : 220) เป็นต้น
3. เป็นแนวคิดการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในลักษณะการวัดแบบหลายมิติ (Multidimensionality) ชนิดหนึ่ง เช่นเดียวกับทฤษฎีการตอบข้อสอบตัวอย่างวิธีการตรวจสอบในกลุ่มนี้ ได้แก่ ไคสแควร์ (Chi – square) แมนเทล – แฮนส์เซล (Mantel – Haenszel) และ ล็อกลิเนียร์ (Log – linear) เป็นต้น

อย่างไรก็ตามนักวิจัยได้พัฒนาเทคนิควิธีใหม่ ๆ ขึ้นมาอย่างต่อเนื่อง เช่น วิธีการทำให้เป็นมาตรฐาน (Standardization Method) แบบจำลองการถดถอยโลจิสติก (Logistic Regression Model) และลัสตัสวิธีชิปเทสต์ (SIBTEST) (Shealy and Stout, 1993)

3. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี Lord 's χ^2

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี Lord 's χ^2 เป็นวิธีการที่อยู่ภายใต้แนวคิดตามทฤษฎีการตอบข้อสอบ ในที่นี้จึงขอลำดับถึงหลักเบื้องต้นของทฤษฎีก่อน ดังต่อไปนี้

3.1 หลักการของทฤษฎีการตอบข้อสอบ (ญาณภัทร สีหะมงคล, 2540 : 20)

ทฤษฎีการตอบข้อสอบ เป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างคุณลักษณะหรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรมการตอบข้อสอบของบุคคลนั้น ๆ ในรูปของฟังก์ชันทางคณิตศาสตร์ กล่าวคือ พฤติกรรมการตอบข้อสอบของบุคคลจะถูกกำหนดโดย คุณลักษณะหรือความสามารถที่มีอยู่ในตัวบุคคล (Lord and Novick, 1968 : 358) ซึ่งไม่สามารถสังเกตได้โดยตรง ซึ่งเขียนในรูปฟังก์ชันทางคณิตศาสตร์ได้ ดังนี้

$$y = f(\theta)$$

เมื่อ	y	แทน	คะแนนที่ได้จากการตอบข้อสอบ ซึ่งเป็นพฤติกรรมการตอบข้อสอบ
	f	แทน	ฟังก์ชัน
	θ	แทน	คุณลักษณะหรือความสามารถภายในตัวบุคคล

ซึ่ง ลอร์ด เรียกฟังก์ชันนี้ว่า ฟังก์ชันการตอบข้อสอบ (Item Response Function) หรือฟังก์ชันคุณลักษณะของข้อสอบ (Item Characteristic Function) (Hambleton and Swaminatan, 1985 : 13)

3.2 ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อสอบ

3.2.1 ความเป็นมิติเดียว (Unidimensionality) ถ้าความสามารถของคน (Ability or Trait) มีอยู่หลายอย่าง ซึ่งความสามารถแต่ละอย่างนี้ ต่างก็ส่งผลต่อการตอบข้อสอบ (Test Performance) ต่าง ๆ ที่รวมกันเป็นแบบทดสอบ ถ้าผลการตอบข้อสอบหรือคะแนนของผู้สอบนั้น สามารถอธิบายได้เพียงความสามารถเดียวก็เพียงพอแล้วเรียกข้อตกลงนี้ว่า ความเป็นมิติเดียว (Hambleton and Swaminathan. 1985 : 16) สำหรับในทางปฏิบัติ ก็คือข้อสอบแต่ละข้อในแบบทดสอบจะต้องวัดในคุณลักษณะเดียว แต่อย่างไรก็ตาม แอมเบิลตัน และสวามินาธาน (Hambleton and Swaminathan. 1985 : 16 - 17) ได้กล่าวว่า ข้อตกลงเบื้องต้นข้อนี้ไม่ได้เข้มงวดนักขอให้มีความเด่น (Dominant) ที่จะวัดในองค์ประกอบใดองค์ประกอบหนึ่งก็ใช้ได้ ซึ่งสอดคล้องกับผลการศึกษาของ ตราสโกส์ พาร์สันส์ และเรคเคส (McCauley and Mendoza. 1985 : 390) ที่พบว่า การประมาณค่าพารามิเตอร์ต่างๆ ยังสามารถทำได้เมื่อการวัดนั้นอาจมีการเบนไปในทางหลายมิติบ้างเล็กน้อย

วิธีการตรวจสอบว่าแบบทดสอบนั้นวัดในมิติเดียวหรือไม่นั้น ทำได้หลายวิธี เช่น การวิเคราะห์องค์ประกอบ (Factor Analysis) แล้วสังเกตค่าไอเกน (Eigen Value) ที่สูงสุดว่าต่างจากค่าอื่น ๆ อย่างชัดเจนหรือไม่ สำหรับ ลอร์ด และ โนวิก (Lord and Novick. 1968 : 382 ; Lord. 1980 : 20) ยังได้แนะนำว่า การวิเคราะห์องค์ประกอบควรจะใช้ค่าสหสัมพันธ์แบบเตตราคอร์ริค ไม่ควรใช้การคำนวณค่าสหสัมพันธ์แบบเพียร์สัน เพราะว่าข้อมูลที่เหมาะ (Fit) กับแบบจำลองของทฤษฎีการตอบข้อสอบไม่ได้สัมพันธ์กันในเชิงเส้นตรง (Hambleton and Swaminathan. 1985 : 306)

3.2.2 ความเป็นอิสระในการตอบข้อสอบ (Local Independence) หมายถึง โอกาสในการตอบข้อสอบแต่ละข้อได้ถูกต้องเป็นอิสระจากกัน นั่นคือ การตอบข้อสอบข้อใด ข้อหนึ่งได้ถูกหรือผิดจะไม่มีผลต่อการตอบข้ออื่น ๆ ด้วย (Lord. 1980 : 19 ; Hambleton and Swaminathan. 1985 : 23) หรือจะกล่าวในเชิงคณิตศาสตร์ได้ว่า ความเป็นอิสระในการตอบข้อสอบ หมายถึง ความน่าจะเป็นที่จะตอบข้อสอบถูกทั้งหมด จะเท่ากับผลคูณของความน่าจะเป็นของการตอบข้อสอบถูกเป็นรายข้อ (Lord. 1980 : 19) และแอมเบิลตันกับสวามินาธาน (Hambleton and Swaminathan. 1985 : 22-23) มีความเห็นตรงกันว่า ถ้ามีความเป็นมิติเดียวแล้วจะมีความเป็นอิสระในการตอบข้อสอบ ซึ่งมีความสำคัญต่อการประมาณค่าพารามิเตอร์ต่าง ๆ มาก ถ้าขาดความเป็นอิสระในการตอบ ข้อสอบแล้ว ค่าพารามิเตอร์ต่าง ๆ ที่ประมาณค่าได้ จะมีความคลาดเคลื่อนจากความเป็นจริง

3.2.3 โค้งแสดงลักษณะข้อสอบ (Item Characteristic Curve : ICC) โค้งแสดงลักษณะข้อสอบ เป็นฟังก์ชันทางคณิตศาสตร์ ที่แสดงถึงความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบข้อนั้นได้ถูกต้อง กับระดับความสามารถที่วัดได้โดยใช้ชุดของข้อสอบหรือแบบทดสอบนั้น (Hambleton and Swaminathan. 1985 : 25) ซึ่งจะเห็นว่า ความน่าจะเป็นที่ผู้สอบตอบข้อสอบถูก จะขึ้นอยู่กับโค้งแสดงลักษณะข้อสอบในแต่ละแบบจำลองที่ใช้ โดยที่รูปร่างของโค้งแสดงลักษณะข้อสอบในแต่ละข้อมีคุณสมบัติไม่แปรเปลี่ยน (Invariant) ไปตามกลุ่มตัวอย่างหรือตัวผู้สอบ (Hambleton and Cook. 1977 : 80) ฉะนั้นจึงทำให้ความน่าจะเป็นในการตอบข้อสอบถูกในแต่ละข้อไม่แปรเปลี่ยนด้วย เนื่องจากมีความเชื่อหลายความเชื่อเกี่ยวกับความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องกับระดับความสามารถที่วัดได้โดยใช้ชุดของข้อสอบหรือแบบทดสอบนั้น จึงทำให้เกิดมีแบบจำลองของโค้งแสดงลักษณะข้อสอบ (Item Response Model) ขึ้นมาหลายแบบจำลองดังที่จะกล่าวต่อไป

3.3 แบบจำลองในทฤษฎีการตอบข้อสอบ (Item Response Model)

แบบจำลองในทฤษฎีการตอบข้อสอบมีการพัฒนาต่อเนื่องกันมานานและมีหลายลักษณะ แต่ในที่นี้จะกล่าวถึงบางลักษณะที่น่าสนใจ และใช้กันแพร่หลาย ดังนี้

3.3.1 แบบจำลองโลจิสติก (Logistic Model)

ผู้มีบทบาทสำคัญในการพัฒนาแบบจำลองนี้ คือ เบิร์นบอม (Hambleton and Swaminathan, 1985 : 36) โดยในครั้งแรก เบิร์นบอมได้พัฒนาแบบจำลองโลจิสติก แบบ 2 พารามิเตอร์ คือ พารามิเตอร์ค่าความยาก (b) และค่าอำนาจจำแนก (a) ซึ่งเป็นแบบจำลองที่มีโค้งแสดงลักษณะข้อสอบ (Item Characteristic Curve : ICC) คล้ายกับโค้งแสดงลักษณะข้อสอบ (ICC) ของแบบจำลองนอร์มัลลอจีฟ (Normal Ogive Model) แบบ 2 พารามิเตอร์ ที่ ลอร์ด ได้เสนอไว้เมื่อปี ค.ศ. 1952 มาก โดยโค้งแสดงลักษณะทั้งสองมีความแตกต่างกันไม่เกิน .01 สำหรับทุก ๆ ค่าของความสามารถ (Lord, 1980 : 14 ; Hambleton and Swaminathan, 1985 : 37)

เนื่องจากแบบจำลองโลจิสติก ได้ใช้แบบจำลองทางคณิตศาสตร์ที่สามารถนำไปประยุกต์ใช้ได้ง่ายกว่าแบบนอร์มัลลอจีฟ จึงทำให้มีคนนำแบบจำลองนี้ไปใช้กันมาก ต่อมา เบิร์นบอม ได้เพิ่มพารามิเตอร์ค่าการเดา (c) เข้าไปอีกตัวหนึ่ง จึงกลายเป็นแบบ 3 พารามิเตอร์ นอกจากนั้นยังได้ตัดพารามิเตอร์ค่าอำนาจจำแนก (a) ออกจากแบบ 2 พารามิเตอร์ จึงเหลือพารามิเตอร์เดียว คือ ค่าความยาก (b) ซึ่งเรียกว่า 1 พารามิเตอร์ สำหรับแบบจำลองทางคณิตศาสตร์ของแต่ละแบบจำลองย่อย มีดังนี้

1. แบบจำลองโลจิสติก แบบ 3 พารามิเตอร์ (Three Parameter Logistic Model)

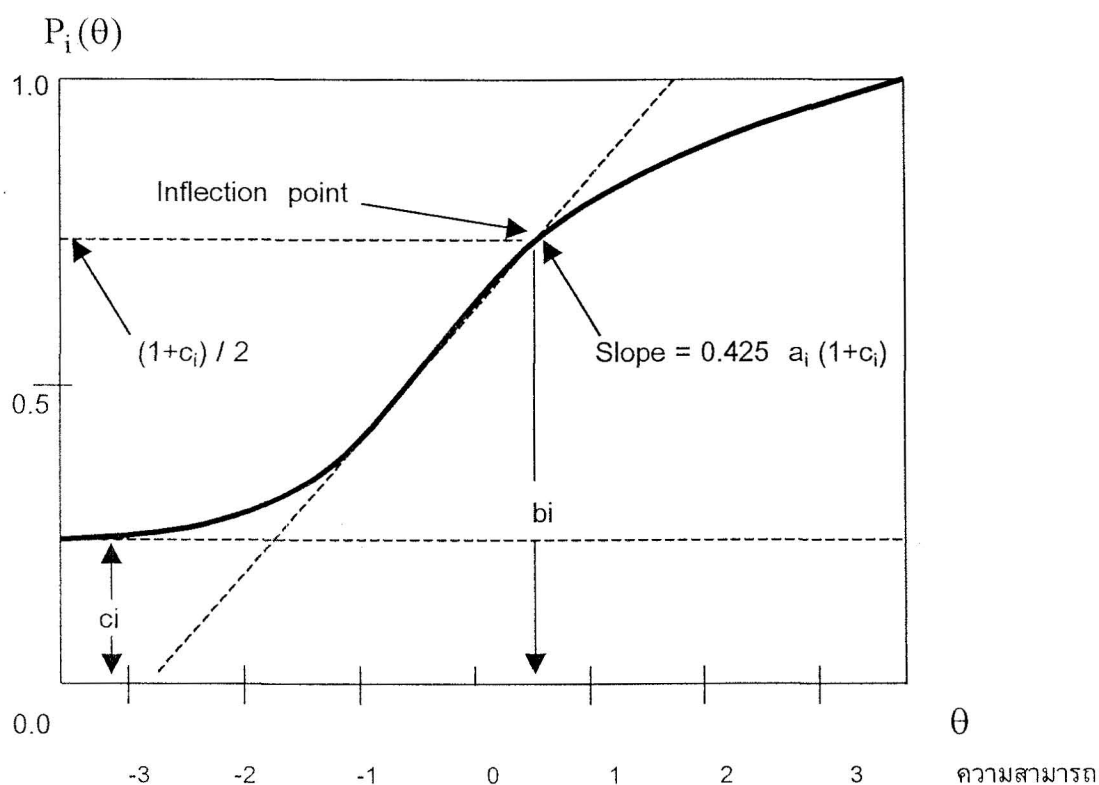
มีแบบจำลองทางคณิตศาสตร์ ดังนี้ (Hambleton Swaminathan and Rogers, 1991 : 17)

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

- เมื่อ $P_i(\theta)$ แทน ความน่าจะเป็นที่ผู้มีความสามารถ (θ) ตอบข้อสอบข้อที่ i ได้ถูกต้อง
- a_i แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ i โดยทฤษฎีแล้ว a จะมีค่าระหว่าง $-\infty$ ถึง $+\infty$ ข้อสอบที่ใช้ได้ควรมีค่า a เป็นบวก โดยปกติจะมีค่าไม่เกิน 2.00
- b_i แทน ค่าความยากของข้อสอบข้อที่ i โดยทฤษฎีแล้ว b จะมีค่าระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติ b จะมีค่าระหว่าง -2.00 ถึง +2.00 ข้อที่มีค่า b ใกล้ -2.00 เป็นข้อที่ง่ายมาก และข้อที่มีค่า b ใกล้ +2.00 เป็นข้อที่ยากมาก (Hambleton and Cook, 1977 : 81)
- c_i แทน ค่าการเดาของข้อสอบ หมายถึง ค่าความน่าจะเป็นที่ผู้สอบมีความสามารถต่ำสุดทำข้อสอบนี้ได้ถูกต้อง ข้อสอบที่มีค่าการเดาสูงเป็นข้อสอบที่ไม่ดีควรมีค่าการเดาไม่เกิน 0.30
- θ แทน ระดับความสามารถของผู้สอบโดยทฤษฎีแล้วจะมีค่าระหว่าง $-\infty$ ถึง $+\infty$ แต่ในทางปฏิบัติจะมีค่าระหว่าง -3.00 ถึง +3.00
- D แทน ค่าคงที่ (Scalling Factor) เพื่อปรับโค้งแสดงลักษณะข้อสอบ (ICC) ให้มีลักษณะใกล้เคียงกับโค้งแสดงลักษณะข้อสอบของแบบจำลองนอร์มัลลอจีฟ
- e แทน ค่าเอ็กซ์โปเนนเชียล (Exponential) เป็นค่าคงที่ซึ่งมีค่าประมาณ 2.7182818 จากแบบจำลอง ทางคณิตศาสตร์ของ 3 พารามิเตอร์ เมื่อนำมาเขียนเป็นกราฟ

แสดงลักษณะข้อสอบได้ดังนี้ (Lord, 1980 : 12 – 14 ; Hambleton and Swaminathan, 1985 : 38 – 39)

ความน่าจะเป็นในการตอบถูก



ภาพประกอบ 2 โค้งแสดงลักษณะข้อสอบ 3 พารามิเตอร์

จากภาพประกอบ 2 เป็นโค้งแสดงลักษณะข้อสอบ 3 พารามิเตอร์ จะเห็นว่า

1. ค่าอำนาจจำแนก (a_i) จะเป็นสัดส่วนกับความชัน (Slope) ของโค้ง ณ จุดเปลี่ยนโค้ง หรือจุดที่มีความชันมากที่สุด (ความชัน ณ จุดนี้มีค่าเท่ากับ $0.425a_i(1 - c_i)$)
2. ค่าความยาก (b_i) จะอยู่บนแกนความสามารถ (θ) ตรงจุดตัดกับเส้นที่ลากมาจากจุดเปลี่ยนโค้งมาตั้งฉากกับแกนความสามารถ
3. ค่าการเดา (c_i) เป็นส่วนตัดบนแกนตั้งของโค้งแสดงลักษณะข้อสอบ

2. แบบจำลองโลจิสติก 2 พารามิเตอร์ (Two Parameter Logistic Model)

มีแบบจำลองทางคณิตศาสตร์ ดังนี้ (Hambleton, Swaminathan and Roger, 1991 : 15)

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

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

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

3.4 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของลอร์ด

เป็นการเปรียบเทียบค่าพารามิเตอร์ของข้อสอบระหว่าง 2 กลุ่มว่ามีความแตกต่างกันหรือไม่ ซึ่งมีผู้ศึกษาค้นคว้ามากพอสมควร ได้แก่ งานวิจัยของ คิมและโคเฮน (Kim and Cohen. 1994) ลอร์ด (Lord. 1980) ธิสเซน สเติร์นเบอร์ก และไวเนอร์ (Thissen, Steinberg and Wainer. 1988, 1993) และไรท์และสโตน (Wright and Stone. 1979) อย่างไรก็ตาม ในกลุ่มนี้วิธีที่ได้รับความนิยม ก็คือ วิธีที่อาศัยแนวคิดของลอร์ด (Lord's χ^2)

สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของลอร์ด สามารถใช้ได้กับแบบจำลองโลจิสติกชนิด 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

$$H_0 : b_1 = b_2, a_1 = a_2, c_1 = c_2$$

สถิติทดสอบ (Hambleton Swaminathan and Rogers. 1991 : 11)

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

เมื่อ

$$a_{\text{diff}} = a_2 - a_1$$

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

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

และ Σ แทน เมตริกซ์ Variance และ Covariance ของความแตกต่างระหว่างพารามิเตอร์ข้อสอบ หากศึกษาเป็นแบบจำลอง 1 พารามิเตอร์ สถิติทดสอบเป็นดังนี้

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

4. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี Mantel-Haenszel (MH)

(พรณี จินตมาศ. 2540 : 27-30)

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี Mantel - Haenszel ได้เสนอขึ้นใช้ตั้งแต่ปี 1959 แต่ฮอลแลนด์และเธเยอร์ (Holland and Thayer) เพิ่งนำมาใช้เพื่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในปี 1986 และในปีนั้นเองดอร์แลนด์(Dorans) แฮมเบิลตัน(Hambleton) โรเจอร์(Rogers) และอาราสมิท (Arrasmith) ฮอลแลนด์ และ เธเยอร์ ไรท์ (Wright) ราชู (Raju) โบต (Bode) และลาเซน (Larsen) ก็ได้นำวิธี Mantel - Haenszel มาใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย

วิธี Mantel - Haenszel จะคล้ายคลึงกับวิธีไคสแควร์ที่เสนอโดย ชูเนแมน (Scheuneman .1979) มาราคุยโลและสลาฟเตอร์ (Marascuilo and Slaughter. 1981) และเมลเลนเบิร์ก (Mellenberg. 1982) ซึ่งเป็นการเปรียบเทียบผลการตอบข้อสอบของกลุ่มผู้สอบ 2 กลุ่ม หรือที่ Holland เรียกว่า กลุ่มอ้างอิงและกลุ่มเปรียบเทียบ (Reference and Focal Group) โดยที่กลุ่มแรกจะใช้อ้างอิงถึงกลุ่มที่คาดว่าจะได้ประโยชน์จากข้อสอบและกลุ่มหลังเป็นกลุ่มเสียประโยชน์จากข้อสอบ โดยมีการตรวจสอบทุก ๆ ระดับคะแนนรวมจากการสอบ ข้อสอบใดที่ผู้สอบทั้งสองกลุ่มทำได้เท่ากัน จะถือว่าข้อสอบข้อนั้นไม่ได้ทำหน้าที่ต่างกัน วิธีการในการตรวจสอบมีดังนี้

ถ้าให้ N_{Rj} แทนจำนวนผู้สอบในกลุ่มอ้างอิง และ N_{Fj} แทนจำนวนผู้สอบในกลุ่มเปรียบเทียบ ที่มีคะแนนรวมในช่วงคะแนน j และให้ N_j ซึ่งเท่ากับ $N_{Rj} + N_{Fj}$ แทนจำนวนผู้สอบรวมทั้งสองกลุ่มที่ได้คะแนนรวมจากการสอบอยู่ในช่วงคะแนน j แล้วนำมาเขียนในรูปตารางสองทาง แสดงผลการตอบถูกและผิด โดยตอบถูกได้ 1 และตอบผิดได้ 0 ดัง ตาราง 1

ตาราง 1 ข้อมูลผลการสอบของกลุ่มตัวอย่างกลุ่มอ้างอิง (R) และกลุ่มเปรียบเทียบ (F) ที่มีคะแนนรวมอยู่ในช่วงคะแนน j

กลุ่มผู้สอบ	คะแนนที่ได้จากข้อสอบที่ต้องการตรวจสอบการทำหน้าที่ต่างกัน			
		1	0	รวม
	R	A_j	B_j	N_{Rj}
	F	C_j	D_j	N_{Fj}
	รวม	N_{1j}	N_{0j}	N_j

เมื่อ A_j B_j C_j D_j เป็นความถี่ที่สังเกตได้ของการตอบถูก (1) และผิด (0) ในช่วงคะแนน j และแสดงในรูปสัดส่วนได้ ดังตาราง 2

ตาราง 2 สัดส่วนการตอบข้อสอบของกลุ่มประชากร 2 กลุ่ม

ผล			
การตอบ กลุ่ม	1	0	รวม
R	P_{Rj}	Q_{Rj}	1
F	P_{Fj}	Q_{Fj}	1

โดยที่ P_{Rj} คือ สัดส่วนของกลุ่มอ้างอิงที่อยู่ในช่วงความสามารถ j ที่ตอบข้อสอบถูก
 P_{Fj} คือ สัดส่วนของกลุ่มเปรียบเทียบที่อยู่ในช่วงความสามารถ j ที่ตอบข้อสอบถูก
 Q_{Rj} คือ $1 - P_{Rj}$
 Q_{Fj} คือ $1 - P_{Fj}$

จากตาราง 1 จะได้ว่าในช่วงระดับคะแนน j A_{Rj} คือ จำนวนผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบข้อนั้นถูกต้องและ B_{Rj} คือ จำนวนผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบข้อนั้นไม่ถูกต้อง ในขณะที่ C_{Fj} คือ จำนวนผู้สอบกลุ่มเปรียบเทียบที่ตอบข้อสอบข้อนั้นถูกต้อง D_{Fj} คือ จำนวนผู้สอบกลุ่มเปรียบเทียบที่ตอบข้อสอบข้อนั้นไม่ถูกต้อง ตามเทคนิค Mantel - Haenszel ที่เสนอโดย Holland และ Thayer นั้น ข้อมูลการตอบข้อสอบแต่ละข้อจะถูกนำมาแสดงในรูปตารางสองทางนี้ เป็นจำนวนตารางเท่ากับคะแนนรวมที่มีผู้สอบทำได้บวกด้วย 1 จากนั้นจะคำนวณค่าความเป็นไปได้ในรูปของค่าสัดส่วนการตอบข้อสอบถูกหรือผิดระหว่างกลุ่ม โดยกำหนดให้เป็นค่า α ดังนี้

$$\alpha_{MH} = \frac{\sum A_j D_j / N_j}{\sum B_j C_j / N_j}$$

ค่า α_{MH} จะมีค่าระหว่าง 0 และ 1 โดยที่ $\alpha_{MH} = 1$ จะแสดงถึงสมมติฐานศูนย์ที่ว่า ไม่มีความแตกต่างหรือข้อสอบไม่ได้ทำหน้าที่ต่างกัน ซึ่งแสดงว่ากลุ่มตัวอย่างทั้งสองกลุ่มทำข้อสอบได้ผลเช่นเดียวกันในแต่ละระดับคะแนนรวม ค่า $A_j D_j / N$ และ $B_j C_j / N$ มีค่าเท่ากันทั้งสองกลุ่มทำให้ $\alpha_{MH} = 1$

ในทางปฏิบัติแล้ว แม้ว่าจะมีการควบคุมให้ผู้สอบสองกลุ่ม มีความสามารถในสิ่งที่ต้องการวัดเท่ากัน แต่เป็นไปได้ที่ผลการตอบข้อสอบของผู้สอบทั้งสองกลุ่มจะเป็นเช่นเดียวกัน Mantel - Haenszel

จึงเสนอค่าสถิติไคสแควร์เพื่อทดสอบค่าที่ได้ว่าจะมีความแตกต่างจาก 1.00 อย่างมีนัยสำคัญหรือไม่ ที่ระดับชั้นของความอิสระเท่ากับ 1 ดังนี้

$$MH - CHISQ = \frac{[\sum A_j - \sum E(A_j) - 0.5]^2}{\sum Var(A_j)}$$

โดยที่ค่าผลรวมเป็นค่าที่ได้จากการรวมทุกชั้นของคะแนนรวม และค่า $E(A_j)$ มีค่าดังนี้

$$E(A_j) = (N_{Rj})(N_{1j})/N_j$$

$$Var(A_j) = \frac{N_{Rj} N_{Fj} N_{1j} N_{0j}}{N_j^2 (N_j - 1)}$$

และสมมติฐานศูนย์แสดงว่าข้อสอบไม่ทำหน้าที่ต่างกัน ได้แก่

$$H_0 : P_{Rj} = P_{Fj} \quad \text{สำหรับทุกชั้นคะแนน } j$$

สมมติฐานศูนย์นี้เป็นสมมติฐานของความเป็นอิสระอย่างมีเงื่อนไข ของสมาชิกกลุ่ม และคะแนนที่ได้จากการตอบข้อสอบที่ต้องการตรวจสอบการทำหน้าที่ต่างกัน และภายใต้สมมติฐานศูนย์ จะได้ค่าคาดหวังของการตอบข้อสอบดังนี้

$$E(A_j) = n_{Rj} m_{1j} / T_j$$

$$E(B_j) = n_{Rj} m_{0j} / T_j$$

$$E(C_j) = n_{Fj} m_{1j} / T_j$$

$$E(D_j) = n_{Fj} m_{0j} / T_j$$

สมมติฐานอื่นของ Mantel - Haenszel ได้แก่

$$H_1 : \frac{P_{Rj}}{Q_{Rj}} \neq \alpha \frac{P_{Fj}}{Q_{Fj}} ; j=1, \dots, k \quad \text{เมื่อ } \alpha \neq 1$$

เมื่อ $\alpha = 1$ ซึ่งสอดคล้องกับสมมติฐานศูนย์ จะได้ว่า

$$H_0 : \frac{P_{Rj}}{Q_{Rj}} = \frac{P_{Fj}}{Q_{Fj}}$$

$$\text{และค่าประมาณของ } \alpha_{MH} = \frac{P_{Rj}}{Q_{Rj}}, \frac{P_{Fj}}{Q_{Fj}} = \frac{P_{Rj} Q_{Fj}}{P_{Fj} Q_{Rj}} \text{ สำหรับทุก } j = 1, \dots, k$$

นอกจากนั้นฮอลแลนด์และเซเยอร์ ให้เสนอแนะให้แปลงค่า α_{MH} ให้เป็นเดลต้า (Δ) เพื่อให้เป็นค่าที่มีค่าเฉลี่ยเป็น 0 ดังนี้

$$\Delta_{MH} = \frac{-4}{1.7} \text{Ln}(\alpha_{MH}) = -2.35 \text{Ln}(\alpha_{MH})$$

ค่า Δ_{MH} มีค่าระหว่าง -2.6 ถึง 2.6 จะใช้เป็นตัวชี้แสดงการทำหน้าที่ต่างกันของข้อสอบ ข้อสอบที่ไม่ทำหน้าที่ต่างกันจะมีค่า $\alpha_{MH} = 1$ หรือ $\Delta_{MH} = 0$

เท่าที่ผ่านมามีผู้นำวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี Mantel - Haenszel มาศึกษา เช่น คลอสเซอร์และคนอื่น ๆ (Clauser and Others, 1991) ใช้วิธี Mantel - Haenszel วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบอย่างสม่ำเสมอ พบว่าวิธี Mantel - Haenszel จะใช้ได้ผลดีในกรณีที่ข้อสอบมีค่าอำนาจจำแนกสูง แต่จะไม่สามารถวิเคราะห์ข้อสอบที่มีค่าความยากมากได้ ซึ่งสอดคล้องกับข้อค้นพบของเมเซอร์และคนอื่น ๆ (Mazor and Others, 1991) ซูดวีคส์และทอลแมน (Sudweeks and Tolman, 1990) ที่พบว่าข้อสอบที่ทำหน้าที่ต่างกันมักเป็นข้อสอบที่ยาก บากิและเฟอรารา (Baghi and Ferrara, 1990) พบว่า เมื่อใช้กลุ่มตัวอย่างขนาด 750 คนขึ้นไปวิธี Mantel - Haenszel ใช้แทนวิธีทฤษฎีการตอบข้อสอบ 3 พารามิเตอร์ได้ สวามินาธานและโรเจอร์ส (Swaminathan and Rogers, 1990) ได้ศึกษาจากข้อมูลจำลองพบว่า วิธี Mantel - Haenszel วิเคราะห์ได้ดีกว่าการถดถอยแบบโลจิสติกน้อย โดยตรวจค้นได้ถูกต้องร้อยละ 75 ในกรณีใช้กลุ่มตัวอย่าง 250 คน และตรวจค้นได้ถูกต้องร้อยละ 100 กรณีกลุ่มตัวอย่าง 500 คน กรณีที่การทำหน้าที่ต่างกันของข้อสอบอย่างสม่ำเสมอและกรณีที่ไม่สม่ำเสมอที่ติดกันที่ปลายข้างใดข้างหนึ่ง แต่วิธี Mantel - Haenszel มีค่าใช้จ่ายน้อยกว่าวิธีโลจิสติกประมาณ 3 - 4 เท่า แฮมเบิลตันและคนอื่น ๆ (Hambleton and Others, 1986) พบว่า วิธี Mantel - Haenszel ให้ค่าใกล้เคียงกับทฤษฎีการตอบข้อสอบ ทั้งที่ใช้ค่าความแตกต่างของค่าเฉลี่ยกำลังสองและการตรวจสอบความแตกต่างของพื้นที่รวมใต้โค้ง แต่วิธี Mantel - Haenszel มีค่าใช้จ่ายต่ำกว่าและใช้เวลาน้อยกว่า ลินเคอร์ (Linacre, 1988) พบว่า วิธี Mantel - Haenszel ใช้ได้ดีในสถานการณ์จำลองทุกสถานการณ์พอ ๆ กับโปรแกรม PROX ของราสช์ (Rasch) แต่ควรใช้เกณฑ์ในการคำนวณการทำหน้าที่ต่างกันของข้อสอบและความคลาดเคลื่อนหลาย ๆ เกณฑ์ และใช้ตัวประมาณค่าความคลาดเคลื่อนมาตรฐานมากกว่า 1 ตัว เพลร์ลแมนและคนอื่น ๆ (Perlman and Others, 1988) พบว่า วิธี Mantel - Haenszel มีปัญหาด้านความเชื่อมั่น เมื่อจำนวนกลุ่มตัวอย่างน้อยกว่า 660 คน ทิสเซน สไตน์เบอร์ก และวายเนอร์ (Thissen, Steinberg and Wainer, 1988) พบว่าวิธี Mantel - Haenszel ให้ผลการวิเคราะห์คล้ายกับวิธีทฤษฎีการตอบข้อสอบแบบกระตดถอยเชิงเส้น (IRT - LR) และอาจใช้วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบก่อนใช้วิธีทฤษฎีการตอบข้อสอบแบบกระตดถอยเชิงเส้น (IRT - LR)

5. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธี SIBTEST

SIBTEST เป็นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เสนอโดยซิลลีและสเตาท์ (Shealy and Stout, 1993:159-194) ซึ่งมีแนวคิดพื้นฐานบนทฤษฎีการตอบข้อสอบแบบพหุมิติ (Multidimensional) แต่ SIBTEST แตกต่างจาก IRT ที่เป็นการใช้การทดสอบค่าสถิติพารามิเตอร์ (non-parametric) โดย SIBTEST มีข้อตกลงว่า มีมิติการวัด 2 มิติ มิติหนึ่งเป็นความสามารถหรือลักษณะแฝงเป้าหมายที่ต้องการวัด และอีกมิติหนึ่งเป็นลักษณะแฝงแทรกซ้อนที่ไม่ต้องการวัดใช้ได้สำหรับ การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบอย่างสม่ำเสมอและมีทิศทางเดียว (unidirectional)

ซิลลีและสเตาท์ เริ่มศึกษาจากกลุ่มผู้สอบ 2 กลุ่ม คือ กลุ่มอ้างอิง (Reference Group) และ กลุ่มเปรียบเทียบ (Focal Group) การตอบข้อสอบของผู้สอบที่สุ่มมาจากกลุ่มผู้สอบคือ U โดยที่ $U = (U_1, \dots, U_n)$ เมื่อ U_i มีค่าเป็น 0 ถ้าตอบข้อสอบข้อนั้นผิด และมีค่าเป็น 1 เมื่อตอบถูก สำหรับโมเดล IRT แล้วโดยทั่วไป U จะเกิดจาก 2 องค์ประกอบ คือ (1) พารามิเตอร์ความสามารถของผู้สอบจำนวน d มิติ และ (2) ฟังก์ชันการตอบข้อสอบแต่ละข้อ ซึ่งเป็นตัวกำหนดความน่าจะเป็นของการตอบข้อสอบได้ถูก

ในการศึกษาครั้งนี้ $d = 2$ เพราะเราจะพิจารณาความสามารถแทรกซ้อนเพิ่มเข้ามาจากความสามารถเป้าหมายที่ต้องการวัด เวกเตอร์ความสามารถคือ (θ, η) เมื่อ θ คือ ความสามารถหรือลักษณะแฝงเป้าหมาย และ η คือความสามารถสอดแทรกหรือความสามารถแฝงแทรกซ้อน ส่วนฟังก์ชันการตอบข้อสอบสำหรับข้อสอบข้อที่ i คือ $P_i(\theta, \eta)$ โดยที่ข้อสอบทุกข้อจะประกอบด้วย θ ในขณะที่บางข้อจะมี η อยู่ด้วย สำหรับข้อที่ประกอบด้วย θ อย่างเดียว ฟังก์ชันการตอบข้อสอบคือ $P_i(\theta)$ ค่า $P_i(\theta, \eta)$ จะสูงขึ้นใน θ ละใน η เมื่อข้อสอบข้อ i ต้องการความสามารถทั้ง 2 อย่าง และจะสูงขึ้นใน θ ถ้าต้องการเฉพาะความสามารถ θ

นิยามการทำหน้าที่ต่างกันของข้อสอบใน SIBTEST คือ “การทำหน้าที่ต่างกันของข้อสอบจะเกิดขึ้นต่อกลุ่มเปรียบเทียบที่ระดับความสามารถ θ ถ้าฟังก์ชันการตอบข้อสอบสำหรับความสามารถเป้าหมาย (Target ability) ของกลุ่มอ้างอิงมีค่ามากกว่าฟังก์ชันการตอบข้อสอบของกลุ่มเปรียบเทียบ” (Shealy and Stout, 1993 : 159-194) ซึ่งแสดงในรูปสัญลักษณ์ได้ดังนี้

$$T_R(\theta) > T_F(\theta)$$

$$\begin{aligned} \text{เมื่อ} \quad T_R(\theta) &= E_R[P(\theta, \eta)/\theta] \\ &= \int p(\theta, \eta) f_R(\eta/\theta) d\eta \end{aligned}$$

$$\begin{aligned} T_F(\theta) &= E_F[P(\theta, \eta)/\theta] \\ &= T_F(\theta) \\ \theta &= (\theta, \eta) \end{aligned}$$

โดยที่	R	คือ	กลุ่มอ้างอิง
	F	คือ	กลุ่มเปรียบเทียบ
	θ	คือ	ความสามารถเป้าหมาย

η	คือ ความสามารถสอบแทรก
$T_R(\theta)$	คือ ฟังก์ชันการตอบข้อสอบของกลุ่มอ้างอิง
$T_F(\theta)$	คือ ฟังก์ชันการตอบข้อสอบของกลุ่มเปรียบเทียบ

ข้อสอบจะทำหน้าที่ต่างกันเมื่อ

$$\eta_F / \theta = \eta_R / \theta$$

นियามการทำหน้าที่ต่างกันของข้อสอบของแบบทดสอบทั้งฉบับ เมื่อให้ U เป็นการตอบข้อสอบแบบสุ่มของผู้สอบคนหนึ่ง $h(U)$ คือ คะแนนจากแบบทดสอบ โดยที่ $h(U) = \sum U_i$ แล้วจะได้ว่าข้อสอบทำหน้าที่ต่างกันต่อกลุ่มเปรียบเทียบในระดับความสามารถ θ ถ้า

$$T_R(\theta) > T_F(\theta)$$

$$T_R(\theta) = \text{def} E_R[h(U)/\theta]$$

$$T_F(\theta) = \text{def} E_F[h(U)/\theta]$$

โดยปริมาณของการทำหน้าที่ต่างกันของข้อสอบที่ระดับ θ คือ

$$\beta(\theta) = \text{def} T_R(\theta) - T_F(\theta)$$

และดัชนีของการทำหน้าที่ต่างกันของข้อสอบ คือ

$$\beta_u = \int B(\theta) f_F(\theta) d\theta$$

ซึ่งเป็นความลำเอียงเฉลี่ยต่อผู้สอบแต่ละคนในกลุ่มที่สนใจศึกษา

สมมติฐานเพื่อการทดสอบการทำหน้าที่ต่างกันของข้อสอบของ SIBTEST ได้แก่

$$H_0 : \beta = 0$$

$$H_1 : |\beta| > 0$$

การทดสอบทำได้โดยการแบ่งแบบทดสอบออกเป็นแบบทดสอบย่อย 2 ฉบับ คือ V เป็นแบบทดสอบที่มีความเที่ยงตรง และ S เป็นแบบทดสอบที่ต้องการศึกษา และคะแนนของแต่ละแบบทดสอบย่อย คือ

$$Y = \frac{\sum u_i}{s}$$

$$X = \frac{\sum u_i}{v}$$

สำหรับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบรายข้อ S คือ ข้อสอบแต่ละข้อที่ตรวจสอบ และ V คือ ข้อสอบที่เหลือในแต่ละรอบของการตรวจสอบ

$\bar{Y}_{Rk}$ คือ คะแนนเฉลี่ยที่ได้จากแบบทดสอบย่อยที่ศึกษาสำหรับผู้เข้าสอบทุกคนที่มีคะแนนรวมในแบบทดสอบย่อยที่ถือว่าเป็นแบบทดสอบที่มีความเที่ยงตรง $(X) = k$ และการทำหน้าที่ต่างกันวัดได้จาก $(\bar{Y}_{Rk} - \bar{Y}_{Fk})S$ ทั้งนี้เพราะ $(\bar{Y}_{Rk} - \bar{Y}_{Fk})$ เป็นความแตกต่างของการตอบของกลุ่มย่อยที่มีความสามารถระดับเดียวกัน $(=k)$ บนความสามารถเป้าหมาย และกรณีที่ข้อสอบไม่ทำหน้าที่ต่างกันค่า $(\bar{Y}_{Rk} - \bar{Y}_{Fk}) = 0$

สำหรับค่าสถิติที่ประมาณค่าความลำเอียง β ได้จาก

$$\beta = \sum_{k=0}^n P_k (\bar{Y}_{Rk} - \bar{Y}_{Fk})$$

เมื่อ $P_k =$ สัดส่วนของกลุ่มผู้สอบกลุ่มที่สนใจศึกษาที่ตอบแบบทดสอบย่อยที่ถือว่าไม่ทำหน้าที่ต่างกันได้ถูกต้อง k ข้อ

สถิติทดสอบของ SIBTEST คือ

$$\beta = \frac{\hat{\beta}}{\sigma(\hat{\beta})} \quad \text{ซึ่ง} = N(0,1) \quad \text{เมื่อ} \quad \beta = 0$$

โดยที่

$$\sigma(\beta) = \left[\sum_{K=0}^n P_k \left[(1/J_{Rk}) \sigma^2(y|k, R) + (1/J_{Fk}) \sigma^2(y|k, F) \right] \right]^{\frac{1}{2}}$$

เมื่อ $\sigma^2(Y|k, g)$ เป็นความแปรปรวนของคะแนนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบในช่วงคะแนน k เมื่อมีความแตกต่างระหว่างความสามารถเป้าหมาย ค่าสถิติ β จะสูงกว่าปกติ ทำให้เกิดการระบุว่าข้อสอบทำหน้าที่ต่างกันทั้ง ๆ ที่ข้อสอบไม่ทำหน้าที่ต่างกัน

การทดสอบนัยสำคัญทางสถิติ เมื่อ $H_0: \beta = 0$ ได้แก่

$$Z_{sib} = \frac{\beta}{SE(\beta)}$$

และจะปฏิเสธสมมติฐานศูนย์ที่ว่า $\beta = 0$

เมื่อ $\beta > Z_\alpha$ โดยที่ค่า Z_α กำหนดไว้จาก $P[N(0,1)] > Z_\alpha = \alpha$

ในการวิจัยครั้งนี้ใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบ 3 วิธี คือ

1. วิธี Lord's χ^2 วิเคราะห์โดยใช้โปรแกรม IRTDIF 1.0 พัฒนาโดยคิมและ โคเฮน (Kim and Cohen. 1991)
2. วิธี Mantel – Haenszel วิเคราะห์โดยใช้โปรแกรม SIBTEST พัฒนาโดย Stout and Rousos (1992)
3. วิธี SIBTEST วิเคราะห์โดยใช้โปรแกรม SIBTEST พัฒนาโดย Stout and Rousos (1992)

6. ทฤษฎีสมรรถภาพทางสมองและความถนัด

ความหมายของความถนัด

นักการศึกษา นักจิตวิทยา และนักวัดผลการศึกษา ได้ให้ความหมายของความถนัด (Aptitude) ไว้แตกต่างกันดังนี้ (ภาวนา วงศ์เพ็ญทักษ์. 2541 : 23)

ครอนบาค (Cronbach. 1970 : 38) กล่าวว่า ความถนัดเป็นประสบการณ์การเรียนรู้ที่กว้างขวาง และอ้างอิงถึงสถานการณ์อนาคต

ซวาล แพร์ตกุล (2517 : 1) กล่าวว่า ความถนัด หมายถึง สมรรถวิสัย (Capacity) และทิศทางแห่งความองกงามของสมอง หรือกล่าวให้ง่ายขึ้นก็หมายถึง ชีตความสามารถสูงสุดของบุคคลที่เขาอาจมีได้ต่อการเรียนรู้และการฝึกฝนในวิทยาการต่าง ๆ และทักษะทั้งปวง ถ้าเขาได้รับการฝึกฝนและประสบการณ์ที่เหมาะสม

สมบุรณ์ ชิตพงศ์และสำเร็จ บุญเรืองรัตน์ (2518 : 17) กล่าวว่า ความถนัด หมายถึง ศักยภาพที่มีอยู่ในตัวบุคคลอันมีผลมาจากการฝึกฝนความรู้และประสบการณ์ทั้งสิ้นทั้งปวง

ล้วน สายยศ และอังคณา สายยศ (2541 : 27) กล่าวว่า ความถนัด หมายถึง ความสามารถที่บุคคลได้รับประสบการณ์ ฝึกฝนตนเอง และสะสมไว้มากจนเกิดเป็นทักษะเด่นชัดด้านใดด้านหนึ่ง พร้อมจะปฏิบัติกิจกรรมด้านนั้นได้อย่างดี

จากความหมายของความถนัดดังกล่าว อาจสรุปได้ดังนี้ ความถนัด หมายถึง ความสามารถสมองของแต่ละบุคคลที่เกิดจากการเรียนรู้ ประสบการณ์ การฝึกฝน จนเกิดเป็นทักษะที่เด่นชัด และสามารถแสดงออกได้อย่างมีประสิทธิภาพ

ทฤษฎีสมรรถภาพสมองและความถนัด

ทฤษฎีเกี่ยวกับสมรรถภาพสมองและความถนัดมีหลายทฤษฎี แต่ที่สำคัญมีดังต่อไปนี้

(ล้วน สายยศ และอังคณา สายยศ. 2541 : 42 – 59)

1. ทฤษฎีองค์ประกอบเดียว (Uni – factor Theory) บางทีทฤษฎีนี้เรียกว่า Global Theory ผู้คิดทฤษฎีนี้คือ บิเนต์และซิมอน (Binet and Simon. 1905) ทฤษฎีนี้เสนอ โครงสร้างเชาวันปัญญาเป็นอันหนึ่งอันเดียวไม่แบ่งเป็นส่วนย่อยคล้ายกับความสามารถทั่วไป (General Ability) นั่นเอง ในปี ค.ศ.

1905 หรือ พ.ศ. 2448 บีเนท์และซิมอน ได้สร้างแบบทดสอบตามแนวคิดของเขาเป็นครั้งแรก ข้อสอบฉบับนี้สร้างวัดเชาห์ปัญญาเป็นแบบ Global Measure คือ วัดออกมาเป็นคะแนนเดียวแล้วแปลความหมายออกมาว่าใครมีปัญญาระดับใด

2. ทฤษฎีสององค์ประกอบ (Bi – Factor Theory) ทฤษฎีนี้ได้นำโดยนักจิตวิทยาชาวอังกฤษชื่อ สเปียร์แมน (Charles Spearman) ในปี ค.ศ. 1927 ทฤษฎีนี้เกิดจากการวิเคราะห์คุณลักษณะโดยกระบวนการทางสถิติ พบว่า สมรรถภาพทางสมองของมนุษย์มีองค์ประกอบอยู่สองประการคือ องค์ประกอบทั่วไป (General Factor) เรียกย่อ ๆ ว่า G – Factor และองค์ประกอบเฉพาะ (Specific Factor) เรียกย่อ ๆ ว่า S – Factor แต่ละองค์ประกอบมี กิจกรรมแบบเฉพาะตัวชนิดหนึ่งของมัน สมรรถภาพทั่วไปที่เรียกว่า G – Factor นั้น จะ สอดแทรกอยู่ในทุกอิริยาบถของความคิดและการกระทำของมนุษย์ แต่ละคนก็มีสมรรถภาพสมองชนิดทั่วไปนี้แตกต่างกันออกไปมากน้อยตามแต่ละบุคคล ส่วนองค์ประกอบเฉพาะหรือ S – Factor นั้นเป็นองค์ประกอบที่สำคัญที่ทำให้มนุษย์แตกต่างกัน และเป็นความสามารถพิเศษที่มีอยู่ในแต่ละบุคคลเช่น ความสามารถพิเศษด้านดนตรี ด้านศิลปะ การวาดเขียน ทางเครื่องดนตรีกลไก และทางช่างต่าง ๆ เป็นต้น

3. ทฤษฎีหลายองค์ประกอบ (Multi – Factor Theory) ทฤษฎีนี้เป็นที่ยอมรับกันอย่างกว้างขวาง ของนักจิตวิทยาชาวอเมริกัน ผู้นำในการสร้างทฤษฎีนี้คือ เทอร์สโตน (L.L. Thurstone) เสนอทฤษฎีนี้เมื่อปี ค.ศ. 1933 โดยทำการวิจัยโครงสร้างทางสมองออกได้หลายอย่าง แต่ที่เห็นเด่นชัดและสำคัญ ๆ มีอยู่ 7 ประการ คือ

3.1 องค์ประกอบด้านภาษา (Verbal Factor ใช้ตัวย่อว่า V.) องค์ประกอบทางสมองส่วนนี้จะส่งผลให้รู้ถึงความสามารถด้านความเข้าใจในภาษาและการสื่อสารทั่วไป ผู้มีองค์ประกอบด้านนี้สูงจะมีความสามารถในการอ่านเรื่อง อ่านแบบเข้าใจความหมาย รู้ความสัมพันธ์ของคำ รู้ความหมายของศัพท์ได้เป็นอย่างดี

3.2 องค์ประกอบด้านความคล่องแคล่วในการใช้คำ (Word Fluency Factor ใช้ตัวย่อว่า W.) เป็นความสามารถที่จะใช้คำได้มากในเวลาจำกัด เช่น ให้หาคำที่ขึ้นต้นด้วย “ต” มากที่สุดในเวลาจำกัด

3.3 องค์ประกอบด้านจำนวน (Number Factor ใช้ตัวย่อว่า N.) เป็นความสามารถที่มองเห็นความสัมพันธ์ และความหมายของจำนวน และมีความแม่นยำคล่องแคล่วในการบวก ลบ คูณหาร ในวิชาเลขคณิตได้เป็นอย่างดี

3.4 องค์ประกอบด้านมิติสัมพันธ์ (Space Factor ใช้ตัวย่อว่า S.) เป็นความสามารถที่จะเข้าใจถึงขนาดและมิติต่าง ๆ ได้แก่ ความสั้น ยาว ไกล ใกล้ และพื้นที่หรือทรวดทรงที่มีขนาดและปริมาตรที่ต่างกันไป สามารถสร้างจินตนาการให้เห็นส่วนย่อยและส่วนผสมของวัตถุต่าง ๆ ที่นำมาซ้อนทับกัน สามารถรู้ความสัมพันธ์ของรูปทรงเลขาคณิตเมื่อเปลี่ยนแปลงที่อยู่

3.5 องค์ประกอบด้านความจำ (Memory Factor ใช้ตัวย่อว่า M.) เป็นความสามารถในด้านความทรงจำเรื่องราว และสตรึงการรู้จนสามารถถ่ายทอดได้

3.6 องค์ประกอบด้านการสังเกตพิจารณา (Perceptual Speed Factor ใช้ตัวย่อว่า P.) เป็นความสามารถในการมองเห็นรายละเอียดความคล้ายคลึง หรือความแตกต่างระหว่างสิ่งของต่าง ๆ อย่างรวดเร็วถูกต้อง

3.7 องค์ประกอบด้านเหตุผล (Reasoning Factor ใช้ตัวย่อว่า R.) เป็นความสามารถในการจัดประเภท อุปมาอุปไมย และสรุปความได้อย่างรวดเร็วและถูกต้อง

4. ทฤษฎีไฮราคิคอล (Hierarchical Theory) ผู้นำทฤษฎีคือ เวอร์นอน (Vernon) เบิร์ต (Burt) ชาวอังกฤษ และฮัมฟรีย์ (Humphreys) ชาวอเมริกา โดยเฉพาะเวอร์นอนได้เสนอโครงสร้างของเชาว์ปัญญาในปี ค.ศ. 1960 โดยเริ่มต้นอธิบายตามแบบของสเปียร์แมน นั่นคือ เวอร์นอนเริ่มจัดแรกด้วย G – Factor ขึ้นต่อไปแบ่งออกเป็น 2 องค์ประกอบใหญ่ ๆ คือ ความถนัดทางภาษา (Verbal Education ; V : ed) ความถนัดทางช่าง (Practical Mechanical ; K : ed) ซึ่งองค์ประกอบใหญ่ทั้งสองรวมเรียกว่า Major Group Factors องค์ประกอบใหญ่ทั้งสองยังแบ่งองค์ประกอบย่อยลงไปอีก เรียกว่า Verbal Education ; V : ed และยังแบ่งย่อยออกเป็นองค์ประกอบด้านภาษา (Verbal) องค์ประกอบด้านตัวเลข (Number) และอื่น ๆ อีก ส่วนด้าน K ; m ได้แบ่งย่อยออกเป็นความรู้เชิงกล (Mechanical Information) มิติสัมพันธ์ (Spatial) ความสามารถในการใช้กลไกของกล้ามเนื้อ (Psychomotor Abilities) และอื่น ๆ อีก ซึ่งแต่ละองค์ประกอบย่อยยังแบ่งเป็นองค์ประกอบย่อย ๆ ลงไปอีกเป็นองค์ประกอบต่ำที่สุด เรียกว่า องค์ประกอบเฉพาะ (Specific Factors)

5. ทฤษฎีโครงสร้างสามมิติของปัญญา (Three Faces of Intellect Model) ทฤษฎีนี้สร้างขึ้นโดยกิลฟอร์ด (Guilford) ได้วิเคราะห์ความสัมพันธ์ของคุณลักษณะของเชาว์ปัญญาเป็น 3 มิติ คือ

มิติที่ 1 ด้านกระบวนการหรือวิธีการของการคิด (Operations) มีองค์ประกอบย่อย 5 ส่วน คือ การรู้การเข้าใจ (Cognition) ความจำ (Memory) การคิดออกนอกนัย (Divergent Production) การคิดแบบเอกนัย (Convergent Production) และการคิดแบบประเมินค่า (Evaluation)

มิติที่ 2 ด้านเนื้อหา (Content) เป็นด้านที่ประกอบด้วยสิ่งเร้าและข้อมูลต่าง ๆ แบ่งออกเป็น 4 อย่าง ดังนี้ ภาพ (Figural) สัญลักษณ์ (Symbolic) ภาษา (Semantic) และพฤติกรรม (Behavioral)

มิติที่ 3 ผลการคิด (Product) เป็นผลของกระบวนการจัดกระทำของความคิดกับข้อมูลจากเนื้อหา แบ่งออกเป็น 6 อย่าง ดังนี้ หน่วย (Units) จำพวก (Classes) ความสัมพันธ์ (Relations) ระบบ (System) การแปลงรูป (Transformations) และการประยุกต์ (Implications)

6. ทฤษฎีความสามารถทางสมองสองระดับ (Two – Level Theory of Mental Ability) ทฤษฎีนี้เสนอโดยเจนเซน (Jensen) เจนเซนเสนอทฤษฎีว่าความสามารถทางสมองมีอยู่สองระดับ ระดับ I (Level I) เป็นความสามารถด้านการเรียนรู้และจำอย่างนกแก้ว นั่นคือ เป็นความสามารถที่จะสะสมหรือเก็บข้อมูลไว้ได้และพร้อมจะระลึกนึกออกได้ ระดับ II (Level II) เป็นระดับของการจัดกระทำทางสมองเป็นขั้นสร้างมโนภาพเหตุผล และแก้ปัญหา ระดับ II นี้ดูไปแล้วก็เหมือนองค์ประกอบทั่วไป (G – Factor) นั่นเอง

7. ทฤษฎีปัญญาของแคทเทิลล์ ทฤษฎีนี้คิดโดย อาร์ บี แคทเทิลล์ (R.B. Cattell) เขาเสนอทฤษฎีปัญญาว่า โครงสร้างเชาว์ปัญญาแบ่งออกเป็น 2 ส่วน คือ

7.1 ส่วนที่เป็นฟลูอิด (Fluid Component) เป็นความสามารถทั่วไป ผู้ที่มีปริมาณความสามารถด้านนี้สูง จะสามารถทำงานชนิดต่าง ๆ ได้ดี เช่น ความสามารถด้านเหตุผลเชิงอุปมาน และอนุมานเหตุผลสัมพันธ์ ความสามารถเข้าใจความเปลี่ยนแปลงของอนุกรมภาพ เป็นต้น

7.2 ส่วนที่เป็นคริสตอลไลซ์ (Crystallized Component) เป็นความสามารถที่เชื่อมโยงกับวัฒนธรรมและสิ่งแวดล้อมอย่างใกล้ชิด พุดง่าย ๆ ว่าความสามารถที่จะเข้าใจภาษา ความสามารถที่จะประเมินคุณค่า เป็นต้น

ประโยชน์ของแบบทดสอบวัดความถนัด

ล้วน สายยศ และอังคณา สายยศ (2541 : 22-27) กล่าวถึงประโยชน์ของแบบทดสอบวัดความถนัดดังนี้

1. ใช้ในการสอบคัดเลือก
2. ใช้ในการแยกประเภทนักเรียน
3. ใช้ในการวินิจฉัยความสามารถ
4. ใช้ในการพยากรณ์ความสำเร็จ
5. ใช้ในการวัดพัฒนาการ
6. ใช้ในการเปรียบเทียบสติปัญญา
7. ใช้ในการประเมินผลการศึกษา
8. ใช้ในการวิจัย

7. งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่ศึกษาเกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ส่วนใหญ่จะศึกษาร่วมกับวิธีอื่น ๆ โดยมีผู้ศึกษาไว้มากพอสมควร ในการศึกษาครั้งนี้แยกกล่าวเป็น 2 ส่วน คือ งานวิจัยในประเทศ และงานวิจัยต่างประเทศ

7.1 งานวิจัยในประเทศ

สำหรับงานวิจัยในประเทศที่ศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบ มีดังนี้

ชัชชัย เผ่าพงศ์ (2527) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบมาตรฐานวัดความถนัดทางการเรียนคณิตศาสตร์และภาษาระดับมัธยมศึกษาตอนต้นจำนวน 2 ฉบับ ซึ่งพัฒนาโดยสำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ระหว่างกลุ่มผู้สอบเพศชายและเพศหญิง โดยใช้วิธีวิเคราะห์ความแตกต่างของโครงสร้างข้อสอบแบบ 3 พารามิเตอร์ กลุ่มตัวอย่างที่ศึกษาเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2524 จำนวน 5,248 คน จากทุกภาคภูมิศาสตร์ทั่วประเทศ ได้แก่ ภาคเหนือ ภาคกลาง ภาคใต้ ภาคตะวันออก และภาคตะวันออกเฉียงเหนือ แบ่งเป็นกลุ่มที่วิเคราะห์ด้วยแบบทดสอบวัดความถนัดทางการเรียนคณิตศาสตร์ 2,947 คน เป็นชาย 1,610 คน และหญิง 1,337 คน ส่วนกลุ่มที่วิเคราะห์ด้วยแบบทดสอบวัดความถนัดทางภาษามีจำนวน 2,310 คน เป็นชาย 1,316 คน หญิง 994 คน ผลการวิจัยพบว่า แบบทดสอบวัดความถนัดทางการเรียนคณิตศาสตร์จำนวน 30 ข้อ พบข้อสอบที่ทำหน้าที่ต่างกันต่อกลุ่มใดกลุ่มหนึ่งโดยเฉพาะจำนวน 9 ข้อ คือ ข้อสอบที่ทำหน้าที่ต่างกันต่อกลุ่มเพศชาย 7 ข้อ และทำหน้าที่ต่างกันต่อกลุ่มเพศหญิง 2 ข้อ ข้อสอบที่ทำหน้าที่ต่างกันต่อกลุ่มเพศชายและหญิงในระดับปานกลางขึ้นไปจำนวน 5 ข้อ ซึ่งวัดในเรื่องร้อยละ การหาปริมาตรและความยาวของเส้นรอบรูปสามเหลี่ยมจำนวนเรื่องละ 1 ข้อ และโจทย์ปัญหาจำนวน 2 ข้อ ข้อสอบจำนวน 5 ข้อนี้เป็นข้อสอบที่ทำหน้าที่ต่างกันต่อกลุ่มเพศหญิงในช่วงความสามารถแรก และทำหน้าที่ต่างกันต่อกลุ่มเพศชายในช่วงความสามารถต่อมาจำนวน 1 ข้อ ส่วนแบบทดสอบวัดความถนัดทางการเรียนด้านภาษาจำนวน 40 ข้อ มีข้อสอบที่ทำหน้าที่ต่างกันเฉพาะกลุ่มเพศชายจำนวน 3 ข้อ และกลุ่มเพศหญิงจำนวน 8 ข้อ ข้อสอบที่ทำหน้าที่ต่างกันระดับปานกลางขึ้นไปจำนวน 9 ข้อ ซึ่งวัดความเข้าใจเกี่ยวกับการอ่านคำประพันธ์ ร้อยกรอง อย่างละ 1 ข้อ และวัดความเข้าใจเกี่ยวกับการอ่านข้อความ 7 ข้อ และ

เป็นข้อสอบที่ทำหน้าที่ต่างกันต่อกลุ่มเพศหญิงจำนวน 6 ข้อ และทำหน้าที่ต่างกันต่อเพศชายจำนวน 2 ข้อในระดับความสามารถต้น ๆ และทำหน้าที่ต่างกันต่อกลุ่มเพศหญิงในระดับความสามารถต่อมา

ทัศนีย์ พิรมนตรี (2530) ได้วิเคราะห์การทำหน้าที่ต่างกันของแบบทดสอบวิชาคณิตศาสตร์ ตามโครงการตรวจสอบคุณภาพทางการศึกษา ของสำนักทดสอบทางการศึกษา กรมวิชาการ กระทรวงศึกษาธิการ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 จำนวน 7,036 คน โดยเปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันระหว่างกลุ่มนักเรียนในกรุงเทพมหานครกับกลุ่มนักเรียนในภาคภูมิศาสตร์ 5 ภาค ได้แก่ ภาคเหนือ ภาคกลาง ภาคใต้ ภาคตะวันออกเฉียงเหนือ และภาคตะวันออก ด้วยวิธีวิเคราะห์ 3 วิธี คือ (1) วิธีกำหนดจุดค่าเฉลี่ย (2) วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ (3) วิธีทดสอบความแตกต่างระหว่างกลุ่มด้วยสถิติไคสแควร์ในโมเดลล็อกลิเนียร์ 2 โมเดล คือ โมเดลที่ไม่มีพารามิเตอร์ผลรวมระหว่างระดับคะแนนกับกลุ่ม และโมเดลที่ไม่มีพารามิเตอร์ของผลหลักที่เกิดจากกลุ่ม ผลการวิเคราะห์พบว่า ข้อสอบที่ทำหน้าที่ต่างกันจากวิธีโค้งลักษณะข้อสอบที่ใช้ 3 พารามิเตอร์มีจำนวนข้อมากที่สุด และในแต่ละวิธีวิเคราะห์ พบว่า มีข้อสอบที่ทำหน้าที่ต่างกันซ้ำกันระหว่างกลุ่มนักเรียนในกรุงเทพมหานครกับกลุ่มนักเรียนในทุกภาค

สุรศักดิ์ อมรรตณศักดิ์ (2531) ได้เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบที่แตกต่างกัน 4 วิธี คือ วิธีวิเคราะห์ความแปรปรวน วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ และวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ โดยหาสัมประสิทธิ์สหสัมพันธ์ของดัชนีการทำหน้าที่ต่างกันระหว่างวิธีวิเคราะห์การทำหน้าที่ต่างกัน 4 วิธี และเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อนและหลังการศึกษาการทำหน้าที่ต่างกันของข้อสอบตามวิธีการคิดคะแนนที่แตกต่างกัน 6 วิธี ในด้านจำนวนผู้ได้รับการคัดเลือก สัดส่วนของ ชาย : หญิง ที่ได้รับการคัดเลือก และความเชื่อมั่นของแบบทดสอบ โดยกลุ่มตัวอย่างคือ ผู้เข้าสอบแข่งขันเพื่อบรรจุเข้ารับราชการเป็นข้าราชการครูในปี พ.ศ. 2529 จำนวน 2,340 คน เป็นชาย 1,170 คน หญิง 1,170 คน เครื่องมือที่ใช้ในการรวบรวมข้อมูล ได้แก่ แบบทดสอบที่ใช้สอบแข่งขันเพื่อบรรจุเข้ารับราชการเป็นข้าราชการครูในปี พ.ศ. 2529 ซึ่งมี 4 ฉบับ คือ ฉบับที่ 1 แบบทดสอบวิชาเอก (ความรู้เกี่ยวกับการประถมศึกษา) ฉบับที่ 2 แบบทดสอบวิชาการศึกษาและกฎหมาย ฉบับที่ 3 แบบทดสอบภาษาไทย ฉบับที่ 4 แบบทดสอบความสามารถทั่วไป ผลการศึกษาพบว่า วิธีวิเคราะห์การทำหน้าที่ต่างกันแต่ละวิธีพบจำนวนข้อสอบที่ทำหน้าที่ต่างกันแตกต่างกัน โดยวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด รองลงมาได้แก่ วิธีวิเคราะห์ความแปรปรวน และวิธีแปลงค่าความยากของข้อสอบพบจำนวนข้อที่ทำหน้าที่ต่างกันน้อยที่สุด ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีวิเคราะห์ 4 วิธี พบว่ามีค่าอยู่ระหว่าง 0.7535 ถึง 0.9921 ในด้านจำนวนผู้ได้รับการคัดเลือกเมื่อเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อนและหลังการศึกษาการทำหน้าที่ต่างกันของข้อสอบตามวิธีการคิดคะแนนกับวิธีรวมคะแนนแบบอื่น ๆ อีก 5 วิธี มีจำนวนผู้ได้รับการคัดเลือกแตกต่างกันประมาณ 4% ถึง 24% ส่วนวิธีแปลงคะแนนเป็นคะแนนมาตรฐานที่ปกติกับวิธีรวมคะแนนแบบอื่น ๆ อีก 4 วิธี มีจำนวนผู้ได้รับการคัดเลือกแตกต่างกันประมาณ 6% ถึง 23% ส่วนในเรื่องความเชื่อมั่นของแบบทดสอบ พบว่า ค่าความเชื่อมั่นของแบบทดสอบลดลงกว่าแบบทดสอบเดิมเล็กน้อย

พัชรี ปิยะภักดิ์ (2531) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 เรื่อง สมการ จำนวน 45 ข้อที่ผู้วิจัยสร้างขึ้น ที่มีต่อกลุ่มผู้สอบเพศชายและเพศหญิง และกลุ่มผู้สอบในกรุงเทพมหานครกับจังหวัดสมุทรสาคร ด้วยวิธี

วิเคราะห์ 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีไคสแควร์ และวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ โดยใช้เทคนิควิเคราะห์ตัวลวงเพื่อตรวจหาแหล่งของการทำหน้าที่ต่างกัน กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2530 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติในกรุงเทพมหานครกับจังหวัดสมุทรสาคร จำนวน 1,422 คน ซึ่งเลือกโดยวิธีสุ่มแบบแบ่งชั้น พบว่า ในแต่ละวิธีพบข้อสอบที่ทำหน้าที่ต่างกันแตกต่างกัน วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ พบจำนวนข้อที่ทำหน้าที่ต่างกันมากที่สุดจากแบบทดสอบ 45 ข้อ พบข้อสอบที่ทำหน้าที่ต่างกันระหว่างกลุ่มนักเรียนในกรุงเทพมหานครกับสมุทรสาคร จำนวน 18 ข้อ ส่วนวิธีแปลงค่าความยากของข้อสอบ และวิธีไคสแควร์ พบข้อสอบที่ทำหน้าที่ต่างกันระหว่างกลุ่มนักเรียนชายกับหญิง จำนวน 2 ข้อและ 3 ข้อ ตามลำดับ และพบข้อสอบที่ทำหน้าที่ต่างกันระหว่างกลุ่มนักเรียนในกรุงเทพมหานครกับสมุทรสาครจำนวน 1 ข้อและ 3 ข้อ ตามลำดับ

ค่าสหสัมพันธ์ของดัชนีการทำหน้าที่ต่างกันของข้อสอบจากการวิเคราะห์ 3 วิธี พบว่าดัชนีการทำหน้าที่ต่างกันของข้อสอบระหว่างกลุ่มนักเรียนชายกับหญิง จากการวิเคราะห์โดยวิธีแปลงค่าความยากของข้อสอบกับวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์มีค่า 0.5618 ($p < .001$) และดัชนีการทำหน้าที่ต่างกันของข้อสอบระหว่างนักเรียนในกรุงเทพมหานครกับจังหวัดสมุทรสาครจากการวิเคราะห์โดยวิธีไคสแควร์กับวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์มีค่าเท่ากับ 0.6009 ($p < .001$) ส่วนวิธีอื่นสัมพันธ์กันแบบไม่มีนัยสำคัญทางสถิติ

ค่าความเชื่อมั่นของแบบทดสอบหลังการคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออกไปแล้วแตกต่างกับค่าความเชื่อมั่นของแบบทดสอบก่อนการคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันอย่างไม่มีนัยสำคัญทางสถิติ โดยความเชื่อมั่นของแบบทดสอบก่อนการคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออกมีค่า 0.9277 และค่าความเชื่อมั่นของแบบทดสอบหลังการคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออกระหว่างกลุ่มนักเรียนชายกับกลุ่มนักเรียนหญิง โดยวิธีแปลงค่าความยากของข้อสอบ วิธีไคสแควร์ และวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์มีค่า 0.9316 , 0.9277 และ 0.9286 ตามลำดับ และค่าความเชื่อมั่นของแบบทดสอบหลังการคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันระหว่างกลุ่มนักเรียนในกรุงเทพมหานครกับจังหวัดสมุทรสาครมีค่า 0.9339 , 0.9311 และ 0.9392 ตามลำดับ เมื่อวิเคราะห์หาแหล่งการทำหน้าที่ต่างกันของข้อสอบในข้อสอบที่ทำหน้าที่ต่างกันทุกข้อ ไม่พบว่าแหล่งการทำหน้าที่ต่างกันของข้อสอบแต่ละข้อเกิดจากการเขียนตัวเลือก

สุวัฒน์ สุกมลสันต์ (2534) ได้วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบภาษาอังกฤษ เพื่อคัดเลือกบุคคลเข้าศึกษาในมหาวิทยาลัย โดยใช้ข้อมูลการตอบแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัย ชุด กข และชุด กขค ปี 2531 – 2533 ซึ่งเป็นข้อสอบชุดละ 100 ข้อ ตัวแปรที่ศึกษาการทำหน้าที่ต่างกันของข้อสอบ คือ เพศ และภาคภูมิศาสตร์ของผู้สอบ ซึ่งแยกออกเป็น 5 ภาค ตามภูมิลำเนาของผู้สอบ ได้แก่ ภาคกลาง ภาคตะวันออก ภาคตะวันออกเฉียงเหนือ ภาคเหนือ และภาคใต้ เทคนิคที่นำมาวิเคราะห์ข้อสอบที่ทำหน้าที่ต่างกัน ได้แก่ วิธีกำหนดจุดเดลต้า (Delta Plot Method) วิธีไคสแควร์ชนิดที่แบ่งความสามารถของผู้สอบเป็น 3 ระดับ ได้แก่ กลุ่มความสามารถระดับต่ำ (ผู้ได้คะแนนรวม 0 – 40 คะแนน) กลุ่มความสามารถระดับปานกลาง (ผู้ได้คะแนนรวม 41 – 70 คะแนน) และระดับสูง (ผู้ได้คะแนนรวม 71 – 100 คะแนน) วิธีที่ 3 ที่นำมาใช้วิเคราะห์ข้อสอบที่ทำหน้าที่ต่างกัน ได้แก่ วิธีการวัดพื้นที่ความแตกต่างระหว่างโค้งลักษณะข้อสอบที่วิเคราะห์ตามทฤษฎีการตอบข้อสอบชนิด 3 พารามิเตอร์

เกณฑ์การตัดสินข้อสอบที่ทำหน้าที่ต่างกันสำหรับวิธีต่าง ๆ ที่สุพัฒน์ สุกมลสันต์ กำหนดขึ้นคือ

1. วิธีกำหนดจุดเดลต้า ข้อสอบที่ทำหน้าที่ต่างกัน คือ ข้อสอบที่มีระยะตั้งฉากจากค่าคู่อันดับเดลต้าไปยังเส้นแกนหลัก (d) มากกว่า ± 1.964 SD
2. วิธีไคสแควร์ ข้อสอบที่ทำหน้าที่ต่างกัน คือ ข้อสอบที่ผู้สอบต่างกลุ่มที่อยู่ในระดับคะแนนเดียวกันมีสัดส่วนการตอบถูกหรือผิดแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ $\alpha = 0.017$
3. วิธีวัดพื้นที่ความแตกต่างระหว่างโค้งลักษณะข้อสอบ ข้อสอบที่ทำหน้าที่ต่างกัน คือ ข้อที่มีพื้นที่ความแตกต่างมากกว่า 0.40 โดยแบ่งเป็น 3 ระดับ ได้แก่
 - 3.1 พื้นที่มากกว่า 0.70 แสดงว่า ข้อสอบทำหน้าที่ต่างกันมาก
 - 3.2 พื้นที่ระหว่าง 0.40 – 0.70 แสดงว่า ข้อสอบทำหน้าที่ต่างกันปานกลาง
 - 3.3 พื้นที่ที่มีค่ามากกว่า 0.00 แต่น้อยกว่า 0.40 แสดงว่า ข้อสอบทำหน้าที่ต่างกันน้อย

ผลการวิเคราะห์พบว่า แบบทดสอบภาษาอังกฤษ กข มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันต่อเพศ 7 – 28 ข้อ และแบบทดสอบภาษาอังกฤษ กขค มีจำนวนข้อสอบที่ทำหน้าที่ต่างกันต่อเพศ 4 – 41 ข้อ โดยมีแนวโน้มทำหน้าที่ต่างกันต่อเพศชายมากกว่าเพศหญิง และมีข้อสอบที่ทำหน้าที่ต่างกันต่อภาคอื่นมากกว่าภาคกลาง วิธีตรวจสอบที่พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด คือ วิธีโค้งลักษณะข้อสอบเมื่อกำหนดขนาดพื้นที่มากกว่า 0.40 ขึ้นไป รองลงมาคือ วิธีไคสแควร์ และวิธีกำหนดจุดเดลต้า

นิรมล ชัยเชาวลิต (2537) ได้ทำการเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของแบบทดสอบความเข้าใจในการอ่านภาษาไทย จำนวน 50 ข้อ ที่ผู้วิจัยสร้างขึ้นเอง ระหว่างกลุ่มผู้สอบที่มีเพศต่างกัน ด้วยวิธีวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบตามทฤษฎีคลาสสิกอล 3 วิธี คือ วิธีแปลงค่าความยาก วิธีไคสแควร์ และวิธีวิเคราะห์ความแปรปรวน โดยศึกษาเปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันและศึกษาเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบ หลังคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออกแล้ว โดยคำนวณด้วยสูตร KR-20 และแบบแบ่งครึ่งฉบับ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชายและหญิง ชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2536 ของโรงเรียนสังกัดกรมสามัญศึกษาส่วนกลาง กลุ่มที่ 5 กรุงเทพมหานคร จำนวน 1,006 คน ซึ่งได้มาโดยการสุ่มอย่างง่าย ผลการวิเคราะห์พบว่า การวิเคราะห์การทำหน้าที่ต่างกันด้วยวิธีแปลงค่าความยากกับวิธีไคสแควร์ และวิธีแปลงค่าความยากกับวิธีวิเคราะห์ความแปรปรวน มีจำนวนข้อที่ทำหน้าที่ต่างกันแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีไคสแควร์กับวิธีวิเคราะห์ความแปรปรวน พบจำนวนข้อที่ทำหน้าที่ต่างกันแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับค่าความเชื่อมั่นของแบบทดสอบหลังคัดข้อที่ทำหน้าที่ต่างกันออกแล้ว ที่คำนวณด้วยสูตร KR-20 และแบบแบ่งครึ่งฉบับ ที่วิเคราะห์ด้วยวิธีทั้ง 3 วิธี พบว่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

กาญจนา วัฒนสุนทร (2538) ได้พัฒนาเกณฑ์ตัดสินข้อสอบที่ทำหน้าที่ต่างกันทางเพศ ด้วยข้อมูลเชิงประจักษ์ โดยใช้ดัชนีการทำหน้าที่ต่างกัน 4 วิธี คือ พื้นที่ระหว่างโค้งการตอบข้อสอบแบบมีเครื่องหมาย (SA) และแบบไม่มีเครื่องหมาย (UA) จากทฤษฎีการตอบข้อสอบแบบจำลอง 2 พารามิเตอร์ วิธี Mantel - Haenszel และวิธี SIBTEST ข้อมูลที่ใช้เป็นข้อมูลจากการตอบข้อสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของทบวงมหาวิทยาลัย ปีการศึกษา 2535 โดยใช้แบบทดสอบวิชาคณิตศาสตร์ที่มีความยาว 20 , 30 , 40 ข้อ และแบบทดสอบวิชาภาษาอังกฤษมีความยาว 50 , 60 , 70 และ 80 ข้อ โดยใช้กลุ่มผู้สอบ 6 ขนาด คือ 100 , 200 , 400 , 600 , 800 และ 1,000 คน

การพัฒนาเกณฑ์การตัดสินข้อสอบที่ทำหน้าที่ต่างกันกระทำโดยคำนวณค่าดัชนีทั้ง 4 ตัว จากข้อมูลการตอบข้อสอบของผู้สอบเพศเดียวกัน เพศละ 50 ค่า สำหรับแต่ละความยาวของแบบทดสอบ และขนาดกลุ่มผู้สอบ จากนั้นนำค่าดัชนีที่ได้ทั้งหมดมาวิเคราะห์ค่าเฉลี่ยและกำหนดเกณฑ์จากค่าเฉลี่ย 2 ลักษณะ คือ เกณฑ์ที่กำหนดโดยพิจารณาค่าเฉลี่ยรวมจากค่าดัชนีทุกข้อ โดยไม่พิจารณาความแตกต่างในด้านความยาวของแบบทดสอบและขนาดของกลุ่มผู้สอบ และเกณฑ์ที่พัฒนาจากค่าเฉลี่ย ซึ่งพิจารณาถึงความยาวของแบบทดสอบและขนาดผู้สอบด้วย จากนั้นนำเกณฑ์ที่กำหนดไว้ไปตัดสินค่าดัชนีที่ได้จากการวิเคราะห์ระหว่างผู้สอบเพศหญิงและชาย พบว่า ไม่ว่าจะเกณฑ์ลักษณะใด ผลด้านจำนวนของข้อสอบในรูปของร้อยละของข้อสอบที่ทำหน้าที่ต่างกัน และความสอดคล้องของการตัดสินมีความไม่คงที่ข้ามความยาวของแบบทดสอบและขนาดกลุ่มผู้สอบ โดยที่เกณฑ์ในลักษณะหลังจะลดความไม่คงที่ลงได้ระดับหนึ่ง ทั้งนี้เนื่องจากค่าดัชนีแต่ละตัวที่ได้จากผู้สอบต่างเพศกัน มีความไม่คงที่ แปรเปลี่ยนไปเมื่อความยาวของแบบทดสอบที่ใช้และขนาดกลุ่มผู้สอบเปลี่ยนไป

นอกจากนี้ยังพบว่า ที่ความยาวของแบบทดสอบ 40 ข้อ สำหรับวิชาคณิตศาสตร์ และ 80 ข้อ สำหรับวิชาภาษาอังกฤษ มีจำนวนข้อที่ได้รับการตัดสินว่าทำหน้าที่ต่างกันมากกว่าในความยาวอื่นมาก แล้ว ยังพบว่าเกณฑ์ที่กำหนดได้จากการศึกษาครั้งนี้มีค่าสูงกว่าเกณฑ์ที่ใช้กันอยู่ในกรณีของค่า UA ในขณะที่ค่า MH มีค่าใกล้เคียงกับเกณฑ์ขั้นต่ำที่ควรใช้ในการตัดสินข้อสอบที่ทำหน้าที่ต่างกันที่พัฒนาได้ในครั้งนี้ คือ $ISAI > .40$ $UA > .50$ เมื่อใช้ความยาวแบบทดสอบ 50 ข้อขึ้นไป $ISAI > .80$ $UA > 1.20$ กรณีที่มีความยาวของแบบทดสอบสั้นกว่า 50 ข้อ ค่า $.60 > MH > 1.40$ และ $ISIBI > .06$ และควรปรับใช้เกณฑ์สูงขึ้น โดยพิจารณาถึงวัตถุประสงค์ที่ต้องการใช้แบบทดสอบ แต่ไม่ควรต่ำกว่านี้เพื่อลดปัญหาการเกิดความคลาดเคลื่อนแบบที่ 1 นอกจากนั้นยังต้องการศึกษาเพิ่มเติมเพื่อหาความยาวของแบบทดสอบและจำนวนผู้สอบที่ให้ค่าดัชนีที่มีความคงที่เพิ่มขึ้น

เจน จันทรแสง (2537) ได้ศึกษาดัชนีการทำหน้าที่ต่างกันของข้อสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ เรื่องจำนวนจริง ที่เกิดจากตัวแปรเพศ และที่ตั้งของโรงเรียน โดยใช้วิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ 4 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีวิเคราะห์ความแปรปรวน วิธีไคสแควร์ และวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ และศึกษาสัมประสิทธิ์สหสัมพันธ์ระหว่างดัชนีการทำหน้าที่ต่างกันของข้อสอบทั้ง 4 วิธี และเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบก่อนและหลังการคัดเลือกข้อสอบที่ทำหน้าที่ต่างกัน เครื่องมือที่ใช้เป็นแบบทดสอบชนิดเลือกตอบจำนวน 40 ข้อ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2536 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดนครสวรรค์ จำนวน 1,000 คน พบว่า

1. วิธีวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแต่ละวิธีตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกันแตกต่างกัน โดยวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด โดยพบข้อสอบที่ทำหน้าที่ต่างกันตามตัวแปรเพศ 10 ข้อ และพบข้อสอบที่ทำหน้าที่ต่างกันตามตัวแปรที่ตั้งของโรงเรียน 15 ข้อ รองลงมาคือ วิธีวิเคราะห์ความแปรปรวน และวิธีไคสแควร์ พบข้อสอบที่ทำหน้าที่ต่างกันตามตัวแปรเพศ 7 ข้อ และพบข้อสอบที่ทำหน้าที่ต่างกันตามตัวแปรที่ตั้งของโรงเรียน 14 ข้อ ส่วนวิธีแปลงค่าความยากของข้อสอบเป็นวิธีที่พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด โดยพบข้อสอบที่ทำหน้าที่ต่างกันตามตัวแปรเพศ 1 ข้อ และพบข้อสอบที่ทำหน้าที่ต่างกันตามตัวแปรที่ตั้งของโรงเรียน 5 ข้อ ซึ่งแสดงว่าเพศและที่ตั้งของโรงเรียน เป็นตัวแปรที่ทำให้ข้อสอบทำหน้าที่ต่างกัน

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

3. แบบทดสอบหลังจากคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออกไปแล้ว มีค่าความเชื่อมั่นลดลงจากเดิม อย่างไม่มีนัยสำคัญทางสถิติ และเมื่อปรับค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออกไปแล้วให้มีจำนวนเท่าเดิมโดยใช้สูตร สเปียร์แมน บราวน์ พบว่า ค่าความเชื่อมั่นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

เรวดี อินทสระ (2539) ได้ศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบ คัดเลือกที่ทำหน้าที่ต่างกันต่อเพศ โดยเปรียบเทียบวิธีวิเคราะห์ 3 วิธี คือ วิธีทฤษฎีการตอบข้อสอบ (IRT) วิธี Mantel - Haenszel และวิธี SIBTEST มีการตัดสินผลการสอบด้วยการคิดคะแนนแบบคะแนนมาตรฐานที่ - ปกติ และคะแนนน้ำหนักความสามารถ และศึกษาสาเหตุของการทำหน้าที่ต่างกันของข้อสอบ ซึ่งได้ทำการศึกษาการทำหน้าที่ต่างกันของข้อสอบคัดเลือกเข้าศึกษาในชั้นปีที่ 1 ปีการศึกษา 2538 ของมหาวิทยาลัยสงขลานครินทร์ ประเภทรับตรง ในวิชาภาษาไทย ก วิชาสังคม ก วิชาภาษาอังกฤษ กข วิชาละ 8,127 คน แบ่งเป็นชาย 2,722 คนและหญิง 5,405 คน วิชาภาษาไทย กข วิชาสังคมศึกษา กข และวิชาภาษาอังกฤษ กขค วิชาละ 5,415 คน แบ่งเป็นชาย 1,454 คน และหญิง 3,961 คน ส่วนความเที่ยงตรงเชิงพยากรณ์ศึกษาจากคะแนนสอบคัดเลือกกับระดับผลการเรียนภาคเรียนที่ 1 ปีการศึกษา 2538 ของนักศึกษาที่ได้รับการคัดเลือกจากประเภทรับตรง สายวิทยาศาสตร์ 763 คน และสายศิลปศาสตร์ 281 คน และสาเหตุของการทำหน้าที่ต่างกันของข้อสอบจากการระบุสาเหตุของนักวัดผลการศึกษาหรืออาจารย์ผู้สอนจำนวน 50 คน และนิสิตนักศึกษาที่เรียนในสาขาวิชานั้น ๆ วิชาละ 50 คน

ผลการศึกษาพบว่า วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้ง 3 วิธี ตัดสินจำนวนข้อที่ทำหน้าที่ต่างกันแตกต่างกันในวิชาภาษาไทย ก ฉบับที่ 2 และวิชาสังคมศึกษา ก ฉบับที่ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 โดยวิธีใช้ทฤษฎีการตอบข้อสอบตัดสินจำนวนข้อสอบที่ทำหน้าที่ต่างกันได้มากที่สุด ความสัมพันธ์ของลำดับที่ของการสอบไม่ว่าจะคิดคะแนนแบบคะแนนมาตรฐานที่-ปกติ หรือคิดคะแนนน้ำหนักความสามารถและใช้ข้อสอบจำนวนทั้งหมดหรือใช้เฉพาะข้อสอบที่ไม่ทำหน้าที่ต่างกัน ต่างมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการเรียนในสายวิทยาศาสตร์ การคิดคะแนนน้ำหนักความสามารถทั้งใช้ข้อสอบจำนวนทั้งหมด และใช้เฉพาะข้อสอบที่ไม่ทำหน้าที่ต่างกัน มีประสิทธิภาพในการทำนายสูงกว่าการคิดคะแนนแบบคะแนนมาตรฐานที่-ปกติ ส่วนสายศิลปศาสตร์ การคิดคะแนนมาตรฐานที่-ปกติทั้งใช้ข้อสอบจำนวนทั้งหมด และใช้เฉพาะ ข้อสอบที่ไม่ทำหน้าที่ต่างกัน มีประสิทธิภาพในการทำนายสูงกว่าการคิดคะแนนแบบน้ำหนักความสามารถ และสาเหตุของการทำหน้าที่ต่างกันของข้อสอบต่อเพศทั้งชายและหญิง เกิดจากข้อสอบเป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนมาเฉพาะ จะมีโอกาสตอบถูกมากกว่าเป็นเรื่องราวที่กลุ่มนั้น ๆ สนใจ และเป็นข้อที่ถามความจำ

พรรณี จินตมาต (2540) ได้ศึกษาผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบต่อเพศจากแบบทดสอบคณิตศาสตร์โจทย์ปัญหา ที่ผู้วิจัยสร้างขึ้นเอง โดยจำแนกแต่ละวิธีวิเคราะห์การทำหน้าที่

ต่างกันของข้อสอบ 3 วิธี คือ วิธีแปลงค่าความยาก วิธี Mantel - Haenszel และวิธี SIBTEST และในแต่ละขนาดกลุ่มผู้สอบ คือขนาดกลุ่มผู้สอบ 500 คน และขนาดกลุ่มผู้สอบ 1,000 คน โดยเปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และเปรียบเทียบค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับของแบบทดสอบหลังคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออกแล้ว ในการศึกษาครั้งนี้ มีกลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2539 ของโรงเรียนสังกัดกรมสามัญศึกษา ส่วนกลาง จำนวน 2,200 คน ซึ่งเลือกมาโดยการสุ่มแบบแบ่งชั้น มีขนาดโรงเรียนเป็นชั้นและโรงเรียนเป็นหน่วยการสุ่ม

ผลการวิจัยพบว่า เมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 500 คน และวิธี SIBTEST พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด และวิธีแปลงค่าความยากพบข้อสอบที่ทำหน้าที่ต่างกันน้อยที่สุด โดยจำนวนข้อที่ทำหน้าที่ต่างกันแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทุกวิธีวิเคราะห์ และเมื่อวิเคราะห์จากกลุ่มผู้สอบ 1,000 คน วิธี Mantel - Haenszel พบข้อสอบที่ทำหน้าที่ต่างกันมากที่สุด วิธีแปลงค่าความยากไม่พบข้อที่ทำหน้าที่ต่างกัน โดยจำนวนข้อสอบที่ทำหน้าที่ต่างกันจากการวิเคราะห์ด้วยวิธีแปลงค่าความยากกับวิธี Mantel - Haenszel และวิธีแปลงค่าความยากกับวิธี SIBTEST แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ สำหรับจำนวนข้อสอบที่ทำหน้าที่ต่างกันจากการวิเคราะห์ด้วยวิธีแปลงค่าความยากระหว่างกลุ่มผู้สอบ 500 คนและขนาดกลุ่มผู้สอบ 1,000 คน จะมีจำนวนข้อสอบ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

ค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออกแล้ว เมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 500 คน จากการวิเคราะห์การทำหน้าที่ต่างกันด้วยวิธีแปลงค่าความยากกับวิธีซิปเทสท์ มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และเมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 1,000 คน ค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อที่ทำหน้าที่ต่างกันออกแล้ว จากการวิเคราะห์การทำหน้าที่ต่างกันทั้ง 3 วิธี มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ สำหรับค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออก จากการวิเคราะห์ด้วยวิธีซิปเทสท์ระหว่างกลุ่มผู้สอบ 500 คน และกลุ่มผู้สอบ 1,000 คน มีค่าแตกต่างกันอย่างมี นัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

ญาณภัทร สีหะมงคล (2540) ได้ทำการศึกษาเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord's χ^2 วิธี Raju's Area Measures และวิธี Closed Interval Area เมื่อขนาดกลุ่มตัวอย่าง ความยาวของแบบทดสอบ และสัดส่วนจำนวนข้อสอบที่ทำหน้าที่ต่างกันแบบทดสอบแตกต่างกัน ข้อมูลที่ใช้ในการศึกษาเป็นผลการสอบประเมินคุณภาพและความก้าวหน้าทางการศึกษาวิชาคณิตศาสตร์ ของนักเรียนชั้นประถมศึกษาปีที่ 4 ปีการศึกษา 2536 สังกัดสำนักงานการศึกษาแห่งชาติ 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 ระหว่างวิธี Lord's χ^2 วิธี Raju's Area Measures และวิธี Closed Interval Area แตกต่างกันเมื่อขนาดกลุ่มตัวอย่างและความยาวของ

แบบทดสอบต่างกัน ยกเว้นกรณีกลุ่มตัวอย่าง 500 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 โดยภาพรวมมีสัมประสิทธิ์สหสัมพันธ์ค่อนข้างสูงมาก และมีนัยสำคัญทางสถิติเกือบทุกเงื่อนไขของการศึกษา

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

เมื่อกลุ่มตัวอย่างขนาดใหญ่ วิธี 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 ส่วนใหญ่ลดลง แต่อัตราความสอดคล้องชนิดที่ 2 ส่วนใหญ่เพิ่มขึ้น

จิตติมา วรรณศรี (2540 : 90 – 99) ได้ศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี Mantel - Haenszel และวิธี SIBTEST โดยใช้ข้อมูลจำลองจากโปรแกรม IRTDATA เงื่อนไขการศึกษา ได้แก่

1. ความยาวแบบทดสอบ 3 ขนาด คือ 30 ข้อ 60 ข้อ และ 90 ข้อ
2. ขนาดกลุ่มตัวอย่าง 3 ขนาด คือ 200 คน 600 คน และ 1,000 คน โดยแต่ละขนาดมีอัตราส่วนระหว่างผู้สอบกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบเป็น 4 อัตราส่วน คือ 1 : 1 , 1 : 0.9 , 1 : 0.75 , 1 : 0.5 รวมเงื่อนไขที่ศึกษาทั้งหมด 36 เงื่อนไข

ผลการศึกษา พบว่า วิธี Mantel - Haenszel และวิธี SIBTEST มีประสิทธิภาพเท่าเทียมกันในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่ทุกขนาดกลุ่มตัวอย่างและทุกอัตราส่วน ภายใต้ความยาวแบบทดสอบเดียวกัน เมื่อขนาดกลุ่มตัวอย่าง 200 และ 600 คน สามารถตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องร้อยละ 50 และเมื่อขนาดกลุ่มตัวอย่าง 1,000 คน สามารถตรวจสอบพบข้อสอบที่ทำหน้าที่ต่างกันได้ถูกต้องร้อยละ 100 โดยส่วนมาก วิธี SIBTEST มีอัตราความคลาดเคลื่อนชนิดที่ 1 มากกว่าวิธี Mantel - Haenszel เล็กน้อย และพบว่าวิธี Mantel - Haenszel และวิธี SIBTEST มีประสิทธิภาพเท่าเทียมกันทุกระดับความยาวแบบทดสอบ โดยพบว่า เมื่อใช้แบบทดสอบที่มีความยาวปานกลาง (60 ข้อ) มีประสิทธิภาพในการตรวจสอบมากที่สุด

อารี วัชรโสทธิกุล (2543) ได้ศึกษาเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยใช้รูปแบบต่างกันคือ รูปแบบคะแนนรวมทั้งฉบับ แยกตามเนื้อหา และแยกตามระดับพฤติกรรม ด้วยวิธีการตรวจสอบต่างกัน คือ วิธี SIBTEST และวิธีถดถอยโลจิสติก แล้วทำการคัดเลือกข้อสอบที่ทำหน้าที่ต่างกันออกจากแบบทดสอบ เพื่อเปรียบเทียบค่าความเชื่อมั่น พบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันโดย

ใช้วิธีการตรวจสอบต่างกันแตกต่างกัน ในรูปแบบรวมทั้งฉบับ ส่วนรูปแบบแยกตามเนื้อหา และแยกตามระดับพฤติกรรมไม่แตกต่างกัน

ทองอยู่ สาระ (2543) ได้ศึกษาเปรียบเทียบอำนาจการตรวจสอบและจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบสมำเสมอและแบบไม่สมำเสมอ ระหว่างวิธี Mantel – Haenszel และวิธีถดถอยโลจิสติก โดยใช้ความยาวแบบทดสอบและขนาดกลุ่มตัวอย่างต่างกัน พบว่า

1. อำนาจการตรวจสอบและการจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันทั้งแบบสมำเสมอและแบบไม่สมำเสมอ ระหว่างวิธี Mantel - Haenszel และวิธีถดถอยโลจิสติก ภายใต้ความยาวแบบทดสอบและขนาดกลุ่มตัวอย่างที่ศึกษา เกือบทุกเงื่อนไขไม่แตกต่างกัน

2. ความยาวของแบบทดสอบไม่มีผลต่ออำนาจการตรวจสอบและการจำแนกผิดพลาดในการตรวจสอบด้วยวิธี Mantel – Haenszel และวิธีถดถอยโลจิสติก

3. ขนาดกลุ่มตัวอย่างมีผลต่ออำนาจการตรวจสอบด้วย วิธี Mantel – Haenszel และวิธีถดถอยโลจิสติก เกือบทุกเงื่อนไขของการศึกษา กล่าวคือ เมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้น อำนาจการตรวจสอบจะมีค่าเพิ่มขึ้น แต่พบว่าขนาดกลุ่มตัวอย่างจะไม่มีผลต่อการจำแนกผิดพลาดในเกือบทุกเงื่อนไขของการศึกษา

7.2 งานวิจัยต่างประเทศ

งานวิจัยต่างประเทศที่ศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบ มีทั้งใช้ ข้อมูลจำลอง (Simulated Data) และข้อมูลจริง (Real Data) ซึ่งจะนำเสนอดังต่อไปนี้

7.2.1 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้ข้อมูลจริง

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยข้อมูลจริง มีผู้ที่ทำการศึกษาดังต่อไปนี้

ผจงจิต อินทสุวรรณ (Intasuwarn. 1979 : 92) ได้ศึกษาวิธีวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ 3 วิธี คือ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ วิธีไคสแควร์ และวิธีราสช์ ของแบบทดสอบ IEA (The International Association for The Evaluation of Education Achivement) กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักเรียนอายุ 14 ปี ชาวอังกฤษ อเมริกัน และนิวซีแลนด์ จำนวน 8,601 คน ผลการวิจัยพบว่า ค่าสหสัมพันธ์ระหว่างวิธีทั้ง 3 มีค่าระหว่าง .51 - .98 และวิธีไคสแควร์กับวิธีราสช์มีค่าสหสัมพันธ์สูงถึง .98

สงบ ลักษณะ (Laksana. 1979) ได้วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบต่อเพศ และสีผิว ในการสอบแบบทดสอบ ITBS (IOWA Test of Basic Skills) ฉบับภาษาด้านคำศัพท์ และฉบับคณิตศาสตร์ ด้วยวิธีการวิเคราะห์ 2 วิธี คือ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และวิธีวิเคราะห์ความแปรปรวน ผลการวิจัยพบว่า วิธีวิเคราะห์ความแปรปรวนเป็นวิธีที่ใช้ได้ดีในทางปฏิบัติ แต่เป็นวิธีที่ไม่ครอบคลุม ส่วนวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ เป็นวิธีที่ครอบคลุมกว่า แต่ต้องใช้กลุ่มตัวอย่างขนาดใหญ่ และคำนวณค่อนข้างยุ่งยากซับซ้อน

สับโคเวียค และคณะ (Subkoviak. 1984 : 49 - 59) ได้เปรียบเทียบวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 3 วิธี คือ (1) วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ (2) วิธีแปลงค่าความยาก (TID) และ (3) วิธีไคสแควร์ซึ่งมี 2 วิธี ได้แก่วิธีที่เสนอโดย Scheuneman (1979) พิจารณาเฉพาะตัวเลือกถูก (CHIS) และที่เสนอโดย Camilli (1979) พิจารณาทั้งตัวเลือกถูกและผิด (CHIC) ของแบบทดสอบวัด

ความรู้เกี่ยวกับคำศัพท์ 50 ข้อ เป็นศัพท์ภาษาอังกฤษมาตรฐานจำนวน 40 ข้อ ศัพท์แสลงชาวผิวดำ จำนวน 10 ข้อ กลุ่มตัวอย่างที่ใช้ศึกษาเป็นนักเรียนผิวขาว 1,021 คน นักเรียนผิวดำจำนวน 1,008 คน ผลจากการวิจัยพบว่า วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์เป็นวิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือวิธีไคสแควร์ ส่วนวิธีแปลงค่าความยากเป็นวิธีที่นำไปใช้ในทางปฏิบัติได้ในกรณีไม่มีเครื่องคอมพิวเตอร์

ดูลิตเติล และ เคลียร์ (Doolittle and Cleary. 1987 : 157 – 166) ได้วิเคราะห์การทำหน้าที่ต่างกันของข้อสอบต่อเพศในแบบทดสอบ ACT (Assessment Mathematic Test) ฉบับการคำนวณทางคณิตศาสตร์ เรขาคณิต และคณิตศาสตร์การใช้เหตุผล จำนวน 8 ฉบับ กลุ่มตัวอย่างที่ใช้เป็นเพศชายและหญิง จำนวน 1,300 – 1,400 คน โดยเป็นเพศหญิงประมาณร้อยละ 55 วิเคราะห์ข้อมูลด้วยดัชนีของลินน์และฮาร์นิสซ์ (Linn and Harnisch. 1981) ซึ่งเป็นโมเดลโลจิสติก 3 พารามิเตอร์ ถ้าค่า Z ที่ได้เป็นบวกแสดงว่าง่ายสำหรับกลุ่มเปรียบเทียบ และถ้าเป็นลบแสดงว่ายากสำหรับกลุ่มเปรียบเทียบ ผลการวิจัยพบว่า ค่าดัชนี Z มีค่าเป็นลบสำหรับข้อสอบที่วัดด้านเรขาคณิต การใช้เหตุผลเชิงพีชคณิต และเลขคณิตในทุกฉบับ แสดงว่าง่ายสำหรับเพศชาย

เพิร์ลแมน และคณะ (Perlman, et al. 1988 : 5 - 9) ได้ศึกษาความคงที่ของวิธีประมาณค่าการทำหน้าที่ต่างกันของข้อสอบ 4 วิธี ได้แก่ วิธีแปลงค่าความยาก วิธีแปลงค่าความยากที่ดัดแปลงโดย Shepard วิธีวิเคราะห์ค่าที่เหลือแบบ 1 พารามิเตอร์ของราสส์ และวิธี Mantel - Haenszel 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 และโรเจอร์ (Hambleton and Rogers. 1989 : 313 - 334) ได้ศึกษาเปรียบเทียบวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี คือ วิธี Mantel - Haenszel และวิธี IRT โดยใช้ข้อมูลการตอบแบบสอบถาม New Mexico High School Proficiency Exam (NMHSPE) จำนวน 150 ข้อ ของนักเรียนชาวอเมริกันผิวขาว 8,000 คน และชาวอเมริกันพื้นเมือง 2,600 คน จากผู้สอบทั้งหมด 23,000 คน

NMHSPE เป็นแบบทดสอบเลือกตอบชนิด 4 ตัวเลือก วัดทักษะการดำเนินชีวิต 5 ด้าน คือ (1) ความรู้เกี่ยวกับแหล่งทรัพยากรของชุมชน (2) ความรู้ด้านผู้บริโภค (3) ความรู้เกี่ยวกับรัฐบาลและกฎหมาย (4) ความรู้ทางด้านสุขภาพร่างกายและจิตใจ และ (5) ความรู้ด้านอาชีพ

กลุ่มตัวอย่างที่นำมาศึกษาได้แก่ ชาวอเมริกันผิวขาวและอเมริกันพื้นเมือง โดยสุ่มมาศึกษา กลุ่มละ 2,000 คน และแบ่งเป็น 2 กลุ่ม ๆ ละ 1,000 คน ตามลำดับคู่-คี่ การศึกษาการทำหน้าที่ต่างกันของข้อสอบ 2 คู่เช่นนี้เพื่อพิจารณาถึงความคงที่ในการระบุข้อสอบที่ทำหน้าที่ต่างกัน ในการศึกษาการทำหน้าที่ต่างกันของข้อสอบใช้ข้อสอบ 75 ข้อ โดยตัดข้อที่ง่ายมาก ($p > .70$) และค่าอำนาจจำแนกต่ำ ($r < .10$) วิธี Mantel - Haenszel วิเคราะห์โดยใช้ 76 ระดับคะแนน ส่วนวิธี IRT ใช้การวัดพื้นที่ระหว่างโค้งลักษณะ ข้อสอบ พิสัยของความสามารถที่นำมาใช้ในการคำนวณพื้นที่ที่กำหนดใช้ 2 ค่าเบี่ยงเบนมาตรฐานสูงกว่าความสามารถเฉลี่ยของชาวอเมริกันผิวขาว และ 2 ค่าเบี่ยงเบนมาตรฐานต่ำกว่าความสามารถเฉลี่ยของชาวอเมริกันพื้นเมือง ความคงที่ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ ร้อยละของข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกัน หรือไม่ทำหน้าที่ต่างกันซ้ำกันในการวิเคราะห์ 2 ครั้ง

เกณฑ์การตัดสินการทำหน้าที่ต่างกันของข้อสอบ สำหรับวิธี IRT- 3 พารามิเตอร์ ได้จากการวิเคราะห์ดัชนี SA ระหว่างผู้สอบที่เป็นอเมริกันพื้นเมือง 2 กลุ่ม และใช้ค่าสูงสุดที่ได้เป็นเกณฑ์ซึ่งมีค่าเท่ากับ .468 และด้วยวิธีเดียวกันเกณฑ์ตัดสินความลำเอียง สำหรับวิธี Mantel - Haenszel คือ $MH - \chi^2$ เท่ากับ 6.64

ผลการวิเคราะห์พบว่า วิธี Mantel - Haenszel มีความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบร้อยละ 80 ส่วน ICC ร้อยละ 73 และพบว่าร้อยละ 47 ของข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันในการวิเคราะห์ครั้งที่ 1 เป็นข้อสอบที่ทำหน้าที่ต่างกันในการวิเคราะห์ครั้งที่ 2 ด้วยสำหรับวิธี Mantel - Haenszel และร้อยละ 61 สำหรับวิธี ICC ในทางกลับกันพบว่า ร้อยละ 64 ของข้อสอบที่ทำหน้าที่ต่างกันในการวิเคราะห์ครั้งที่ 2 เป็นข้อสอบที่ทำหน้าที่ต่างกันในการวิเคราะห์ครั้งที่ 1 สำหรับวิธี Mantel - Haenszel และร้อยละ 56 สำหรับวิธี ICC ผลดังกล่าวแสดงให้เห็นว่าความคงที่ของการวิเคราะห์ทั้งวิธี Mantel - Haenszel และวิธี ICC อยู่ในระดับปานกลาง ซึ่งอาจเป็นเพราะจำนวนตัวอย่างค่อนข้างน้อย (วิธี Mantel - Haenszel มีข้อสอบที่ระบุว่าทำหน้าที่ต่างกันในการวิเคราะห์ทั้งสองครั้ง 7 ข้อ ICC 14 ข้อ และพื้นที่แสดงความแตกต่างของ ICC มีค่าระหว่าง -2.7 และ 1.5)

ซูดวิก และ ทอลแมน (Sudweeks and Tolman, 1990 : 8-11) ได้เปรียบเทียบผลจากการใช้วิธี Mantel - Haenszel และการตัดสินด้วยผู้เชี่ยวชาญในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ วิชาวิทยาศาสตร์ ระหว่างนักเรียนชายและนักเรียนหญิงระดับ 5 จำนวน 926 คน เครื่องมือที่ใช้เป็นแบบทดสอบอิงเกณฑ์ชนิดเลือกตอบที่สร้างขึ้นตามมาตรฐาน และวัตถุประสงค์ของหลักสูตรวิทยาศาสตร์ร่วมในรัฐยูทาห์ ผลจากการวิเคราะห์พบว่า วิธีการทั้งสองให้ผลต่างกัน และข้อสอบหลายข้อที่ถูกระบุว่าทำหน้าที่ต่างกันทั้ง 2 วิธี จะเป็นข้อสอบที่ยาก

คิม และโคเฮน (Kim and Cohen, 1991 : 269 - 278) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบที่สร้างขึ้นอย่างจงใจให้ทำหน้าที่ต่างกัน โดยการเปรียบเทียบผลการคำนวณพื้นที่ความแตกต่างระหว่างโค้งลักษณะข้อสอบของกลุ่มตัวอย่างคนผิวขาว และผิวดำ 2 วิธี ประกอบด้วย การคำนวณพื้นที่ภายใต้ความสามารถที่มีขอบเขตจำกัด (ปิด) และไม่มีขอบเขตจำกัด (เปิด) ขนาดกลุ่มตัวอย่างที่ใช้ในการศึกษากลุ่มละ 1,000 คน สุ่มมาจากคนผิวขาว 1,021 คน และผิวดำ 1,008 คน เครื่องมือที่ใช้ในการศึกษาเป็นแบบทดสอบคำศัพท์แบบเลือกตอบชนิด 4 ตัวเลือก จำนวน 50 ข้อ เป็นคำศัพท์ภาษาอังกฤษมาตรฐาน 40 ข้อ อีก 10 ข้อ เป็นคำศัพท์ที่มีความง่ายสำหรับกลุ่มผิวดำมากกว่ากลุ่มผิวขาว

การคำนวณพื้นที่ระหว่างฟังก์ชันการตอบ ทั้งภายใต้ช่วงความสามารถแบบเปิด และแบบปิด จะพิจารณา 2 ลักษณะ คือ พื้นที่ที่มีเครื่องหมาย (Signed Area) และพื้นที่ชนิดไม่มีเครื่องหมาย (Unsigned Area) แบบจำลองที่ใช้ในการศึกษาเป็นแบบจำลอง 3 พารามิเตอร์ และ 3 พารามิเตอร์ ที่กำหนดให้ c คงที่ = .23

ผลการศึกษา พบว่า การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างการคำนวณพื้นที่แบบช่วงเปิด และแบบช่วงปิด แทบไม่มีความแตกต่างกัน และพื้นที่แบบไม่มีเครื่องหมายไม่ว่าจะเป็นแบบช่วงเปิดหรือปิด มีความไวต่อการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมากกว่าแบบมีเครื่องหมาย นอกจากนั้นยังพบอีกว่าพื้นที่แบบมีเครื่องหมาย และไม่มีเครื่องหมายของการคำนวณแบบช่วงปิด

สำหรับความคลาดเคลื่อนในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน พบว่าเมื่อใช้แบบจำลอง 3 พารามิเตอร์ พื้นที่แบบมีเครื่องหมาย มีอัตราความคลาดเคลื่อนต่ำกว่าวิธีอื่น ๆ และมีความคลาดเคลื่อน

ในการยอมรับผิดพลาด จำนวน 2 ข้อ ที่ .05 ส่วนพื้นที่แบบไม่มีเครื่องหมายมีอัตราความคลาดเคลื่อนสูงกว่าวิธีอื่น ๆ เล็กน้อย

โคเฮน คิม และสับโคเวียค (Cohen, Kim and Subkoviak. 1991 : 49 - 59) ได้ศึกษาอิทธิพลของการทราบการกระจายในการตรวจค้นการทำหน้าที่เบี่ยงเบนของข้อสอบ สำหรับกลุ่มผู้สอบผิวดำและผิวขาว ที่สร้างขึ้นอย่างมีเจตนาให้ข้อสอบทำหน้าที่ต่างกัน โดยใช้โปรแกรม LOGIST ซึ่งใช้วิธี Joint Maximum Likelihood Estimation (JMLE) และโปรแกรม BILOG โดยที่โปรแกรม BILOG ใช้การประมาณค่าพารามิเตอร์ 3 สถานการณ์ คือ

1. MMLE (Marginal Maximum Likelihood Estimation) เป็น BILOG ที่ไม่มีการกำหนดค่าพารามิเตอร์ ค่าอำนาจจำแนก (a) จะอยู่ในช่วง $0 - M = 1.13$ a = .604 เป็น BILOG A
2. MMAPE (Marginal Maximum a Posteriori Estimation) กำหนดค่า a ในรูป lognormal ค่าเฉลี่ย = 0 (ไม่มี FLOAT option) เป็น BILOG B
3. MMAPE กำหนดค่า a ในรูป lognormal เช่นกัน แต่มีข้อตกลงว่าการกระจายของค่า (ค่าแปลงของ a) มีการกระจายเป็นปกติ มีค่าความเบี่ยงเบนมาตรฐาน = 0.5 (มี FLOAT option) เป็น BILOG C

ส่วนการวัดความแตกต่างของพื้นที่ระหว่างโค้งการตอบข้อสอบนั้น ทำ 2 ลักษณะ คือ อย่างคิดเครื่องหมาย (SA) และไม่คิดเครื่องหมาย (UA) การประมาณค่าพารามิเตอร์ใช้แบบ 2 พารามิเตอร์ โดยการกำหนดค่า C ให้เป็นค่าคงที่

พบจากค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันว่า การวัดพื้นที่ โดยใช้ BILOG B และ BILOG C มีความสัมพันธ์กันสูงมาก ทั้งในดัชนีแบบคิดเครื่องหมาย และไม่มีเครื่องหมาย ส่วนระหว่าง LOGIST และ BILOG ก็สูงเช่นกัน แต่ความคงที่ในรูปแบบของความคล้ายคลึงจะน้อยกว่า ส่วนความคล้ายคลึงกันระหว่าง LOGIST กับ BILOG B และ BILOG C ลดลงกว่าเล็กน้อย เป็นผลที่สอดคล้องกับการศึกษาที่คนอื่น ๆ เคยพบ ในการวัดพื้นที่ของความเบี่ยงเบนของข้อสอบ

BILOG B และ C จะดีกว่าวิธีอื่น ๆ และมีความไวในการตรวจค้นมากเมื่อค่า b อยู่ในช่วง +5 มากกว่าในช่วง +3 ยกเว้น BILOG B ที่เท่า ๆ กันทั้ง 2 ช่วงส่วน MMLE จะดีกว่า JMLE มาก ในการตรวจค้นความเบี่ยงเบน ในกรณีที่ค่าสัมบูรณ์ของ b อยู่ในช่วง +3 และ ดีกว่าเล็กน้อยในกรณีที่ค่าสัมบูรณ์ของ b น้อยกว่า หรือเท่ากับ 5 ในการวัดแบบมีเครื่องหมาย จำนวนตัวอย่างที่ใช้ก็ไม่มีอิทธิพลมากนักต่อการตรวจค้นความเบี่ยงเบนของข้อสอบสรุปแล้ว การใช้ BILOG แบบ B และ C ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่สร้างขึ้นมาโดยทราบค่า หรือ กำหนดค่าการทำหน้าที่ต่างกันไว้ก่อนแล้วดีกว่าวิธีอื่น

ไรอัน (Ryan. 1991 : 325 - 337) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการที่เสนอโดยแมนเทลล์ และแฮนส์เซล (Mantel and Haenszel. 1959) โดยใช้ตัวอย่างกลุ่มอ้างอิง 500 คน และกลุ่มเปรียบเทียบ 100 คน เพื่อตรวจสอบความคงที่ของค่าประมาณการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Mantel - Haenszel ข้ามกลุ่มตัวอย่างและขนาดตัวอย่างต่าง ๆ กัน และเพื่อศึกษาว่าวิธี Mantel - Haenszel มีความแข็งแกร่งต่อรบกวนของข้อสอบเพียงไร โดยมีข้อสอบรวมจำนวน 40 ข้อ และข้อสอบที่ใช้สลับกันอีกจำนวน 35 ข้อ กลุ่มตัวอย่างเป็นชาวผิวขาวและผิวดำจำนวน 5,015 คน และ 670 คน ตามลำดับโดยมีเกณฑ์การแยกประเภทข้อสอบจากผลการวิเคราะห์เป็น 3 ประเภท

ประเภท A ข้อสอบที่ทำหน้าที่ได้ตรง มีค่า $\alpha_{MH} < 1.00$ และมีค่าไม่ต่างจาก 0 อย่างมีนัยสำคัญ

ประเภท B ข้อสอบที่พอจะนำมาใช้ได้ แต่ควรเลือกข้อที่มีค่า α_{MH} ต่ำ คือ $1 = \alpha_{MH} < 1.5$ และไม่ต่างจาก 1 อย่างมีนัยสำคัญ

ประเภท C เลือกมาใช้ถ้าตรงตามวัตถุประสงค์เฉพาะที่ระบุไว้อย่างมีนัยสำคัญว่าต้องการข้อสอบลักษณะนี้ ได้แก่ ข้อที่มีค่า $\alpha_{MH} > 1.5$ และมากกว่า 1 อย่างมีนัยสำคัญ

ขั้นตอนในการวิเคราะห์ขั้นแรกมีการวิเคราะห์ 5 แบบ ในการสร้างฐานเรียกว่า ชุด 1 ทำโดยการสุ่มกลุ่มผิวขาวมาอีก 4 กลุ่ม เป็นกลุ่มอ้างอิง ได้แก่ W_1, W_2, W_3, W_4 และวิเคราะห์ข้อสอบรวม 40 ข้อด้วยวิธี Mantel - Haenszel จับคู่การวิเคราะห์เป็น $W/W, W/W_1, W/W_2, W/W_3, W/W_4$ คะแนนที่ใช้แบ่งระดับความสามารถ คือ คะแนนรวมจากข้อสอบ 40 ข้อ

การวิเคราะห์ชุด 2 ทำในลักษณะเดียวกับ ชุด 1 แต่ใช้กลุ่มเป้าหมายเป็นผิวดำวิเคราะห์ด้วยวิธี MH 5 คู่ ได้แก่ $BW, BW_1, BW_2, BW_3, BW_4$ เป็นการวิเคราะห์เพื่อดูความคงที่ของค่า α_{MH} ข้ามกลุ่มและขนาดตัวอย่าง

ชุด 3 เป็นการวิเคราะห์ผลของบริบทของข้อสอบ BW_1, BW_2, BW_3 และ BW_4 โดยใช้ข้อสอบ 75 ข้อ จากฟอร์ม 1 ฟอร์ม 2 ฟอร์ม 3 ฟอร์ม 4 ใช้คะแนนรวมเป็นเกณฑ์แบ่งช่วงความสามารถ (แต่ละฟอร์มมีข้อสอบรวม 40 ข้อ)

ชุด 4 เป็นการวิเคราะห์เช่นเดียวกับชุดที่ 3 แต่ใช้ข้อสอบ 4 ฟอร์มที่ไม่มีข้อสอบรวม (มีข้อสอบ 35 ข้อ) เป็นเกณฑ์แบ่งระดับความสามารถ

ผลการวิเคราะห์พบว่า การทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มผิวขาวด้วยกัน 5 กลุ่มนี้มีความสัมพันธ์กันต่ำมาก ไม่มีความแตกต่างอันเนื่องมาจากการทำหน้าที่ต่างกันของข้อสอบ แต่มีความแตกต่างจากความคลาดเคลื่อนในการสุ่มกลุ่มตัวอย่างที่มีความซับซ้อนมากกว่า จึงอาจเป็นข้อสอบที่ทำหน้าที่ต่างกันที่ไม่แท้ก็ได้ เกณฑ์การแบ่งช่วงระดับความสามารถ ซึ่งเป็นบริบทที่ศึกษาไม่มีผลต่อค่า α_{MH} ไม่ว่าจะใช้เกณฑ์คะแนนรวม 35 ข้อหรือ 75 ข้อ ก็ได้ค่าสหสัมพันธ์ของการเปรียบเทียบแต่ละคู่เท่า ๆ กัน ผลจากการจำแนกประเภทข้อสอบมีเพียง 1 ข้อ ที่มีความสอดคล้องกันทุกวิธีวิเคราะห์ คือเป็นประเภท B ทุกครั้ง นอกนั้นมีความแปรปรวนไปตามการวิเคราะห์

เมเซอร์และคณะ (Mazor, et. al. 1991 : 443 – 451) ศึกษาผลกระทบของขนาดกลุ่มตัวอย่างที่มีต่อการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Mantel - Haenszel โดยใช้ข้อมูลที่จำลองโดยโปรแกรม DATAGEN แบบ 3 พารามิเตอร์ ขนาดกลุ่มตัวอย่างที่ศึกษามี 5 ขนาดคือ 2,000 คน 1,000 คน 500 คน 200 คน และ 100 คน ความยาวของแบบทดสอบชุดละ 75 ข้อ ผลการวิจัยพบว่า เมื่อกลุ่มตัวอย่างขนาด 2,000 คน วิธี Mantel - Haenszel จะตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ผิดพลาดร้อยละ 30 และเมื่อใช้จำนวนผู้สอบน้อยกว่า 500 คน พบว่าจะตรวจสอบการทำหน้าที่ต่างกันของข้อสอบผิดพลาดร้อยละ 50 และข้อสอบที่ไม่สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้เป็นข้อสอบที่ยากมาก ข้อสอบที่มีค่าอำนาจจำแนกต่ำและข้อสอบที่มีค่าความยากต่างกันเล็กน้อยสำหรับ 2 กลุ่ม

มิลเลอร์ และโอชิม่า (Miller and Oshima. 1992 : 381 – 388) ได้ศึกษาผลของขนาดกลุ่มตัวอย่าง จำนวนข้อสอบที่ทำหน้าที่ต่างกัน และขนาดของการทำหน้าที่ต่างกันที่ประมาณค่าด้วยวิธีสองขั้นตอน ข้อมูลที่ใช้เป็นข้อมูลที่จำลองขึ้นแบบ 2 พารามิเตอร์ จำนวน 40 ข้อ มีค่าอำนาจจำแนกเฉลี่ย

เป็น 1.13 ซึ่งอยู่ระหว่าง .41 ถึง 3.82 ส่วนเบี่ยงเบนมาตรฐานเป็น .63 ค่าความยากเฉลี่ยเป็น .25 มีค่าระหว่าง -1.99 ถึง 2.10 ส่วนเบี่ยงเบนมาตรฐานเป็น 1.05 เงื่อนไขการทำหน้าที่ต่างกันของข้อสอบที่ทำการศึกษามี 24 เงื่อนไข ดังนี้ ใช้พารามิเตอร์ของข้อสอบแบบ 2 พารามิเตอร์ ขนาดการทำหน้าที่ต่างกัน 3 ขนาดคือ น้อย ปานกลาง และแบบผสม โดยที่เงื่อนไขขนาดการทำหน้าที่ต่างกันแบบน้อย จะมีค่าความยากเพิ่มขึ้น .20 ในขนาดการทำหน้าที่ต่างกันแบบปานกลางค่าความยากเพิ่มขึ้น .35 และขนาดการทำหน้าที่ต่างกันแบบผสมเป็นแบบผสมระหว่าง 2 แบบ และเงื่อนไขจำนวนข้อที่ทำหน้าที่ต่างกัน 4 แบบ คือ ร้อยละ 5 ร้อยละ 10 ร้อยละ 20 และร้อยละ 40 ของแบบทดสอบ ($2 \times 3 \times 4$) ดัชนีการทำหน้าที่ต่างกันของข้อสอบคำนวณจากวิธีที่ใช้ทฤษฎีการตอบข้อสอบ 5 วิธี คือ วิธี SA, UA, SSOS, USOS, WSSOS และวิธี Mantel - Haenszel ขั้นที่ 2 ของวิธีการเป็นการหาจำนวนข้อสอบที่ผิดพลาดทางบวก (วิเคราะห์ข้อสอบที่ไม่ทำหน้าที่ต่างกันว่าเป็นข้อสอบที่ทำหน้าที่ต่างกัน) ในขั้นตอนนี้จะมีประโยชน์เมื่อขนาดการทำหน้าที่ต่างกันมีมาก (ร้อยละ 20 ถึงร้อยละ 40 ของแบบทดสอบ) วิธี Mantel - Haenszel จะตรวจสอบความผิดพลาดทางบวกได้น้อย และวิธี Mantel - Haenszel จะวิเคราะห์การทำหน้าที่ต่างกันได้ดีพอ ๆ กับวิธีทฤษฎีการตอบข้อสอบ

โรเจอร์ และ สวามินาธาน (Roger and Swaminathan. 1993 : 105 - 116) ได้ศึกษาเปรียบเทียบผลการวิเคราะห์ข้อสอบที่ทำหน้าที่ต่างกัน ด้วยวิธีถดถอยโลจิสติก (วิธี LR) กับวิธี Mantel - Haenszel โดยใช้ข้อมูลที่จำลองขึ้น เงื่อนไขการศึกษาเป็น 4 แบบ คือ ขนาดกลุ่มตัวอย่าง 2 กลุ่ม ขนาด 250 คน และ 500 คนต่อกลุ่ม ระดับความเหมาะสมของโมเดลข้อมูลเป็น 2 ระดับ คือ ระดับที่เหมาะสมใช้รูปแบบ 2 พารามิเตอร์ ระดับที่ไม่เหมาะสมใช้รูปแบบ 3 พารามิเตอร์ โดยค่าการเดา (c) จะถูกกำหนดที่ .2 เครื่องมือที่ใช้เป็นข้อสอบจำนวน 40 ข้อที่จำลองขึ้นมา 7 แบบ ดังนี้ แบบที่ 1 ค่าความยากและค่าอำนาจจำแนกต่ำ แบบที่ 2 ค่าความยากสูงและค่าอำนาจจำแนกต่ำ แบบที่ 3 ค่าความยากปานกลางค่าอำนาจจำแนกต่ำ แบบที่ 4 ค่าความยากปานกลางค่าอำนาจจำแนกสูง แบบที่ 5 ค่าความยากสูงค่าอำนาจจำแนกต่ำ แบบที่ 6 ค่าความยากและค่าอำนาจจำแนกสูง แบบที่ 7 เป็นแบบผสม คือค่าอำนาจจำแนกและค่าความยากต่างกันสำหรับ 2 กลุ่ม โดยวิเคราะห์ร้อยละของข้อสอบที่ทำหน้าที่ต่างกันจากการตรวจค้นแบบยูนิฟอร์มและนอนยูนิฟอร์ม ผลการวิจัยพบว่า ร้อยละของข้อสอบที่ทำหน้าที่ต่างกัน จากการตรวจค้นแบบยูนิฟอร์มสูงกว่าแบบนอนยูนิฟอร์มทั้ง 2 วิธีและในทุกเงื่อนไข โดยวิธี LR จะตรวจพบร้อยละของข้อสอบที่ทำหน้าที่ต่างกันสูง ที่ระดับความยากปานกลางค่าอำนาจจำแนกสูง สำหรับข้อสอบที่ทำหน้าที่ต่างกันแบบ นอน-ยูนิฟอร์ม พบว่า วิธี LR จะตรวจค้นข้อสอบที่ทำหน้าที่ต่างกันลดลงร้อยละ 14 ในโมเดล ข้อมูลที่ไม่เหมาะสม ส่วนวิธี Mantel - Haenszel จะตรวจค้นพบลดลงร้อยละ 7 และในขนาดกลุ่มตัวอย่างที่เพิ่มขึ้น วิธี LR ตรวจค้นข้อสอบที่ทำหน้าที่ต่างกันได้เพิ่มขึ้นร้อยละ 19 ส่วนวิธี Mantel - Haenszel จะตรวจพบเพิ่มขึ้น ร้อยละ 11 และที่ความยากต่ำวิธี Mantel - Haenszel ตรวจค้นพบได้น้อยกว่าวิธี LR แต่ที่ความยากสูงทั้ง 2 วิธีให้ผลใกล้เคียงกันแต่วิธี Mantel - Haenszel ไม่สามารถตรวจสอบได้ในระดับความยากปานกลาง และวิธี LR จะตรวจค้นข้อสอบที่ทำหน้าที่ต่างกันได้ดีกว่าวิธี Mantel - Haenszel ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 แบบ แต่ค่าใช้จ่ายที่ใช้ในการวิเคราะห์จะสูงกว่าวิธี Mantel - Haenszel

ราชู และเจฟฟรี (Raju and Jeffrey. 1993 : 301 - 314) ได้เปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของ ข้อสอบระหว่าง วิธีลอร์ด (χ^2) วิธีราชู (Z) และวิธี Mantel - Haenszel เครื่องมือที่ใช้ในการศึกษาเป็น แบบทดสอบวัดคำศัพท์ จำนวน 45 ข้อ ชนิด 5 ตัวเลือก กลุ่มตัวอย่างเป็นนักเรียนระดับ 10

และ 12 จำนวน 839 คน แยกเป็นชาย 399 คน หญิง 440 คน ผิวขาว 426 คน และผิวดำ 245 คน (อีก 158 คน เป็นเชื้อชาติอื่น)

ผลการศึกษาพบว่า การตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างนักเรียนผิวดำและผิวขาว มีข้อสอบที่ทำหน้าที่ต่างกัน 1 ข้อ สำหรับการทดสอบด้วยวิธีลอร์ด และวิธีราชู ทั้งพื้นที่ชนิดที่มีเครื่องหมาย และไม่มีเครื่องหมาย ได้แก่ ข้อ 41 ทั้ง 3 วิธี ส่วนวิธี Mantel - Haenszel พบข้อสอบที่ทำหน้าที่ต่างกัน 2 ข้อ ได้แก่ ข้อ 14 และ 27 ทั้ง 3 ข้อ เป็นข้อที่มีความยากต่างกันมากสำหรับนักเรียนผิวดำและผิวขาว

ส่วนผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างชายหญิง ด้วยวิธีลอร์ด และวิธีราชูพบ ข้อสอบที่ทำหน้าที่ต่างกันซ้ำกัน 4 ข้อ และวิธี 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 ซึ่งแย้งกับการศึกษาของ สวามินาธาน และโรเจอร์ (1990) ที่พบว่า วิธีทฤษฎีการ ตอบข้อสอบตรวจพบข้อสอบที่ทำหน้าที่ต่างกันได้มากกว่าวิธี Mantel - Haenszel

นารายานาน และสวามินาธาน (Narayanan and Swaminathan . 1994 : 315-328) ได้ ศึกษาเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบแบบยูนิฟอร์ม ด้วยวิธี Mantel - Haenszel และวิธี SIBTEST โดยใช้ข้อมูลที่จำลองภายใต้เงื่อนไข 1,296 เงื่อนไข (9X3X2X4X6) แล้วทำการ คำนวณซ้ำ 100 ครั้งแต่ละเงื่อนไขที่ทำการศึกษาประกอบด้วย

1. ขนาดกลุ่มตัวอย่างที่ต่างกัน 3 ขนาด คือกลุ่มเปรียบเทียบใช้ขนาด 100,200 และ 300 คน กลุ่มอ้างอิงใช้ขนาด 300,500 และ 1,000 คน วางไขว้กันได้กลุ่มตัวอย่างทั้งหมด 9 กลุ่ม
2. ความแตกต่างในการแจกแจงความสามารถ 3 ระดับ ระดับที่ 1 ให้ทั้ง 2 กลุ่มมีค่าเฉลี่ย การแจกแจงความสามารถอยู่ที่ 0.0 ระดับที่ 2 ให้ค่าเฉลี่ยของความสามารถในกลุ่ม อ้างอิงและกลุ่ม เปรียบเทียบเป็น 0.0 และ 0.5 ตามลำดับ และระดับที่ 3 ค่าเฉลี่ยความสามารถในกลุ่มอ้างอิงและกลุ่ม เปรียบเทียบเป็น 0.0 และ -1.0 ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานของทั้ง 3 ระดับเป็น 1.0
3. ร้อยละของข้อสอบที่ทำหน้าที่ต่างกันมี 2 ขนาด คือ ร้อยละ 10 และร้อยละ 20 จาก แบบทดสอบ จำนวน 40 ข้อ
4. ขนาดของการทำหน้าที่ต่างกันของข้อสอบ 4 ระดับ คือ 0.4, 0.6, 0.8, และ 1.0
5. ชนิดของแบบทดสอบ 6 ชนิด คือ ชนิดของแบบทดสอบที่ประกอบด้วยค่าความยาก และ ค่าอำนาจจำแนก 3 ระดับคือ สูง ปานกลาง ต่ำ วางไขว้กัน แต่ตัดแบบทดสอบที่ประกอบด้วยค่าความยากและ ค่าอำนาจจำแนกสูง ปานกลาง ต่ำ ทั้ง 2 ค่าออก จำนวน 3 ฉบับ

ผลการวิจัยพบว่า วิธี Mantel - Haenszel และวิธี SIBTEST สามารถตรวจค้นข้อสอบที่ทำ หน้าที่ต่างกันได้ใกล้เคียงกันโดยจะตรวจค้นได้ดีในขนาดกลุ่มตัวอย่าง 600 คน (กลุ่มอ้างอิง และกลุ่ม เปรียบเทียบกลุ่มละ 300 คน) ขนาดของความลำเอียงทั้ง 4 ขนาดมีผลต่อการตรวจค้นการทำหน้าที่ต่างกัน ของข้อสอบทั้ง 2 วิธีอย่างมี นัยสำคัญทางสถิติ ความแตกต่างในการแจกแจงความสามารถไม่มีผลต่อวิธี การวิเคราะห์ความลำเอียงแบบ SIBTEST แต่มีผลต่อวิธี Mantel - Haenszel ที่เป็นเช่นนี้อาจเป็นเพราะวิธี SIBTEST ได้มีการปรับแก้การถดถอย ขนาดของแบบทดสอบและร้อยละของข้อสอบที่ทำหน้าที่ต่างกัน ไม่มีผลต่อวิธีทั้งสอง สำหรับการเกิดความคลาดเคลื่อนแบบที่ 1 วิธี Mantel - Haenszel จะมีอัตราการเกิด อยู่ในระดับปกติแต่วิธี SIBTEST จะเพิ่มขึ้นเล็กน้อยในกลุ่มที่มีความสามารถเท่ากัน แต่ในกลุ่มที่มีความ

สามารถไม่เท่ากันการเกิดความคลาดเคลื่อนแบบที่ 1 จะมีอัตราการเกิดสูงขึ้นทั้ง 2 วิธี ซึ่งการเพิ่มความยาวของแบบทดสอบจะทำให้เกิดความเชื่อมั่นสูงขึ้นซึ่งมีผลให้ความคลาดเคลื่อนแบบที่ 1 น่าจะมีอัตราการเกิดลดลง

คิม และโคเฮน (Kim and Cohen. 1995 : 291 - 312) ได้ศึกษาเปรียบเทียบการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยทฤษฎีการตอบข้อสอบ ด้วยวิธีลอร์ด (χ^2) วิธีราชู (Raju's Area Measures) และวิธีทดสอบอัตราส่วนไลค์ลิฮูด (Likelihood Ratio Test) โดยใช้ข้อมูลจากแบบทดสอบจัดชั้นเรียนของนักศึกษาระดับมหาวิทยาลัย วิชาคณิตศาสตร์ 2 ฟอรัม ๆ ละ 28 ข้อ ฟอรัมแรกแบ่งเป็น 2 ส่วน สำหรับฟอรัมหลังเป็นข้อสอบชุดเดียวกัน แต่เรียงสลับกันระหว่างส่วนที่ 1 และส่วนที่ 2 ผลการศึกษาพบว่าทั้ง 3 วิธี ให้ผลการตรวจสอบสอดคล้องกันมาก

รูสโซ และสตาท์ (Roussos and Stout. 1996 : 215 - 230) ได้ศึกษาผลการใช้จำนวนกลุ่มตัวอย่างขนาดเล็ก ความแตกต่างกันของค่าพารามิเตอร์ข้อสอบรวมทั้งความแตกต่างในค่าเฉลี่ยของความสามารถของกลุ่มอ้างอิง และกลุ่มเปรียบเทียบ เพื่อศึกษาถึงผลที่มีต่อความคลาดเคลื่อนชนิดที่ 1 (Type I error) ซึ่งเกิดจากการปฏิเสธสมมุติฐานศูนย์ที่เป็นจริงหรือในกรณีนี้ คือ การยอมรับว่าข้อสอบทำหน้าที่ต่างกันทั้ง ๆ ที่ข้อสอบไม่ทำหน้าที่ต่างกัน

การศึกษาที่ 1 ใช้กลุ่มตัวอย่าง และกลุ่มเปรียบเทียบกลุ่มละ 100 200 500 และ 1000 คน เลือกข้อสอบ 1 ข้อ จากแบบสอบ ASVAB เลือกข้อที่มีค่า 0.03 และ 0.25 (ค่าเฉลี่ยคือ 1.22 0.09 และ 0.02) และเป็นข้อสอบที่ไม่ทำหน้าที่ต่างกัน ความสามารถของกลุ่มตัวอย่างทั้ง 2 กลุ่มถูกจำลองให้มีการกระจายแบบปกติ มีค่าความแปรปรวนเป็น 1.00 ความแตกต่างระหว่างค่าเฉลี่ยของการแจกแจงของ 2 กลุ่ม ถูกจำลองให้มีค่า 0.0 0.5 และ 1.0 แล้วคำนวณค่าความสามารถเฉลี่ยของแต่ละกลุ่มให้จุดกลางของค่าเฉลี่ยอยู่บนระดับความยากเฉลี่ยของแบบสอบ จำลองการตอบ 400 ครั้ง และวิเคราะห์ 400 ครั้ง ในแต่ละสถานการณ์

ผลการใช้วิธี SIBTEST และวิธี Mantel - Haenszel ตรวจสอบข้อสอบข้อสอบที่ทำหน้าที่ต่างกัน ที่เลือกขึ้นมาดังกล่าว พบว่า นอกจากกรณีกลุ่มตัวอย่างขนาด 100 คน และค่าความแตกต่างของความสามารถ (d_T) = 0.0 แล้ว อัตราการปฏิเสธสมมุติฐานศูนย์ หรืออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีทั้งสองแตกต่างกันอย่างไม่มีนัยสำคัญ ทั้งยังมีความแตกต่างที่พบในจำนวนตัวอย่าง 100 คน และ $d_T = 0.0$ นั้น อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสองวิธีต่ำกว่า 0.05 อย่างไรก็ตาม SIBTEST มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่า วิธี Mantel - Haenszel เล็กน้อย สำหรับการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยใช้กลุ่มตัวอย่าง ไม่เกิน 1,000 คน และมีแนวโน้มว่าอัตราความคลาดเคลื่อนจะสูงขึ้นเมื่อกลุ่มตัวอย่างมากขึ้นทั้ง 2 วิธี

การศึกษาที่ 2 ใช้แบบสอบเดิม แต่เลือกค่า a , b และ c สำหรับข้อสอบที่นำมาวิเคราะห์ให้แตกต่างกันโดย

$$a = 0.4 \quad 1.0 \quad \text{และ} \quad 2.5$$

$$b = -1.5 \quad -0.5 \quad 0.0 \quad 0.5 \quad \text{และ} \quad 1.5$$

$$c = 0.20$$

จำลองการกระจายของความสามารถเป้าหมายให้เป็นปกติ d_T มีค่า 0.0 และ 1.0 นอกจากนั้นการจำลองทุกอย่างเป็นเช่นเดียวกับการศึกษาที่ 1 แต่ใช้จำนวนผู้สอบขนาดใหญ่ คือ 500 1,000 และ 3,000 คน และแต่ละสถานการณ์จำลองการตอบ 100 ครั้ง และตรวจสอบโดยใช้วิธี SIBTEST และวิธี Mantel - Haenszel พบว่าเมื่อค่า a สูงขึ้นและ b ต่ำ อัตราการปฏิเสธสมมุติฐานจะสูงกว่าระดับที่กำหนดไว้คือ 0.05

ทั้ง 2 วิธี และวิธี Mantel - Haenszel จะมีอัตราการปฏิเสธสูงกว่าวิธี SIBTEST และยิ่งจำนวนกลุ่มตัวอย่างใหญ่ขึ้น อัตราการปฏิเสธจะยิ่งสูงขึ้น ในขนาดขนาดกลุ่มตัวอย่าง 3,000 คน ที่ค่า $a = 2.5$ และ $b = -1.5$ อัตราการปฏิเสธของวิธี Mantel - Haenszel สูงถึง 1.00 และวิธี SIBTEST = 0.91 ในกรณีที่ $d_T = 1.0$ และ $c = 0.20$ ในภาพรวมแล้ววิธี SIBTEST มีอัตราการปฏิเสธต่ำกว่าวิธี Mantel - Haenszel แต่เมื่อใช้ค่า $d_T = 0.0$ ไม่มีค่า c ผลปรากฏว่า อัตราการปฏิเสธของทั้ง 2 วิธีอยู่ในระดับใกล้เคียงกับ 0.05 แม้ว่า SIBTEST จะสูงกว่าวิธี Mantel - Haenszel แต่ไม่มีนัยสำคัญ

สรุปว่าเมื่อใช้จำนวนตัวอย่างน้อย ไม่เกิน 1,000 คน วิธี Mantel - Haenszel และ วิธี SIBTEST ให้ผลในด้านความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกัน และอยู่ภายใต้ระดับ .05 ในขณะที่พบว่าวิธี SIBTEST มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธี Mantel - Haenszel ในกรณีที่จำนวนตัวอย่างมีขนาดใหญ่ (มากกว่า 1,000 คน) และค่าเฉลี่ยของการแจกแจงของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีความแตกต่างกัน รวมทั้งข้อสอบมีความยากต่ำ และค่าอำนาจจำแนกสูง นอกจากผลทางด้านอัตราการปฏิเสธแล้ว ยังพบว่าตัวประมาณค่าการทำหน้าที่ต่างกันของข้อสอบ คือ β และ Δ ที่มีค่าต่ำกว่าเกณฑ์หรือมีค่าแสดงการทำหน้าที่ต่างกันเพียงเล็กน้อยก็อาจมีนัยสำคัญได้ ($\Delta = 1.5$ และ 1.0 และ $\beta = .10$ หรือ $.05$)

7.2.2 การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยข้อมูลจำลอง

การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยข้อมูลจำลอง ส่วนใหญ่เป็นการศึกษาเพื่อเปรียบเทียบวิธีการตรวจสอบต่าง ๆ ที่มีผู้เสนอขึ้นมาใช้ ดังนี้

ลินเคอร์ (Linacre. 1988) ได้ตรวจสอบความถูกต้องของวิธี Mantel - Haenszel และวิธี PROX ของ Rasch ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่ได้จากการจำลองข้อมูลขึ้น โดยใช้ตัวประมาณค่าความคลาดเคลื่อนมาตรฐานหลาย ๆ วิธี สำหรับวิธี Mantel - Haenszel ผลการวิเคราะห์พบว่า ทั้งสองวิธีใช้ได้ดีในสถานการณ์จำลองทุกสถานการณ์ ส่วนตัวประมาณค่าความคลาดเคลื่อนมาตรฐานของ Mantel - Haenszel มีคุณสมบัติแตกต่างกันไปและควรมีการแปลผลอย่างระมัดระวัง เมื่อใช้วิธี Mantel - Haenszel ผู้วิเคราะห์ควรจะคำนวณขนาดของการทำหน้าที่ต่างกันของข้อสอบ และความแปรปรวนของความคลาดเคลื่อน ด้วยการใส่เกณฑ์หลาย ๆ เกณฑ์ และใช้ตัวประมาณค่าความคลาดเคลื่อนมาตรฐานมากกว่า 1 ตัว สำหรับวิธี PROX จะให้การตรวจสอบการประมาณที่ได้จากวิธี Mantel - Haenszel

ธีสเซน สเตนเบิร์ก และไวเนอร์ (Thissen, Steinberg and Wainer. 1988 : 147 - 170) ได้ศึกษาความแตกต่างของโค้งการตอบข้อสอบระหว่างกลุ่ม ด้วยทฤษฎีการตอบข้อสอบ เพื่อศึกษาถึงสภาพการณ์ในการทดสอบสมมุติฐาน ความเท่ากันของค่าพารามิเตอร์ ด้วยค่าไคสแควร์แบบอัตราส่วน likelihood (IRT-LR) โดยใช้ข้อมูลจำลองและได้ทำการเปรียบเทียบการทดสอบสมมุติฐานความเท่ากันของค่าพารามิเตอร์ด้วย IRT-LR และวิธี Mantel - Haenszel พบว่า IRT-LR สามารถตรวจค้นการทำหน้าที่ต่างกันของข้อสอบได้ เมื่อค่าความแตกต่างของ ความยากระหว่างกลุ่มมีค่า 0.3 และความแตกต่างในด้านการเดามีค่า 0.1 จำนวนกลุ่มตัวอย่างน้อย (500คน) โดยจำนวนข้อสอบรวมจะน้อย หรือมากก็ไม่มีผลกระทบ ค่าความแตกต่างที่น้อยกว่านี้ไม่สามารถใช้ IRT-LR ตรวจค้นได้ว่าข้อสอบทำหน้าที่ต่างกันหรือไม่ แต่อาจทำได้ถ้าจำนวนกลุ่มตัวอย่าง มากกว่านี้

การระบุการทำหน้าที่ต่างกันของข้อสอบ ใช้วิธีเปรียบเทียบความแตกต่างระหว่างกลุ่มในข้อสอบที่ต้องการตรวจค้นการทำหน้าที่ต่างกับข้อสอบรวม ซึ่งมีข้อตกลงว่าไม่มีการทำหน้าที่ต่างกัน

อย่างไรก็ตาม ควรได้ศึกษาเพื่อหาวิธีเลือกข้อสอบรวม และการทดสอบการทำหน้าที่ต่างกันของแบบทดสอบทั้งฉบับ

สำหรับผลการตรวจค้นการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี IRT-LR และวิธี Mantel - Haenszel ให้ผลคล้ายคลึงกัน ข้อเสนอแนะ คือวิธี Mantel - Haenszel อาจเป็นประโยชน์ในการประเมินแบบสอบยาว ๆ และอาจใช้เป็นเครื่องมือกลั่นกรองก่อนจะใช้ IRT-LR ต่อไป

สวามินาธาน และโรเจอร์ (Swaminathan และ Rogers. 1990 : 361 - 370) ได้ตรวจค้นการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการถดถอยแบบโลจิส และวิธี Mantel - Haenszel โดยมีขนาดของตัวอย่าง 2 ขนาด ได้แก่ 250 และ 500 คน และความยาวของแบบสอบ 3 ขนาด ได้แก่ 40 60 และ 80 ข้อ จำลองข้อมูลการตอบข้อสอบด้วยโปรแกรม DATAGEN แบบ 3 พารามิเตอร์ ข้อมูลจำลองการทำหน้าที่ต่างกันของข้อสอบใช้ 2 พารามิเตอร์โดยให้อำนาจจำแนกของ 2 กลุ่มเท่ากัน ส่วนค่าความยากแปรผันไปเพื่อให้เกิดระดับการทำหน้าที่ต่างกันของข้อสอบต่าง ๆ ตามต้องการ แต่ในลักษณะการทำหน้าที่ต่างกันของข้อสอบแบบไม่สม่ำเสมอ (Nonuniform) นั้นให้ค่าความยากของ 2 กลุ่มเท่ากัน แต่ค่าอำนาจจำแนกแปรไปผลการวิเคราะห์พบว่าการตรวจค้นการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี โดยการใช้การทดสอบ χ^2 มีชั้นความเป็นอิสระ (df) เท่ากับ 2 สำหรับวิธีทดสอบโลจิสและการทดสอบ MH - χ^2 ให้ผลใกล้เคียงกันโดยวิธี Mantel - Haenszel ดีกว่าเล็กน้อย คือ มีการตรวจค้นได้ถูกต้องร้อยละ 75 กรณีใช้ตัวอย่าง 250 คน และตรวจค้นได้ร้อยละ 100 ในตัวอย่าง 500 คน ในทุกความยาวแบบทดสอบและในกรณีการทำหน้าที่ต่างกันของข้อสอบแบบสม่ำเสมอ (Uniform) ในกรณีของการทำหน้าที่ต่างกันของข้อสอบแบบไม่สม่ำเสมอนั้นวิธี Mantel - Haenszel ไม่สามารถตรวจค้นได้ ส่วนการถดถอยแบบโลจิส ตรวจค้นได้ถูกต้องประมาณร้อยละ 50 ในกรณีตัวอย่างน้อยและข้อสอบสั้นและถูกต้องประมาณร้อยละ 75 ในแบบทดสอบยาวและกลุ่มตัวอย่างขนาดใหญ่ แต่กรณีการทำหน้าที่ต่างกันของข้อสอบแบบไม่สม่ำเสมอ โดยที่โค้งตัดกันที่ปลายสุดของโค้งไม่ว่าจะเป็นด้านความสามารถสูงหรือต่ำก็ตามวิธี Mantel - Haenszel จะดีกว่า

ในกรณีตรวจค้นผิดคือระบุว่ามีค่าเฉลี่ยทั้ง ๆ ที่ไม่มีค่าเฉลี่ยนั้นวิธี Mantel - Haenszel ดีกว่า โดยตรวจค้นผิดประมาณร้อยละ 1 ส่วนการถดถอยแบบโลจิสผิดประมาณร้อยละ 1-6

วิธีการถดถอยแบบโลจิสแม้จะยืดหยุ่นกว่า และมีความเป็นไปได้มากกว่าแต่ค่าใช้จ่ายมากกว่าประมาณ 3-4 เท่า

ประโยชน์ของการถดถอยแบบโลจิส อยู่ตรงที่เป็นวิธีการที่มีพื้นฐานเป็นแบบจำลองที่สามารถเพิ่มตัวแปรความสามารถเข้าไปในแบบจำลองได้ เช่นเดียวกับองค์ประกอบอื่น ๆ ที่มีความเกี่ยวข้องและสำคัญซึ่งทำให้เข้าใจในธรรมชาติของการทำหน้าที่ต่างกันของข้อสอบได้ดีกว่า

ปาร์ค และ ลอเทินชเลเจอร์ (Park and Lautenschlager. 1990 : 163 - 173) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้ทฤษฎีการตอบข้อสอบที่ได้ดัดแปลงจากวิธีของลอว์ (1990) ที่เรียกว่าวิธี "ปราศจากการทำหน้าที่ต่างกัน" (Purification) ซึ่งมีวิธีการดังต่อไปนี้

1. วิเคราะห์แบบสอบตามทั้งฉบับ โดยการประมาณค่าพารามิเตอร์รวม แล้วแปลงค่าความยากให้เป็นค่ามาตรฐาน จากนั้นกำหนดค่าการเดา (c) ตามที่ได้จากการวิเคราะห์แล้วประมาณค่าความยากและอำนาจจำแนก ใหม่อีกครั้ง โดยแยกตามกลุ่มที่ศึกษาแล้วแปลงค่าความยากเป็นค่ามาตรฐานอีกครั้ง จากนั้นเปรียบเทียบฟังก์ชันการตอบข้อสอบเพื่อดูการทำหน้าที่ต่างกันของข้อสอบ

2. ประมาณค่าระดับความสามารถโดยรวมทุกกลุ่มเข้าด้วยกันใหม่ ใช้เฉพาะข้อสอบที่ไม่ทำหน้าที่ต่างกัน จะได้ข้อสอบที่เป็นเอกพันธ์

3. กำหนดค่า θ เท่ากับที่ประมาณได้ในข้อ 2 และแล้วประมาณค่า a_i และ b_i ของข้อสอบทุกข้อ รวมข้อสอบที่ทำหน้าที่ต่างกันไว้ด้วย

4. เปรียบเทียบฟังก์ชันการตรวจข้อสอบ หรือค่าพารามิเตอร์อีกครั้งด้วยค่า ไคสแควร์ วิธีนี้เรียกว่า ปราศจากการทำหน้าที่ต่างกัน เพราะพยายามที่จะประมาณค่าพารามิเตอร์โดยตัดข้อสอบที่ทำหน้าที่ต่างกันออกก่อน

วิธีที่นำมาใช้ในการศึกษาค้างนี้เป็นวิธีที่ปรับปรุงจากวิธีของลอร์ด ได้แก่

1. วิธีที่ดัดแปลงจาก Drasgow ที่ได้เสนอไว้ เป็นการตรวจค้นการทำหน้าที่ต่างกันของข้อสอบโดยการทำซ้ำหลาย ๆ รอบ จนกว่าจะได้ชุดข้อสอบที่ทำหน้าที่ต่างกันซ้ำกันในการทำซ้ำ 2 รอบสุดท้าย วิธีนี้ใช้ค่าไคสแควร์ต่ำสุดของดิฟกี (Divgi) แทนวิธีของสต็อกกิงและลอร์ด (Stocking and Lord, 1983 : 201 - 210) วิธีการนี้เรียกว่า M-DIL (Modified-Drasgow's Iterative Linking Procedure) มีวิธีการดังนี้

1) ใช้วิธีการเชื่อมโยงมาตราของสต็อกกิงและลอร์ด ประมาณค่าพารามิเตอร์บนมาตราที่เท่ากันโดยประมาณ

2) คำนวณค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบ

3) เชื่อมโยงค่าพารามิเตอร์ที่ประมาณได้อีกครั้ง โดยใช้แต่ข้อที่ไม่ทำหน้าที่ต่างกัน

4) คำนวณค่าดัชนีการทำหน้าที่ต่างกันของทุกข้ออีกครั้ง

5) ทำซ้ำ ๆ จนกว่าจะได้ข้อสอบที่ทำหน้าที่ต่างกันชุดเดียวกันใน 2 รอบสุดท้าย

2. วิธีที่ดัดแปลงจากวิธี "ปราศจากการทำหน้าที่ต่างกัน" ของ ลอร์ด (1980) ซึ่งพาร์ค(Park, 1988) ได้เสนอขึ้นเรียกว่า M-LTP (Modified-Lord Test Purification) ซึ่งมีขั้นตอนดังนี้

1) คำนวณค่า θ จากทุกกลุ่มรวมกัน

2) จากค่า θ ของแต่ละคนที่คำนวณได้ในข้อ 1) นำมาคำนวณค่าพารามิเตอร์ของข้อสอบ แล้วคำนวณค่าการทำหน้าที่ต่างกันของข้อสอบ

3) คำนวณค่า θ ใหม่อีกครั้ง โดยใช้เฉพาะข้อสอบที่ไม่ทำหน้าที่ต่างกัน

4) คำนวณค่าพารามิเตอร์ของข้อสอบอีกครั้งโดยใช้ค่า θ ที่คำนวณได้จาก 3)

5) คำนวณค่าการทำหน้าที่ต่างกันของข้อสอบ

6) ทำซ้ำจนกว่าจะได้ข้อสอบซ้ำกัน ในการทำซ้ำ 2 รอบสุดท้าย

3. ใช้วิธีผสมระหว่าง M-DIL และวิธีปราศจากการทำหน้าที่ต่างกันของข้อสอบ เรียกว่า ILAP (Iterative Parameter Linking and θ Scale Purification) ซึ่งจะเริ่มด้วย M-DIL จนถึงรอบสุดท้ายแล้วเริ่มประมาณค่า θ ใหม่อีกครั้งแยกตามกลุ่ม เฉพาะข้อสอบที่ไม่ทำหน้าที่ต่างกัน จากนั้น จึงคำนวณค่าพารามิเตอร์ของข้อสอบแยกตามกลุ่ม แล้วใช้วิธีของดิฟกี ประมาณค่าคงที่ของการเชื่อมโยง โดยใช้เฉพาะข้อสอบที่ไม่ทำหน้าที่ต่างกัน แล้วคำนวณค่าดัชนีการทำหน้าที่ต่างกันของข้อสอบใหม่ แล้วตามด้วย M-DIL ต่อ

ข้อมูลที่ศึกษา เป็นข้อมูลที่จำลองด้วยวิธีของซิมป์สัน (Simpson) ภายใต้สถานการณ์การทำหน้าที่ต่างกันชนิดทิศทางเดียวกัน และทิศทางผสม พบว่าการวิเคราะห์ 2 วิธีแรก ให้ผลเท่ากัน แต่ M-DIL ต้องทำซ้ำในจำนวนรอบที่มากกว่า M-LTP และถ้าใช้วิธีสุดท้ายแล้วจะสามารถลดจำนวนข้อสอบที่ทำหน้าที่ต่างกัน แต่ถูกระบุจากการตรวจค้นว่าไม่ทำหน้าที่ต่างกันลงมากกว่า 2 วิธีแรก

คลอสเซอร์ และคณะ (Clauser and Others. 1991) ได้ศึกษาถึงอิทธิพลต่าง ๆ ที่มีผลต่อวิธีการตรวจค้นการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี Mantel - Haenszel ข้อมูลที่ใช้เป็นข้อมูลที่จำลองขึ้นเพื่อใช้ในการศึกษาการตรวจค้นการทำหน้าที่ต่างกันแบบสมำเสมอ ข้อมูลที่จำลองขึ้นเป็นแบบสอบถามจำนวน 75 ข้อ เป็นข้อที่ทำหน้าที่ต่างกันอยู่ 16 ข้อ มีระดับความแตกต่างในค่าความยากหลายระดับระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ซึ่งมีจำนวนกลุ่มละ 1,500 คน มีการวิเคราะห์ระดับความยากที่ต่างกันหลายระดับกับระดับความยากของแบบทดสอบ 5 ระดับ อำนาจจำแนก 4 ระดับ และการกระจายของความสามารถ 2 ลักษณะ ผลการวิเคราะห์แสดงอิทธิพลจากความแตกต่างในระดับความยาก รวมทั้งค่าอำนาจจำแนกและค่าความยาก ซึ่งทำให้ได้ข้อเสนอนี้ว่า ค่าสถิติวิธี Mantel - Haenszel นี้ใช้ได้ดีสำหรับข้อสอบที่มีค่าอำนาจจำแนกสูงและมีแนวโน้มจะไม่สามารถตรวจค้นการทำหน้าที่ต่างกันในข้อสอบที่มีค่าความยากสูง

เมเซอร์ และคณะ (Mazor, et. al. 1991 : 443-451) ได้ใช้วิธีการตรวจค้นการทำหน้าที่ต่างกันของข้อสอบวิธี Mantel - Haenszel ศึกษาข้อมูลที่จำลองขึ้นโดยใช้กลุ่มตัวอย่างขนาด 2,000 1,000 500 200 และ 100 คน เพื่อเปรียบเทียบอัตราการตรวจค้นโดยใช้ข้อสอบที่จำลองขึ้นที่จำลองขึ้น 5 ชุด ชุดละ 75 ข้อ ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยการทดสอบค่า $MH - \chi^2$ $df = 1$ ผลการศึกษพบว่า เมื่อใช้จำนวนผู้สอบ 2,000 คน วิธีนี้ตรวจค้นการทำหน้าที่ต่างกันของข้อสอบได้ผิดพลาดร้อยละ 30 และเมื่อใช้จำนวนผู้สอบ 500 คนลงไป พบว่าตรวจค้นการทำหน้าที่ต่างกันของข้อสอบผิดพลาดร้อยละ 50 ข้อสอบที่ไม่สามารถตรวจค้นการทำหน้าที่ต่างกันของข้อสอบได้ ได้แก่ ข้อสอบที่ยากมาก ข้อสอบที่มีค่าความยากต่างกันเล็กน้อยสำหรับคน 2 กลุ่ม และเป็นข้อสอบที่มีค่าอำนาจจำแนกต่ำ

โคเฮน และคิม (Cohen and Kim. 1993: 39 - 52) ศึกษาเปรียบเทียบการใช้ทฤษฎีการตอบข้อสอบ 2 พารามิเตอร์ โดยใช้สถิติทดสอบ χ^2 ของลอร์ด และสถิติทดสอบพื้นที่ความแตกต่างระหว่างโค้งลักษณะข้อสอบชนิดคิดเครื่องหมายและไม่คิดเครื่องหมาย อันได้แก่ สถิติ Z ของราชู โดยใช้ข้อมูลจำลองตามทฤษฎีการตอบข้อสอบ 2 พารามิเตอร์ จำลองการตอบข้อสอบ 2 ขนาด ข้อสอบ 20 และ 60 ข้อ จำนวนผู้สอบ 100 และ 500 คน และมีจำนวนข้อสอบที่ทำหน้าที่ต่างกันแบบทดสอบจำนวนร้อยละ 0 10 และ 20 ผลการวิเคราะห์พบว่า

1) แบบทดสอบที่ยาวการระบุข้อสอบที่ทำหน้าที่ต่างกันผิดพลาด (FP) จะมากขึ้น และถ้ามีจำนวนข้อสอบที่ทำหน้าที่ต่างกันอยู่ในแบบทดสอบน้อย จำนวนข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันผิดพลาดจะยิ่งมากขึ้น

องค์ประกอบที่มีอิทธิพลต่อการระบุว่าข้อสอบทำหน้าที่ต่างกันผิดพลาด(FP) อีกประการคือ ระดับนัยสำคัญ (α) ค่า α ยิ่งมาก FP จะเกิดมาก (ที่ .05 จะเกิด FP มากกว่าที่ .01) ผลที่เกิดขึ้นเกิดในทุกสถานการณ์ที่ศึกษา สำหรับ Z แต่เกิดในบางสถานการณ์สำหรับ χ^2

2) แบบทดสอบที่ยาวกว่า มีร้อยละของข้อสอบที่ทำหน้าที่ต่างกันมากกว่า จะมีข้อสอบที่ถูกระบุว่าไม่ทำหน้าที่ต่างกัน ผิดพลาด (FN) มากกว่า แต่จะเกิดขึ้นกับการทดสอบด้วย Z มากกว่า

กรณีที่จำนวนผู้สอบเท่ากับ 100 คนพบว่าความแตกต่างในการระบุว่าข้อสอบไม่ทำหน้าที่ต่างกันผิดพลาด สำหรับ Z และ χ^2 ไม่แตกต่างกันมากเท่ากรณีใช้กลุ่มตัวอย่าง 500 คน ซึ่งมีการระบุว่าไม่ทำหน้าที่ต่างกันผิดพลาด สำหรับ χ^2 น้อยกว่า Z โดยเฉพาะอย่างยิ่งในกรณีที่มีข้อสอบที่ทำหน้าที่ต่างกันแบบทดสอบมากที่สุดคือร้อยละ 20

โคเฮน และ คิม วิจารณ์ว่าการที่ค่าสถิติตรวจสอบการทำหน้าที่แบบ χ^2 ให้ผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันดีกว่า Z ของพื้นที่ที่ติดเครื่องหมายอาจเป็นเพราะ Z (ESA) ใช้การเปรียบเทียบเฉพาะค่าความยาก ในขณะที่ χ^2 ใช้ทั้งค่าความยากและค่าอำนาจจำแนก แต่ไม่สามารถอธิบายความแตกต่างได้ในกรณีใช้ Z (H) ซึ่งทดสอบพื้นที่ที่ไม่ติดเครื่องหมายเพราะ Z (H) ก็ใช้การทดสอบค่าพารามิเตอร์ทั้งสองตัวเช่นกัน

คิม และโคเฮน (Kim and Cohen. 1994 : 217 - 228) ได้ศึกษาวิธีการของลอร์ด (Lord's χ^2 Test) ในการตรวจสอบฟังก์ชันข้อสอบที่ทำหน้าที่ต่างกัน โดยพิจารณาอัตราการเกิดความคลาดเคลื่อนชนิดที่ 1 (Type I error rates) ใช้การจำลองข้อมูล กลุ่มตัวอย่าง 250 และ 1,000 คน จำนวนข้อสอบ 50 ข้อ แบบจำลอง 2, 3 และ 3 พารามิเตอร์ ที่กำหนดให้ค่าการเดาคงที่ (Fixed Guessing Parameter) และประมาณค่าพารามิเตอร์ด้วยวิธี มาร์จินัล แมกซ์ิมัม โลลิฮูด (Marginal Maximum Likelihood Estimation) และ มาร์จินัล เบย์เซียน (Marginal Bayesian Estimation) พบว่าอัตราการเกิดความคลาดเคลื่อนชนิดที่ 1 (Type I error rates) ที่ใช้วิธีการประมาณค่า 2 วิธี ดังกล่าว เมื่อเทียบกับวิธีประมาณค่าแบบ จอยท์ แมกซ์ิมัม โลลิฮูด (Joint Maximum Likelihood Estimation)

นารายานานและสวามินาทาน (Narayanan and Swaminathan. 1996) ได้ศึกษาเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบไม่สม่ำเสมอ ด้วยวิธี Mantel - Haenszel วิธีถดถอยโลจิสติก และวิธี SIBTEST โดยศึกษาอำนาจการตรวจสอบและการจำแนกผิดพลาด โดยการจำลองข้อมูล ตัวแปรที่ศึกษา ได้แก่ ขนาดกลุ่มตัวอย่าง ความแตกต่างของการกระจายความสามารถระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ สัดส่วนของข้อสอบที่ทำหน้าที่ต่างกันที่มีภายในแบบทดสอบ ขนาดของพื้นที่ระหว่างโดเมนลักษณะข้อสอบของผู้สอบ 2 กลุ่ม ค่าความยากและค่าอำนาจจำแนกของแบบทดสอบ

ผลการศึกษา พบว่า วิธี SIBTEST และวิธีถดถอยโลจิสติก มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบไม่สม่ำเสมอได้เท่าเทียมกันในทุกเงื่อนไขที่ศึกษา ส่วนวิธี Mantel - Haenszel ไม่สามารถตรวจพบข้อสอบที่ทำหน้าที่ต่างกันแบบไม่สม่ำเสมอและการจำแนกผิดพลาด วิธี Mantel - Haenszel จะสูงกว่าวิธี SIBTEST และวิธีถดถอยโลจิสติก

จากการศึกษาในระยะประมาณ 10 ปี ที่ผ่านมา วิธีที่นิยมใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ได้แก่ วิธี Lord's χ^2 ตามแนวคิดทฤษฎีการตอบข้อสอบ วิธี Mantel - Haenszel วิธี SIBTEST ซึ่งมีการศึกษาทั้งข้อมูลจำลองและข้อมูลจริง รวมทั้งกำหนดเงื่อนไขในการศึกษาแตกต่างกันไป เช่น ความยาวของแบบทดสอบต่างกัน ขนาดของกลุ่มตัวอย่างต่างกัน และจำนวนข้อสอบที่ทำหน้าที่ต่างกันแบบทดสอบต่างกัน แม้ว่าจะมีการศึกษาอยู่บ้าง แต่ยังไม่มียุทธวิธีที่ชัดเจนว่าวิธีใดมีประสิทธิภาพดีกว่ากันในเงื่อนไขของการศึกษาที่กำหนด

ดังนั้นผู้วิจัยจึงสนใจจะศึกษาผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน ระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST เมื่อกำหนดเงื่อนไขของการศึกษาแตกต่างกันดังกล่าว

บทที่ 3

วิธีดำเนินการศึกษา

การศึกษาค้างนี้ มีความมุ่งหมายเพื่อศึกษาผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันว่ามี ความ สอดคล้องหรือไม่ โดยเปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST เมื่อขนาดของกลุ่มตัวอย่าง ความยาวของแบบทดสอบต่างกัน มีขั้นตอนการ ศึกษาดังนี้

ประชากร

ประชากรที่ใช้ในการศึกษาค้างนี้ ได้แก่ นักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2543 สังกัดกรมสามัญศึกษา จังหวัดกรุงเทพมหานคร ทั้งหมด 100 โรงเรียน จำนวนนักเรียน 49,708 คน เป็น นักเรียนชาย 25,872 คน และนักเรียนหญิง 23,830 คน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้างนี้ ได้แก่ นักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2543 สังกัดกรมสามัญศึกษา จังหวัดกรุงเทพมหานคร จำนวน 10 โรงเรียน จำนวน 790 คน แบ่งเป็น นักเรียนชาย 378 คน และนักเรียนหญิง 412 คน ซึ่งได้มาโดยการสุ่มแบบสองขั้นตอน (Two Stage Random Sampling) โดยมีขั้นตอนดำเนินการดังนี้

ประมาณกลุ่มตัวอย่างที่ใช้ในการศึกษา ด้วยความเชื่อมั่น 95% ($\alpha = .05$) เมื่อเทียบจาก ตารางขนาดกลุ่มตัวอย่างที่ควรเลือกจากประชากร พบว่า ต้องใช้กลุ่มตัวอย่างอย่างน้อย 397 คน (ล้วน สายยศและอังคณา สายยศ. 2538 ; อ้างอิงมาจาก Yamane. 1967) แต่ในการศึกษาค้างนี้ใช้ กลุ่มตัวอย่าง 790 คน ซึ่งมีขั้นตอนการสุ่มดังนี้

ขั้นที่ 1 สุ่มโรงเรียนตามขนาดโรงเรียน ด้วยวิธีการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) มีขนาดโรงเรียนเป็นชั้น (Strata) มีโรงเรียนเป็นหน่วยการสุ่ม (Sampling Unit) โดยแบ่ง โรงเรียนออกเป็น 3 ขนาด คือ โรงเรียนขนาดใหญ่พิเศษ โรงเรียนขนาดใหญ่ และโรงเรียนขนาดกลาง ดังนี้

1. โรงเรียนขนาดใหญ่พิเศษ ที่มีนักเรียนมากกว่า 2,500 คน มี 47 โรงเรียน มีจำนวน ห้องเรียน 631 ห้องเรียน
2. โรงเรียนขนาดใหญ่ ที่มีนักเรียนมากกว่า 1,500 ถึง 2,500 คน มี 32 โรงเรียน มี จำนวนห้องเรียน 373 ห้องเรียน
3. โรงเรียนขนาดกลาง ที่มีนักเรียน 500 ถึง 1,500 คน มี 21 โรงเรียน มีจำนวนห้องเรียน 151 ห้องเรียน

สุ่มโรงเรียนมาประมาณ 10% ของแต่ละขนาด ได้โรงเรียนขนาดใหญ่พิเศษ 5 โรงเรียน โรงเรียน ขนาดใหญ่ 3 โรงเรียน และโรงเรียนขนาดกลาง 2 โรงเรียน

ขั้นที่ 2 สุ่มห้องเรียนมาประมาณ 15% ของแต่ละโรงเรียนจากโรงเรียนแต่ละขนาดในขั้นที่ 1 ด้วยวิธีการสุ่มอย่างง่าย โดยสุ่มโรงเรียนขนาดใหญ่พิเศษ 10 ห้องเรียน จำนวนนักเรียน 454 คน เป็นนัก

เรียนชาย 250 คนและนักเรียนหญิง 204 คน โรงเรียนขนาดใหญ่ 6 ห้องเรียน จำนวนนักเรียน 262 คน เป็นนักเรียนชาย 74 คน นักเรียนหญิง 188 คน โรงเรียนขนาดกลาง 2 ห้องเรียน จำนวนนักเรียน 74 คน เป็นนักเรียนชาย 54 คน นักเรียนหญิง 20 คน รวมนักเรียนที่ใช้เป็นกลุ่มตัวอย่าง 790 คน ดังแสดงในตาราง 3

ตาราง 3 จำนวนกลุ่มตัวอย่าง แยกตามขนาดโรงเรียน

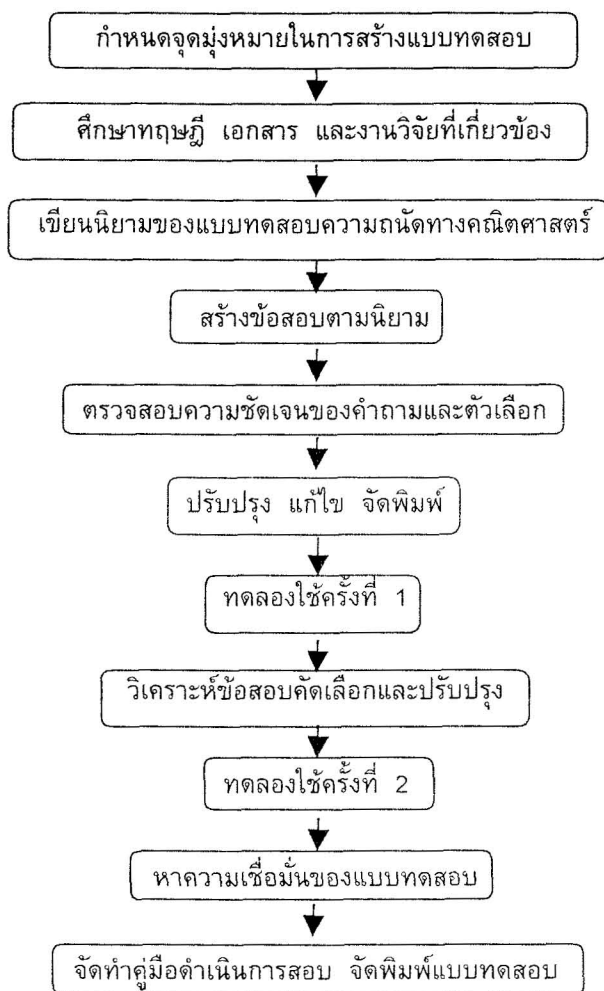
โรงเรียน	จำนวนนักเรียนทั้งหมด (ห้องเรียน)	จำนวนกลุ่มตัวอย่าง (ห้องเรียน)
<u>ขนาดใหญ่พิเศษ</u>		
พระโขนงพิทยาล	714(18)	124(3)
วัดสุทธวราราม	689(12)	110(2)
นนทรีวิทยา	560(14)	80(2)
ปทุมคงคา	541(14)	92(2)
สตรีวัดมหาพฤฒาราม	480(11)	48(1)
<u>ขนาดใหญ่</u>		
สันติราษฎร์วิทยาลัย	466(12)	78(2)
มัธยมวัดธาตุทอง	465(10)	94(2)
สตรีศรีสุริโยทัย	446(10)	90(2)
<u>ขนาดกลาง</u>		
มัธยมวัดเบญจมบพิตร	240(6)	40(1)
มักกะสันพิทยาล	202(6)	34(1)
รวม	4,803(113)	790(18)

เครื่องมือที่ใช้ในการศึกษา

เครื่องมือที่ใช้ในการศึกษาครั้งนี้ ได้แก่ แบบทดสอบความถนัดทางคณิตศาสตร์ ซึ่งสร้างตามแนวทางวิชของเทอร์สโตน เป็นแบบปรนัยชนิดเลือกตอบ 5 ตัวเลือก จำนวน 60 ข้อ ที่ผู้วิจัยสร้างขึ้น

วิธีดำเนินการสร้างเครื่องมือ

การสร้างเครื่องมือสำหรับการศึกษาค้นคว้าได้ดำเนินการสร้างตามขั้นตอนต่อไปนี้



ภาพประกอบ 3 แสดงลำดับขั้นตอนในการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า

รายละเอียดขั้นตอนการสร้างเครื่องมือ

เครื่องมือที่ใช้ในการวิจัยครั้งนี้มีขั้นตอนการสร้าง ดังนี้

1. กำหนดจุดมุ่งหมายในการสร้างแบบทดสอบความถนัดทางคณิตศาสตร์ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 1
2. ศึกษาทฤษฎีและเอกสารงานวิจัยที่เกี่ยวข้องคือ แบบทดสอบความถนัดตามแนวทฤษฎีของเทอร์สตัน เพื่อนำมาเป็นแนวทางในการสร้างแบบทดสอบวัดความถนัดทางคณิตศาสตร์
3. เขียนนิยามของแบบทดสอบความถนัดทางคณิตศาสตร์
4. สร้างข้อสอบตามนิยามที่เขียนไว้ ซึ่งแบ่งเป็น 4 ประเภท ประเภทละ 20 ข้อ จำนวนทั้งหมด 80 ข้อ
5. นำแบบทดสอบความถนัดทางคณิตศาสตร์ที่สร้างขึ้น ให้ผู้เชี่ยวชาญทางด้านวัดผลการศึกษ จำนวน 5 คน ตรวจสอบคุณภาพเบื้องต้นเพื่อหาค่าความเที่ยงตรงเชิงพินิจ (Face Validity) และวิจารณ์เกี่ยวกับความชัดเจนของคำถามและตัวเลือก แล้วคัดเลือกข้อที่มีค่า IOC ตั้งแต่ 0.5 ขึ้นไปเพื่อนำไปทดลองใช้ ปรากฏว่าได้ 80 ข้อ ประกอบด้วย ประเภททักษะ 20 ข้อ ประเภทความคิดรวบยอด 20 ข้อ ประเภทเรียงอันดับ 20 ข้อ และประเภทโจทย์ปัญหา 20 ข้อ
6. นำแบบทดสอบที่จัดพิมพ์แล้วไปทดลองใช้กับนักเรียนในโรงเรียนซึ่งไม่ใช่กลุ่มตัวอย่าง ในการศึกษาครั้งนี้ใช้นักเรียนโรงเรียนศรีอยุธยา จำนวน 100 คน
7. วิเคราะห์ข้อสอบหาค่าความยาก และค่าอำนาจจำแนกของแบบทดสอบ แล้วคัดเลือกข้อที่มีค่าความยากระหว่าง .20 - .80 และค่าอำนาจจำแนกตั้งแต่ .20 ขึ้นไป ไว้ใช้ในการทดลองใช้ครั้งที่ 2 จำนวน 60 ข้อ
8. นำแบบทดสอบที่ได้จำนวน 60 ข้อ ไปทดลองใช้ครั้งที่ 2 กับนักเรียนในโรงเรียนที่ไม่ใช่กลุ่มตัวอย่าง ในการศึกษาครั้งนี้ใช้นักเรียนโรงเรียนยานนาวา จำนวน 100 คน จากนั้นนำผลที่ได้จากการทดลองใช้ มาวิเคราะห์หาค่าความเชื่อมั่นของแบบทดสอบ โดยใช้สูตร KR-20 ได้ค่าความเชื่อมั่นเท่ากับ 0.895 จึงได้แบบทดสอบไว้ใช้ในการวิจัย 60 ข้อ
9. จัดทำคู่มือดำเนินการสอบและพิมพ์แบบทดสอบเป็นรูปเล่ม เพื่อใช้ในการวิจัย

ตัวอย่างแบบทดสอบ

แบบทดสอบความถนัดทางคณิตศาสตร์

คำชี้แจง

1. แบบทดสอบฉบับนี้เป็นแบบปรนัยชนิดเลือกตอบ มีจำนวน 60 ข้อ ใช้เวลา 60 นาที ฉะนั้น นักเรียนควรรีบตอบโดยเร็วและให้ครบทุกข้อจึงจะได้คะแนนดี
2. แบบทดสอบฉบับนี้เป็นแบบทดสอบวัดความถนัดด้านคณิตศาสตร์ การตอบของนักเรียนไม่มีผลต่อการเรียนวิชาคณิตศาสตร์แต่อย่างใด โปรดทำให้เต็มความสามารถ

ประเภททักษะ(Skill) เป็นคำถามประเภทการคำนวณด้านตัวเลขขั้นพื้นฐานทั่วไป เช่น

(0) $0.001 \times 0.01 = ?$

ก 0.001

ข 0.0001

- ค 0.0011
- (ง) 0.00001
- จ 0.00011

ประเภทความคิดรวบยอด (Concept) เป็นคำถามประเภทความเข้าใจเกี่ยวกับหลักการ วิธีการ รวมทั้งการแปลความ ตีความ ขยายความ เช่น

(00) ซื้อส้มมา 1 ถูง กับ 20 ผล เป็นเงิน 200 บาท เมื่อขายไปได้กำไร 20 บาท ขายส้มไปผลละเท่าไร จากโจทย์หาคำตอบได้หรือไม่เพราะเหตุใด

- ก หาไม่ได้เพราะไม่ทราบราคาทุ่น
- ข หาไม่ได้เพราะไม่ทราบราคาขาย
- (ค) หาไม่ได้เพราะไม่ทราบจำนวนส้มที่ซื้อมา
- ง หาไม่ได้เพราะไม่ทราบจำนวนส้มที่ขาย
- จ หาไม่ได้

ประเภทโจทย์ปัญหา(Problem Solving) เป็นคำถามประเภทโจทย์ปัญหาทั่วไป เช่น

(000) นายขวามีเงิน 32 บาท ถ้าให้นายแดง 4 บาท นายแดงจะมีเงินเป็นครึ่งหนึ่งของ นายขาว เดิมนายแดงมีเงินกี่บาท

- ก 16
- ข 14
- (ค) 12
- ง 10
- จ 6

ประเภทเรียงอันดับ (N – series) เป็นคำถามประเภทให้หาคำตอบที่อยู่ถัดไป โดยพิจารณาจาก โจทย์ที่กำหนดให้ เช่น

(0000) 20 15 11 8 6?.....

- ก 1
- ข 2
- ค 3
- ง 4
- (จ) 5

วิธีดำเนินการรวบรวมข้อมูล

1. ขอนหนังสือจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อขอความ อนุเคราะห์ในการเก็บรวบรวมข้อมูล

2. ติดต่อโรงเรียนที่ใช้เป็นกลุ่มตัวอย่าง ขออนุญาตผู้บริหารโรงเรียน เพื่อกำหนด วัน เวลา ที่ทำการทดสอบ

3. วางแผนในการดำเนินการสอบ โดยผู้วิจัยเป็นผู้ดำเนินการสอบเอง จัดเตรียมแบบทดสอบให้เพียงพอกับจำนวนนักเรียนในแต่ละครั้ง

4. นำแบบทดสอบไปทดสอบกับนักเรียนที่เป็นกลุ่มตัวอย่างทั้งหมด ตามวัน เวลา ที่นัดหมาย

5. ชี้แจงให้นักเรียนในกลุ่มตัวอย่างเข้าใจ และทราบถึงวัตถุประสงค์ในการสอบ และขอความร่วมมือในการสอบ

6. อธิบายให้นักเรียนทุกคนเข้าใจวิธีการทำแบบทดสอบ

7. หลังจากดำเนินการสอบกลุ่มตัวอย่างครบทุกโรงเรียนแล้ว นำกระดาษคำตอบมาตรวจให้คะแนน ดังนี้ ถ้าตอบถูกให้ 1 คะแนน ถ้าตอบผิดหรือว่างไว้ให้ 0 คะแนน และนำผลมาวิเคราะห์ทางสถิติ

การวิเคราะห์ข้อมูลและสถิติที่ใช้ในการวิเคราะห์ข้อมูล

1. การจัดเตรียมข้อมูลให้สอดคล้องกับจุดมุ่งหมายของการศึกษา เพื่อการตรวจสอบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน และความสอดคล้องในการตรวจสอบทั้ง 3 วิธี มีขั้นตอนดังนี้

1.1. สุ่มข้อมูลจากกลุ่มตัวอย่างให้มีขนาดขนาด 300, 500 และ 700 คน ตามลำดับแยกตามเพศชาย - หญิง อย่างละครึ่ง

1.2 สุ่มข้อสอบ เพื่อจัดชุดแบบทดสอบใหม่ ขนาด 20, 30, และ 40 ข้อ ตามลำดับ โดยใช้การสุ่มอย่างง่ายแบบใส่คืน (จับสลากหมายเลขข้อสอบที่เขียนไว้) ได้กรอบหรือแบบแผนของการศึกษาดังนี้

I	N	วิธีตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน		
		LC	MH	SIB
20	300			
	500			
	700			
30	300			
	500			
	700			
40	300			
	500			
	700			

* I = จำนวนข้อสอบ

N = ขนาดกลุ่มตัวอย่าง

ได้แบบทดสอบที่ใช้ในการศึกษา 3 ชุด ดังนี้

แบบทดสอบชุดที่ 1 จำนวน 20 ข้อ เพื่อนำไปตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสำหรับกลุ่มผู้สอบ 300 คน 500 คน และ 700 คน

แบบทดสอบชุดที่ 2 จำนวน 30 ข้อ เพื่อนำไปตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มผู้สอบ 300 คน 500 คน และ 700 คน

แบบทดสอบชุดที่ 3 จำนวน 40 ข้อ เพื่อนำไปตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับกลุ่มผู้สอบ 300 คน 500 คน และ 700 คน

2. การตรวจสอบความเป็นมิติเดียวของแบบทดสอบ ใช้การวิเคราะห์องค์ประกอบ (Factor Analysis) ด้วยการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis) หมุนแกนด้วยวิธีวาริแมกซ์ (Varimax) ตามข้อเสนอของลอร์ด สำหรับการตรวจสอบความเป็นมิติเดียว ครั้งนี้ใช้โปรแกรมสำเร็จรูป SPSS FOR WINDOWS Version 7.5

3. การประมาณค่าพารามิเตอร์ของข้อสอบ ใช้โปรแกรม BILOG version 3.04 (Mislevy and Bock, 1989) ซึ่งเป็นโปรแกรมที่สามารถใช้กับไมโครคอมพิวเตอร์ เพื่อประมาณค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าอำนาจจำแนก (a) ค่าความยาก (b) และค่าการเดา (c) ตามแนวคิดทฤษฎีการตอบข้อสอบ (Item Response Theory) วิธีประมาณค่ามีทั้งวิธีความน่าจะเป็นสูงสุดรวม (Marginal Maximum Likelihood Method) และวิธีเบย์ (Bayesian Method) ข้อมูลที่ใช้เป็นแบบ 0 – 1 (Binary) แบ่งเป็นโปรแกรมย่อย ดังนี้

INPUT เป็นโปรแกรมย่อย ที่ใช้ในขั้นตอนการกำหนดลักษณะของข้อมูล การแปลข้อมูลดิบ และแบบจำลองที่ใช้ในการวิเคราะห์

CALIBRATE เป็นโปรแกรมย่อยที่ใช้ในการประมาณค่าพารามิเตอร์ของข้อสอบ สำหรับผลของการวิเคราะห์ ประกอบด้วย 3 ส่วน (Phase) คือ

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

แบบที่ 2 เป็นการประมาณค่าพารามิเตอร์ของข้อสอบ ตามแนวคิดทฤษฎีการตอบข้อสอบ ได้แก่ ค่าอำนาจจำแนก (a) ค่าความยาก (b) ค่าการเดา (c) เมตริกซ์ของความแปรปรวน และความแปรปรวนร่วม รวมทั้งจะมีผลการทดสอบความเหมาะสมระหว่างแบบจำลองกับข้อมูลในการศึกษาด้วย ซึ่งการคำนวณเพื่อการเปรียบเทียบค่าพารามิเตอร์ของข้อสอบ จะต้องใช้ไฟล์พารามิเตอร์ของข้อสอบ และไฟล์เมตริกซ์ของความแปรปรวนและความแปรปรวนร่วม ในโปรแกรม EQUATE และโปรแกรม IRTDIF ต่อไป

แบบที่ 3 เป็นผลการวิเคราะห์เกี่ยวกับฟังก์ชันสารสนเทศของข้อสอบหรือแบบทดสอบ (Test Information Function)

4. การแปลงหน่วยของพารามิเตอร์ข้อสอบ (Metric Transformation) ถือเป็นขั้นตอนที่สำคัญ 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. การตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันโดยวิธี Lord's χ^2

ใช้โปรแกรม IRTDIF version 1.0 พัฒนาโดยคิมและโคเฮน (Kim and Cohen. 1991) เพื่อวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบตามวิธีของลอร์ด

6. การตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันโดยวิธี Mantel - Haenszel และวิธี SIBTEST

ใช้โปรแกรม SIBTEST ของ Stout และ Rousos (1992) ซึ่งมีโปรแกรมย่อยอยู่ 2 โปรแกรม ดังนี้

- SIBIN ใช้ในการตั้งค่าพารามิเตอร์
- SIBTEST ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบและการทดสอบสมมติฐานทางสถิติ โดย SIBTEST จะนำค่าพารามิเตอร์ที่เก็บไว้ใน SIBIN มาใช้ในการประมวลผลโปรแกรม

7. การหาความสอดคล้องในการตรวจสอบระหว่างวิธีตรวจสอบ

เป็นการวิเคราะห์ความสอดคล้องของผลการตัดสินการทำหน้าที่ต่างกันของข้อสอบ จากสูตร (กาญจนา วัฒนสุนทร. 2537 : 104)

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

เมื่อ R แทน จำนวนข้อสอบที่ตรวจสอบได้ตรงกันในการตัดสิน

T แทน จำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในการเปรียบเทียบแต่ละคู่

8. การทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธีการตรวจสอบ โดยใช้สถิติทดสอบ คือ Cochran Q Test (Siegel. 1956 : 162)

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

เมื่อ Q แทน การทดสอบแบบคิว

K แทน จำนวนกลุ่มตัวอย่าง (จำนวนวิธีการตรวจสอบ)

G_j แทน ผลรวมในแนวตั้ง (จำนวนข้อสอบที่ทำหน้าที่ต่างกันแต่ละวิธี)

L_i แทน ผลรวมในแนวนอน

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การนำเสนอผลการวิเคราะห์ข้อมูล ได้เรียงตามลำดับขั้นตอนดังนี้

ตอนที่ 1 การตรวจสอบความเป็นมิติเดียวของข้อสอบ

ตอนที่ 2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

ตอนที่ 3 ผลการทดสอบความแตกต่างของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST แบ่งเป็น

3.1 ผลการทดสอบความแตกต่างของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
เมื่อกลุ่มตัวอย่างเท่ากัน วิธีการตรวจสอบต่างกัน

3.2 ผลการทดสอบความแตกต่างของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
เมื่อกลุ่มตัวอย่างต่างกัน วิธีการตรวจสอบเหมือนกัน

ตอนที่ 4 ความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

เพื่อให้การนำเสนอข้อมูลเข้าใจตรงกันและง่ายขึ้น จึงกำหนดสัญลักษณ์แทนดังนี้

N แทน จำนวนคน

I แทน จำนวนข้อสอบ

DIF แทน การทำหน้าที่ต่างกันของข้อสอบ

LC แทน การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธี Lord 's χ^2

MH แทน การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธี Mantel - Haenszel

SIB แทน การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธี SIBTEST

* แทน $p < .05$

** แทน $p < .01$

*** แทน $p < .001$

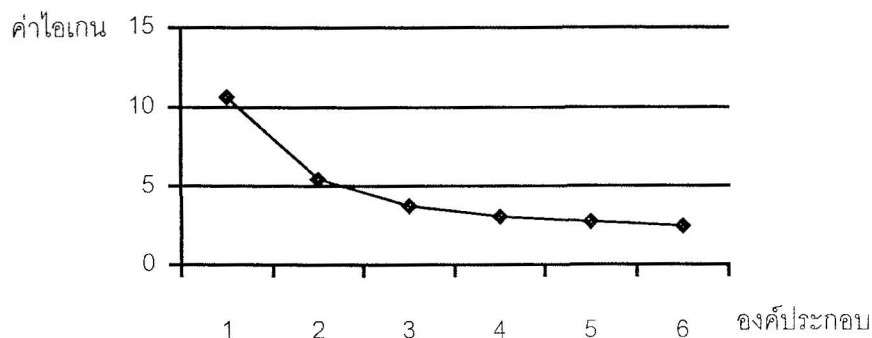
 แทน ข้อสอบที่ตรวจสอบพบว่าทำหน้าที่ต่างกันตรงกัน

ตอนที่ 1 การตรวจสอบความเป็นมิติเดียวกันของข้อสอบ

การวิเคราะห์ตอนที่ 1 ผู้วิจัยได้นำคะแนนจากการสอบ มาตรวจสอบความเป็นมิติเดียวกันของข้อสอบโดยใช้การวิเคราะห์องค์ประกอบ (Factor Analysis) ด้วยการวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis) หมุนแกนด้วยวิธีแวนริแมกซ์ (Varimax) โดยใช้โปรแกรมสำเร็จรูป SPSS For Windows Version 7.5 และหาค่าไอเกน แล้วนำค่าไอเกนมาลงจุดกราฟ แสดงผลดังตาราง 4 และภาพประกอบ 4

ตาราง 4 ค่าไอเกนและร้อยละของความแปรปรวนขององค์ประกอบ

องค์ประกอบ	ค่าไอเกน	ร้อยละความแปรปรวน
1	10.63	17.43
2	5.45	8.93
3	3.75	6.15
4	3.05	4.99
5	2.73	4.47
6	2.44	4.00



ภาพประกอบ 4 ผลการตรวจสอบความเป็นมิติเดียวกันของข้อสอบ

ผลการวิเคราะห์ตามตาราง 4 และภาพประกอบ 4 พบว่า ค่าไอเกนขององค์ประกอบที่ 1 มีค่าสูงกว่าค่าไอเกนขององค์ประกอบที่ 2 ประมาณ 2 เท่า ส่วนเส้นกราฟจากองค์ประกอบที่ 1 มา 2 จะลาดชัน และตั้งแต่องค์ประกอบที่ 2 เป็นต้นไปจะใกล้เคียงกัน นับว่าองค์ประกอบที่ 1 มีค่าเด่นกว่าองค์ประกอบอื่น แม้ว่าค่าร้อยละของความแปรปรวนจะไม่สูงนัก (ร้อยละ 17.43) หากพิจารณาตามเกณฑ์ที่เรคเคส (Reckase. 1979 : 3, citing Raju. 1993 : 304) เสนอค่าร้อยละของความแปรปรวนควรมีค่าไม่น้อยกว่าร้อยละ 20 จึงจะถือว่าแบบทดสอบมีความเป็นมิติเดียว ซึ่งในทางปฏิบัติหาพบได้ยาก เนื่องจากแบบทดสอบประกอบด้วยเนื้อหาหลายอย่างและวัดหลายสมรรถภาพ

ตอนที่ 2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธี Lord 's χ^2

วิธี Mantel - Haenszel และวิธี SIBTEST

การวิเคราะห์ตอนที่ 2 ผู้วิจัยได้นำผลที่ได้จากการสอบมาสุ่มกลุ่มตัวอย่างเป็นสามขนาดคือ ขนาด 300 คน 500 คน และ 700 คน ตามลำดับ แยกตามเพศชาย - หญิง อย่างละครึ่ง แล้วสุ่มข้อสอบ เพื่อจัดชุดแบบทดสอบใหม่ โดยใช้การสุ่มอย่างง่ายแบบใส่คืน (จับสลากหมายเลขข้อที่เขียนไว้) ได้แบบทดสอบสามชุดคือ แบบทดสอบที่มีจำนวน 20 ข้อ 30 ข้อ และ 40 ข้อ ตามลำดับ หลังจากการสุ่ม จะได้กลุ่มที่จะทำการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 9 กลุ่ม ได้แก่ กลุ่มที่หนึ่งแบบทดสอบ 20 ข้อ กลุ่มตัวอย่างขนาด 300 คน กลุ่มที่สองแบบทดสอบ 20 ข้อ กลุ่มตัวอย่างขนาด 500 คน กลุ่มที่สามแบบทดสอบ 20 ข้อ กลุ่มตัวอย่างขนาด 700 คน กลุ่มที่สี่แบบทดสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 300 คน กลุ่มที่ห้าแบบทดสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 500 คน กลุ่มที่หกแบบทดสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 700 คน กลุ่มที่เจ็ดแบบทดสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 300 คน กลุ่มที่แปดแบบทดสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 500 คน กลุ่มที่เก้า แบบทดสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 700 คน จากนั้นทำการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยวิธี Lord 's χ^2 ใช้โปรแกรม IRTDIF Version 1.0 ส่วนวิธี Mantel - Haenszel และวิธี SIBTEST ใช้โปรแกรม SIBTEST ของ Stout และ Rousos (1992) แล้วนำเสนอผลการวิเคราะห์ข้อมูลเฉพาะข้อที่ตรวจสอบพบว่าทำหน้าที่ต่างกัน เท่านั้น แสดงผลดังตาราง 5 - 7

ตาราง 5 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST เมื่อจำนวนข้อสอบ 20 ข้อ กลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ข้อที่	300			500			700		
	LC	MH	SIB	LC	MH	SIB	LC	MH	SIB
1	6.23 *	2.96 ***	.22 ***	6.15 *	3.12 ***	.25 **	6.47 *	3.33 ***	.25 ***
2						.10 *	6.92 *	1.62 **	2.17 **
3				10.55 **	2.98 ***	.23 ***	7.26 *	1.99 ***	.15 ***
4	6.60 *	.38 ***	-.20 ***	5.99 *	.63 *	-.13 **	6.98 *	.32 **	-.11 **
5				6.00 *		.10 *			.08 *
6				6.25 *	1.69 *	.09 *	6.57 *	1.70 **	.11 **
9	5.33 *	.48 *	-.17 *	7.52 *	.32 ***	-.20 ***		.34 ***	-.18 ***
11								.67 *	-.09 *
15	5.07 *	2.44 ***	.16 *	7.21 *	1.87 **	.13 **	8.23 *	1.91 **	.12 **
17	6.50 *	2.83 ***	.218 ***	6.54 *	2.13 **	.16 ***	10.52 **	2.49 ***	.20 ***
18	6.24 *		-.17 *	7.41 *	.54 **	-.12 **	6.06 *	.68 *	-.08 *
19	7.12 *	.36 ***	-.27 ***	6.99 *	.31 **	-.23 ***	9.18 **	.27 ***	-.28 ***
20	6.52 *	.46 *		6.78 *	.38 **	-.22 ***	7.36 *	.34 ***	-.23 ***
รวม	8	7	7	11	10	12	10	12	13

ผลการวิเคราะห์ตามตาราง 5 พบว่า เมื่อกลุ่มตัวอย่าง 300 คน วิธี Lord's χ^2 ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 8 ข้อ วิธี Mantel - Haenszel 7 ข้อ และวิธี SIBTEST 7 ข้อ โดยทั้งสามวิธีตรวจพบข้อที่ทำหน้าที่ต่างกันตรงกัน 6 ข้อ ได้แก่ ข้อ 1, 4, 9, 15, 17, 19 ตามลำดับ ส่วนสองข้อที่เหลือคือ ข้อ 18 วิธี Lord's χ^2 กับ วิธี SIBTEST ตรวจพบตรงกัน และข้อ 20 วิธี Lord's χ^2 กับ วิธี Mantel - Haenszel ตรวจพบตรงกัน

เมื่อกลุ่มตัวอย่าง 500 คน พบว่า วิธี Lord's χ^2 ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 11 ข้อ วิธี Mantel - Haenszel 10 ข้อ และวิธี SIBTEST 12 ข้อ โดยทั้งสามวิธีตรวจพบข้อที่ทำหน้าที่ต่างกันตรงกัน 10 ข้อ ได้แก่ ข้อ 1, 3, 4, 6, 9, 15, 17, 18, 19, 20 ตามลำดับ ส่วนสองข้อที่เหลือคือ ข้อ 2 วิธี SIBTEST ตรวจพบ และข้อ 5 วิธี Lord's χ^2 กับวิธี SIBTEST ตรวจพบตรงกัน

เมื่อกลุ่มตัวอย่าง 700 คน พบว่า วิธี Lord's χ^2 ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 10 ข้อ วิธี Mantel - Haenszel 12 ข้อ และวิธี SIBTEST 13 ข้อ โดยทั้งสามวิธีตรวจพบข้อที่ทำหน้าที่ต่างกันตรงกัน 10 ข้อ ได้แก่ ข้อ 1, 2, 3, 4, 6, 15, 17, 18, 19, 20 ตามลำดับ ส่วนสามข้อที่เหลือคือ ข้อ 5 วิธี SIBTEST ตรวจพบ และข้อ 9 กับ 11 วิธี Mantel - Haenszel กับ วิธี SIBTEST ตรวจพบตรงกัน

ตาราง 6 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST เมื่อจำนวนข้อสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ข้อที่	300			500			700		
	LC	MH	SIB	LC	MH	SIB	LC	MH	SIB
1	6.42*	2.87**	.243**	9.58**	2.53***	.277***	7.59*	2.88***	.239***
2	6.21*		.167*			.104*	6.54*	1.69**	.118**
3				6.72*	2.75***	.201***	6.12*	1.96***	.133**
4			-.159*					.70*	-.079*
6	6.11*	2.03*	.177*	6.11*	1.71*	.094*	6.44*	1.84**	.109**
9				7.23*	.42***	-1.62**	10.62**	.44***	-.153***
15	6.57*	2.27*	.149**		1.60*		6.98*	1.85**	.127**
17	7.25*	3.02**	.183*	6.54*	1.86**	.138**	9.82**	2.40***	.178***
18	6.54*		-.168*	7.23*	.45**	-.193***	6.54*	.61**	-.089*
19	7.89*	.22**	-.261***	7.58*	.18***	-.290***	9.66**	.21***	-.314***
20	6.75*	.33**	-.193**	8.21*	.34***	-.235***	7.85*	.31***	-.245***
21	10.13**	5.10***	.272***	9.69**	3.64***	.246***	9.37**	3.66***	.250***
22				7.54*	.48**	-.170***	6.21*	.55**	-.114**
23	6.48*	.26***	-.188*	7.15*	.27***	-.288***	10.28**	.31***	-.237***
24		.34*	-.205**	6.52*	.50**	-.161***	11.24**	.48***	-.155***
25			-.133*			-.112*			
27					1.70*	.125*			
28	6.21*	2.04*	.174*	7.53*	2.14***	.205***	9.82**	2.16***	.200***
30				6.25*	1.83**	.109*			
รวม	11	10	14	14	16	18	15	16	16

ผลการวิเคราะห์ตามตาราง 6 พบว่า เมื่อกลุ่มตัวอย่าง 300 คน วิธี Lord's χ^2 ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 11 ข้อ วิธี Mantel - Haenszel 10 ข้อ และวิธี SIBTEST 14 ข้อ โดยทั้งสามวิธีตรวจพบข้อที่ทำหน้าที่ต่างกันตรงกัน 9 ข้อ ได้แก่ ข้อ 1, 6, 15, 17, 19, 20, 21, 23, 28 ตามลำดับ ส่วนหัวข้อที่เหลือคือ ข้อ 2 กับ 18 วิธี Lord's χ^2 กับ วิธี SIBTEST ตรวจพบตรงกัน และข้อ 24 วิธี Mantel - Haenszel กับ วิธี SIBTEST ตรวจพบตรงกัน และ ข้อ 4 กับ 25 วิธี SIBTEST ตรวจพบ

เมื่อกลุ่มตัวอย่าง 500 คน พบว่า วิธี Lord's χ^2 ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 14 ข้อ วิธี Mantel - Haenszel 16 ข้อ และวิธี SIBTEST 18 ข้อ โดยทั้งสามวิธีตรวจพบข้อที่ทำหน้าที่ต่างกันตรงกัน 14 ข้อ ได้แก่ ข้อ 1, 3, 6, 9, 17, 18, 19, 20, 21, 22, 23, 24, 28, 30 ตามลำดับ

ส่วนข้อที่เหลือคือ ข้อ 2 กับ 25 วิธี SIBTEST ตรวจพบ ข้อ 15 วิธี Mantel – Haenszel ตรวจพบ และข้อ 27 วิธี Mantel – Haenszel กับวิธี SIBTEST ตรวจพบตรงกัน

เมื่อกุ่มตัวอย่าง 700 คน พบว่า วิธี Lord's χ^2 ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 15 ข้อ วิธี Mantel - Haenszel 16 ข้อ และวิธี SIBTEST 16 ข้อ โดยทั้งสามวิธีตรวจพบข้อที่ทำหน้าที่ต่างกันตรงกัน 15 ข้อ ได้แก่ ข้อ 1, 2, 3, 6, 9, 15, 17, 18, 19, 20, 21, 22, 23, 24, 28 ตามลำดับ ส่วนหนึ่งข้อที่เหลือคือ ข้อ 4 วิธี Mantel – Haenszel กับ วิธี SIBTEST ตรวจพบตรงกัน

ตาราง 7 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST เมื่อจำนวนข้อสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ข้อที่	300			500			700		
	LC	MH	SIB	LC	MH	SIB	LC	MH	SIB
1	7.52*	2.69**	.194*	7.25*	2.70***	.226**	10.22**	2.93***	.224***
2									.082*
3				6.12*	2.62***	.174***	7.48*	1.83**	.113***
4	6.62*	.37**	-.276***	6.62*	.57**	-.132**	6.13*	.59**	-.116***
6					1.65*		6.14*	1.66*	.084*
9				7.62*	.31***	-.244***	7.62*	.33***	-.219***
15	6.98**	2.30**	.138*	6.95**	1.59*		6.16**	1.75**	.099**
17	9.42**	3.45***	.255**	9.12**	2.19***	.158**	9.78**	2.65***	.214***
18	6.74*		-.205*	6.34*	.53**	-.153**	6.29*	.68*	
19	9.16**	.29***	-.268***	9.37**	.25***	-.245***	9.06**	.24***	-.295***
20	7.25*	.39**	-.189**	7.32*	.36***	-.223***	7.11*	.33***	-.252***
21	9.17**	4.32***	.270***	9.86**	3.38***	.255***	8.24**	3.18***	.262***
22	6.11*		-.098	6.41*	.53**	-.146**	6.03*	.59**	-.123**
23	10.48*	.36**	-.219**	10.08**	.31***	-.260***	10.19**	.38***	-.201***
24	7.24*	.40*	-.173*	7.02*	.52**	-.143**	6.21*	.45***	-.147***
25			-.165*						
27					1.73*	.138**			
28	9.42**	2.20*	.225**	9.32**	2.20***	.232***	8.23**	2.15***	.207***
30					1.56*	.135**			
31	6.52*	2.69**	.194*	6.31*	2.70***	.226***		2.93***	.224***
32									.082*
33				6.13**	2.62***	.174***		1.83**	.113**
34	6.98*	.37**	-.276***	6.08*	.57**	-.132**		.59**	-.116**
36					1.65*			1.66*	.084*
39	6.02*	.47*		6.82*	.31***	-.244***	6.98*	.33***	-.219***
รวม	15	14	14	18	22	19	16	21	21

ผลการวิเคราะห์ตามตาราง 7 พบว่า เมื่อกลุ่มตัวอย่าง 300 คนวิธี Lord's χ^2 ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 15 ข้อ วิธี Mantel - Haenszel 14 ข้อ และวิธี SIBTEST 14 ข้อ โดยทั้งสามวิธีตรวจพบข้อที่ทำหน้าที่ต่างกันตรงกัน 12 ข้อ ได้แก่ ข้อ 1, 4, 15, 17, 19, 20, 21, 23, 24, 28, 31, 34 ตามลำดับ ส่วนสี่ข้อที่เหลือคือ ข้อ 18 กับ 22 วิธี Lord's χ^2 กับ วิธี SIBTEST ตรวจพบตรงกัน ข้อ 25 วิธี SIBTEST ตรวจพบ และข้อ 39 วิธี Lord's χ^2 กับ วิธี Mantel - Haenszel ตรวจพบตรงกัน

เมื่อกลุ่มตัวอย่าง 500 คน พบว่า วิธี Lord's χ^2 ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 18 ข้อ วิธี Mantel - Haenszel 22 ข้อ และวิธี SIBTEST 19 ข้อ โดยทั้งสามวิธีตรวจพบข้อที่ทำหน้าที่ต่างกันตรงกัน 17 ข้อ ได้แก่ ข้อ 1, 3, 4, 9, 17, 18, 19, 20, 21, 22, 23, 24, 28, 31, 33, 34, 39 ตามลำดับ ส่วนห้าข้อที่เหลือคือ ข้อ 6 กับ 36 วิธี Mantel - Haenszel ตรวจพบ ข้อ 15 วิธี Lord's χ^2

กับ วิธี Mantel – Haenszel ตรวจพบตรงกัน และข้อ 27 กับ 30 วิธี Mantel – Haenszel กับวิธี SIBTEST ตรวจพบตรงกัน

เมื่อกลุ่มตัวอย่าง 700 คน พบว่า วิธี Lord 's χ^2 ตรวจพบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน 16 ข้อ วิธี Mantel - Haenszel 21 ข้อ และวิธี SIBTEST 21 ข้อ โดยทั้งสามวิธีตรวจพบข้อที่ทำหน้าที่ต่างกันตรงกัน 15 ข้อ ได้แก่ ข้อ 1, 3, 4, 6, 9, 15, 17, 19, 20, 21, 22, 23, 24, 28, 39 ตามลำดับ ส่วน เจ็ดข้อที่เหลือคือ ข้อ 2 กับ 32 วิธี SIBTEST ตรวจพบ ข้อ 18 วิธี Lord 's χ^2 กับ วิธี Mantel – Haenszel ตรวจพบตรงกัน และข้อ 31, 33, 34, 36 วิธี Mantel – Haenszel กับวิธี SIBTEST ตรวจพบตรงกัน

ตอนที่ 3 ผลการทดสอบความแตกต่างของการตรวจสอบการทำหน้าที่ต่างกันของ
ข้อสอบ ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

การวิเคราะห์ตอนที่ 3 ผู้วิจัยได้นำผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST ที่ตรวจพบในตอนที่ 2 มาทดสอบความแตกต่าง ซึ่งสถิติที่ใช้ทดสอบคือ Cochran Q - Test (Siegel. 1956 : 162) วิเคราะห์โดยใช้โปรแกรมสำเร็จรูป Microsoft Excel 2000 โดยแบ่งการทดสอบเป็นสองกรณีตามสมมติฐานการวิจัย ตามลำดับดังนี้

3.1 ผลการทดสอบความแตกต่างของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
เมื่อกลุ่มตัวอย่างเท่ากัน วิธีการตรวจสอบต่างกัน แสดงผลดังตาราง 8 - 20

ตาราง 8 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 20 ข้อ
กลุ่มตัวอย่าง 300 คน ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

ข้อสอบที่ถูก ตรวจพบว่า (ข้อ)	LC	MH	SIB	Q
ทำหน้าที่ต่างกัน	8	7	7	1.00
ไม่ทำหน้าที่ต่างกัน	12	13	13	

ผลการวิเคราะห์ตามตาราง 8 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Lord 's χ^2 ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 8 ข้อ ไม่ทำหน้าที่ต่างกัน 12 ข้อ วิธี Mantel - Haenszel ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 7 ข้อ ไม่ทำหน้าที่ต่างกัน 13 ข้อ และวิธี SIBTEST ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 7 ข้อ ไม่ทำหน้าที่ต่างกัน 13 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจาก 3 วิธี ไม่แตกต่างกัน

ตาราง 9 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 20 ข้อ

กลุ่มตัวอย่าง 500 คน ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

ข้อสอบที่ถูก ตรวจพบว่า (ข้อ)	LC	MH	SIB	Q
ทำหน้าที่ต่างกัน	11	10	12	3.00
ไม่ทำหน้าที่ต่างกัน	9	10	8	

ผลการวิเคราะห์ตามตาราง 9 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Lord 's χ^2 ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 11 ข้อ ไม่ทำหน้าที่ต่างกัน 9 ข้อ วิธี Mantel - Haenszel ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 10 ข้อ ไม่ทำหน้าที่ต่างกัน 10 ข้อ และวิธี SIBTEST ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 12 ข้อ ไม่ทำหน้าที่ต่างกัน 8 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจาก 3 วิธี ไม่แตกต่างกัน

ตาราง 10 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 20 ข้อ

กลุ่มตัวอย่าง 700 คน ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

ข้อสอบที่ถูก ตรวจพบว่า (ข้อ)	LC	MH	SIB	Q
ทำหน้าที่ต่างกัน	10	12	13	4.67
ไม่ทำหน้าที่ต่างกัน	10	8	7	

ผลการวิเคราะห์ตามตาราง 10 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Lord 's χ^2 ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 10 ข้อ ไม่ทำหน้าที่ต่างกัน 10 ข้อ วิธี Mantel - Haenszel ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 12 ข้อ ไม่ทำหน้าที่ต่างกัน 8 ข้อ และวิธี SIBTEST ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 13 ข้อ ไม่ทำหน้าที่ต่างกัน 7 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจาก 3 วิธี ไม่แตกต่างกัน

ตาราง 11 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 30 ข้อ
กลุ่มตัวอย่าง 300 คน ระหว่างวิธี Lord ' s χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

ข้อสอบที่ถูก ตรวจพบว่า (ข้อ)	LC	MH	SIB	Q
ทำหน้าที่ต่างกัน	11	10	14	5.20*
ไม่ทำหน้าที่ต่างกัน	19	20	16	

ผลการวิเคราะห์ตามตาราง 11 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Lord ' s χ^2 ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 11 ข้อ ไม่ทำหน้าที่ต่างกัน 19 ข้อ วิธี Mantel - Haenszel ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 10 ข้อ ไม่ทำหน้าที่ต่างกัน 20 ข้อ และวิธี SIBTEST ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 14 ข้อ ไม่ทำหน้าที่ต่างกัน 16 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจาก 3 วิธี แตกต่างกัน และได้ทำการเปรียบเทียบความแตกต่างของจำนวนข้อสอบที่ตรวจพบเป็นรายคู่ แสดงผลดังตาราง 12

ตาราง 12 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่ เมื่อจำนวนข้อสอบ 30 ข้อ
กลุ่มตัวอย่าง 300 คน ระหว่างวิธี Lord ' s χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

วิธีตรวจสอบ	การทดสอบแบบคี่เป็นรายคู่ (Q-test)		
	LC	MH	SIB
LC	-	0.33	3.00
MH		-	4.00*
SIB			-

การวิเคราะห์ตามตาราง 12 พบว่า เมื่อนำจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ จากข้อสอบจำนวน 20 ข้อกับกลุ่มตัวอย่าง 300 คนมาทดสอบความแตกต่างเป็นรายคู่ด้วยสูตร Q - test พบว่าจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบโดยวิธี Mantel - Haenszel กับวิธี SIBTEST แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนสองคู่ที่เหลือคือ วิธี Lord ' s χ^2 กับวิธี Mantel - Haenszel และ วิธี Lord ' s χ^2 กับวิธี SIBTEST แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

ตาราง 13 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 30 ข้อ
กลุ่มตัวอย่าง 500 คน ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

ข้อสอบที่ถูก ตรวจพบว่า (ข้อ)	LC	MH	SIB	Q
ทำหน้าที่ต่างกัน	14	16	18	6.00*
ไม่ทำหน้าที่ต่างกัน	16	14	12	

ผลการวิเคราะห์ตามตาราง 13 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Lord 's χ^2 ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 14 ข้อ ไม่ทำหน้าที่ต่างกัน 16 ข้อ วิธี Mantel - Haenszel ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 16 ข้อ ไม่ทำหน้าที่ต่างกัน 14 ข้อ และวิธี SIBTEST ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 18 ข้อ ไม่ทำหน้าที่ต่างกัน 12 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจาก 3 วิธี แตกต่างกัน และได้ทำการเปรียบเทียบความแตกต่างของจำนวนข้อสอบที่ตรวจพบเป็นรายคู่ แสดงผลดังตาราง 14

ตาราง 14 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่ เมื่อจำนวนข้อสอบ 30 ข้อ
กลุ่มตัวอย่าง 500 คน ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

วิธีตรวจสอบ	การทดสอบแบบคิเป็นรายคู่ (Q-test)		
	LC	MH	SIB
LC	-	2.00	4.00*
MH		-	2.00
SIB			-

การวิเคราะห์ตามตาราง 14 พบว่า เมื่อนำจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ จากข้อสอบจำนวน 30 ข้อกับกลุ่มตัวอย่าง 500 คนมาทดสอบความแตกต่างเป็นรายคู่ด้วยสูตร Q - test พบว่าจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบโดย วิธี Lord 's χ^2 กับวิธี SIBTEST แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนสองคู่ที่เหลือคือ วิธี Lord 's χ^2 กับวิธี Mantel - Haenszel และ วิธี Mantel - Haenszel กับวิธี SIBTEST แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

ตาราง 15 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 30 ข้อ
กลุ่มตัวอย่าง 700 คน ระหว่างวิธี Lord ' s χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

ข้อสอบที่ถูก ตรวจพบว่า (ข้อ)	LC	MH	SIB	Q
ทำหน้าที่ต่างกัน	15	16	16	2.00
ไม่ทำหน้าที่ต่างกัน	15	14	14	

ผลการวิเคราะห์ตามตาราง 15 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Lord ' s χ^2 ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 15 ข้อ ไม่ทำหน้าที่ต่างกัน 15 ข้อ วิธี Mantel - Haenszel ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 16 ข้อ ไม่ทำหน้าที่ต่างกัน 14 ข้อ และวิธี SIBTEST ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 16 ข้อ ไม่ทำหน้าที่ต่างกัน 14 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจาก 3 วิธี ไม่แตกต่างกัน

ตาราง 16 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 40 ข้อ
กลุ่มตัวอย่าง 300 คน ระหว่างวิธี Lord ' s χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

ข้อสอบที่ถูก ตรวจพบว่า (ข้อ)	LC	MH	SIB	Q
ทำหน้าที่ต่างกัน	15	14	14	0.40
ไม่ทำหน้าที่ต่างกัน	25	26	26	

ผลการวิเคราะห์ตามตาราง 16 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Lord ' s χ^2 ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 15 ข้อ ไม่ทำหน้าที่ต่างกัน 25 ข้อ วิธี Mantel - Haenszel ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 14 ข้อ ไม่ทำหน้าที่ต่างกัน 26 ข้อ และวิธี SIBTEST ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 14 ข้อ ไม่ทำหน้าที่ต่างกัน 26 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจาก 3 วิธี ไม่แตกต่างกัน

ตาราง 17 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 40 ข้อ
กลุ่มตัวอย่าง 500 คน ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

ข้อสอบที่ถูก ตรวจพบว่า (ข้อ)	LC	MH	SIB	Q
ทำหน้าที่ต่างกัน	18	22	19	6.50*
ไม่ทำหน้าที่ต่างกัน	22	18	21	

ผลการวิเคราะห์ตามตาราง 17 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Lord 's χ^2 ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 18 ข้อ ไม่ทำหน้าที่ต่างกัน 22 ข้อ วิธี Mantel - Haenszel ทำหน้าที่ต่างกัน 22 ข้อ ไม่ทำหน้าที่ต่างกัน 18 ข้อ และวิธี SIBTEST ทำหน้าที่ต่างกัน 19 ข้อ ไม่ทำหน้าที่ต่างกัน 21 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างมี นัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจาก 3 วิธี แตกต่างกัน และได้ทำการเปรียบเทียบความแตกต่างของจำนวนข้อสอบที่ตรวจพบเป็นรายคู่ แสดงผลดังตาราง 18

ตาราง 18 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่ เมื่อจำนวนข้อสอบ 40 ข้อ
กลุ่มตัวอย่าง 500 คน ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

วิธีตรวจสอบ	การทดสอบแบบคิวเป็นรายคู่ (Q-test)		
	LC	MH	SIB
LC	-	4.00*	1.00
MH		-	3.00
SIB			-

การวิเคราะห์ตามตาราง 18 พบว่า เมื่อนำจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ จากข้อสอบจำนวน 40 ข้อกับกลุ่มตัวอย่าง 500 คนมาทดสอบความแตกต่างเป็นรายคู่ด้วยสูตร Q - test พบว่าจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบโดยวิธี Lord 's χ^2 กับวิธี Mantel - Haenszel แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนสองคู่ที่เหลือคือวิธี Lord 's χ^2 กับวิธี SIBTEST และวิธี Mantel - Haenszel กับวิธี SIBTEST แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

ตาราง 19 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 40 ข้อ
กลุ่มตัวอย่าง 700 คน ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

ข้อสอบที่ถูก ตรวจพบว่า (ข้อ)	LC	MH	SIB	Q
ทำหน้าที่ต่างกัน	16	21	21	6.25*
ไม่ทำหน้าที่ต่างกัน	24	19	19	

ผลการวิเคราะห์ตามตาราง 19 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Lord 's χ^2 ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 16 ข้อ ไม่ทำหน้าที่ต่างกัน 24 ข้อ วิธี Mantel - Haenszel ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 21 ข้อ ไม่ทำหน้าที่ต่างกัน 19 ข้อ และวิธี SIBTEST ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 21 ข้อ ไม่ทำหน้าที่ต่างกัน 19 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจาก 3 วิธี แตกต่างกัน และได้ทำการเปรียบเทียบความแตกต่างของจำนวนข้อสอบที่ตรวจพบเป็นรายคู่ แสดงผลดังตาราง 20

ตาราง 20 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่ เมื่อจำนวนข้อสอบ 40 ข้อ
กลุ่มตัวอย่าง 700 คน ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

วิธีตรวจสอบ	การทดสอบแบบคี่เป็นรายคู่ (Q-test)		
	LC	MH	SIB
LC	-	5.00*	3.57
MH		-	0.00
SIB			-

การวิเคราะห์ตามตาราง 18 พบว่า เมื่อนำจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ จากข้อสอบจำนวน 40 ข้อกับกลุ่มตัวอย่าง 700 คนมาทดสอบความแตกต่างเป็นรายคู่ด้วยสูตร Q - test แล้วพบว่าจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบโดย วิธี Lord 's χ^2 กับวิธี Mantel - Haenszel แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนสองคู่ที่เหลือคือวิธี Lord 's χ^2 กับวิธี SIBTEST และวิธี Mantel - Haenszel กับวิธี SIBTEST แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

3.2 ผลการทดสอบความแตกต่างของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
เมื่อกลุ่มตัวอย่างต่างกัน วิธีการตรวจสอบเหมือนกัน แสดงผลดังตาราง 21 - 34

ตาราง 21 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 20 ข้อ
ตรวจสอบด้วยวิธี Lord ' s χ^2 ระหว่างกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ข้อสอบที่ถูก ตรวจพบว่า (ข้อ)	300	500	700	Q
ทำหน้าที่ต่างกัน	8	11	10	2.80
ไม่ทำหน้าที่ต่างกัน	12	9	10	

ผลการวิเคราะห์ตามตาราง 21 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Lord ' s χ^2 จากกลุ่มตัวอย่างขนาด 300 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 8 ข้อ ไม่ทำหน้าที่ต่างกัน 12 ข้อ กลุ่มตัวอย่างขนาด 500 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 11 ข้อ ไม่ทำหน้าที่ต่างกัน 9 ข้อ และกลุ่มตัวอย่างขนาด 700 คนตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 10 ข้อ ไม่ทำหน้าที่ต่างกัน 10 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 3 ขนาด ไม่แตกต่างกัน

ตาราง 22 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 20 ข้อ
ตรวจสอบด้วย วิธี Mantel - Haenszel ระหว่างกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ข้อสอบที่ถูก ตรวจพบว่า (ข้อ)	300	500	700	Q
ทำหน้าที่ต่างกัน	7	10	12	7.60*
ไม่ทำหน้าที่ต่างกัน	13	10	8	

ผลการวิเคราะห์ตามตาราง 22 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Mantel - Haenszel จากกลุ่มตัวอย่างขนาด 300 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 7 ข้อ ไม่ทำหน้าที่ต่างกัน 13 ข้อ กลุ่มตัวอย่างขนาด 500 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 10 ข้อ ไม่ทำหน้าที่ต่างกัน 10 ข้อ และกลุ่มตัวอย่างขนาด 700 คนตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 12 ข้อ ไม่ทำหน้าที่ต่างกัน 8 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 3 ขนาด แตกต่างกัน และได้ทำการเปรียบเทียบความแตกต่างของจำนวนข้อสอบที่ตรวจพบเป็นรายคู่ แสดงผลดังตาราง 23

ตาราง 23 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่ เมื่อจำนวนข้อสอบ 20 ข้อ
วิธี Mantel - Haenszel ระหว่างกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ขนาดกลุ่มตัวอย่าง	การทดสอบแบบคิวเป็นรายคู่ (Q-test)		
	300	500	700
300	-	3.00	5.00*
500		-	2.00
700			-

ผลการวิเคราะห์ตามตาราง 23 พบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ จากกลุ่มตัวอย่างขนาด 300 คนกับ 700 คน แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนสองคู่ที่เหลือคือ จากกลุ่มตัวอย่างขนาด 300 คนกับ 500 คน และกลุ่มตัวอย่างขนาด 500 คนกับ 700 คน แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ.05

ตาราง 24 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 20 ข้อ
ตรวจสอบด้วยวิธี SIBTEST ระหว่างกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ข้อสอบที่ถูกตรวจพบว่า (ข้อ)	300	500	700	Q
ทำหน้าที่ต่างกัน	7	12	13	10.33*
ไม่ทำหน้าที่ต่างกัน	13	8	7	

ผลการวิเคราะห์ตามตาราง 24 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี SIBTESTจากกลุ่มตัวอย่างขนาด 300 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 7 ข้อ ไม่ทำหน้าที่ต่างกัน 13 ข้อ กลุ่มตัวอย่างขนาด 500 คน ตรวจพบข้อสอบที่ ทำหน้าที่ต่างกัน 12 ข้อ ไม่ทำหน้าที่ต่างกัน 8 ข้อ และกลุ่มตัวอย่างขนาด 700 คน ตรวจพบข้อสอบที่ ทำหน้าที่ต่างกัน 13 ข้อ ไม่ทำหน้าที่ต่างกัน 7 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่า จำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 3 ขนาด แตกต่างกัน และได้ทำการเปรียบเทียบความแตกต่างของจำนวนข้อสอบที่ตรวจพบเป็นรายคู่ แสดงผลดังตาราง 25

ตาราง 25 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่ เมื่อจำนวนข้อสอบ 20 ข้อ
วิธี SIBTEST ระหว่างกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ขนาดกลุ่มตัวอย่าง	การทดสอบแบบคิวเป็นรายคู่ (Q-test)		
	300	500	700
300	-	5.00*	6.00*
500		-	1.00
700			-

ผลการวิเคราะห์ตามตาราง 25 พบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ จากกลุ่มตัวอย่างขนาด 300 คนกับ 700 คน และกลุ่มตัวอย่างขนาด 300 คนกับ 500 คน แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนหนึ่งคู่ที่เหลือคือ กลุ่มตัวอย่างขนาด 500 คนกับ 700 คน แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

ตาราง 26 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 30 ข้อ
ตรวจสอบด้วยวิธี Lord 's χ^2 ระหว่างกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ข้อสอบที่ถูกตรวจพบว่า (ข้อ)	300	500	700	Q
ทำหน้าที่ต่างกัน	11	14	15	3.71
ไม่ทำหน้าที่ต่างกัน	19	16	15	

ผลการวิเคราะห์ตามตาราง 26 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Lord 's χ^2 จากกลุ่มตัวอย่างขนาด 300 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 11 ข้อ ไม่ทำหน้าที่ต่างกัน 19 ข้อ กลุ่มตัวอย่างขนาด 500 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 14 ข้อ ไม่ทำหน้าที่ต่างกัน 16 ข้อ และกลุ่มตัวอย่างขนาด 700 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 15 ข้อ ไม่ทำหน้าที่ต่างกัน 15 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่า จำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 3 ขนาด ไม่แตกต่างกัน

ตาราง 27 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 30 ข้อ
ตรวจสอบด้วย วิธี Mantel - Haenszel ระหว่างกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ข้อสอบที่ถูก ตรวจพบว่า (ข้อ)	300	500	700	Q
ทำหน้าที่ต่างกัน	10	16	16	9.00*
ไม่ทำหน้าที่ต่างกัน	20	14	14	

ผลการวิเคราะห์ตามตาราง 27 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Mantel - Haenszel จากกลุ่มตัวอย่างขนาด 300 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 10 ข้อ ไม่ทำหน้าที่ต่างกัน 20 ข้อ กลุ่มตัวอย่างขนาด 500 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 16 ข้อ ไม่ทำหน้าที่ต่างกัน 14 ข้อ และ กลุ่มตัวอย่างขนาด 700 คนตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 16 ข้อ ไม่ทำหน้าที่ต่างกัน 14 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 3 ขนาด แตกต่างกันได้ และได้ทำการเปรียบเทียบความแตกต่างของจำนวนข้อสอบที่ตรวจพบเป็นรายคู่ แสดงผลดังตาราง 28

ตาราง 28 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่ เมื่อจำนวนข้อสอบ 30 ข้อ
วิธี Mantel - Haenszel ระหว่างกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ขนาดกลุ่มตัวอย่าง	การทดสอบแบบคี่เป็นรายคู่ (Q-test)		
	300	500	700
300	-	6.00*	6.00*
500		-	0.00
700			-

ผลการวิเคราะห์ตามตาราง 28 พบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ จากกลุ่มตัวอย่างขนาด 300 คนกับ 500 คน และกลุ่มตัวอย่างขนาด 300 คนกับ 700 คน แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนหนึ่งคู่ที่เหลือคือ กลุ่มตัวอย่างขนาด 500 คนกับ 700 คน แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ.05

ตาราง 29 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 30 ข้อ
ตรวจสอบด้วยวิธี SIBTEST ระหว่างกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ข้อสอบที่ถูกตรวจพบว่า (ข้อ)	300	500	700	Q
ทำหน้าที่ต่างกัน	14	18	16	3.42
ไม่ทำหน้าที่ต่างกัน	16	12	14	

ผลการวิเคราะห์ตามตาราง 29 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี SIBTESTจากกลุ่มตัวอย่างขนาด 300 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 14 ข้อ ไม่ทำหน้าที่ต่างกัน 16 ข้อ กลุ่มตัวอย่างขนาด 500 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 18 ข้อ ไม่ทำหน้าที่ต่างกัน 12 ข้อ และกลุ่มตัวอย่างขนาด 700 คนตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 16 ข้อ ไม่ทำหน้าที่ต่างกัน 14 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 3 ขนาด ไม่แตกต่างกัน

ตาราง 30 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 40 ข้อ
ตรวจสอบด้วยวิธี Lord ' s χ^2 ระหว่างกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ข้อสอบที่ถูกตรวจพบว่า (ข้อ)	300	500	700	Q
ทำหน้าที่ต่างกัน	15	18	16	2.33
ไม่ทำหน้าที่ต่างกัน	25	22	24	

ผลการวิเคราะห์ตามตาราง 30 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Lord ' s χ^2 จากกลุ่มตัวอย่างขนาด 300 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 15 ข้อ ไม่ทำหน้าที่ต่างกัน 25 ข้อ กลุ่มตัวอย่างขนาด 500 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 18 ข้อ ไม่ทำหน้าที่ต่างกัน 22 ข้อ และกลุ่มตัวอย่างขนาด 700 คนตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 16 ข้อ ไม่ทำหน้าที่ต่างกัน 24 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 3 ขนาด ไม่แตกต่างกัน

ตาราง 31 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 40 ข้อ
ตรวจสอบด้วยวิธี Mantel - Haenszel ระหว่างกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ข้อสอบที่ถูกตรวจพบว่า (ข้อ)	300	500	700	Q
ทำหน้าที่ต่างกัน	14	22	21	12.67**
ไม่ทำหน้าที่ต่างกัน	26	18	19	

ผลการวิเคราะห์ตามตาราง 31 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี Mantel - Haenszel จากกลุ่มตัวอย่างขนาด 300 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 14 ข้อ ไม่ทำหน้าที่ต่างกัน 26 ข้อ กลุ่มตัวอย่างขนาด 500 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 22 ข้อ ไม่ทำหน้าที่ต่างกัน 18 ข้อ และกลุ่มตัวอย่างขนาด 700 คนตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 21 ข้อ ไม่ทำหน้าที่ต่างกัน 19 ข้อ เมื่อทดสอบความแตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 3 ขนาด แตกต่างกัน และได้ทำการเปรียบเทียบความแตกต่างของจำนวนข้อสอบที่ตรวจพบเป็นรายคู่ แสดงผลดังตาราง 32

ตาราง 32 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่ เมื่อจำนวนข้อสอบ 40 ข้อ
วิธี Mantel - Haenszel ระหว่างกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ขนาดกลุ่มตัวอย่าง	การทดสอบแบบคิวนเป็นรายคู่ (Q-test)		
	300	500	700
300	-	8.00*	7.00*
500		-	0.33
700			-

ผลการวิเคราะห์ตามตาราง 32 พบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ จากกลุ่มตัวอย่างขนาด 300 คนกับ 500 คน และกลุ่มตัวอย่างขนาด 300 คนกับ 700 คน แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนหนึ่งคู่ที่เหลือคือ กลุ่มตัวอย่างขนาด 500 คนกับ 700 คน แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ.05

ตาราง 33 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อจำนวนข้อสอบ 40 ข้อ
ตรวจสอบด้วยวิธี SIBTEST ระหว่างกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ข้อสอบที่ถูก ตรวจพบว่า (ข้อ)	300	500	700	Q
ทำหน้าที่ต่างกัน	14	19	21	6.50*
ไม่ทำหน้าที่ต่างกัน	26	21	19	

ผลการวิเคราะห์ตามตาราง 33 พบว่า จำนวนข้อสอบที่ตรวจสอบด้วยวิธี SIBTESTจากกลุ่มตัวอย่าง
ขนาด 300 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 14 ข้อ ไม่ทำหน้าที่ต่างกัน 26 ข้อ กลุ่มตัวอย่างขนาด
500 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 19 ข้อ ไม่ทำหน้าที่ต่างกัน 21 ข้อ และกลุ่มตัวอย่างขนาด
700 คน ตรวจพบข้อสอบที่ทำหน้าที่ต่างกัน 21 ข้อ ไม่ทำหน้าที่ต่างกัน 19 ข้อ เมื่อทดสอบความ
แตกต่างของจำนวนข้อสอบที่ทำหน้าที่ต่างกันด้วยสูตร Q - test พบว่า แตกต่างกันอย่างมีนัยสำคัญทาง
สถิติที่ระดับ .05 แสดงว่าจำนวนข้อสอบที่ตรวจพบจากกลุ่มตัวอย่าง 3 ขนาด แตกต่างกัน และได้ทำการ
เปรียบเทียบความแตกต่างของจำนวนข้อสอบที่ตรวจพบเป็นรายคู่ แสดงผลดังตาราง 34

ตาราง 34 เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเป็นรายคู่ เมื่อจำนวนข้อสอบ 40 ข้อ
วิธี SIBTEST ระหว่างกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน

ขนาดกลุ่มตัวอย่าง	การทดสอบแบบคี่เป็นรายคู่ (Q-test)		
	300	500	700
300	-	3.57	4.45*
500		-	0.66
700			-

ผลการวิเคราะห์ตามตาราง 34 พบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบจากกลุ่มตัวอย่าง
ขนาด 300 คนกับ 700 คน แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนสองคู่ที่เหลือคือ จาก
กลุ่มตัวอย่างขนาด 300 คนกับ 500 คน และกลุ่มตัวอย่างขนาด 500 คนกับ 700 คน แตกต่างกัน
อย่างไม่มีนัยสำคัญทางสถิติที่ระดับ.05

ตอนที่ 4 ความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST

การวิเคราะห์ตอนที่ 4 ผู้วิจัยได้นำผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ระหว่างวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST จากตอนที่ 2 มาหาค่าความสอดคล้องของทั้งสามวิธี โดยหารจำนวนข้อที่ตรงกันในการตัดสินด้วยจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในการเปรียบเทียบแต่ละคู่แล้วคูณด้วย 100 (กาญจนา วันสุนทร. 2537 : 104) แสดงผลดังตาราง 35 - 36

ตาราง 35 ร้อยละของความสอดคล้องในการตรวจสอบเมื่อกลุ่มตัวอย่างเท่ากัน วิธีตรวจสอบแตกต่างกัน

I	N	N	300			500			700		
		วิธี	LC	MH	SIB	LC	MH	SIB	LC	MH	SIB
20	300	LC	-	87.50	87.50						
		MH		-	75.00						
		SIB			-						
	500	LC				-	90.90	91.67			
		MH					-	83.33			
		SIB						-			
	700	LC							-	83.33	76.92
		MH								-	92.31
		SIB									-
30	300	LC	-	75.00	78.57						
		MH		-	71.43						
		SIB			-						
	500	LC				-	87.50	77.78			
		MH					-	88.89			
		SIB						-			
	700	LC							-	93.75	93.75
		MH								-	100.00
		SIB									-
40	300	LC	-	81.25	81.25						
		MH		-	75.00						
		SIB			-						
	500	LC				-	81.82	94.74			
		MH					-	86.36			
		SIB						-			
	700	LC							-	76.19	68.18
		MH								-	82.61
		SIB									-

ผลการวิเคราะห์ตามตาราง 35 พบว่า ความสอดคล้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ อยู่ระหว่างร้อยละ 71.43 ถึง ร้อยละ 100

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

จุดมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อเปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกัน ระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST เมื่อขนาดกลุ่มตัวอย่าง และจำนวนข้อสอบในแบบทดสอบ ต่างกัน
2. เพื่อศึกษาความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน ระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST เมื่อขนาดกลุ่มตัวอย่าง และจำนวนข้อสอบในแบบทดสอบ ต่างกัน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้า ได้แก่ นักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2543 สังกัดกรมสามัญศึกษา จังหวัดกรุงเทพมหานคร โดยการสุ่มแบบสองขั้นตอน (Two Stage Random Sampling) โดยสุ่มโรงเรียนตามขนาดโรงเรียน ด้วยวิธีการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) ได้ทั้งหมด 10 โรงเรียน จำนวน 113 ห้องเรียน แบ่งเป็น นักเรียนชาย 378 คน และนักเรียนหญิง 412 คน

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้า ได้แก่ แบบทดสอบความถนัดทางคณิตศาสตร์ ซึ่งสร้างตามแนวทางทฤษฎีของเทอร์สโตน เป็นแบบปรนัยชนิดเลือกตอบ 5 ตัวเลือก จำนวน 60 ข้อ ที่ผู้วิจัยสร้างขึ้น

วิธีดำเนินการรวบรวมข้อมูล

1. ขอลหนังสือจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อขอความอนุเคราะห์ในการเก็บรวบรวมข้อมูล
2. ติดต่อโรงเรียนที่ใช้เป็นกลุ่มตัวอย่าง ขออนุญาตผู้บริหารโรงเรียน เพื่อกำหนด วัน เวลา ที่ทำการทดสอบ
3. วางแผนในการดำเนินการสอบ โดยผู้วิจัยเป็นผู้ดำเนินการสอบเอง จัดเตรียมแบบทดสอบให้เพียงพอ กับจำนวนนักเรียนในแต่ละครั้ง
4. นำแบบทดสอบไปทดสอบกับนักเรียนที่เป็นกลุ่มตัวอย่างทั้งหมด ตามวัน เวลา ที่นัดหมาย
5. ชี้แจงให้นักเรียนในกลุ่มตัวอย่างเข้าใจ และทราบถึงวัตถุประสงค์ในการสอบและขอความร่วมมือในการสอบ
6. อธิบายให้นักเรียนทุกคนเข้าใจวิธีการทำแบบทดสอบ
7. หลังจากดำเนินการสอบกลุ่มตัวอย่างครบทุกโรงเรียนแล้วนำกระดาษคำตอบมาตรวจให้คะแนน ดังนี้ ถ้าตอบถูกให้ 1 คะแนน ถ้าตอบผิดหรือว่างไว้ให้ 0 คะแนน และนำผลมาวิเคราะห์ทางสถิติ

สรุปผลการวิจัย

ผลการวิจัยสรุปได้ดังต่อไปนี้

1. จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจสอบพบ จากแบบทดสอบ 3 ชุด คือ แบบทดสอบที่มีจำนวน 20 ข้อ 30 ข้อ และ 40 ข้อ กับกลุ่มตัวอย่าง 3 ขนาด คือ กลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน ตรวจสอบด้วยวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST สรุปได้ดังตาราง 37

ตาราง 37 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจสอบพบจากวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST จากจำนวนข้อสอบ 3 ชุด และกลุ่มตัวอย่าง 3 ขนาด

จำนวนข้อสอบ (ข้อ)	ขนาดกลุ่มตัวอย่าง (คน)	จำนวนข้อสอบที่ตรวจสอบพบในแต่ละวิธี (ข้อ)			จำนวนข้อสอบที่ตรวจสอบพบตรงกันทั้ง 3 วิธี (ข้อ)
		Lord 's χ^2	Mantel-Haenszel	SIBTEST	
20	300	8	7	7	6
	500	11	10	12	10
	700	10	12	13	10
30	300	11	10	14	9
	500	14	16	18	14
	700	15	16	16	15
40	300	15	14	14	12
	500	18	22	19	17
	700	16	21	21	15

2. เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจสอบพบเมื่อขนาดกลุ่มตัวอย่างเท่ากัน แต่วิธีการตรวจสอบต่างกัน จากจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ ตามลำดับ ตรวจสอบกับกลุ่มตัวอย่างขนาดเดียวกัน ในแต่ละวิธีตรวจสอบ คือ วิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST โดยใช้ Q - test ผลทดสอบ สรุปว่า เมื่อจากจำนวนข้อสอบ 30 ข้อกับกลุ่มตัวอย่างขนาด 300 คน 500 คน และจากจำนวนข้อสอบ 40 ข้อ กับกลุ่มตัวอย่างขนาด 500 คน 700 คน จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจสอบพบโดยวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST แตกต่างกัน ส่วนจากจำนวนข้อสอบ 20 ข้อกับกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน จากจำนวนข้อสอบ 30 ข้อ กับกลุ่มตัวอย่างขนาด 700 คน และจากจำนวนข้อสอบ 40 ข้อ กับกลุ่มตัวอย่างขนาด 300 คน จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจสอบพบโดยวิธี Lord 's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST ไม่แตกต่างกัน

3. เปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อขนาดกลุ่มตัวอย่างต่างกัน แต่วิธีการตรวจสอบเหมือนกัน จากจำนวนข้อสอบ 20 ข้อ 30 ข้อ และ 40 ข้อ ตามลำดับ จากวิธีการตรวจสอบเดียวกัน ในแต่ละขนาดกลุ่มตัวอย่าง คือ 300 คน 500 คน และ 700 คน โดยใช้ Q - test ผลทดสอบ สรุปว่า จากจำนวนข้อสอบ 20 ตรวจสอบโดยวิธี Mantel - Haenszel และวิธี SIBTEST จากจำนวนข้อสอบ 30 ข้อ ตรวจสอบโดยวิธี Mantel - Haenszel และจากจำนวนข้อสอบ 40 ข้อ ตรวจสอบโดยวิธี Mantel - Haenszel และวิธี SIBTEST จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบจากกลุ่มตัวอย่างสามขนาดแตกต่างกัน ส่วนจากจำนวนข้อสอบ 20 ข้อ ตรวจสอบโดยวิธี Lord 's χ^2 จากจำนวนข้อสอบ 30 ข้อ ตรวจสอบโดยวิธี Lord 's χ^2 กับวิธี SIBTEST และจากจำนวนข้อสอบ 40 ข้อ ตรวจสอบโดยวิธี Lord 's χ^2 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบจากกลุ่มตัวอย่างสามขนาดไม่แตกต่างกัน

4. ความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อกลุ่มตัวอย่างเท่ากัน แต่วิธีการตรวจสอบต่างกัน สรุปผลดังตาราง 38

ตาราง 38 สรุปผลการศึกษาความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
เมื่อขนาดกลุ่มตัวอย่างเท่ากัน วิธีการตรวจสอบต่างกัน

จำนวน ข้อสอบ	ขนาด กลุ่มตัวอย่าง (คน)	ความสอดคล้องระหว่างวิธีการตรวจสอบ	
		สอดคล้องสูงสุด	สอดคล้องต่ำสุด
20	300	LC กับ MH และ LC กับ SIB	MH กับ SIB
	500	LC กับ SIB	MH กับ SIB
	700	MH กับ SIB	LC กับ SIB
30	300	LC กับ SIB	MH กับ SIB
	500	MH กับ SIB	LC กับ SIB
	700	MH กับ SIB	LC กับ MH และ LC กับ SIB
40	300	LC กับ MH และ LC กับ SIB	MH กับ SIB
	500	LC กับ SIB	LC กับ MH
	700	MH กับ SIB	LC กับ SIB

* LC แทน วิธี Lord 's χ^2

MH แทน วิธี Mantel - Haenszel

SIB แทน วิธี SIBTEST

5. ความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อกลุ่มตัวอย่างต่างกัน แต่วิธีการตรวจสอบเหมือนกัน สรุปผลดังตาราง 39

ตาราง 39 สรุปผลการศึกษาความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อขนาดกลุ่มตัวอย่างต่างกัน วิธีการตรวจสอบเหมือนกัน

จำนวนข้อสอบ	วิธีการตรวจสอบ	ความสอดคล้องระหว่างขนาดกลุ่มตัวอย่าง (คน)	
		สอดคล้องสูงสุด	สอดคล้องต่ำสุด
20	LC	500 กับ 700	300 กับ 700
	MH	500 กับ 700	300 กับ 700
	SIB	500 กับ 700	300 กับ 700
30	LC	500 กับ 700	300 กับ 500
	MH	500 กับ 700	300 กับ 500 และ 300 กับ 700
	SIB	500 กับ 700	300 กับ 500
40	LC	300 กับ 500	300 กับ 700
	MH	500 กับ 700	300 กับ 500
	SIB	500 กับ 700	300 กับ 700

* LC แทน วิธี Lord 's χ^2
MH แทน วิธี Mantel - Haenszel
SIB แทน วิธี SIBTEST

อภิปรายผล

จากการวิจัยสามารถอภิปรายผลได้ดังนี้

1. จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อขนาดกลุ่มตัวอย่างเท่ากัน วิธีการตรวจสอบต่างกัน พบว่า เมื่อจำนวนข้อสอบ 30 ข้อ ขนาดกลุ่มตัวอย่าง 300 คน จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบระหว่างวิธี Mantel - Haenszel กับวิธี SIBTEST แตกต่างกันอย่างมีนัยสำคัญทางสถิติ เมื่อจำนวนข้อสอบ 30 ข้อ ขนาดกลุ่มตัวอย่าง 500 คน จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบระหว่างวิธี Lord 's χ^2 กับวิธี SIBTEST แตกต่างกันอย่างมีนัยสำคัญทางสถิติ เมื่อจำนวนข้อสอบ 40 ข้อ ขนาดกลุ่มตัวอย่าง 500 คน จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบระหว่าง วิธี Lord 's χ^2 กับวิธี Mantel - Haenszel แตกต่างกันอย่างมีนัยสำคัญทางสถิติ เมื่อจำนวนข้อสอบ 40 ข้อ ขนาดกลุ่มตัวอย่าง 700 คน จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบระหว่างวิธี Lord 's χ^2 กับวิธี Mantel - Haenszel แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ซึ่งสอดคล้องกับสมมติฐานการวิจัยที่ตั้งไว้ ส่วนกรณีจำนวนข้อสอบ 20 ข้อ กลุ่มตัวอย่างขนาด 300 คน 500 คนและ 700 คน จำนวนข้อสอบ 30 ข้อ กลุ่มตัวอย่างขนาด 700 คน จำนวนข้อสอบ 40 ข้อ กลุ่มตัวอย่างขนาด 300 คน แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ซึ่งไม่สอดคล้องกับสมมติฐานการวิจัยการวิจัยที่ตั้งไว้ ทั้งนี้อาจเนื่องมาจาก จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบทั้ง 3 วิธี ใกล้เคียงกัน เมื่อนำจำนวนข้อสอบที่ทำหน้าที่ต่างกันมาทดสอบนัยสำคัญทางสถิติ จึง

แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และเมื่อพิจารณาเป็นรายข้อจะพบว่า ทั้ง 3 วิธี ส่วนใหญ่จะตรวจพบข้อสอบที่ทำหน้าที่ต่างกันตรงกัน ซึ่งสอดคล้องกับงานวิจัยของ พรณี จินตมาศ (2540 : 92) ศึกษาเปรียบเทียบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจากการใช้วิธีการตรวจสอบ 3 วิธี คือ วิธีแปลงค่าความยาก วิธี Mantel - Haenszel และ วิธี SIBTEST เมื่อขนาดกลุ่มตัวอย่าง 500 คนและ 1,000 คน พบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ นอกจากนี้ยังสอดคล้องกับงานวิจัยของ นารายานานและสวามินาทาน (Narayanan and Swaminathan. 1994 : 315 – 328) ได้ศึกษาเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบยูนิฟอร์มด้วยวิธี Mantel - Haenszel และวิธี SIBTEST พบว่า วิธี Mantel - Haenszel และ วิธี SIBTEST สามารถตรวจสอบการทำหน้าที่ต่างกันได้ใกล้เคียงกัน

2. จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อขนาดกลุ่มตัวอย่างต่างกัน วิธีการตรวจสอบเหมือนกัน พบว่า เมื่อจำนวนข้อสอบ 20 ข้อ ตรวจสอบโดยวิธี Mantel - Haenszel จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบระหว่างกลุ่มตัวอย่างขนาด 300 คน กับ 700 คน แตกต่างกันอย่างมีนัยสำคัญทางสถิติ เมื่อจำนวนข้อสอบ 20 ข้อ ตรวจสอบโดยวิธี SIBTEST จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบระหว่างกลุ่มตัวอย่างขนาด 300 คน กับ 500 คน และกลุ่มตัวอย่างขนาด 300 คน กับ 700 คน แตกต่างกันอย่างมีนัยสำคัญทางสถิติ เมื่อจำนวนข้อสอบ 30 ข้อ ตรวจสอบโดยวิธี Mantel - Haenszel จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบระหว่างกลุ่มตัวอย่างขนาด 300 คน กับ 500 คน และกลุ่มตัวอย่างขนาด 300 คน กับ 700 คน แตกต่างกันอย่างมีนัยสำคัญทางสถิติ เมื่อจำนวนข้อสอบ 40 ข้อ ตรวจสอบโดยวิธี Mantel - Haenszel จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบระหว่างกลุ่มตัวอย่างขนาด 300 คน กับ 500 คน และกลุ่มตัวอย่างขนาด 300 คน กับ 700 คน แตกต่างกันอย่างมีนัยสำคัญทางสถิติ เมื่อจำนวนข้อสอบ 40 ข้อ ตรวจสอบโดยวิธี SIBTEST จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบระหว่างกลุ่มตัวอย่างขนาด 300 คน กับ 700 คน แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ซึ่งสอดคล้องกับสมมติฐานการวิจัยที่ตั้งไว้ ส่วนกรณีจำนวนข้อสอบ 20 ข้อ ตรวจสอบโดยวิธี Lord's χ^2 กลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน จำนวนข้อสอบ 30 ข้อ ตรวจสอบโดยวิธี Lord's χ^2 และวิธี SIBTEST กลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน จำนวนข้อสอบ 40 ข้อ ตรวจสอบโดยวิธี Lord's χ^2 กลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ซึ่งไม่สอดคล้องกับสมมติฐานการวิจัยที่ตั้งไว้ ทั้งนี้อาจเนื่องมาจากจำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบเมื่อขนาดกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน ทั้ง 3 วิธี สามารถตรวจสอบได้ใกล้เคียงกัน ดังนั้นเมื่อนำจำนวนข้อสอบที่ทำหน้าที่ต่างกันมาทดสอบนัยสำคัญทางสถิติ จึง แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และเมื่อพิจารณาเป็นรายข้อแล้ว พบว่า วิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST เมื่อขนาดกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน ส่วนใหญ่จะตรวจพบข้อสอบที่ทำหน้าที่ต่างกันซ้ำกัน

นอกจากนี้จากการศึกษายังพบว่า เมื่อขนาดกลุ่มตัวอย่างใหญ่ขึ้นสำหรับวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของลอร์ด โดยใช้การเปรียบเทียบค่าพารามิเตอร์ของข้อสอบแล้วทดสอบนัยสำคัญทางสถิติด้วยสถิติทดสอบไคกำลังสอง (χ^2) สามารถตรวจพบข้อสอบที่ทำหน้าที่ต่างกันได้มากขึ้น ซึ่งสอดคล้องกับข้อเสนอแนะของ มิลแซป และอีเวอร์สัน (Millsap and Everson. 1993 : 314) และแฮมเบิลตัน, สวามินาทาน และโรเจอร์ (Hambleton, Swaminathan and Rogers. 1991 : 113) ว่า วิธีนี้มีความไวต่อการ

ตัดสินข้อสอบที่ ทำหน้าที่ต่างกันมาก โดยเฉพาะเมื่อกลุ่มตัวอย่างขนาดใหญ่ สำหรับวิธี SIBTEST ใช้สถิติทดสอบ "Z" ในการทดสอบความมีนัยสำคัญทางสถิติ ซึ่งก็พบว่าเมื่อกลุ่มตัวอย่างขนาดใหญ่ และจำนวนข้อสอบปานกลางถึงมาก จำนวนข้อสอบที่ตรวจพบว่าทำหน้าที่ต่างกันมีจำนวนใกล้เคียงกับวิธี Lord's χ^2 ทั้งนี้ เพราะสถิติทดสอบ Z และ χ^2 จะไวต่อการทดสอบใกล้เคียงกันในเงื่อนไขการทดสอบดังกล่าว

2. ความสอดคล้องในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีลอร์ด Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST พบว่าส่วนใหญ่มีค่าปานกลางถึงสูง และมีความไม่คงที่ในเกือบทุกเงื่อนไขของการศึกษา ซึ่งสอดคล้องกับการศึกษาของกาญจนา วัฒนสุนทร (2537) ที่พบความสอดคล้องระหว่าง วิธี Mantel - Haenszel และวิธี SIBTEST มีความไม่คงที่ในแบบทดสอบที่มีจำนวนข้อสอบต่างกัน และขนาดของกลุ่มตัวอย่างต่างกัน เนื่องจากเกณฑ์ที่ใช้ตัดสินข้อสอบที่ทำหน้าที่ ต่างกันยังไม่สามารถกำหนดแน่นอน และความไวต่อการตัดสินของสถิติทดสอบที่ใช้ในการศึกษาต่างกัน

ข้อเสนอแนะ

1. งานวิจัยครั้งนี้ได้ข้อค้นพบว่า เมื่อใช้จำนวนข้อสอบ 20 ข้อ กับกลุ่มตัวอย่างขนาด 300 คน 500 คน และ 700 คน ใช้จำนวนข้อสอบ 30 ข้อกับกลุ่มตัวอย่างขนาด 700 คน และใช้จำนวนข้อสอบ 40 ข้อกับกลุ่มตัวอย่างขนาด 300 คน ผลการทดสอบระหว่างวิธี Lord's χ^2 วิธี Mantel - Haenszel และวิธี SIBTEST ไม่แตกต่างกัน ดังนั้นควรเลือกใช้ วิธี Mantel - Haenszel หรือวิธี SIBTEST เพราะคำนวณง่ายกว่า
2. งานวิจัยครั้งนี้ได้ข้อค้นพบว่า เมื่อใช้จำนวนข้อสอบ 20 ข้อ และ 30 ข้อ ทั้ง 3 วิธีตรวจสอบกับกลุ่มตัวอย่างขนาด 500 กับ 700 คน ให้ผลสอดคล้องกัน ดังนั้น ถ้าต้องการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้ง 3 วิธี ไม่จำเป็นต้องใช้กลุ่มตัวอย่างถึง 700 คนใช้เพียง 500 คน ก็สามารถตรวจสอบได้และให้ผลสอดคล้องกันด้วย
3. งานวิจัยนี้ได้ข้อค้นพบว่า เมื่อจำนวนข้อสอบมากขึ้น ข้อสอบที่ตรวจสอบพบว่าทำหน้าที่ต่างกัน แต่ละวิธีจะเพิ่มมากขึ้น

ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

ในการทำวิจัยต่อไปควรมีการศึกษาความคลาดเคลื่อนของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ทั้งความคลาดเคลื่อนชนิดที่ 1 และความคลาดเคลื่อนชนิดที่ 2

บรรณานุกรม

บรรณานุกรม

- กาญจนา วัธนสุนทร. (2538) การพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ. วิทยานิพนธ์ ค.ต. (การวัดและประเมินผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย, อุดลำนเา.
- เจน จันทรแสง.(2537) การวิเคราะห์ความลำเอียงของข้อสอบของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา) พิษณุโลก : บัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร. อุดลำนเา.
- จิตติมา วรรณศรี. (2540). การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีแมนเทิล – แอนส์กับวิธีชิปเทสต์ เมื่อขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบ ต่างกัน. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย . อุดลำนเา .
- ชัชชัย เผ่าพงศ์. (2527) การวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบวัดความถนัดทางการเรียน ด้านคณิตศาสตร์และภาษา ระดับชั้นมัธยมศึกษาตอนต้น. วิทยานิพนธ์ ศศ.ม. (การวัดและ ประเมินผลการศึกษา) ขอนแก่น : บัณฑิตวิทยาลัย มหาวิทยาลัยขอนแก่น. อุดลำนเา.
- ชวาล แพรัตกุล. (2517) "แบบทดสอบเพื่อจำแนกอาชีพ," พัฒนาวัดผล 10. กรุงเทพฯ : สำนักทดสอบทาง การศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ.
- ญาณภัทร สีหะมงคล. (2540) การเปรียบเทียบผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน ระหว่างวิธี Lord's χ^2 วิธี Raju's Area Measurement และวิธี Closed Interval Area . ปรินญานินพนธ์ กศ.ต. (การทดสอบและการวัดผลการศึกษา) กรุงเทพฯ :บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อุดลำนเา.
- ทองอยู่ สาระ.(2543) การเปรียบเทียบอำนาจการตรวจสอบและการจำแนกผิดพลาดในการตรวจสอบข้อสอบ ที่ทำหน้าที่ต่างกันแบบสมำเสมอและแบบไม่สมำเสมอ ระหว่างวิธีแมนเทิล – แอนส์เชล และวิธี ถดถอยโลจิสติก โดยใช้ความยาวแบบทดสอบและขนาดกลุ่มตัวอย่างต่างกัน ปรินญานินพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อุดลำนเา.
- ทัศนีย์ พิรมนตรี.(2530)การวิเคราะห์ความลำเอียงของแบบสอบวิชาคณิตศาสตร์ โครงการตรวจสอบคุณภาพ การ ศึกษาชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการ ศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. อุดลำนเา.
- นิรมล ชัยเชาวลิต.(2537) การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของแบบทดสอบความเข้าใจในการ อ่านภาษาไทยตามทฤษฎีคลาสสิกคอลที่ใช้วิธีวิเคราะห์ต่างกัน. ปรินญานินพนธ์ กศ.ม. (การวัดผล การศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ . อุดลำนเา.
- พัชรื ปิยะภักดิ์.(2531) การวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน. ปรินญานินพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัย ศรีนครินทรวิโรฒ. อุดลำนเา.

- พรณี จินตมาศ. (2540) การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบ โดยใช้ขนาดกลุ่มผู้สอบ และวิธีวิเคราะห์ต่างกัน. ปรินทิพนิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อุดลำนเา.
- ภาวนา วงศ์เพ็ญทักษ์. (2541) จำนวนข้อและความเชื่อมั่นของแบบทดสอบที่ได้จากการวิเคราะห์ข้อสอบด้วยวิธีต่างกัน. ปรินทิพนิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อุดลำนเา.
- เรวดี อินทสระ. (2539) ผลการตรวจสอบความลำเอียงของข้อสอบต่อการศึกษาคความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่คิดคะแนนต่างกัน. ปรินทิพนิพนธ์ กศ.ด. (การทดสอบและการวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อุดลำนเา.
- ล้วน สายยศ และอังคณา สายยศ.(2538)เทคนิคการวิจัยทางการศึกษา. กรุงเทพฯ : ชมรมเด็ก.
- ล้วน สายยศ และอังคณา สายยศ. (2541) เทคนิคการสร้างและสอบข้อสอบความถนัดทางการเรียน. กรุงเทพฯ : ชมรมเด็ก.
- สมบุญ ชิตพงศ์ และสำเริง บุญเรืองรัตน์.(2518) การวัดความถนัด. กรุงเทพฯ : ไทยวัฒนาพานิช.
- สุพัฒน์ สุกมลสันต์. (2534) การวิเคราะห์ความลำเอียงของข้อสอบภาษาอังกฤษเข้ามหาวิทยาลัยปี 2531 – 2533. กรุงเทพฯ : สถาบันภาษาจุฬาลงกรณ์มหาวิทยาลัย. อุดลำนเา.
- เสรี ชัดแฉ้ม. (2540) การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันแบบไม่สมำเสมอของข้อสอบระหว่างวิธีแมนเทล แชนเซล แบบปกติกับวิธีแมนเทล แชนเซล แบบแบ่งกลุ่มตามความสามารถและความยาวของข้อสอบ. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. (อุดลำนเา).
- สุรศักดิ์ อมรรัตนศักดิ์. (2531) การศึกษาเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่ต่างกัน 4 วิธี. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. อุดลำนเา.
- อารี วัชรโสติกุล. (2543) การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้รูปแบบและวิธีการต่างกัน. ปรินทิพนิพนธ์ กศ.ม. (การวัดผลการศึกษา) กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อุดลำนเา.
- อุทุมพร จามรมาน. (2537) ทฤษฎีการวัดทางจิตวิทยา. กรุงเทพฯ : ฟิ้นนี่พับลิชชิ่ง.
- Allen, Mary J. and Wendy M. Yen. (1979) *Introduction to Measurement Theory*. California : Brooks/Cole Publishing Company.
- Baghi, Heibatollah and 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) : 14-17 .
- Baker, F.B., A Al – Karni and I.M.AI-Dosary.(1991) "EQUATE : A Computer Program for the test Characteristic method of IRT equating," *Applied Psychological Measurement*. 15 (3) : 78.
- Berk, R.A. (1982) *Handbook of Methods for Detecting item Bias*. Baltimore, MD : John Hopkins University Press.

- Burrill, L.E. (1982) "Comparative Studies of Item Bias Method," *Handbook of Methods for Detecting Test Bias*. Baltimore and London : The John Hopkins University Press.
- Camilli, Gregory and 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 the American Educational Research Association*. Chicago, Illinois 11 (4) : 3-7.
- Cohen, Allan S. and Seock-Ho. Kim. (1993, Spring) "A Comparison of Lord's χ^2 and Raju's Area Measures In Detection of Differential Item Functioning," *Applied Psychological Measurement*. 17(1) : 39-52 .
- Cohen, Allan S. and Seock-Ho. Kim and Michael J. Subkoviak. (1991, Winter) " Influence of Distributions on Detection of DIF," *Journal of Educational Measurement*. 28(1) : 49-59 .
- Cohen, Allan S. and Seock-Ho. Kim and Wollack, James A.(1995, September) "An Investigation of the Likelihood Ratio Test for Detection of Differential Item Functioning," *Applied Psychological Measurement*. 19(7) : 45 - 57 .
- Conbach, Lee J. (1970) *Essentials of Psychological Testing*. 3 rd ed. New York : Harper and Row Publrishers.
- Doolittle, Allen E. and T. Anne. Cleary.(1987, Summer) "Gender-Based Differential Item Performance In Mathematics Achievement Items," *Journal of Educational Measurement*. 24(2) :157-166 .
- Guilford J.P. and B. Fruchter. (1978) *Fundamental Statistics in Psychology and Education*. McGraw-Hill Singapore.
- Hambleton, Ronald K. and L.L. Cook. (1977, Summer) "Latent Trait Model and their Use in the Analysis of Educational Test Data," *Journal of Educational Measurement*. 14 (2) : 75-96 .
- Hambleton, Ronald K. and H. Jane. Rogers. (1989, April) "Detecting Potential Biased Test Items.Comparison of IRT Area and Mantel-Haenszel Methods," *Applied Measurement in Education*. 2(4) : 313-334.
- Hambleton, Ronald K. and Hariharan Swaminathan. (1985) *Item Response Theory Principles and Applications*. Boston : Kluwer-Nijhoff Publishing.
- Hambleton, Pankaja and Hariharan Swaminathan. (1996, September) "Identification of item that show nonuniform DIF," *Applied Psychological Measuerment*. 20(3) : 257-274 .
- Hambleton, Ronald K., H. Swaminathan and H. Jane. Rogers. (1991) *Fundamentals of Item Response Theory*. Newbury Park, California : Sage Publications.

- Hambleton, Ronald K. and Others. (1986, April) "Identifying Potentially Biased Test Items: A 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) : 16-20 .
- Harnisch, Delwyn I. (1991) *Techniques for Assessing Differential Item Performance on Achievement Tests*. Paper presented at the proceeding of the sixteenth Annual SAS User Group International Conference. N. C : SAS Institute, Inc.
- Holland, P.W. and H. Wainer. (1993) *Differential Item Functioning*. Hillsdale NJ : Erlbaum.
- Intasuwan, P.A. (1979) *Comparison of Three Approaches for Determining Item Bias in Gross - National Testing*. Doctor 's Thesis The University of Pittsburgh .photocopied.
- Kim, Seock-Ho and Allan S. Cohen. (1991, September) "A Comparison of Two Area Measures for Detecting Differential Item Functioning," *Applied Psychological Measurement*. 15(3) : 269 – 278 .
- Kim, Seock-Ho and Allan S. Cohen. (1994, April) "An Investigation of Lord's χ^2 Procedure for the Detection of Differential Item Functioning," *Applied Psychological Measurement*. 18(3) : 217-228 .
- Kim, Seock-Ho and Allan S. Cohen. (1995, June) " A Comparision of Lord's χ^2 Chi-Square, Raju's Area Measures and the Likelihood Ratio Test on Detection Of Differential Item Functioning," *Applied Measurement in Education*. 8(4) : 291-312 .
- Laksana, S. (1979) *Application of Analysis of Variance Approach and ITBS From 7*. The University of Iowa.
- Linacre, John Michael. (1988) *The Practical of the Standard Error of the Mantel – Haenszel Statistics*. Paper present at the Annual Meeting of the American Educational Research Association. New Orleans, Los angeles, 5 – 9 , April .
- Lord, F.M. and M.R. Novick. (1968) *Statistical Theories of Mental Test Score*. Massachusetts : Addison Wesley Publishing Company.
- Lord, F.M. (1980) *Application of Item Response Theory to Practical Testing Problems*. Hillsdale, N.J. : Lawrence Erlbaum Associates.
- Mazor, Kathleen M. et. al. (1991, September) "The effect of small sample size on the Mantel -Haenszel statistics," *Educational and Psychological Measurement*, 52 (3) : 443 – 451.
- McCauley, Cynthia D. and Jorge. Mendoza. (1985, December) "A Simulation Study of Item Bias Using a Two-Parameter Item Response Model," *Applied Psychological Measurement*. 9(4) : 389-400.

- Miller, M.D. and T.C. Oshima. (1992, December) "Effect of Sample Size, Number of Biased Item and Magnitude of Bias on a Two – Stage Item Bias Estimation Method," *Applied Psychological Measurement*. 16(4) : 381 – 388 .
- Millsap, R.E. and H.T. Everson.(1993, February) "Methodology Review : Statistical Approaches for Assessing Measurement Bias," *Applied Psychological Measurement*. 17 (1) : 297—334 .
- Mislevy, R.J. and 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) : 57-75 .
- Narayanan, Pankaja and H. Swaminathan. (1994, December) "Performance of the Mantel – Haenszel and Simultaneous Item Bias Procedures for Detecting Differential Item Function," *Applied Psychological Measurement*. 18(4) : 315 - 328.
- Park, Dong-Gun and Lautenschlager, Gary J. (1990, June) "Improving IRT Item Bias Detection With Iterative Linking and Ability Scale Purification," *Applied Psychological Measurement*. 14(2) : 163-173.
- 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, April 5-9.
- Popham, W. James. (1990) *Modern Educational Measurement A Practitioner's Perspective*.2nd ed. Englewood Cliffs New Jersey : Prentice Hall.
- Raju, Drasgow Fritz and Jeffrey A. Slinde. (1993, March) "An Empirical Comparison of the Area Methods, Lord 's Chi-Square Test and the Mantel-Haenszel Technique for Assessing Differential Item Functioning," *Educational and Psychological Measurement*, 53(3) : 301 – 314.
- Roger, H. Jane and H. Swaminathan. (1993, April) "A Comparison of Logistic Regression and Mantel – Haenszel Procedures for Detecting Differential Item Functioning," *Applied Psychological Measurement*. 17(2) : 105 - 116 .
- Roussos, Louis A. and 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) : 215 – 230.
- Rudner, L.M. P.R. Getson and D.L. Knight. (1980, Winter) "A Monte Carlo Comparison of Seven Biased Item Detection Techniques," *Journal of Educational Measurement*. 17(1) : 1-10.
- Ryan, Katherine. (1991, Winter) The Performance of the Mantel-Haenszel Procedure Across Samples and Matching Criteria. *Journal of Educational Measurement*. 28(4) : 325-337.

- Seigel, Sidney. (1956) *Nonparametric Statistics for the Behavioral Sciences*. Kogakusha Tokyo : McGraw-Hill .
- Shealy, R. and W. Stout. (1993, May) "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) : 159-194.
- Shepard, L.A., G. Camilli and D.M. Williams. (1985, Summer) "Validity of Approximation Techniques for Detecting Item Bias," *Journal of Educational Measurement*. 22(2) : 77-105 .
- Stocking, M.L. and Lord, F.M. (1983, December) "Developing a common metric in item response theory," *Applied Psychological Measurement*. 7(3) : 201 – 210 .
- Stout, William and Rousos, Louis. (1992) *SIBTEST USER MANUAL*. Illinois : Department of Statistics, University of Illinois, December.
- Subkoviak, Micheal J. and others. (1984, Spring) "Empirical Comparison of Detected Item Bias Detection Procedures with Bias Manipulation," *Journal of Educational Measurement*. 21(3) : 49-59 .
- Sudweeks, Richard, R. and 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) : 8-11.
- Swaminathan, Hariharan and H. Jane, Rogers. (1990, Winter) "Detecting Differential Item Functioning Using Logistic Regression Procedures," *Journal of Educational Measurement*. 27(1) : 361-370 .
- Thissen, David, Lynne Steinberg and Howard. Wainer. (1988) *Use of Item Response Theory In The Study of Group Differences in Trace Test Validity*. p147-170 in Wainer, Howard and Braun, Henry I. *Test Validity*. Hillsdale, New Jersey : Lawrence Erlbaum Associates.
- _____. (1993) *Detection of Differential Item Functioning using the Parameters of Item Response Models*. P67-113 In P.W. Holland and H. Wainer *Differential Item Functioning*. Hillsdale NJ : Erlbaum.
- Wright, B.D. and M.H. Stone. (1979) *Best Test Design*. Chicago : Mesa Press.

ภาคผนวก

ภาคผนวก ก

รายชื่อผู้เชี่ยวชาญในการตรวจสอบคุณภาพเครื่องมือ

รายชื่อผู้เชี่ยวชาญในการตรวจสอบคุณภาพเครื่องมือ

- | | |
|-----------------------------|--|
| 1. ผศ.ดร.สมทรง อัสวกุล | ภาควิชาทดสอบและวิจัยการศึกษา
สถาบันราชภัฏนครราชสีมา |
| 2. ผศ.ดร.สมบูรณ์ ตันยะ | ภาควิชาทดสอบและวิจัยการศึกษา
สถาบันราชภัฏนครราชสีมา |
| 3. ผศ.จิรภา นาคสมพงษ์ | ภาควิชาทดสอบและวิจัยการศึกษา
สถาบันราชภัฏนครราชสีมา |
| 4. ผศ.วนิดา เตชตานนท์ | ภาควิชาทดสอบและวิจัยการศึกษา
สถาบันราชภัฏนครราชสีมา |
| 5. ผศ.อาภรณ์ บางเจริญพรพงศ์ | ภาควิชาวิจัยการศึกษา
สถาบันเทคโนโลยีราชมงคล |

ภาคผนวก ข

ค่าพารามิเตอร์ของข้อสอบและผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกัน

ตาราง 40 ค่าพารามิเตอร์ของข้อสอบ แยกตามเพศ ชาย - หญิง

ข้อที่	พารามิเตอร์ข้อสอบ					
	เพศชาย		เพศหญิง		รวม	
	ค่าอำนาจ จำแนก (a_1)	ค่าความยาก (b_1)	ค่าอำนาจ จำแนก (a_2)	ค่าความยาก (b_2)	ค่าอำนาจ จำแนก (a_i)	ค่าความยาก (b_i)
1	.380	-.977	.547	1.503	.513	-.005
2	1.315	.451	.858	1.656	1.307	.708
3	.734	.325	.628	1.183	.788	.441
4	.723	.769	.380	.004	.472	.324
5	.936	.648	.505	.969	.729	.482
6	.472	-.950	.466	.403	.659	-.377
7	.842	.491	.463	1.319	.785	.500
8	1.126	.110	.468	.840	.944	.268
9	1.726	.361	.314	.219	1.382	.673
10	.825	-.560	.332	.422	.802	-.086
11	.801	1.276	.457	1.141	.724	.729
12	1.033	-.368	.438	.423	1.018	-.016
13	.808	.747	.636	.639	.848	.536
14	1.144	.264	.631	.478	1.074	.376
15	1.858	.084	.838	.567	1.744	.267
16	4.067	.764	1.206	.516	1.326	.388
17	.627	-.697	1.359	.574	.790	-.026
18	1.524	.540	2.068	.497	1.736	.683
19	1.063	1.470	2.466	.445	.873	1.332
20	1.328	.907	1.820	.341	1.223	.873
21	.894	-.597	4.208	.597	.885	.145
22	1.335	.857	5.238	.618	2.276	1.000
23	1.924	1.106	5.897	.654	2.674	1.200
24	1.552	1.261	3.767	.704	2.044	1.300
25	2.549	.377	3.750	.676	5.306	.750
26	.855	-.261	2.277	.930	2.162	.868
27	.835	1.113	1.052	.753	.758	1.205
28	.799	.080	1.729	.958	1.221	.955
29	.988	2.177	1.010	.904	.800	1.959
30	1.768	.613	.815	1.395	1.661	1.102

ตาราง 40 (ต่อ)

ข้อที่	พารามิเตอร์ข้อสอบ					
	เพศชาย		เพศหญิง		รวม	
	ค่าอำนาจ จำแนก (a_1)	ค่าความยาก (b_1)	ค่าอำนาจ จำแนก (a_2)	ค่าความยาก (b_2)	ค่าอำนาจ จำแนก (a_i)	ค่าความยาก (b_i)
31	.380	-.977	.547	1.503	.513	-.005
32	1.315	.451	.858	1.656	1.307	.708
33	.734	.325	.628	1.183	.788	.441
34	.723	.769	.380	.004	.472	.324
35	.936	.648	.505	.969	.729	.482
36	.472	-.950	.466	.403	.659	-.377
37	.842	.491	.463	1.319	.785	.500
38	1.126	.110	.468	.840	.944	.268
39	1.726	.361	.314	.219	1.382	.673
40	.825	-.560	.332	.422	.802	-.086
41	.801	1.276	.457	1.141	.724	.729
42	1.033	-.368	.438	.423	1.018	-.016
43	.808	.474	.636	.639	.848	.536
44	1.144	.264	.631	.478	1.074	.376
45	1.858	.084	.838	.567	1.744	.267
46	4.067	.764	1.206	.516	1.326	.388
47	.627	-.697	1.359	.574	.790	-.026
48	1.524	.540	2.068	.497	1.736	.683
49	1.063	1.470	2.466	.445	.873	1.322
50	1.328	.907	1.820	.341	1.223	.873
51	.894	-.597	4.208	.579	.885	.145
52	1.335	.857	5.238	.618	2.276	1.000
53	1.924	1.106	5.879	.654	2.674	1.200
54	1.502	1.261	3.767	.704	2.044	1.300
55	2.549	.377	3.750	.676	5.306	.750
56	.855	-.261	2.277	.935	2.162	.868
57	.835	1.113	1.052	.753	.758	1.205
58	.799	.080	1.729	.958	1.221	.955
59	.988	2.177	1.010	.904	.800	1.959
60	1.786	.613	.815	1.395	1.661	1.102

ตาราง 41 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่างขนาด 300 คน และจำนวนข้อสอบ 20 ข้อ

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ		
	เพศชาย		เพศหญิง		LC	MH	SIB
	ค่าอำนาจ จำแนก (a_1)	ค่าความยาก (b_1)	ค่าอำนาจ จำแนก (a_2)	ค่าความยาก (b_2)			
1	.508	-1.108	.755	1.288	6.23*	2.96***	.221***
2	1.199	.254	1.279	1.445	0.13	1.47	.114
3	.739	.288	1.631	1.212	0.81	1.69	.066
4	1.183	.359	.647	.335	6.60*	.38***	-.201***
5	.863	.769	1.581	1.150	3.79	1.27	.096
6	.954	-.950	1.362	.767	3.09	1.74	.068
7	1.306	.108	1.997	1.036	0.45	.79	-.007
8	1.026	.064	1.655	1.049	2.17	1.23	.101
9	1.403	.154	.812	.624	5.33*	.48*	-.165*
10	.670	-.584	.811	.247	3.55	.69	-.080
11	.662	1.282	2.280	.927	0.95	.64	-.030
12	.634	-1.054	.959	.337	1.22	1.23	.029
13	.744	.356	2.578	.969	0.47	.97	-.022
14	1.423	-.004	1.692	1.209	1.35	1.04	.050
15	1.398	-.271	1.169	1.079	5.07*	2.44***	.162*
16	1.276	.162	1.351	1.059	1.97	1.27	.039
17	.656	-1.005	.672	1.268	6.50*	2.83***	.218***
18	1.200	.166	.728	.633	6.24*	.54	-.171*
19	.818	2.115	.799	1.613	7.12*	.36***	-.269***
20	1.265	.850	.791	.625	6.52*	.46*	-.117
รวม					8	7	7

ตาราง 42 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่างขนาด 500 คน และจำนวนข้อสอบ 20 ข้อ

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ		
	เพศชาย		เพศหญิง		LC	MH	SIB
	ค่าอำนาจจำแนก (a_1)	ค่าความยาก (b_1)	ค่าอำนาจจำแนก (a_2)	ค่าความยาก (b_2)			
1	.413	-.831	.501	1.703	6.15*	3.12***	.250**
2	1.319	.235	.685	1.423	4.23	1.47	.102*
3	.630	.142	.934	1.235	10.55**	2.98***	.227***
4	2.039	.835	.508	.304	5.99*	.63*	-.132**
5	.785	.573	.950	1.033	6.00*	1.46	.102*
6	.684	-.761	1.113	.453	6.25*	1.69*	.091*
7	1.727	.537	1.394	.793	2.23	.80	-.039
8	1.129	-.028	1.456	.729	1.48	1.15	.043
9	1.787	.318	1.434	.579	7.52*	.32***	-.201***
10	.771	-.465	1.068	.263	0.47	1.00	.001
11	.643	1.391	1.751	.729	0.26	.67	-.081
12	.956	-.426	.946	.424	1.58	1.24	.029
13	.802	.832	1.276	.872	2.56	.79	-.040
14	1.250	.141	.884	1.112	3.24	.96	.004
15	1.768	.156	.768	.994	7.21*	1.87**	.132**
16	1.586	.470	.830	1.161	3.51	1.29	.066
17	.720	-.402	.498	2.105	6.54*	2.13**	.159***
18	1.734	.398	.582	2.201	7.41*	.54**	-.122**
19	.796	1.689	.589	2.146	6.99*	.31**	-.232***
20	.820	.889	.462	.628	6.78*	.38**	-.216***
รวม					11	10	12

ตาราง 43 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่างขนาด 700 คน และจำนวนข้อสอบ 20 ข้อ

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ		
	เพศชาย		เพศหญิง		LC	MH	SIB
	ค่าอำนาจจำแนก (a_1)	ค่าความยาก (b_1)	ค่าอำนาจจำแนก (a_2)	ค่าความยาก (b_2)			
1	.385	-.794	.517	1.313	6.47*	3.33***	.252***
2	1.120	.353	.713	1.110	6.92*	1.62**	.116**
3	.679	.366	.969	.959	7.26*	1.99***	.151***
4	1.103	.862	.485	-.091	6.98*	.62**	-.109**
5	1.067	.741	1.031	.754	5.42	.37	.078*
6	.495	-.752	1.184	.314	6.57*	1.70**	.110**
7	.965	.438	1.147	.605	4.23	.99	-.009
8	.994	.051	1.238	.526	0.19	1.16	.034
9	2.075	.372	1.561	.485	5.43	.34***	-.184***
10	.734	-.391	1.285	.307	2.26	1.30	.045
11	.694	1.241	1.700	.533	5.48	.67*	-.089*
12	1.037	-.276	1.155	.314	0.38	1.27	.030
13	.895	.735	1.103	.616	0.89	.69	-.068
14	1.180	.272	.700	.585	1.63	.74	-.033
15	2.029	.265	.785	.608	8.23*	1.91**	.121**
16	2.474	.638	.686	.750	4.13	1.23	.054
17	.660	-.537	.418	1.196	10.52**	2.49***	.195***
18	1.331	.538	.364	.827	6.06*	.68*	-.079*
19	.998	1.523	.418	.947	9.18**	.27***	-.282***
20	.966	1.059	.401	.131	7.36*	.34***	-.233***
รวม					10	12	13

ตาราง 44 จำนวนข้อสอบที่กำหนดที่ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่างขนาด 300 คน และจำนวนข้อสอบ 30 ข้อ

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ		
	เพศชาย		เพศหญิง		LC	MH	SIB
	ค่าอำนาจจำแนก (a_1)	ค่าความยาก (b_1)	ค่าอำนาจจำแนก (a_2)	ค่าความยาก (b_2)			
1	.498	-1.310	.551	1.554	6.42*	2.87**	243**
2	1.168	.204	.739	1.990	6.21*	1.58	.167*
3	.765	.098	.818	1.656	3.44	1.62	.099
4	.752	.077	.456	.214	4.17	.57	-.159*
5	.727	.743	.759	1.507	0.58	1.33	.076
6	.805	-1.168	.520	.880	6.11*	2.03*	.177*
7	1.019	-.011	.639	1.569	1.43	.96	-.005
8	1.072	-.056	.528	1.348	2.11	1.26	.085
9	1.073	.056	.317	.325	2.35	.59	-.032
10	.744	-.732	.435	.046	1.44	.57	-.034
11	.652	1.190	.736	1.329	3.29	.71	-.095
12	.604	-1.217	.601	.249	0.89	1.37	-.001
13	.683	.271	1.169	1.172	0.67	.81	-.009
14	1.279	-.167	1.581	1.159	1.49	1.25	.051
15	1.166	-.505	1.018	1.145	6.57*	2.27*	.149**
16	1.085	.084	1.595	.851	2.64	1.19	.014
17	.558	-1.289	.885	.963	7.25*	3.02**	.183*
18	1.436	.089	.947	.580	6.54*	.57	-.168*
19	.730	1.813	1.757	.981	7.89*	.22**	-.261***
20	1.300	.545	.775	.490	6.75*	.33**	-.193**
21	1.113	-.673	1.858	1.278	10.13**	5.10***	.272***
22	1.057	.616	1.060	.991	3.21	.67	-.082
23	1.066	1.077	.764	.846	6.48*	.26***	-.188*
24	1.241	.999	1.538	1.225	5.11	.34*	-.205**
25	.845	-.566	.957	-.170	4.69	.49	-.133*
26	.830	-.701	.576	.364	0.92	.87	-.034
27	.706	.743	.99	1.761	0.84	1.67	.099
28	.605	-.387	.830	1.417	6.21*	2.04*	.174*
29	.882	2.220	.736	2.030	0.53	1.05	.039
30	1.229	.266	.734	1.773	1.85	1.29	-.014
รวม					11	10	14

ตาราง 45 จำนวนข้อสอบที่กำหนดให้ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่างขนาด 500 คน และจำนวนข้อสอบ 30 ข้อ

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ		
	เพศชาย		เพศหญิง		LC	MH	SIB
	ค่าอำนาจจำแนก (a_1)	ค่าความยาก (b_1)	ค่าอำนาจจำแนก (a_2)	ค่าความยาก (b_2)			
1	.336	-1.109	2.344	1.937	9.58**	2.53***	.277***
2	1.334	.276	3.044	1.661	5.47	1.44	.104*
3	.676	.107	3.203	1.359	6.72*	2.75***	.201***
4	.550	.413	.892	1.636	0.87	.69	-.092
5	.663	.481	1.855	2.488	1.25	1.38	.070
6	.704	-.767	2.693	1.819	6.11*	1.71*	.094*
7	.904	.488	2.070	2.221	2.34	.93	-.036
8	1.359	.007	1.616	2.300	0.49	1.37	.037
9	1.768	.360	.490	1.794	7.23*	.42***	-1.62**
10	.833	-.492	.434	3.006	1.52	1.03	.024
11	.831	1.524	1.787	2.761	0.58	.89	-.031
12	.826	-.459	1.126	4.001	0.96	1.30	.061
13	.688	.865	.658	1.446	0.64	.92	-.027
14	1.381	.154	.675	1.296	1.43	1.02	.006
15	1.618	-.028	.745	1.220	5.41	1.60*	.123
16	2.206	.589	.889	.861	1.65	1.15	.014
17	.574	-.597	1.190	.855	6.54*	1.86**	.138**
18	1.760	.360	1.367	.637	7.23*	.45**	-.193***
19	.815	1.593	2.192	.712	7.58*	.18***	-.290***
20	1.087	.744	1.643	.558	8.21*	.34***	-.235***
21	.949	-.506	3.469	.892	9.69**	3.64***	.246***
22	1.035	.828	3.183	.737	7.54*	.48**	-.170***
23	1.432	1.189	3.515	.857	7.15*	.27***	-.288***
24	1.476	1.179	3.050	.951	6.52*	.50**	-.161***
25	1.233	-.153	1.372	.678	4.58	.64	-.112*
26	.883	-.404	1.287	1.128	0.39	.96	.002
27	.614	1.304	1.101	1.435	5.98	1.70*	.125*
28	.785	-.169	1.339	1.345	7.53*	2.14***	.205***
29	.941	2.256	.925	1.435	1.20	1.00	.005
30	1.363	.493	.825	1.896	6.25*	1.83**	.109*
รวม					14	16	18

ตาราง 46 จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่างขนาด 700 คน และจำนวนข้อสอบ 30 ข้อ

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ		
	เพศชาย		เพศหญิง		LC	MH	SIB
	ค่าอำนาจจำแนก (a_1)	ค่าความยาก (b_1)	ค่าอำนาจจำแนก (a_2)	ค่าความยาก (b_2)			
1	.359	-1.012	.497	2.137	7.59*	2.88***	.239***
2	1.098	.335	.600	2.341	6.54*	1.69**	.118**
3	.739	.238	.481	1.981	6.12*	1.96***	.133**
4	.562	.443	.296	.379	5.64	.70*	-.079*
5	.853	.556	.425	1.652	2.11	1.16	.045
6	.478	-.887	.3451	.912	6.44*	1.84**	.109**
7	.778	.400	.411	2.678	2.02	1.15	.019
8	1.075	.019	.358	1.690	1.35	1.33	.060
9	1.868	.348	.233	.660	10.62**	.44***	-.153***
10	.831	-.517	.240	1.205	2.36	1.32	.075
11	.786	1.270	.364	2.276	0.87	.81	-.060
12	.951	-.371	.321	1.070	1.46	1.33	.074
13	.817	.724	.498	1.062	0.59	.77	-.061
14	1.263	.281	.500	.822	1.28	.90	-.011
15	1.628	.029	.602	.934	6.98*	1.85**	.127**
16	3.506	.586	.762	.741	2.13	1.11	.027
17	.601	-.701	.838	.754	9.82**	2.40***	.178***
18	1.291	.479	1.003	.561	6.54*	.61**	-.089*
19	.976	1.442	1.269	.533	9.66**	.21***	-.314***
20	1.143	.830	.978	.261	7.85*	.31***	-.245***
21	.843	-.620	2.640	.716	9.37**	3.66***	.250***
22	1.196	.732	2.613	.754	6.21*	.55**	-.114**
23	1.502	1.040	3.259	.827	10.28**	.31***	-.237***
24	1.337	1.204	2.727	.884	11.24**	.48***	-.155***
25	1.616	.175	2.575	.817	1.58	.77	-.047
26	.851	-.266	1.597	1.152	0.98	1.18	.059
27	.649	1.027	1.106	.958	2.16	.95	-.018
28	.727	-.063	1.575	1.142	9.82**	2.16***	.200***
29	.817	2.051	.952	1.117	1.26	.74	-.073
30	1.499	.524	.789	1.590	2.19	1.32	.061
รวม					15	16	16

ตาราง 47 จำนวนข้อสอบที่กำหนดให้ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่างขนาด 300 คน และจำนวนข้อสอบ 40 ข้อ

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ		
	เพศชาย		เพศหญิง		LC	MH	SIB
	ค่าอำนาจจำแนก (a_1)	ค่าความยาก (b_1)	ค่าอำนาจจำแนก (a_2)	ค่าความยาก (b_2)			
1	.600	-1.021	1.615	1.317	7.52*	2.69**	.194*
2	1.520	.367	2.064	1.155	2.54	1.28	.037
3	.748	.225	2.842	1.044	1.12	1.35	.057
4	1.636	.425	1.344	.749	6.62*	.37**	-.276***
5	.869	.863	2.475	1.028	2.11	1.03	.032
6	.908	-.970	2.917	.594	2.44	1.91	.106
7	1.665	.215	2.104	1.025	2.02	.79	-.036
8	1.456	.202	1.453	.793	1.35	1.11	.005
9	1.551	.225	1.010	.269	0.62	.47	-.131
10	.782	-.590	1.519	-.111	2.36	.64	-.058
11	.614	1.569	1.673	.974	0.87	.74	-.001
12	.626	-1.017	1.061	.810	1.46	1.36	.065
13	.751	.458	2.232	1.152	0.59	.89	.019
14	1.297	.061	1.881	1.362	1.28	1.09	.005
15	1.234	-.200	1.230	1.401	6.98**	2.30**	.138*
16	1.599	.451	1.854	1.475	2.13	1.28	.064
17	.607	-.951	1.601	3.677	9.42**	3.45***	.255**
18	1.366	.323	.921	2.096	6.74*	.61	-.205*
19	.746	2.069	1.728	2.509	9.16**	.29***	-.268***
20	1.385	.818	.722	.857	7.25*	.39**	-.189**
21	1.065	-.453	2.184	2.853	9.17**	4.32***	.270***
22	.911	.747	1.763	2.585	6.11*	.72	-.098
23	.918	1.493	1.734	1.921	10.48**	.36**	-.219**
24	1.067	1.238	2.131	2.814	7.24*	.40*	-.173*
25	.875	-.275	.951	.169	1.58	.53	-.165*
26	.707	-.569	1.685	2.264	0.98	1.04	-.037
27	.682	.950	1.736	3.376	2.16	1.90	.083
28	.548	-.143	2.092	2.706	9.42**	2.20*	.225**
29	1.074	2.374	.873	2.145	1.24	1.27	.067
30	1.364	.471	1.702	3...5	2.12	1.35	-.032
31	.600	-1.021	1.615	1.317	6.52*	2.69**	.194*
32	1.520	.367	2.064	1.155	1.28	1.28	.037
33	.748	.225	2.842	1.044	2.13	1.35	.057
34	1.636	.425	1.344	.749	6.98*	.37**	-.276***
35	.869	.863	2.745	1.028	0.89	1.03	.032
36	.908	-.970	2.917	.594	1.77	1.91	.106
37	1.655	.215	2.104	1.025	2.01	.79	-.036
38	1.456	.202	1.453	.796	1.55	1.11	.005
39	1.551	.225	1.010	.269	6.02*	.47*	-.131
40	.782	-.590	1.519	-.111	0.22	.64	-.058
รวม					15	14	14

ตาราง 48 จำนวนข้อสอบที่กำหนดที่ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่างขนาด 500 คน และจำนวนข้อสอบ 40 ข้อ

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ		
	เพศชาย		เพศหญิง		LC	MH	SIB
	ค่าอำนาจจำแนก (a_1)	ค่าความยาก (b_1)	ค่าอำนาจจำแนก (a_2)	ค่าความยาก (b_2)			
1	.391	-1.052	1.411	1.118	7.25*	2.70***	.226**
2	1.846	.339	2.330	1.065	2.54	1.22	.051
3	.658	.010	5.860	.729	6.12*	2.62***	.174***
4	.693	.438	2.376	.791	6.62*	.57**	-.132**
5	.740	.491	2.197	.767	2.13	1.25	.047
6	.661	-.870	3.160	.215	2.45	1.65*	.066
7	1.103	.529	1.308	.592	2.42	.72	-.074
8	1.685	.088	1.093	.397	1.45	1.06	.039
9	2.571	.394	.744	.127	7.62*	.31***	-.244***
10	.915	-.548	.861	-.173	1.26	.97	-.001
11	.626	1.840	.844	.969	2.87	.70	-.051
12	.785	-.390	.535	.734	1.46	1.18	.063
13	.751	.903	.823	1.754	1.59	.87	-.028
14	1.109	.204	1.000	2.503	1.22	1.01	-.032
15	1.661	.167	2.006	2.015	6.95**	1.59*	.086
16	1.880	.666	2.101	1.873	2.13	1.24	.043
17	.576	-.535	1.471	3.182	9.12**	2.19***	.158**
18	1.576	.459	1.364	3.019	6.34*	.53**	-.153**
19	.826	1.671	1.878	2.306	9.37**	.25***	-.245***
20	.759	.846	1.031	1.763	7.32*	.36***	-.223***
21	.899	-.417	1.810	2.628	9.86**	3.38***	.255***
22	.790	.929	1.736	2.540	6.41*	.53**	-.146**
23	1.047	1.551	.476	3.005	10.08**	.31***	-.260***
24	.897	1.439	1.831	2.780	7.02*	.52**	-.143**
25	1.040	-.132	.778	1.314	1.52	.78	-.076
26	.711	-.418	.494	2.748	1.98	1.13	.029
27	.664	1.524	1.069	3.326	2.15	1.73*	.138**
28	.681	-.054	1.155	1.864	9.32**	2.20***	.232***
29	.937	2.641	1.259	4.000	1.04	1.02	-.004
30	1.619	.599	1.527	3.116	2.19	1.56*	.125**
31	.391	-1.052	1.411	1.118	6.31*	2.70***	.226***
32	1.846	.399	2.330	1.065	1.96	1.22	.051
33	.658	.010	5.860	.729	6.13**	2.62***	.174***
34	.693	.438	2.376	.791	6.08*	.57**	-.132**
35	.740	.491	2.197	.767	0.89	1.25	.047
36	.661	-.870	3.160	.215	1.77	1.65*	.066
37	1.103	.529	1.308	.592	2.01	.72	-.074
38	1.685	.088	1.093	.397	1.25	1.06	.039
39	2.571	.394	.744	.127	6.82*	.31***	-.244***
40	.915	-.548	.861	-.173	0.29	.97	-.001
รวม					18	22	19

ตาราง 49 จำนวนข้อสอบที่กำหนดหน้าที่ต่างกันที่ตรวจพบ เมื่อกลุ่มตัวอย่างขนาด 700 คน และจำนวนข้อสอบ 40 ข้อ

ข้อที่	ค่าพารามิเตอร์				วิธีตรวจสอบ		
	เพศชาย		เพศหญิง		LC	MH	SIB
	ค่าอำนาจจำแนก (a_1)	ค่าความยาก (b_1)	ค่าอำนาจจำแนก (a_2)	ค่าความยาก (b_2)			
1	.389	-1.056	1.357	1.022	10.22**	2.93***	224***
2	1.238	.3771	2.485	1.067	0.54	1.43	.082*
3	.699	.140	4.185	.881	7.48*	1.83**	.113***
4	.648	.324	2.384	.891	6.13*	.59**	-.116***
5	1.172	.690	2.526	.810	2.01	1.10	.027
6	.489	-.947	1.785	.287	6.14*	1.66*	.084*
7	.873	.333	1.158	.662	1.02	.96	-.022
8	1.159	-.043	1.076	.415	0.35	1.11	.034
9	3.476	.380	.819	.220	7.62*	.33***	-.219***
10	.876	-.634	.913	.045	3.36	1.25	.044
11	.634	1.455	.890	.848	1.87	.68*	-.062
12	.884	-.279	.615	.531	1.46	1.22	.064
13	.855	.761	.716	1.403	3.59	.72	-.050
14	1.071	.326	.549	1.519	0.28	.82	-.054
15	1.967	.235	.661	1.422	6.16**	1.75**	.099**
16	2.637	.721	2.006	1.652	1.53	1.20	.044
17	.550	-.716	2.624	2.092	9.78**	2.65***	.214***
18	1.202	.579	2.243	2.085	6.29*	.68*	-.078
19	1.038	1.470	2.225	1.893	9.06**	.24***	-.295***
20	.943	.994	.892	1.657	7.11*	.33***	-.252***
21	.744	-.578	1.927	1.943	8.24**	3.18***	.262***
22	.936	.803	2.745	1.582	6.03*	.59**	-.123**
23	1.124	1.240	1.850	1.991	10.19**	.38***	-.201***
24	.955	1.401	2.047	1.930	6.21*	.45***	-.147***
25	1.266	.157	.990	1.564	2.49	.83	-.033
26	.702	-.288	1.136	2.037	0.98	1.32	.073
27	.630	1.216	1.918	2.040	2.92	.95	-.005
28	.672	.168	2.782	2.176	8.23**	2.15***	.207***
29	.713	2.393	1.758	2.507	1.24	.77	-.058
30	1.592	.619	2.139	2.327	2.12	1.21	.064
31	.389	-1.056	1.357	1.022	5.52	2.93***	224***
32	1.238	.371	2.485	1.067	1.28	1.43	.082*
33	.699	.140	4.185	.881	2.13	1.83**	.113**
34	.648	.324	2.384	.891	4.98	.59**	-.116**
35	1.172	.690	2.526	.810	1.48	1.10	.027
36	.489	-.947	1.785	.287	1.67	1.66*	.084*
37	.873	.333	1.158	.662	3.27	.96	-.022
38	1.159	-.043	1.076	.415	1.65	1.11	.034
39	3.476	.380	.819	.220	6.98*	.33***	-.219***
40	.876	-.634	.913	.045	0.19	1.25	.044
รวม					16	21	21

ภาคผนวก ค
เครื่องมือที่ใช้ในการศึกษาค้นคว้า

แบบทดสอบความถนัดทางคณิตศาสตร์

คำชี้แจง

1. แบบทดสอบฉบับนี้เป็นแบบปรนัยชนิดเลือกตอบ มีจำนวน 60 ข้อ ใช้เวลา 60 นาที
ฉะนั้น นักเรียนควรรีบตอบโดยเร็วและให้ครบทุกข้อจึงจะได้คะแนนดี
2. แบบทดสอบฉบับนี้เป็นแบบทดสอบวัดความถนัดด้านคณิตศาสตร์ การตอบของนักเรียนไม่มี
ผลต่อการเรียนวิชาคณิตศาสตร์แต่อย่างใด โปรดทำให้เต็มความสามารถ

ตัวอย่างที่ 1

(0) $0.001 \times 0.01 = ?$

- ก 0.001
- ข 0.0001
- ค 0.0011
- (ง) 0.00001
- จ 0.00011

ตัวอย่างที่ 2

(0000) 20 15 11 8 6?.....

- ก 1
- ข 2
- ค 3
- ง 4
- (จ) 5

3. ให้กาเครื่องหมาย X ลงในช่องที่ท่านเห็นว่าถูกต้องที่สุดลงในกระตามคำตอบ

1. ซื้อสินค้าราคา A บาท ขายไปราคา B บาท
จะได้กำไรกี่เปอร์เซ็นต์ ?

ก. $\frac{100(B + A)}{B}$

ข. $\frac{100(B + A)}{A}$

ค. $\frac{100(B - A)}{B}$

ง. $\frac{100(B - A)}{A}$

จ. $\frac{100B}{A}$

2. เลขจำนวนหนึ่งเมื่อนำไปหาร A จะได้ผล

ลัพธ์เป็น $\frac{3}{5}$ เลขจำนวนนั้นตรงกับข้อใด ?

ก. $5A$

ข. $3A$

ค. $\frac{5A}{3}$

ง. $\frac{3A}{5}$

จ. $\frac{5}{3}$

จ. ไม่มีข้อถูก

3. แม่ค้าขายมะม่วง y ผล ราคา 10 บาท
สมชายมีเงินอยู่ 2y บาท ซื้อมะม่วงได้
125 ผล แม่ค้าขายมะม่วงผลละเท่าไร ?

ก. $\frac{2y}{10}$ บาท

ข. $\frac{2}{y}$ บาท

ค. $\frac{10}{y}$ บาท

ง. 2 บาท

จ. 5 บาท

4. สามเหลี่ยมมุมฉากรูปหนึ่ง มีด้านยาว A, B
และ C เซนติเมตร ถ้า A เป็นด้านยาวที่สุด
พื้นที่ของสามเหลี่ยมนี้เท่ากับข้อใด ?

ก. $\frac{AB}{2}$

ข. $\frac{BC}{2}$

ค. $2AB$

ง. $\frac{B}{2}$

จ. คำนวณไม่ได้

5. พ่อมีเงิน A บาท แบ่งให้ลูกคนโต 50%
และคนกลาง 30% ลูกคนเล็กจะได้ส่วนแบ่ง
เท่าไร ?

ก. $\frac{A}{5}$ บาท

ข. $5A$ บาท

ค. $0.5A$ บาท

ง. $\frac{A}{2}$ บาท

จ. ไม่มีข้อถูก

6. ด้านของรูปสามเหลี่ยม เป็นอัตราส่วน
2 : 3 : 4 และความยาวรอบรูปเท่ากับ
3X นิ้ว อยากทราบว่าด้านที่ยาวที่สุดยาว
เท่าใด ?

ก. X นิ้ว

ข. $\frac{X}{2}$ นิ้ว

ค. $\frac{4X}{3}$ นิ้ว

ง. $\frac{X}{4}$ นิ้ว

จ. ไม่มีข้อถูก

7. $X\%$ ของ 125 เท่ากับกี่บาท ? $125X$ บาท

$$\begin{array}{r} \text{ก. } \frac{12500}{X} \text{ บาท} \\ \text{ข. } \frac{4}{5X} \text{ บาท} \\ \text{ง. } \frac{4X}{5} \text{ บาท} \\ \text{จ. } \frac{5X}{4} \end{array}$$

8. สามเหลี่ยมรูปหนึ่งมีพื้นที่เท่ากับ AB ตารางหน่วย ถ้าความสูงเป็น B หน่วย ฐานจะยาวกี่หน่วย ?

$$\begin{array}{l} \text{ก. } \frac{1}{2}A \text{ หน่วย} \\ \text{ข. } 2A \text{ หน่วย} \\ \text{ค. } A \text{ หน่วย} \\ \text{ง. } 2 + A \text{ หน่วย} \\ \text{จ. } A - 2 \text{ หน่วย} \end{array}$$

9. สามเหลี่ยมหน้าจั่วมีด้านทั้งสามรวมกัน 12 A นิ้ว ถ้าฐานยาวเท่ากับ $2A$ นิ้ว จะมีพื้นที่เท่ากับข้อใด ?

$$\begin{array}{l} \text{ก. } 24A^4 \text{ ตารางนิ้ว} \\ \text{ข. } A\sqrt{24} \text{ ตารางนิ้ว} \\ \text{ค. } 12\sqrt{A} \text{ ตารางนิ้ว} \\ \text{ง. } A^2\sqrt{24} \text{ ตารางนิ้ว} \\ \text{จ. } \sqrt{12A} \text{ ตารางนิ้ว} \end{array}$$

10. ผู้เข้าสอบภาษาไทยมี $2A$ คน ผู้เข้าสอบคณิตศาสตร์มี $4B$ คน ผู้เข้าสอบภาษาไทยเป็นร้อยละเท่าไรของคณิตศาสตร์ ?

$$\begin{array}{ll} \text{ก. } \frac{50A}{B} & \text{ข. } \frac{25A}{B} \\ \text{ค. } \frac{50B}{A} & \text{ง. } \frac{25B}{A} \\ \text{จ. } \frac{15A}{2B} & \end{array}$$

11. แดงทำงาน $2A$ ของงานในเวลา $3B$ วัน ถ้าทำงาน $4C$ วัน จะได้งานเท่าไร ?

$$\begin{array}{l} \text{ก. } \frac{8AC}{3B} \text{ ของงาน} \\ \text{ข. } \frac{2AB}{3C} \text{ ของงาน} \\ \text{ค. } \frac{2C}{3BA} \text{ ของงาน} \\ \text{ง. } \frac{3C}{2AB} \text{ ของงาน} \\ \text{จ. ไม่มีข้อถูก} \end{array}$$

12. จากสมการ $k = (x - c)g$ และ $k = g$ ดังนั้น x มีค่าเท่ากับข้อใด ?

$$\begin{array}{ll} \text{ก. } 2k - c & \text{ข. } 1 - c \\ \text{ค. } 1 + c & \text{ง. } 2g - c \\ \text{จ. } kg - 2c & \end{array}$$

13. ถ้า $2k = x$, $3c = 4g$ และ x เป็นสองเท่าของ g แล้วข้อใดถูกต้อง ?

$$\begin{array}{ll} \text{ก. } 2k = 3g & \text{ข. } 4k = 6c \\ \text{ค. } k = 2c & \text{ง. } 4k = 3c \\ \text{จ. } 4k = 5c & \end{array}$$

14. $\sqrt{\frac{x}{y}} = ?$

ก. $\sqrt{xy}$

ค. $\frac{\sqrt{x}}{\sqrt{y}}$

จ. $\frac{x}{\sqrt{y}}$

ข. $\frac{1}{\sqrt{xy}}$

ง. $\frac{\sqrt{x}}{y}$

15. $\frac{A}{5}$ ของ 120 น้อยกว่า $\frac{A}{3}$ ของ 75

อยู่เท่าใด ?

ก. 5

ค. 15

จ. 5A

ข. 10

ง. 25

16. เลขจำนวนใด ? หาดด้วย A แล้วได้ผลลัพธ์เท่ากับ 2A

ก. 2

ค. A^2

จ. $3A^2$

ข. A

ง. $2A^2$

17. $21 + 18 - 21 + 30 - 18 + 15 - 30$ เท่ากับ
ข้อใด ?

ก. 15

ค. 21

จ. 32

ข. 18

ง. 30

18. $P^3 \div P^6 \times C^3 = ?$

ก. $(CP)^{3-3}$

ค. P^{-3}

จ. CP^3

ข. $\frac{C^3}{P^3}$

ง. CP^{-3}

19. เลขจำนวนใดที่หารด้วย 100 ได้ผลลัพธ์เท่ากับ 1,000

ก. 10

ค. 1,000

จ. 100,000

ข. 100

ง. 10,000

20. เลขจำนวนใด ? เมื่อนำมาคูณ 100 แล้วได้เท่ากับ 3.6

ก. 0.036

ค. 36

จ. 3600

ข. 0.36

ง. 360

21. $3(10)^4 + 2(10)^3 + 5(10)^2 + 1(10)^1 + 2$
มีค่าเท่ากับข้อใด ?

ก. 31412

ค. 325012

จ. 352012

ข. 32512

ง. 35212

22. เศษส่วนต่อไปนี้จำนวนใดมีค่ามากที่สุด ?

ก. $\frac{N}{0.01}$

ค. $\frac{N}{0.0001}$

จ. $\frac{N}{0.000001}$

ข. $\frac{N}{0.001}$

ง. $\frac{N}{0.00001}$

23. $6^3 + 0^5 - 2 = ?$

ก. 0

ค. 3

จ. 5

ข. 2

ง. 4

24. $\frac{5}{0.5} \times \frac{5}{0.5} = ?$

ก. 10

ข. 20

ค. 25

ง. 50

จ. 100

25. ถ้า $\frac{A}{2} = \frac{32}{A^2}$ แล้ว $A+5$ มีค่าเท่ากับ

เท่าใด ?

ก. 4

ข. 5

ค. 9

ง. 13

จ. 25

26. $48 + 12$ ของ $100 - 80 = ?$

ก. 320

ข. 3,200

ค. 6,400

ง. 8,400

จ. 11,168

27. $(9 \times 2) + (9 \times 8) = 9 \times d$ แล้ว d มีค่าเท่ากับจำนวนใด ?

ก. 4

ข. 5

ค. 6

ง. 9

จ. 10

28. $1110 + 11111 + 10001 = ?$

ก. 22221

ข. 22222

ค. 22201

ง. 22122

จ. 21012

29. $(7 \times \frac{1}{10}) + (4 \times \frac{1}{10^2}) + (6 \times \frac{1}{10^3}) =$

เท่าใด ?

ก. 0.742

ข. 0.74

ค. 7.46

ง. 0.746

จ. 74.6

30. 30 1 2 35 3 4 40?....

ก. 3 4

ข. 3 5

ค. 4 5

ง. 5 6

จ. 6 7

31. 18 3 6 24 4 6 15 3

....?....

ก. 6

ข. 5

ค. 4

ง. 3

จ. 2

32. 2 0 2 1 2?....

ก. 0

ข. 1

ค. 2

ง. 3

จ. 4

33. 15 20 17 22 19 24

....?.....

ก. 21

ข. 23

ค. 25

ง. 27

จ. 29

34. 7 14 9 18 11 22?....

ก. 10

ข. 11

ค. 12

ง. 13

จ. 14

35. -6 3 -5 4 -4 5?....

ก. 6

ข. -6

ค. 3

ง. -3

จ. 5

36. 2 4 8 3 6 12 4 8

.....?

ก. 12 ข. 16

ค. 24 ง. 32

จ. 42

37. 3 2 6 4 2 8 5 2

.....?

ก. 9 ข. 10

ค. 11 ง. 12

จ. 13

38. 4 5 8 6 7 10 8 9

.....?

ก. 10 ข. 11

ค. 12 ง. 13

จ. 14

39. 8 9 11 13 14 16 18 19

.....?

ก. 23 ข. 22

ค. 21 ง. 20

จ. 19

40. 24 12 22 11 20 10 18

.....?

ก. 12 ข. 11

ค. 10 ง. 9

จ. 8

41. 36 6 64 8 100 10

.....?

ก. 121 ข. 132

ค. 144 ง. 156

จ. 184

42. 7 3 5 6 2 4 5?

ก. 5 ข. 4

ค. 3 ง. 2

จ. 1

43. 3 7 8 4 7 6 5 7

.....?

ก. 4 ข. 5

ค. 6 ง. 7

จ. 8

44. 15 16 14 17 13 18

.....?

ก. 16 ข. 15

ค. 14 ง. 13

จ. 12

45. 8 9 4 8 10 16 8 11 5 8

12?

ก. 20 ข. 25

ค. 30 ง. 35

จ. 40

46. 6 1 6 5 2 7 4 3?

ก. 6 ข. 7

ค. 8 ง. 9

จ. 10

57. ร้านค้าแห่งหนึ่ง ขายแปลงสีฟันอันละ 2.80 บาท ถ้าซื้อเป็นโหลคิดโหลละ 24 บาท
อยากทราบว่า ถ้าขายเป็นโหล เขาลดราคา
ไปที่เปอร์เซ็นต์
- ก. 3% ข. $3\frac{1}{3}\%$
ค. $9\frac{1}{11}\%$ ง. 10%
จ. ผิดทุกข้อ
58. หลังเวลา 10.56 นาฬิกา อีก 3 ชั่วโมง
58 นาที เป็นเวลาเท่าไร
- ก. 16.54 ข. 14.54
ค. 16.15 ง. 14.15
จ. ไม่มีข้อถูก
59. เชือกไหมชนิดหนึ่งราคาหลาละ 12 บาท
ถ้าซื้อ 63 นิ้ว จะต้องจ่ายเงินเท่าไร
- ก. 20 บาท ข. 21 บาท
ค. 22 บาท ง. 23 บาท
จ. ไม่มีข้อถูก
60. พ่อค้าลดราคาสินค้าต้อนรับปีใหม่ โดยลด
40% ของราคาที่ปิดป้าย อยากทราบว่าราคา
ที่ลดเป็นกี่เปอร์เซ็นต์ของราคาขาย
- ก. 30% ข. $33\frac{1}{3}\%$
ค. 50% ง. 150%
จ. ไม่มีข้อถูก

ประวัติย่อผู้วิจัย

ประวัติย่อผู้วิจัย

ชื่อ	นางสาวนุชริน ไบโพธิ์
เกิดวันที่	2 เดือน มิถุนายน พุทธศักราช 2515
สถานที่เกิด	อำเภอเมือง จังหวัดนครราชสีมา
สถานที่อยู่ปัจจุบัน	770 ซอยสวนพลู 3 ทุ่งมหาเมฆ สาทร กรุงเทพฯ
ตำแหน่งหน้าที่การงานปัจจุบัน	อาจารย์ 1 ระดับ 4
สถานที่ทำงานปัจจุบัน	สถาบันเทคโนโลยีราชมงคล วิทยาเขตบพิตรพิมุข มหาเมฆ
ประวัติการศึกษา	
พ.ศ. 2533	มัธยมศึกษาตอนปลาย โรงเรียนบุญเหลือวิทยานุสรณ์ จ.นครราชสีมา
พ.ศ.2537	ค.บ. (การวัดผลการศึกษา) (เกียรตินิยมอันดับ 1) สถาบันราชภัฏนครราชสีมา
พ.ศ.2544	กศ.ม. (การวัดผลการศึกษา) มหาวิทยาลัยศรีนครินทรวิโรฒ