

การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
ในข้อสอบแบบต่อเนื่องสองขั้นตอน ระหว่างวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก
กับวิธีแมนเทิล-แฮนส์เซลทั่วไป



ปริญญานิพนธ์
ของ
อัชฌาภรณ์ ตันติดำรงกุล

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวิจัยและพัฒนาศักยภาพมนุษย์

สิงหาคม 2560

การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
ในข้อสอบแบบต่อเนื่องสองขั้นตอน ระหว่างวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก
กับวิธีแมนเทิล-แฮนส์เซลทั่วไป



เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวิจัยและพัฒนาศักยภาพมนุษย์

สิงหาคม 2560

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
ในข้อสอบแบบต่อเนื่องสองขั้นตอน ระหว่างวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก
กับวิธีแมนเทิล-แฮนส์เซลทั่วไป



เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต สาขาวิชาการวิจัยและพัฒนาศักยภาพมนุษย์
สิงหาคม 2560

อรรถาภิธานศัพท์ (2560). การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่
ต่างกันของข้อสอบในข้อสอบแบบต่อเนื่องสองขั้นตอน ระหว่างวิธีวิเคราะห์ ฟังก์ชัน
การจำแนกแบบโลจิสติกกับวิธีแมนเทิล-แฮนส์เซลทั่วไป. ปรินซิเพิล กศ.ม. (การวิจัย
และพัฒนาศักยภาพมนุษย์ แผนวิชาการทดสอบและวัดผลการศึกษา). กรุงเทพฯ:
บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. อาจารย์ที่ปรึกษาปรินซิเพิล:
ผู้ช่วยศาสตราจารย์ ดร.เรืองเดช ศิริกิจ , อาจารย์ ดร.วิไลลักษณ์ ลังกา.

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อการศึกษาและเปรียบเทียบประสิทธิภาพในการตรวจสอบ
การทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกกับวิธีแมนเทิล-
แฮนส์เซลทั่วไป ในข้อสอบแบบต่อเนื่องสองขั้นตอนที่มีการตรวจให้คะแนนหลายค่า ในกลุ่มผู้สอบที่
มีเพศต่างกัน กลุ่มตัวอย่างที่ใช้ในการศึกษาครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา
2559 ของโรงเรียนขนาดใหญ่ในจังหวัดกรุงเทพมหานคร สังกัดสำนักงานเขตพื้นที่การศึกษา
มัธยมศึกษาเขต 1 ซึ่งได้มาจากการสุ่มแบบแบ่งกลุ่ม (Cluster Random Sampling) จากโรงเรียน 9
โรงเรียน จำนวน 1,021 คน เป็นนักเรียนชาย 508 คน และเป็นนักเรียนหญิง 513 คน เครื่องมือที่ใช้
ในการวิจัยเป็นข้อสอบแบบต่อเนื่องสองขั้นตอนเรื่องความเท่ากันทุกประการ จำนวน 20 ข้อ มีความ
ยากง่ายอยู่ระหว่าง 0.29–0.61 และมีค่าอำนาจจำแนกอยู่ระหว่าง 0.16–0.45

ผลการวิจัยพบว่า

วิธีการวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก พบข้อสอบที่ทำหน้าที่ต่างกัน 4 ข้อ ได้แก่
ข้อ 3, 8, 16 และ 20 วิธีแมนเทิล-แฮนส์เซลทั่วไป พบข้อสอบที่ทำหน้าที่ต่างกัน 3 ข้อ ได้แก่ข้อ 3 ,
8 และ 16

วิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกมีความถูกต้องของการตรวจสอบเท่ากับ
33.33% สูงกว่าวิธีแมนเทิล-แฮนส์เซลทั่วไปซึ่งมีความถูกต้องเท่ากับ 22.22% และทั้งสองวิธีมีความ
คลาดเคลื่อนของการตรวจสอบเท่ากับ 9.09%

COMPARISON OF THE EFFICIENCY OF LOGISTIC DISCRIMINANT FUNCTION
ANALYSIS AND GENERALIZED MANTEL–HAENSZEL PROCEDURES FOR DETECTING
DIFFERENTIAL ITEM FUNCTIONING ON TWO–STEPS CONTINUOUS TESTS.



Presented in Partial Fulfillment of the Requirements for the
Master of Education Degree in Research and Development on Human Potentials
at Srinakharinwirot University
August 2017

Achaporn Tantidumrongkul. (2017). *Comparison of the Efficiency of Logistic Discriminant Function Analysis and Generalized Mantel–Haenszel Procedures for Detecting Differential Item Functioning on Two–Steps Continuous Tests*. Master's thesis. M.Ed. (Research and development on Human Potential: Educational Testing and Measurement). Bangkok: Graduate School, Srinakarinwirot University.
Advisor Committee: Asst. Prof. Dr. Ruangdach Sirikit, Dr. Wilailak Langka.

The objectives of this research were to study and to compare of the efficiency of The Logistic Discriminant Function Analysis and Generalized Mantel-Haenszel Procedures on detecting differential item functioning (DIF) based on polytomously scores on two-steps continuous tests for different gender groups. The sample consisted of one thousand and twenty one Mathayom Suksa Two students, consisting of five hundred and eight males and five hundred and thirteen females studying in the semester of the 2016 academic year, in nine schools under the authority of the authority of The Secondary Educational Service Area, Office one The instrument was two steps continuous tests on the congruence of twenty items, difficulty index ($p=0.29-0.61$), discrimination index ($r=0.16-0.45$).

The results were revealed as follows;

The Logistic Discriminant Function Analysis found Differential Item Functioning (DIF) with four items (3,8,16 and 20)

The Generalized Mantel-Haenszel found Differential Item Functioning (DIF) with three items (3,8 and 16)

The Logistic Discriminant Function Analysis is more efficient the Generalized Mantel-Haenszel in terms of detecting differential item functioning in polytomously scored items and the power rate of the Logistic Discriminant Function Analysis 33.33% higher than Generalized Mantel-Haenszel 22.22 % and both of instruments has The Type I error rate in the of Logistic Discriminant Function Analysis, while the Generalized Mantel-Haenszel is 9.09%.

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี เพราะผู้วิจัยได้รับความกรุณาอย่างยิ่งจาก ผู้ช่วยศาสตราจารย์ ดร. เรืองเดช ศิริกิจ ประธานกรรมการควบคุมปริญญานิพนธ์ อาจารย์ ดร.วิไลลักษณ์ ลังกา กรรมการควบคุมปริญญานิพนธ์ ที่ได้เสียสละเวลาอันมีค่าในการให้คำปรึกษา แนะนำ ในการทำงานวิจัย อีกทั้งยังช่วยเหลือและแก้ไขข้อบกพร่องต่างๆ อีกทั้งยังให้กำลังใจในการทำงาน จนวิจัยฉบับนี้สำเร็จลุล่วงไปได้ด้วยดี ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูงไว้ ณ ที่นี้

ขอกราบขอบพระคุณคณะกรรมการสอบปริญญานิพนธ์ผู้ช่วยศาสตราจารย์ ดร. สุวิมล กฤษณฤาสน์ ผู้ช่วยศาสตราจารย์ ดร. เรืองเดช ศิริกิจ อาจารย์ ดร.วิไลลักษณ์ ลังกา และ ผู้ช่วยศาสตราจารย์ ดร. สุรัชย์ มีชาญ ที่ท่านให้ความกรุณาแนะนำแนวทางและข้อเสนอแนะต่างๆ ในการทำปริญญานิพนธ์ฉบับนี้ให้มีความสมบูรณ์มากยิ่งขึ้น

ขอกราบขอบพระคุณคณาจารย์ในภาควิชาการวัดผลและวิจัยการศึกษาทุกท่านที่ให้ความรู้ แก่ผู้วิจัย ในการศึกษาตามหลักสูตรการวิจัยและพัฒนาศักยภาพมนุษย์ (แขนงการทดสอบและวัดผล การศึกษา) ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง

ขอกราบขอบพระคุณ ดร. มนตา ตูลย์เมธากาญจนา ดร. นวรินทร์ ตาก้อนทอง อาจารย์ ปานवासน์ มหาลวเลิศ อาจารย์ชัยพร นาควิเชียร อาจารย์พรวิวัฒน์ ศรียะพันธุ์ ที่ได้ให้ความ อนุเคราะห์รับเป็นผู้เชี่ยวชาญในการตรวจสอบแก้ไขเครื่องมืองานวิจัย โดยได้ให้คำแนะนำที่เป็น ประโยชน์ในการแก้ไขข้อบกพร่องต่างๆเป็นอย่างดี

ขอกราบขอบพระคุณท่านผู้อำนวยการ หัวหน้ากลุ่มสาระการเรียนรู้คณิตศาสตร์และคุณครู ทุกท่านและนักเรียนของโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 1 ที่เป็น กลุ่ม ตัวอย่าง ที่ได้ให้ความร่วมมือในการจัดการสอบและเก็บรวบรวมข้อมูลให้กับผู้วิจัยทำให้การทำวิจัย ในครั้งนี้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบคุณเพื่อน ๆ พี่ ๆ น้องๆ ทุกคนที่ได้ให้คำแนะนำ ช่วยเหลือและเป็นกำลังใจในการทำ วิจัยในครั้งนี้ให้สำเร็จลุล่วงไปได้ด้วยดี

ขอขอบพระคุณบิดามารดา และครอบครัว ที่ให้การสนับสนุนด้านการศึกษา คอยเป็นกำลังใจที่ ดีตลอดระยะเวลาที่ศึกษาและทำงานวิจัยจนประสบความสำเร็จ

คุณค่าและประโยชน์ใดๆ อันพึงบังเกิดขึ้นจากปริญญานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบเป็นเครื่อง บูชาพระคุณบิดา-มารดาครูอาจารย์ตลอดจนผู้มีพระคุณทุกท่าน ที่ให้การอบรมสั่งสอน ประสิทธิ์ประสาท ความรู้แก่ผู้วิจัย

อัชฌาภรณ์ ตันติดำรงกุล

สารบัญ

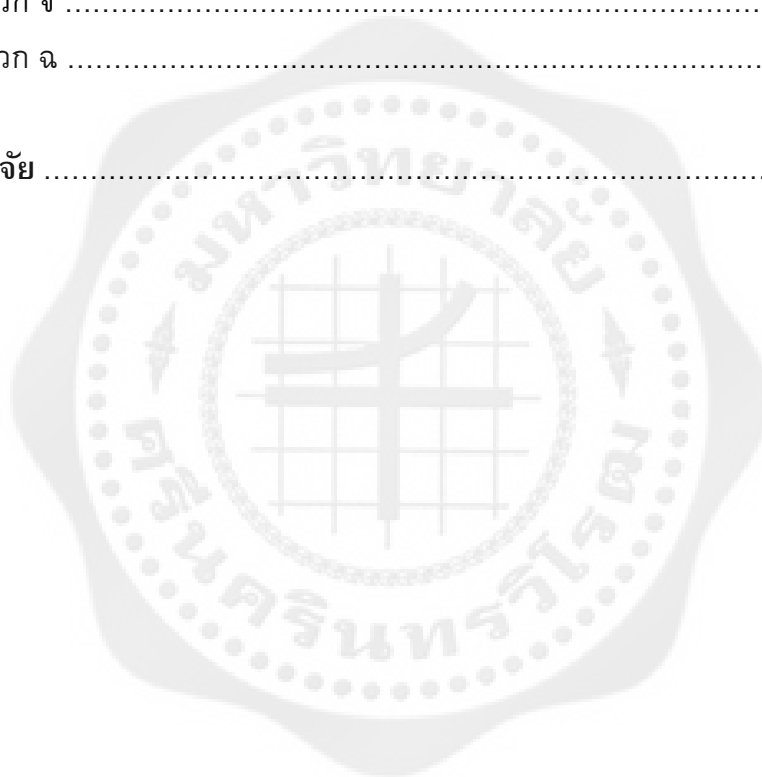
บทที่	หน้า
1 บทนำ	1
ภูมิหลัง	1
ความมุ่งหมายของการวิจัย	6
ความสำคัญของการวิจัย	6
ขอบเขตของการวิจัย	7
นิยามศัพท์เฉพาะ	8
กรอบแนวคิดในการวิจัย	10
2 เอกสารและงานวิจัยที่เกี่ยวข้อง	12
เอกสารที่เกี่ยวข้องกับข้อสอบแบบต่อเนื่องสองขั้นตอน	13
ข้อสอบแบบต่อเนื่องสองขั้นตอน	13
เกณฑ์การให้คะแนนข้อสอบแบบต่อเนื่องสองขั้นตอน	13
ข้อดีและข้อจำกัดของข้อสอบแบบต่อเนื่องสองขั้นตอน	14
เอกสารและงานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ	15
ความหมายของการทำหน้าที่ต่างกันของข้อสอบ	15
สาเหตุที่ก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบ	15
ลักษณะของข้อสอบที่แสดงการทำหน้าที่ต่างกัน	17
ประเภทของการทำหน้าที่ต่างกันของข้อสอบ	18
หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ	21
วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ	24
เกณฑ์เปรียบเทียบคุณภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ	26
งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ	27
เอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธี	
Polytomous SIBTEST	32
แนวคิดและหลักการของวิธี Polytomous SIBTEST	32
งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันโดยใช้วิธี Polytomous SIBTEST	39

สารบัญ (ต่อ)

บทที่	หน้า
2 (ต่อ)	
เอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธี	
GMH	41
แนวคิดและหลักการของวิธี MH	41
แนวคิดและหลักการของวิธี GMH	46
เอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธี	
LDFA	47
แนวคิดและหลักการของวิธี LDFA	51
กระบวนการตรวจสอบ	52
การทดสอบสมมติฐาน	53
งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันโดยใช้วิธี LDFA	54
3 วิธีดำเนินการวิจัย	56
การกำหนดประชากรและกลุ่มตัวอย่าง	56
การสร้างเครื่องมือที่ใช้ในการวิจัย	58
การเก็บรวบรวมข้อมูล	63
การวิเคราะห์ข้อมูล	64
สถิติที่ใช้ในการวิจัย	64
4 ผลการวิเคราะห์ข้อมูล	69
สัญลักษณ์ที่ใช้ในการวิเคราะห์ข้อมูล	69
ผลการวิเคราะห์ข้อมูล	70
5 สรุปผล อภิปรายผลและข้อเสนอแนะ	79
สรุปผลการวิจัย	80
อภิปรายผลการวิจัย	81
ข้อเสนอแนะ	83
บรรณานุกรม	85

สารบัญ (ต่อ)

บทที่	หน้า
ภาคผนวก	91
ภาคผนวก ก	92
ภาคผนวก ข	94
ภาคผนวก ค	97
ภาคผนวก ง	100
ภาคผนวก จ	125
ภาคผนวก ฉ	140
ประวัติย่อผู้วิจัย	143



บัญชีตาราง

ตาราง	หน้า
1 ตัวอย่างเกณฑ์การให้คะแนนแบบรวมของข้อสอบแบบต่อเนื่องสองขั้นตอน	14
2 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีการตรวจให้คะแนนแบบ ทวิภาค (Dichotomous DIF) และพหุวิภาค (Polytomous DIF)	25
3 คุณภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ	26
4 สรุปการศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบพหุ วิภาคในประเทศไทย	31
5 ผลการตอบข้อสอบข้อหนึ่งระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มี คะแนนรวม j	42
6 สัดส่วนของผู้สอบระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ณ ระดับ คะแนน j	42
7 แสดงผลการตอบข้อสอบของกลุ่มผู้สอบ ในแต่ละลำดับชั้นคะแนน	46
8 รายชื่อโรงเรียน จำนวนห้องเรียน และจำนวนนักเรียน	57
9 ตารางวิเคราะห์เนื้อหาวิชาคณิตศาสตร์ เรื่องความเท่ากันทุกประการ	60
10 เกณฑ์การให้คะแนนแบบรวมของแบบทดสอบแบบต่อเนื่องสองขั้นตอน	63
11 ค่าสถิติพื้นฐานของแบบทดสอบวิชาคณิตศาสตร์ เรื่องความเท่ากันทุกประการ 20 ข้อ	70
12 ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบ ของแบบทดสอบ จำนวน 20 ข้อ	71
13 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชัน เชิง จำแนกแบบโลจิสติก	73
14 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทิล-แฮนส์เซลทั่วไป	74
15 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโพลีโทมัสชิปเทสต์	75
16 การเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี วิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก วิธีแมนเทิล-แฮนส์เซลทั่วไปและวิธีโพลี โทมัสชิปเทสต์	77
17 อัตราความถูกต้องและอัตราความคลาดเคลื่อนของการตรวจสอบโดยมีวิธีโพลี โทมัสชิปเทสต์เป็นเกณฑ์	78
18 ค่าดัชนีความสอดคล้อง(Index of item-objective congruence:IOC) ของ แบบทดสอบแบบต่อเนื่องสองขั้นตอนเรื่องความเท่ากันทุกประการ	95

บัญชีตาราง

ตาราง

หน้า

- 19 ผลการวิเคราะห์ค่าความยากง่าย (p) ค่าอำนาจจำแนก (r) และค่าความเชื่อมั่น
ของแบบทดสอบแบบต่อเนื่องสองขั้นตอน (Cronbach's Alpha) 98



บัญชีภาพประกอบ

ภาพประกอบ	หน้า
1 กรอบแนวคิดในการวิจัย	11
2 ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF)	19
3 ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (Non-Unidirectional DIF)	20
4 ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียวกัน (Unidirectional DIF)	21
5 ขั้นตอนการสร้างเครื่องมือที่ใช้ในการวิจัย	58
6 ผลการตรวจสอบความเป็นมิติเดียวกันของข้อสอบ	72



บทที่ 1

บทนำ

ภูมิหลัง

แบบทดสอบ (Test) เป็นเครื่องมือที่นิยมใช้ในการทดสอบ การทดสอบเป็นวิธีการที่สำคัญอย่างหนึ่งสำหรับการวัดผลการเรียนรู้ของผู้เรียน ผู้เรียนจะเกิดการเรียนรู้เมื่อได้มีการเปลี่ยนแปลงปริมาณหรือคุณภาพของความรู้ ความสามารถ พฤติกรรม หรือลักษณะทางจิตใจ ถ้าการเปลี่ยนแปลงเป็นไปในทิศทางที่พึงประสงค์ตามจุดมุ่งหมายของหลักสูตร อันเป็นผลมาจากการประสบการณ์การเรียนการสอนที่ผู้สอนจัดขึ้น เราเรียกว่าผู้เรียนเกิดผลสัมฤทธิ์ทางการเรียนรู้ตามหลักสูตร (ศิริชัย กาญจนวาสี. 2556: 165) แบบทดสอบจึงเป็นเครื่องมือวัดผลที่ผู้สอนสร้างขึ้นเพื่อใช้ทดสอบผู้เรียนซึ่งประกอบด้วยแบบทดสอบประเภทต่าง ๆ ได้แก่ แบบเลือกตอบ แบบถูกผิด แบบจับคู่ แบบเติมคำ แบบเขียนตอบ แบบต่อเนื่อง และแบบแสดงวิธีทำ เป็นต้น (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. 2546: 21)

ข้อสอบแบบหลายตัวเลือก (Multiple Choice) มีลักษณะเป็นข้อสอบแบบเลือกตอบที่นิยมใช้กันอย่างกว้างขวาง เพราะสามารถใช้วัดผลการเรียนรู้ ทั้งความรู้ ความเข้าใจ การนำไปใช้ และผลการเรียนรู้ขั้นสูงได้ สามารถใช้วัดได้ครอบคลุมเนื้อเรื่องตามโครงสร้างอย่างมีประสิทธิภาพ และนำไปพัฒนาเป็นแบบทดสอบมาตรฐานได้ ข้อสอบแบบหลายตัวเลือกเป็นข้อสอบที่ให้ผู้สอบเลือกคำตอบจากตัวเลือกที่กำหนดให้ โดยจะประกอบด้วย 2 ส่วน ได้แก่ ตัวคำถาม (Stem) และตัวเลือก (Alternatives หรือ Options) ซึ่งนิยมใช้ 3-6 ตัวเลือก ในส่วนของตัวเลือกประกอบด้วยตัวเลือกที่เป็นคำตอบที่ถูกเรียกว่า ตัวคำตอบ (Answer หรือ Key) 1 ตัว ส่วนที่เหลือเป็นตัวเลือกที่ผิด เรียกว่า ตัวลวง (Distracters) (ศิริชัย กาญจนวาสี. 2556: 196-197) โดยมีหลักการให้คะแนนโดยพิจารณาจากการตอบถูกผิด เช่น ตอบผิดได้ 0 คะแนน และตอบถูกได้ 1 คะแนน

แม้ว่าข้อสอบแบบหลายตัวเลือกที่มีการให้คะแนนเป็น 0 และ 1 จะเป็นที่ยอมรับในปัจจุบันแต่จะเห็นได้ว่าการสอบระดับชาติจะมีรูปแบบข้อสอบที่หลากหลายมากขึ้น ซึ่งเป็นข้อสอบที่มีการให้คะแนนแบบหลายค่า คือมีการให้คะแนนมากกว่า 0 และ 1 ข้อสอบที่มีการให้คะแนนแบบหลายค่า (Polytomously scored items) มีหลายประเภท เช่น ข้อสอบความเรียง (Essay items) ข้อสอบแบบเลือกตอบจากแต่ละหมวดที่สัมพันธ์กัน ข้อสอบแบบต่อเนื่องสองขั้นตอนหรือเป็นการประเมินงานปฏิบัติ (Performance tasks assessment) ข้อสอบแบบต่อเนื่องเป็นข้อสอบประเภทหนึ่งที่มีการให้คะแนนแบบหลายค่า เพราะเป็นข้อสอบที่ผสมผสานหลายรูปแบบเข้าด้วยกัน เช่น ข้อสอบแบบเลือกตอบกับแบบเลือกตอบ ข้อสอบแบบเลือกตอบกับแบบถูกผิด ข้อสอบแบบเลือกตอบกับแบบเขียนตอบ และข้อสอบแบบถูกผิดกับแบบเขียนตอบ ตัวอย่างข้อสอบแบบต่อเนื่องจึงมีได้หลายลักษณะ ในที่นี้จะกล่าวถึงข้อสอบแบบต่อเนื่องสองขั้นตอน ซึ่งข้อสอบแบบต่อเนื่องสองขั้นตอน จะมีทั้งหมด 2 ตอน ตอนที่ 1 จะเป็นคำถามและคำตอบที่มีลักษณะเป็นตัวเลือกในแบบเป็นจริงหรือเท็จ

และในตอนที่ 2 จะเป็นการให้เหตุผลหรือให้คำอธิบายประกอบตัวเลือกในขั้นตอนที่ 1 มีลักษณะเป็นแบบเลือกตอบซึ่งข้อสอบแบบต่อเนื่องสองขั้นตอนนี้จะสามารถที่จะวินิจฉัยข้อบกพร่องของผู้สอบ และสามารถวัดกระบวนการให้เหตุผลได้ ทำให้ทราบว่าผู้สอบมีความรู้ ความเข้าใจและสามารถให้เหตุผลประกอบการอธิบายได้ถูกต้องหรือไม่ ถ้าผู้สอบไม่สามารถที่จะให้เหตุผลในการตอบได้อย่างถูกต้อง ก็แสดงว่าผู้สอบนั้นไม่มีความรู้ความเข้าใจในเนื้อหาสาระที่ต้องการวัด(สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. 2555: 61)

ในการสร้างแบบทดสอบนั้น นักวัดผลทั้งหลายต่างก็พยายามอย่างยิ่งที่จะอธิบายคำว่า “แบบทดสอบที่ดี” ว่ามีลักษณะอย่างไร ทั้งนี้เพราะแบบทดสอบมีความสำคัญเกี่ยวข้องกับชีวิตและอนาคตของคนเป็นอย่างมากดังนั้นถ้าหากแบบทดสอบด้อยคุณภาพ ผลลัพธ์ที่ออกมาก็จะด้อยคุณภาพและเชื่อถือไม่ได้ ซึ่งแบบทดสอบที่ดีควรมีลักษณะดังนี้ ความเที่ยงตรง (Validity) ความเชื่อมั่น (Reliability) ความเป็นปรนัย (Objectivity) ความยากง่ายของข้อสอบ (Item Difficulty) อำนาจจำแนก (Discriminating Power) ความยุติธรรม (Fairness) งามลึก (Searching) จำเพาะเจาะจง (Definite) ยั่วยุ (Exemplary) และประสิทธิภาพ (Efficiency) (สุรศักดิ์ อมรรัตนศักดิ์; และอนุสรณ์ สกุลคุ. 2549: 31) การพัฒนาแบบทดสอบขึ้นมาจะต้องคำนึงถึงคุณภาพของข้อสอบว่าข้อสอบนั้นสามารถวัดได้ตรงตามความสามารถที่ต้องการหรือไม่ แบบทดสอบที่ดีจะต้องเป็นแบบทดสอบที่มีความเที่ยงตรง วัดได้ตรงตามสิ่งที่ต้องการวัด มีความยุติธรรมกับผู้สอบทุกกลุ่มไม่เข้าข้างกลุ่มใดกลุ่มหนึ่งหรือทำให้กลุ่มใดกลุ่มหนึ่งเสียประโยชน์ ข้อสอบที่เอื้อต่อผู้สอบกลุ่มใดกลุ่มหนึ่งอาจทำให้ผู้สอบกลุ่มดังกล่าวมีความน่าจะเป็นในการตอบถูกต้องมากกว่ากลุ่มอื่นที่ระดับความสามารถเดียวกัน ลักษณะการทำหน้าที่ของข้อสอบประเภทนี้เรียกว่า “การทำหน้าที่ต่างกันของข้อสอบ” (ชวาล แพรัตกุล. 2516: 87-88)

การทำหน้าที่ต่างกันของข้อสอบหรือความลำเอียงของข้อสอบมักเกิดจากปฏิสัมพันธ์ (Interaction) ระหว่างลักษณะของข้อสอบกับลักษณะของกลุ่มผู้สอบ ที่มีผลต่อคะแนนสอบ ลักษณะข้อสอบเช่น การใช้ภาษา ความเฉพาะของเนื้อหา โครงสร้างการวัด วิธีการที่เลือกใช้ ลักษณะของกลุ่มผู้สอบ เช่น เพศ เชื้อชาติ วัฒนธรรม ภูมิหลัง ภูมิภาค เป็นต้น เมื่อทั้งสองปัจจัยมีปฏิสัมพันธ์ต่อกัน สำหรับข้อสอบข้อใด ทำให้เกิดการทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning: DIF) จึงมีผลต่อคะแนนของข้อสอบนั้น อันอาจนำไปสู่ความลำเอียงของข้อสอบด้านต่างๆ เช่น ความลำเอียงด้านเพศ (Gender Bias) ความลำเอียงด้านเชื้อชาติ (Race Bias) ความลำเอียงทางวัฒนธรรม (Culture Bias) เป็นต้น (ศิริชัย กาญจนวาสี. 2556: 250)

การทำหน้าที่ต่างกันของข้อสอบนั้นอาจเกิดได้จากลักษณะของกลุ่มผู้สอบที่ต่างกัน ในลักษณะต่างๆ ความแตกต่างทางเพศถือว่าเป็นเรื่องที่ได้รับการสนใจมานาน โดยเฉพาะการศึกษาความแตกต่างทางเพศกับวิชาคณิตศาสตร์ โดยผลการศึกษาที่ผ่านมาจะพบว่าเพศชายมีความสามารถทางคณิตศาสตร์มากกว่าเพศหญิง (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. 2557: 92) รายงานการศึกษาผลการสอบ PISA ในปี 2012 พบว่าเกิดความแตกต่างกันระหว่างเพศชายและเพศหญิง โดยที่ความแตกต่างทางเพศในการรู้เรื่องคณิตศาสตร์นั้นระบุว่า เพศ

ชายมีความสามารถและทำได้ดีกว่าเพศหญิงในเรื่องของปริภูมิและรูปทรงสามมิติ และเพศหญิงจะมีความสามารถและทำได้ดีกว่าเพศชายในเรื่องของพีชคณิต ซึ่งสอดคล้องกับงานวิจัยของโอ ลีเดีย ลู และมาร์ค วิลสัน (Ou Lydia Liu; & Mark Wilson. 2009: 164-184) ที่ศึกษาการประเมิน PISA วิชาคณิตศาสตร์ในปี 2000 และปี 2003 พบว่าโดยรวมเพศชายมีความได้เปรียบกว่าเพศหญิงในเรื่องของปริภูมิและรูปทรงสามมิติและได้เปรียบในการทำแบบทดสอบแบบเลือกตอบที่มีความซับซ้อนและสอดคล้องกับงานวิจัยของนุชริน ไบโพธิ์ (2544: 80-87) ทำการเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord's χ^2 วิธี Mantel-Haenszel และวิธี SIBTEST พบว่าเพศชายมีความสามารถทางด้านคณิตศาสตร์มากกว่าเพศหญิง และสอดคล้องกับงานวิจัยของแคทเทอร์ลีน เทเลอร์ (Catherine S Taylor. 2012: 246-280) ที่ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้แบบทดสอบรูปแบบผสม ผลการศึกษาพบว่าเพศชายมีความสามารถด้านการระบุ อธิบายและวิเคราะห์ข้อมูลได้ดีกว่าเพศหญิงและมีความถนัดเนื้อหาเรื่องเรขาคณิต การวัดและความน่าจะเป็น จากที่กล่าวมาข้างต้นจะพบว่าความแตกต่างทางเพศนั้นมีผลต่อการประเมินความสามารถทางคณิตศาสตร์เป็นอย่างมาก ซึ่งทำให้แบบทดสอบนั้นเกิดความลำเอียงเข้าข้างเพศใดเพศหนึ่ง ทำให้เกิดความไม่เท่าเทียมกันและส่งผลให้แบบทดสอบนั้นด้อยคุณภาพ

จากการศึกษางานวิจัยที่ผ่านมาจะพบว่าเป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบสองค่าเป็นส่วนใหญ่ แต่เนื่องจากการพัฒนาแบบทดสอบทั้งวิธีการตอบและการให้คะแนน ดังนั้นวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบจึงได้มีการปรับเปลี่ยนวิธีการ และพัฒนาให้สามารถใช้ได้กับข้อสอบที่มีการให้คะแนนแบบหลายค่าได้อย่างเหมาะสมและรัดกุมยิ่งขึ้น ซึ่งโปเทนซ่า; และดูรานส์ (อุทัยวรรณ สายพัฒนา. 2547: 3; อ้างอิงจาก Potenza; & Durans.1995) ได้รวบรวมวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ใช้กับข้อสอบที่มีการให้คะแนนแบบหลายค่า และนำเสนอไว้ดังนี้ วิธีการตรวจสอบในกลุ่มวิธี IRT ได้แก่ วิธีการทดสอบอัตราส่วนความน่าจะเป็น (General IRT LR), วิธีพาเชียลเครดิตโมเดล (Partial Credit Model: PCM), วิธีซิปเทสต์แบบหลายค่า (Polytomous SIBTEST) และวิธีพาเชียลเครดิตโมเดลทั่วไป (Generalized Partial Credit Model: GPCM) เป็นต้น ส่วนวิธีการตรวจสอบในกลุ่มวิธีที่ไม่ใช่ IRT ได้แก่ วิธีการทำให้เป็นมาตรฐานแบบหลายค่า (Polytomous STND) วิธีแมนเทล-แฮนส์เซลทั่วไป (Generalized Mantel-Haenszel: GMH) และวิธีการถดถอยโลจิสติกแบบหลายค่า (Polytomous LRDIF) นอกจากนั้นยังมี วิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติกแบบหลายค่า (Logistic Discriminant Function Analysis: LDFA) ซึ่งพัฒนาจากแนวคิดของวิธีการถดถอยโลจิสติกและวิธีการวิเคราะห์เชิงจำแนกกลุ่ม (Discriminant Analysis) เป็นต้น

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่ายังมีผู้ที่ศึกษาไม่มากนัก งานวิจัยของอุทัยวรรณ สายพัฒนา (2547: 97-112) ได้ศึกษาประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี GMH และวิธี Polytomous SIBTEST โดยใช้แบบทดสอบแบบถูกผิด 5 ตัวเลือก ซึ่งเป็นแบบทดสอบที่มีการให้คะแนนแบบหลายค่า ผลการวิจัยพบว่า ในลักษณะของข้อสอบรูปแบบนี้เงื่อนไขความยาวของแบบทดสอบมีผลต่อความถูกต้องและ

ความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ กลุ่มตัวอย่างทุกขนาดส่งผลต่อประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบไม่แตกต่างกัน โดยความถูกต้องในการตรวจสอบและความผิดพลาดในการตรวจสอบด้วยวิธี GMH และวิธี Polytomous SIBTEST มีความผันแปรอย่างเป็นระบบเมื่อขนาดของกลุ่มตัวอย่างหรือความยาวของแบบทดสอบเปลี่ยนไป และมีค่าใกล้เคียงกันในแต่ละขนาดของกลุ่มตัวอย่างและความยาวของแบบทดสอบ งานวิจัยของ บุษวรรษ์ แสนปลื้ม (2550: 67-73) ได้ศึกษาการเปรียบเทียบวิธีชิปเทสต์แบบหลายค่าที่ใช้รูปแบบการคิดคะแนนต่างกันในการตรวจสอบการทำหน้าที่ต่างกันของข้อคำถามวัดโน้ตส์แห่งตน โดยใช้มาตรวัดโน้ตส์แห่งตน ที่มีลักษณะเป็นมาตรประเมินค่า (Rating-Scale) ชนิด 5 ระดับ ผลการวิจัยพบว่า รูปแบบการคิดคะแนนรวมทั้งฉบับมีแนวโน้มที่จะระบุข้อคำถามที่ทำหน้าที่ต่างกันได้น้อยกว่ารูปแบบการคิดคะแนนแยกองค์ประกอบ และทั้งสองรูปแบบสามารถระบุข้อคำถามที่ทำหน้าที่ต่างกันสอดคล้องสัมพันธ์กันสูง โดยที่รูปแบบการคิดคะแนนแยกองค์ประกอบมีอัตราความถูกต้องสูงกว่ารูปแบบการคิดคะแนนรวมทั้งฉบับ แต่ก็มียัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่ารูปแบบการคิดคะแนนรวมทั้งฉบับด้วย งานวิจัยของวริษา ชะม้อย (2550: 115-129) ได้เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย (MACS) กับวิธีวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก (LDFA) เมื่อมีเงื่อนไขความยาวของแบบสอบและขนาดของกลุ่มตัวอย่างที่แตกต่างกัน ผลวิจัยพบว่า วิธี LDFA มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงตั้งแต่ร้อยละ 0 ถึงร้อยละ 15.6 และวิธี MACS มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงตั้งแต่ร้อยละ 8.3 ถึงร้อยละ 21.9 โดยวิธี LDFA เกือบทุกเงื่อนไขอยู่ในเกณฑ์ที่กำหนด ส่วนวิธี MACS พบว่าส่วนใหญ่อยู่เกินกว่าเกณฑ์ที่กำหนด เมื่อพิจารณาการเปรียบเทียบอำนาจการทดสอบพบว่า วิธี LDFA มีอำนาจการทดสอบอยู่ในช่วงตั้งแต่ร้อยละ 62.5 ถึงร้อยละ 100 สำหรับวิธี MACS พบว่ามีอำนาจการทดสอบอยู่ในช่วงตั้งแต่ร้อยละ 37.5 ถึงร้อยละ 100 และพบว่าทั้งสองวิธีมีอำนาจการทดสอบในเงื่อนไขความยาวแบบสอบ 30 ข้อสูงกว่าเงื่อนไขความยาวแบบสอบ 40 ข้อทุกเงื่อนไข และงานวิจัยของสุชาติ สิริมินันท์ (2554: 137-157) ได้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า โดยใช้วิธีโพลีโตมัสชิปเทสต์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติก โดยใช้ข้อมูลจำลองภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย คือ รูปแบบของข้อสอบ ความยาวของแบบทดสอบ สัดส่วนของข้อสอบและขนาดตัวอย่าง ผลวิจัยพบว่า ในภาพรวมภายใต้ทุกเงื่อนไข วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความเหมาะสมมากกว่าอีกสองวิธี เพราะมีผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีอัตราความถูกต้องสูงและมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำ

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่ามีวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่าที่น่าสนใจ คือวิธีแมนเทล-แฮนส์เซลทั่วไป (ศิริชัย กาญจนวาสี 2555: 126-127) วิธีแมนเทล-แฮนส์เซลพัฒนาโดย ฮอลแลนด์และทาเยอร์ วิธีนี้เป็นที่ยอมรับว่าเป็นวิธีที่ง่าย สะดวกและประหยัดโดยใช้หลักการของตารางการถ่วง (Contingency

table) แบบประเพณีนิยม (Traditional) ไม่มีการคำนวณทวนซ้ำจึงสามารถนำไปใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ง่าย มีขั้นตอนที่ไม่สลับซับซ้อน ประหยัดเวลาและค่าใช้จ่าย สามารถใช้กับกลุ่มตัวอย่างที่มีขนาดเล็กได้ มีการทดสอบทางสถิติแบบนันทพาราเมตริก (Nonparametric) ซึ่งไม่จำเป็นต้องใช้โมเดลประมาณค่า จากที่กล่าวมาวิธีแมนเทิล-แฮนส์เซลจึงเป็นที่นิยมนำมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบสองค่า

แสมเบอตัน (กาญจนา วันสุนทร. 2538; อ้างอิงจาก Hambleton. 1986) วิธีแมนเทิล-แฮนส์เซลเดิมจึงถูกปรับขยายเป็นวิธีแมนเทิล-แฮนส์เซลทั่วไป (Generalized Mantel-Haenszel: GMH) ที่สามารถนำมาใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ลำดับขั้นของคะแนนผลการตอบมีมากกว่า 2 ลำดับขั้น และลำดับขั้นการตอบไม่จำเป็นต้องมีการจัดเรียง โดยปรับตารางการแจกแจงของคะแนนผลการตอบแบบ $2 \times 2 \times k$ เป็นแบบ $2 \times T \times k$ โดยที่ T เป็นจำนวนลำดับขั้นการตอบ และปรับสถิติที่ใช้ในการทดสอบนัยสำคัญจาก MH_2^2 ที่มี $df = 1$ เป็น GMH_2^2 ที่มี $df = T - 1$ เพื่อให้ใช้ได้กับข้อสอบที่มีการให้คะแนนแบบหลายค่า

วิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก (Logistic Discriminant Function Analysis: LDFA) (สุชาติ สิริมินันท์. 2554: 23) ซึ่งพัฒนาโดยมิลเลอร์และสเปร์ย์ จากแนวคิดของวิธีการถดถอยโลจิสติกและวิธีการวิเคราะห์เชิงจำแนกกลุ่ม (Discriminant Analysis) เพื่อใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบหลายค่า วิธีดังกล่าวอยู่บนโมเดลพื้นฐานที่ใช้สำหรับคัดแยกคนออกเป็นกลุ่มๆ ที่มีส่วนเหมือนกันบนพื้นฐานของการตอบข้อสอบและคะแนนแบบทดสอบรวม โมเดลดังกล่าวสามารถตรวจสอบอิทธิพลของคะแนนแบบทดสอบรวมและคะแนนข้อสอบทำให้มีประสิทธิภาพในการตรวจสอบข้อสอบทั้งข้อสอบที่ทำหน้าที่ต่างกันที่เป็นรูปแบบเดียวกันและข้อสอบทำหน้าที่ต่างกันที่ไม่เป็นรูปแบบเดียวกัน วิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก มีความสะดวกในการแปรผล สามารถใช้กับกลุ่มตัวอย่างที่มีขนาดเล็กได้

วิธีชิปเทสต์แบบหลายค่า (Polytomous SIBTEST) เป็นวิธีที่ปรับขยายมาจากวิธี SIBTEST ที่พัฒนาโดยเชียลีและสเตาร์ท ใช้สถิติการทดสอบแบบนันทพาราเมตริก (Nonparametric) ซึ่งต่างจากวิธี IRT ทั่วไป เนื่องจากไม่ต้องใช้ค่าประมาณความสามารถแฝง วิธีการนี้สามารถใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ทั้งข้อสอบที่เป็นการให้คะแนนแบบสองค่าและคะแนนแบบหลายค่า สามารถตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF) และการทำหน้าที่ต่างกันของกลุ่มข้อสอบ (Differential Bundle Functioning ; DBF) วิธีนี้สามารถวิเคราะห์ได้ทั้งในแบบทดสอบมิติเดียว (unidimensional test) และแบบทดสอบที่เป็นหลายมิติ (multidimensional test) จุดเด่นของวิธีนี้คือ สามารถคำนวณได้ง่าย ไม่ซับซ้อน ประหยัดค่าใช้จ่าย และไม่จำเป็นต้องใช้กลุ่มตัวอย่างขนาดใหญ่ เป็นวิธีที่ยอมรับว่ามีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบสูง บนพื้นฐานทฤษฎี IRT จากการศึกษาของนารายานันและสวามินาธาน (Narayanan; & Swaminathan. 1994: 315-327) ที่เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี MH และวิธี SIBTEST พบว่า ทั้งสองวิธีมีความถูกต้องในการตรวจสอบเท่าเทียมกัน แต่วิธี SIBTEST มีความผิดพลาดประเภทที่ 1 สูงกว่า ส่วนการศึกษาของรุสโซ; และ

สเตาท์ (อุทัยวรรณ สายพัฒนา. 2547: 4-5; อ้างอิงจาก Roussos and Stout. 1996) ที่ศึกษาผลกระทบของกลุ่มตัวอย่างที่มีผลต่อความผิดพลาดประเภทที่ 1 ของวิธี MH และวิธี SIBTEST พบว่า วิธี MH มีความผิดพลาดประเภทที่ 1 สูงกว่า และจิตติมา วรรณศรี (2539: 99-113) ได้เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี MH และ SIBTEST พบว่าความสามารถในการตรวจสอบทั้งสองวิธีเท่าเทียมกันและอัตราความคลาดเคลื่อนชนิดที่ 1 พบว่าวิธีวิธี SIBTEST มีอัตราความคลาดเคลื่อนมากกว่าวิธี MH เล็กน้อย และการศึกษาของยา-ฮุย ซู; และเว่ย-ชาง หวัง (Ya-Hui Su & Wen-Chung Wang. 2005: 313-350) ที่ศึกษาประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย 3 วิธี คือวิธีแมนเทล วิธีแมนเทล-แฮนส์เซลทั่วไป และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก พบว่า วิธีแมนเทล และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกให้อำนาจการตรวจสอบที่สูงกว่าวิธีแมนเทล-แฮนส์เซลทั่วไป

จากที่กล่าวมาข้างต้นและจากการศึกษางานวิจัยที่ผ่านมา พบว่าความแตกต่างทางเพศนั้นเป็นสิ่งที่ส่งผลให้เกิดการทำหน้าที่ต่างกันของข้อสอบในวิชาคณิตศาสตร์โดยเพศชายจะมีความสามารถด้านเรขาคณิต การวัดและการทำข้อสอบในลักษณะของการอธิบายและซับซ้อนได้ดีกว่าเพศหญิง อีกทั้งการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบที่อยู่ในรูปแบบการให้คะแนนแบบหลายค่าโดยใช้ข้อสอบแบบต่อเนื่องสองขั้นตอนนั้นยังไม่มีผู้ศึกษา

ดังนั้นผู้วิจัยจึงสนใจที่จะศึกษาและเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อสอบแบบต่อเนื่องสองขั้นตอน โดยใช้เนื้อหาเรื่องความเท่ากันทุกประการในระดับชั้นมัธยมศึกษาปีที่ 2 เนื่องจากเนื้อหาเรื่องความเท่ากันทุกประการอยู่ในสาระที่ 3 เรขาคณิต และเป็นเนื้อหาที่เหมาะสมกับการวัดกระบวนการให้เหตุผล โดยใช้วิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก กับวิธีแมนเทล-แฮนส์เซลทั่วไป และมีตัวแปรแบ่งกลุ่มคือนักเรียนเพศชายและเพศหญิง ที่ส่งผลต่อการทำหน้าที่ต่างกันของข้อสอบ เพื่อที่จะเป็นแนวทางในการพัฒนาแบบทดสอบให้มีความยุติธรรมต่อไป

ความมุ่งหมายของการวิจัย

การวิจัยครั้งนี้ผู้วิจัยมีจุดมุ่งหมายเพื่อการศึกษาและเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกกับวิธีแมนเทล-แฮนส์เซลทั่วไป ในข้อสอบแบบต่อเนื่องสองขั้นตอนที่มีการตรวจให้คะแนนหลายค่า ในกลุ่มผู้สอบที่มีเพศต่างกัน

ความสำคัญของการวิจัย

การศึกษาครั้งนี้เป็นการเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยใช้ข้อสอบแบบต่อเนื่องสองขั้นตอนที่มีการตรวจให้คะแนนแบบหลายค่า โดยใช้วิธีการตรวจสอบ 2 วิธี คือวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก กับวิธีแมนเทล-แฮนส์เซล

ทั่วไป ซึ่งการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เป็นการให้คะแนนแบบหลายค่ายังมีไม่มากนัก และผลจากการศึกษาครั้งนี้ได้ข้อสรุปว่าวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบวิธีใดที่เหมาะสมกับข้อสอบแบบต่อเนื่องสองขั้นตอน และสามารถนำมาเป็นแนวทางในการพัฒนาแบบทดสอบให้มีมาตรฐาน คุณภาพและความยุติธรรมต่อผู้สอบในทุกกลุ่ม ซึ่งจะทำให้สามารถวัดความรู้ ความสามารถของผู้สอบได้อย่างแท้จริง

ขอบเขตของการวิจัย

ประชากรที่ใช้ในการวิจัย

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2559 ของโรงเรียนขนาดใหญ่ในจังหวัดกรุงเทพมหานคร สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาทั้งหมด 39 โรงเรียน จำนวน 14,299 คน เป็นนักเรียนชาย 6,603 คน และเป็นนักเรียนหญิง 7,696 คน

กลุ่มตัวอย่างที่ใช้ในการวิจัย

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ที่กำลังศึกษาอยู่ในภาคเรียนที่ 2 ปีการศึกษา 2559 ของโรงเรียนขนาดใหญ่ สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 1 ซึ่งได้มาจากการสุ่มแบบแบ่งกลุ่ม (Cluster Random Sampling) จากจำนวน 9 โรงเรียน นักเรียนทั้งหมด 1,021 คน เป็นนักเรียนชาย 508 คน และเป็นนักเรียนหญิง 513 คน

ตัวแปรที่ศึกษา

1. ตัวแปรอิสระ ได้แก่

1.1 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ได้แก่

1.1.1 วิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก (Logistic Discriminant

Function Analysis:LDFA)

1,1,2 วิธีแมนเทิล-แฮนส์เซลทั่วไป (Generalized Mantel-Haenszel: GMH)

2. ตัวแปรตาม ได้แก่

2.1 ประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

2.1.1 ความถูกต้องของการตรวจสอบ

2.1.2 ความคลาดเคลื่อนของการตรวจสอบ

เนื้อหาที่ใช้ในการวิจัย

เนื้อหาที่ใช้ในการวิจัยครั้งนี้เป็นเนื้อหารายวิชาคณิตศาสตร์พื้นฐาน กลุ่มสาระการเรียนรู้คณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 2 เรื่อง ความเท่ากันทุกประการ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551

ข้อจำกัดในงานวิจัย

ในการวิจัยครั้งนี้ พบว่าในการตรวจสอบความเป็นเอกมิตีของแบบทดสอบเรื่องความเท่ากันทุกประการจำนวน 20 ข้อ ไม่สามารถบ่งบอกได้ชัดเจนว่าแบบทดสอบมีความเป็นมิตีเดียว ซึ่งการจะวิเคราะห์โดยใช้ทฤษฎีการตอบสนองข้อสอบต่อไปนั้นถือว่าเป็นการละเมิดข้อตกลงเบื้องต้น แต่จากการศึกษางานวิจัยที่เกี่ยวข้องพบว่า อ้อมใจ คงเสาร์ (2540: 56-58) ที่ตรวจสอบความเป็นเอกมิตีของแบบทดสอบคณิตศาสตร์จำนวน 17 ข้อ และพบว่าไม่เป็นเอกมิตี ซึ่งสอดคล้องกับงานวิจัยของ วริษา ชะม้อย (2550: 79-80) ที่ตรวจสอบความเป็นเอกมิตีของแบบทดสอบคณิตศาสตร์จำนวน 40 ข้อ และพบว่าไม่เป็นเอกมิตี ซึ่งสอดคล้องกับงานวิจัยของ จันทนา เปรมฤดี ปรีชาชาญ (2551: 63-64) ที่ตรวจสอบความเป็นเอกมิตีของแบบทดสอบคณิตศาสตร์จำนวน 60 ข้อ และพบว่าไม่เป็นเอกมิตี และสอดคล้องกับงานวิจัยของนุชริน ไบโพธิ์ (2551: 56) ที่ตรวจสอบความเป็นเอกมิตีของแบบทดสอบวัดความถนัดทางคณิตศาสตร์ 60 ข้อ และพบว่าไม่เป็นเอกมิตี อาจกล่าวได้ว่าการตรวจสอบความเป็นเอกมิตีของแบบทดสอบเป็นไปได้ยากเนื่องจากแบบทดสอบมีเนื้อหาที่หลากหลายในแบบทดสอบฉบับเดียว

นิยามศัพท์เฉพาะ

1. **ข้อสอบแบบต่อเนื่องสองขั้นตอน** หมายถึงข้อสอบที่มีทั้งหมด 2 ตอน โดยตอนที่ 1 มีลักษณะเป็นตัวเลือกในแบบจริงหรือเท็จ และในตอนที่ 2 เป็นการให้เหตุผลหรือคำอธิบายในตอนที่ 1 มีลักษณะเป็นแบบเลือกตอบ และมีการกำหนดการให้คะแนนที่มีมากกว่าสองค่า ซึ่งในงานวิจัยนี้กำหนดให้ไว้ 3 ระดับคะแนน คือ 0 , 1 และ 2 คะแนน

2. **ผลสัมฤทธิ์ทางการเรียนเรื่องความเท่ากันทุกประการ** หมายถึง คะแนนที่ได้จากการทำข้อสอบแบบต่อเนื่องสองขั้นตอนเรื่องความเท่ากันทุกประการ ของนักเรียนชั้นมัธยมศึกษาปีที่ 2 ซึ่งประกอบด้วยเนื้อหาเรื่องความเท่ากันทุกประการของรูปสามเหลี่ยมในความสัมพันธ์แบบต่างๆ

3. **การทำหน้าที่ต่างกันของข้อสอบ** หมายถึง ข้อสอบที่นำไปทดสอบกับผู้สอบที่มีความสามารถเท่าเทียมกัน แต่มีโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกัน เนื่องจากมีลักษณะบางอย่างที่แตกต่างกัน เช่น เพศ เชื้อชาติ ศาสนา วัฒนธรรม ภูมิภาค ทำให้ได้ผลการตอบข้อสอบที่ถูกต้องไม่เท่ากันทำให้เกิดการได้เปรียบเสียเปรียบจากการตอบข้อสอบเดียวกัน ซึ่งในงานวิจัยนี้มีตัวแปรเพศเป็นตัวแปรแบ่งกลุ่มในการศึกษา

4. **กลุ่มอ้างอิง(Reference Group)** หมายถึง กลุ่มผู้สอบที่คาดว่าจะเป็กลุ่มที่ได้เปรียบในการตอบข้อสอบ คือกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องสูงกว่าผู้สอบอีกกลุ่มหนึ่ง สำหรับกลุ่มอ้างอิงในการวิจัยในครั้งนี้คือ กลุ่มนักเรียนเพศชาย

5. กลุ่มเปรียบเทียบ (Focal Group) หมายถึง กลุ่มผู้สอบที่คาดว่าจะจะเป็นกลุ่มที่เสียเปรียบในการตอบข้อสอบ คือกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องต่ำกว่าผู้สอบอีกกลุ่มหนึ่ง สำหรับกลุ่มเปรียบเทียบในการวิจัยครั้งนี้คือ กลุ่มนักเรียนเพศหญิง

6. วิธีแมนเทล-แฮนส์เซลทั่วไป (Generalized Mantel-Haenszel:GMH) หมายถึง วิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ ตามแนวคิดของวิธี Mantel-Haenszel เดิมมาทำการปรับขยายใช้กับข้อสอบที่มีการให้คะแนนแบบหลายค่า โดยนำมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ลำดับขั้นของคะแนนผลการตอบมีมากกว่า 2 ลำดับขั้น ระหว่างผู้สอบ 2 กลุ่มที่มีระดับความสามารถในระดับเดียวกัน

7. วิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก (Logistic Discriminant Function Analysis:LDFA) หมายถึง วิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ จากแนวคิดของวิธีการถดถอยโลจิสติกและวิธีการวิเคราะห์เชิงจำแนกกลุ่ม (Discriminant Analysis) โดยอยู่บนโมเดลพื้นฐานที่ใช้สำหรับคัดแยกคนออกเป็นกลุ่มๆ ที่มีส่วนเหมือนกันบนพื้นฐานของการตอบข้อสอบและคะแนนแบบทดสอบรวม โมเดลดังกล่าวสามารถตรวจสอบอิทธิพลของคะแนนแบบทดสอบรวมและคะแนนข้อสอบทำให้มีประสิทธิภาพในการตรวจสอบข้อสอบทั้งข้อสอบที่ทำหน้าที่ต่างกันที่เป็นรูปแบบเดียวกันและข้อสอบทำหน้าที่ต่างกันที่ไม่เป็นรูปแบบเดียวกัน

8. วิธี โพลีโทมัสซิปเทสต์ (Polytomous SIBTEST) หมายถึง วิธีการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบตามแนวคิดของวิธี SIBTEST เดิมมาทำการปรับปรุงใหม่เพื่อใช้กับข้อสอบที่มีการให้คะแนนแบบหลายค่า ใช้สถิติการทดสอบแบบนันทพาราเมตริก (Nonparametric) โดยพิจารณาจากความแตกต่างของคะแนนจริงระหว่างผู้สอบกลุ่มย่อย 2 กลุ่ม ที่มีความสามารถระดับเดียวกัน

9. ประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ หมายถึง ความถูกต้องของการระบุการทำหน้าที่ต่างกันของข้อสอบ จากการตรวจสอบด้วยวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก และวิธีแมนเทล-แฮนส์เซลทั่วไป ซึ่งพิจารณาได้จากความถูกต้องของการตรวจพบข้อที่ทำหน้าที่ต่างกันและความคลาดเคลื่อนของการตรวจสอบ

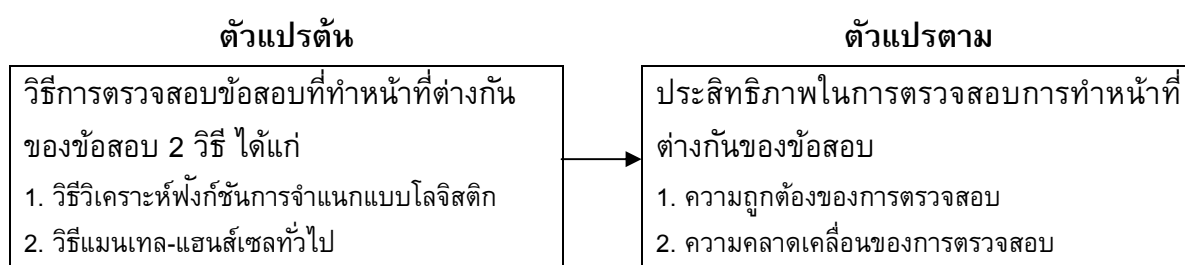
9.1 ความถูกต้องของการตรวจสอบ หมายถึง อัตราส่วนหรือร้อยละของจำนวนข้อสอบที่วิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกและวิธีแมนเทล-แฮนส์เซลทั่วไป ที่ตรวจสอบพบว่าทำหน้าที่ต่างกันได้อย่างถูกต้อง โดยคำนวณจากค่าอัตราส่วนหรือร้อยละของจำนวนข้อสอบที่ตรวจสอบพบว่าทำหน้าที่ต่างกันได้ถูกต้องต่อจำนวนข้อสอบที่ทำหน้าที่ต่างกันทั้งหมดในแบบทดสอบ โดยในงานวิจัยนี้ข้อสอบที่แสดงการทำหน้าที่ต่างกันของข้อสอบได้มาจากการใช้วิธีการตรวจสอบโพลีโทมัสซิปเทสต์ เป็นเกณฑ์มาตรฐานตัดสิน

9.2 ความคลาดเคลื่อนของการตรวจสอบ หมายถึง อัตราส่วนหรือร้อยละของจำนวนข้อสอบที่ระบุผิดพลาด ซึ่งมี 2 ประเภท ซึ่งในงานวิจัยนี้จะพิจารณาเพียงอัตราความคลาดเคลื่อนประเภทที่ 1

9.2.1 อัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) หมายถึง การระบุผิดพลาดว่าข้อสอบทำหน้าที่ต่างกันทั้งๆที่ความจริงข้อสอบไม่ได้ทำหน้าที่ต่างกัน (False positive) ซึ่งในงานวิจัยนี้ สามารถคำนวณได้จากอัตราส่วนของจำนวนข้อสอบที่ระบุผิดพลาดว่าทำหน้าที่ต่างกันต่อจำนวนข้อสอบที่ทำหน้าที่ไม่ต่างกันทั้งหมดในแบบทดสอบ

กรอบแนวคิดในการวิจัย

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ โอ ลีเดีย ลุยและมาร์ค วิลสัน (Ou Lydia Liu; & Mark Wilson. 2009: 164-184) และ แคทเทอร์ลีน เทเลอร์ (Catherine S Taylor. 2012: 246-280) พบว่า เกิดความแตกต่างระหว่างเพศชายและเพศหญิงต่อความสามารถทางด้านคณิตศาสตร์ โดยเพศชายจะมีความสามารถในเนื้อหาสาระเรขาคณิต การวัด การมองภาพรูปทรงต่างๆ และมีความสามารถในการทำแบบทดสอบแบบเลือกตอบ ที่มีความซับซ้อน มากกว่าเพศหญิงและเพศหญิงจะมีความสามารถทางด้านเนื้อหาสาระพีชคณิตมากกว่า ซึ่งจากการศึกษางานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า จะเห็นว่าวิธีโพลีโทมัสซิปเทสท์ วิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก และวิธีแมนเทล-แฮนส์เซลทั่วไป เป็นวิธีที่น่าสนใจโดยงานวิจัยของวริชชา ชะม้อย (2550: 115-129) พบว่าวิธีวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก (LDFA) มีอำนาจการทดสอบที่อยู่ในเกณฑ์ที่กำหนด สุชาติ สิริมีนันทน์ (2554: 137-157) ตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า โดยใช้วิธีโพลีโทมัสซิปเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติก พบว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความเหมาะสมมากกว่าอีกสองวิธี เพราะมีผลการตรวจสอบการทำหน้าที่เบี่ยงเบนมีอัตราความถูกต้องสูงและมีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำ และยา-ฮุย ซู; และเว่ย-ชาง หวัง (Ya-Hui Su & Wen-Chung Wang. 2005: 313-350) ได้ตรวจสอบประสิทธิภาพของการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล วิธีแมนเทล-แฮนส์เซลทั่วไป และวิธีวิเคราะห์ฟังก์ชันเชิงจำแนกแบบ โลจิสติกพบว่าทั้ง 3 วิธีมีการควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดี วิธีแมนเทล และวิธีวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก ให้ผลการตรวจสอบที่ดีกว่าวิธีแมนเทล-แฮนส์เซลทั่วไปภายใต้เงื่อนไขทุกข้อ จึงสรุปเขียนเป็นแผนภาพแสดงกรอบแนวคิดในการวิจัย ดังภาพประกอบ 1



ภาพประกอบ 1 กรอบแนวคิดในการวิจัย



บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

ในงานวิจัยครั้งนี้ ผู้วิจัยได้ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง และได้นำเสนอตามหัวข้อต่อไปนี้

1. เอกสารที่เกี่ยวข้องกับข้อสอบแบบต่อเนื่องที่สองขั้นตอน
 - 1.1 ข้อสอบแบบต่อเนื่องสองขั้นตอน
 - 1.2 เกณฑ์การให้คะแนนข้อสอบแบบต่อเนื่องสองขั้นตอน
 - 1.3 ข้อดีและข้อจำกัดของข้อสอบแบบต่อเนื่องสองขั้นตอน
2. เอกสารและงานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ
 - 2.1 ความหมายของการทำหน้าที่ต่างกันของข้อสอบ
 - 2.2 สาเหตุที่ก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบ
 - 2.3 ลักษณะของข้อสอบที่แสดงการทำหน้าที่ต่างกัน
 - 2.4 ประเภทของการทำหน้าที่ต่างกันของข้อสอบ
 - 2.5 หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
 - 2.6 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
 - 2.7 เกณฑ์เปรียบเทียบคุณภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ
3. งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ
4. เอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธี

Polytomous SIBTEST

- 4.1 แนวคิดและหลักการของวิธี Polytomous SIBTEST
- 4.2 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันโดยใช้วิธี Polytomous SIBTEST
5. เอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธี GMH
 - 5.1 แนวคิดและหลักการของวิธี MH
 - 5.2 แนวคิดและหลักการของวิธี GMH
 - 5.3 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันโดยใช้วิธี GMH
6. เอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธี LDFA
 - 6.1 แนวคิดและหลักการของวิธี LDFA
 - 6.2 กระบวนการตรวจสอบ
 - 6.3 การทดสอบสมมติฐาน
 - 6.4 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันโดยใช้วิธี LDFA

1. เอกสารที่เกี่ยวข้องกับข้อสอบแบบต่อเนื่องสองขั้นตอน

จากการศึกษาเอกสารที่เกี่ยวข้องกับข้อสอบแบบต่อเนื่อง พบว่า (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. 2555: 61–73) ได้กล่าวว่า ข้อสอบแบบต่อเนื่องเป็นข้อสอบที่ให้ผู้เรียนมีการคิดอย่างต่อเนื่องเป็นชุด โดยมีการผสมผสานข้อสอบหลายรูปแบบไว้ด้วยกัน เช่น ข้อสอบแบบเลือกตอบกับแบบเขียนตอบ ข้อสอบแบบเลือกตอบกับแบบเลือกตอบ ข้อสอบแบบต่อเนื่องจึงมีได้หลายลักษณะ ในที่นี้จะกล่าวถึงข้อสอบแบบต่อเนื่องสองขั้นตอน

1.1 ข้อสอบแบบต่อเนื่องสองขั้นตอน

ข้อสอบแบบต่อเนื่องสองขั้นตอน (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. 2555: 67) มีลักษณะมุ่งให้ผู้เรียนพิจารณาเลือกคำตอบของคำถามที่กำหนดให้ แล้วใช้การคิดหาเหตุผลเพื่ออธิบายคำตอบที่เลือกนั้น เป็นข้อสอบที่เน้นกระบวนการคิดที่ต่อเนื่อง ผู้เรียนจะต้องทำงานตามขั้นตอนที่กำหนดให้ในข้อสอบประกอบด้วย 2 ตอน ดังนี้

ตอนที่ 1 มีคำถามและตัวเลือกที่ให้ผู้เรียนเลือกคำตอบที่ถูกต้อง หรือตัวเลือกแบบถูกผิดก็ได้

ตอนที่ 2 เป็นการให้เหตุผลหรือให้คำอธิบายประกอบตัวเลือกจากขั้นตอนที่ 1 ซึ่งมีได้ 2 ลักษณะดังนี้

1. มีตัวเลือกที่ให้ผู้เรียนเลือกคำตอบ แต่ละตัวเลือกจะเป็นการแสดงผลของคำตอบในตอนที่ 1 ทั้งนี้เหตุผลที่ใช้เป็นตัวลงอาจสร้างมาจากข้อผิดพลาดต่างๆที่พบจากการตอบของผู้เรียน
2. มีลักษณะเป็นคำถามปลายเปิดให้ผู้เรียนได้เขียนอธิบายเหตุผลที่เลือกคำตอบในตอนที่

1

1.2 เกณฑ์การให้คะแนนข้อสอบแบบต่อเนื่องสองขั้นตอน

ข้อสอบแบบต่อเนื่องสองขั้นตอนจะมีเกณฑ์การให้คะแนนได้หลายวิธี แต่จะต้องมีเกณฑ์ที่ครอบคลุมอยู่ 2 ประเด็น ดังนี้ (สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. 2555: 67- 73)

1. ความถูกต้องของคำตอบในตอนต้นที่ 1
2. ความสอดคล้องระหว่างคำตอบของตอนที่ 1 กับตอนที่ 2

การตรวจให้คะแนนควรพิจารณาสาระสำคัญดังนี้

2.1 ถ้าผู้ตอบเลือกคำตอบในตอนต้นที่ 1 ถูกต้องและตอบเหตุผลในตอนต้นที่ 2 ได้ถูกต้อง จะได้รับการประเมินว่าเป็นมีความรู้ความเข้าใจในสาระการเรียนรู้ที่ต้องการวัดครบถ้วนทุกผลการเรียนรู้ที่คาดหวัง

2.2 ถ้าผู้ตอบเลือกคำตอบในตอนต้นที่ 1 ไม่ถูกต้อง แสดงว่าผู้ตอบไม่มีความรู้ความเข้าใจในสาระการเรียนรู้ที่ต้องการวัดผล แต่จากผลการตอบจะทำให้เห็นถึงความเข้าใจของผู้สอบว่ามีความรู้ความเข้าใจอย่างไร

2.3 ถ้าผู้ตอบเลือกคำตอบในตอนที่ 1 ไม่ถูกต้อง และตอบในตอนที่ 2 ไม่สอดคล้องกับตัวเลือกในตอนที่ 1 ผู้ตอบควรได้รับการประเมินว่าไม่มีความรู้ความเข้าใจในสาระการเรียนรู้ที่ต้องการวัดครบถ้วนทุกผลการเรียนรู้ที่คาดหวัง

การให้คะแนนข้อสอบแบบต่อเนื่องสองขั้นตอนสามารถทำได้หลายวิธีโดยอาจกำหนดเกณฑ์การให้คะแนนที่แบ่งเป็น 3 ระดับ คือ 0 , 1 และ 2 ดังนี้

ตาราง 1 ตัวอย่างเกณฑ์การให้คะแนนแบบรวมของข้อสอบแบบต่อเนื่องสองขั้นตอน

ระดับคะแนน	เกณฑ์การให้คะแนน
2	เลือกคำตอบและบอกเหตุผลประกอบถูกต้อง
1	เลือกคำตอบถูกต้อง แต่บอกเหตุผลประกอบไม่ถูกต้อง
0	เลือกคำตอบไม่ถูกต้อง

1.3 ข้อดีและข้อจำกัดของข้อสอบแบบต่อเนื่องสองขั้นตอน

แบบทดสอบแบบต่อเนื่องสองขั้นตอนมีข้อดีและข้อจำกัดดังนี้

ข้อดี

1. ใช้เป็นข้อสอบเพื่อวินิจฉัยข้อบกพร่องของผู้เรียนได้
2. ใช้วัดทักษะและกระบวนการด้านการให้เหตุผลและการเชื่อมโยงได้
3. ใช้วัดได้ตรงกับสาระการเรียนรู้ที่ต้องการวัด

ข้อจำกัด

1. ตรวจให้คะแนนยาก
2. ข้อสอบบางข้ออาจต้องใช้เวลามากในการตรวจให้คะแนน
3. ความเที่ยงตรงของการให้คะแนนขึ้นอยู่กับข้อกำหนดเกณฑ์การให้คะแนนที่เหมาะสม

เหมาะสม

จากการศึกษาความหมายของข้อสอบแบบต่อเนื่องสองขั้นตอนที่กล่าวมาสรุปได้ว่าข้อสอบแบบต่อเนื่องสองขั้นตอนเป็นข้อสอบที่สามารถให้คะแนนได้มากกว่าสองค่าและสามารถนำมาวิเคราะห์ผู้เรียนได้เนื่องจากข้อสอบแบบต่อเนื่องสองขั้นตอนจะประกอบด้วยตอนที่ 1 ที่ผู้เรียนจะต้องระบุว่าข้อความที่กำหนดให้ถูกหรือผิดและให้เหตุผลประกอบคำตอบในตอนที่ 2 ซึ่งจะสามารถทำให้วินิจฉัยข้อบกพร่องของผู้เรียนได้อย่างแท้จริง

2. เอกสารและงานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ

2.1 ความหมายของการทำหน้าที่ต่างกันของข้อสอบ

นักวิจัยทางการวัดหลายท่านได้ให้ความหมายของความลำเอียงของข้อสอบและการทำหน้าที่ต่างกันของข้อสอบ ดังนี้

อเลน; และเยน (Allen; & Yen. 1979: 129) ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบว่า เป็นข้อสอบที่วัดความสามารถหรือคุณลักษณะทางจิตวิทยาของผู้สอบแต่ละกลุ่มไม่ตรงกัน

พอร์ฟแฮม (Popham. 1981: 78) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ความชอบหรือความโอเนียงที่ทำให้การพิจารณาตัดสินเป็นไปอย่างไม่ยุติธรรม

คามิลลี; และเชพเพิร์ด (Camilli; & Shepard. 1994: 8) ได้ให้ความหมายของการทำหน้าที่ต่างกันของข้อสอบว่า เป็นความคลาดเคลื่อนอย่างเป็นระบบ (Systematic Error) หรือความไม่เที่ยงตรง (Invalidity) ของการวัดชนิดหนึ่ง ทำให้สมาชิกบางกลุ่มได้ประโยชน์หรือเสียประโยชน์ เช่น ข้อสอบที่เข้าข้างกลุ่มนักเรียนชายมากกว่านักเรียนหญิง หรือคนผิวขาวกับคนผิวดำ หรือระหว่างคนที่อยู่ต่างภูมิภาคกัน เป็นต้น

ดูรานส์; และคูลิค (อุทัยวรรณ สายพัฒนา. 2547: 24; อ้างอิงจาก Dorans; & Kulick. 1986: 355) กล่าวว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง โอกาสในการตอบข้อสอบได้ถูกต้องของผู้สอบกลุ่มหนึ่งที่มีค่าต่ำกว่าหรือสูงกว่า ผู้สอบอีกกลุ่มหนึ่งที่มีระดับความสามารถเดียวกัน

ฮอลแลนด์; และวายเนอร์ (จินทนา เปรมฤดีปรีชาชาญ. 2551: 11; อ้างอิงจาก Holland; & Wainer. 1993) การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่ผู้เข้าสอบ ณ ความสามารถเดียวกัน แต่อยู่ต่างกลุ่มกัน มีโอกาสในการตอบถูกไม่เท่ากัน กลุ่มที่ให้ความสนใจหรือกลุ่มที่เสียเปรียบ เรียกว่า “Focal group” และกลุ่มที่ใช้เปรียบเทียบกับทั่วไปจะเป็นกลุ่มได้เปรียบ เรียกว่า “Reference group”

จากความหมายของการทำหน้าที่ต่างกันของข้อสอบที่กล่าวมาสรุปได้ว่า การทำหน้าที่ต่างกันของข้อสอบ หมายถึง ข้อสอบที่นำไปทดสอบกับผู้สอบที่มีความสามารถเท่าเทียมกัน แต่จะมีลักษณะบางอย่างที่แตกต่างกัน เช่น เพศ เชื้อชาติ ศาสนา วัฒนธรรม ภูมิภาค แล้วผลในการตอบข้อสอบที่ถูกต้องไม่เท่ากันทำให้เกิดการได้เปรียบเสียเปรียบจากการตอบข้อสอบเดียวกัน

2.2 สาเหตุที่ก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบ

จากการนำข้อสอบที่มีการทำหน้าที่ต่างกันของข้อสอบมาศึกษาภายหลังจากการวิเคราะห์ การทำหน้าที่ต่างกันของข้อสอบด้วยค่าสถิติแบบต่างๆ ในแบบทดสอบหลายๆฉบับ เพื่อศึกษาถึงสาเหตุหรือแหล่งที่ก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบในตัวคำถามเหล่านั้น พบว่า สาเหตุที่น่าจะก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบมีหลายสาเหตุด้วยกัน ซึ่งสามารถสรุปสาเหตุที่

ก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบมากที่สุด ได้ดังนี้ ชูนิแมน (จันทนา เปรมฤดีปรีชาชาญ. 2551: 12; อ้างอิงจาก Scheuneman. 1982: 180 - 195)

1. การเดา (guessing) อาจเกิดจากข้อสอบนั้นยากเกินไปหรือเวลาไม่เพียงพอจะก่อให้เกิดความไม่เท่ากันในโอกาสการตอบข้อสอบถูกของผู้สอบแต่ละคน

2. ความเร็ว (Speed) หรือเวลาในการสอบ จะทำให้เกิดการเดาหรือกรณีในเวลาอันน้อย อาจจะทำข้อสอบไม่ทัน ซึ่งจะมีผลกับข้อสอบข้อหลังๆ โดยเฉพาะในการศึกษาการทำหน้าที่ต่างกันของข้อสอบวัดความถนัด

3. ความกำกวมหรือความไม่ชัดเจน (Unclear) ของข้อคำถาม นั่นคือข้อคำถามขาดความเป็นปรนัย การใช้ภาษาถิ่นหรือใช้คำไม่เป็นภาษากลางในการสื่อความหมาย ซึ่งก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบกับกลุ่มภาษาใดภาษาหนึ่งขึ้นได้

4 ลำดับขั้นของคำถาม (Series) อาจจะเป็นสิ่งที่ก่อให้เกิดความสับสน หรือชี้แนะคำตอบของข้อสอบบางข้อได้

5 สถานภาพทางสังคมหรือความเกี่ยวข้องทางสังคม (Social implication) ก็เป็นสิ่งที่ก่อให้เกิดความแตกต่างระหว่างกลุ่มผู้สอบแต่ละกลุ่มได้

6 ประสิทธิภาพหรือการฝึกฝนของแต่ละกลุ่มย่อย เป็นสิ่งที่ก่อให้เกิดการได้เปรียบเสียเปรียบของแต่ละกลุ่มค่อนข้างจะชัดเจน

7. องค์ประกอบทางวัฒนธรรม ความเป็นอยู่ ขนบธรรมเนียม เชื้อชาติ ศาสนา ก็จะมีอิทธิพลให้กับบางกลุ่มย่อย จึงทำให้เกิดการได้เปรียบเสียเปรียบในบางเนื้อหาวิชาได้

นอกจากนั้นความลำเอียงของข้อสอบอาจเกิดจากสาเหตุหรือแหล่งที่สำคัญ 2 แหล่งคือ

1. การเลือกเนื้อหา (bias in selection) คือ ผู้สร้างข้อสอบเลือกเนื้อหาเฉพาะส่วนใดส่วนหนึ่งมาสร้างข้อสอบ ทำให้ได้ข้อสอบที่มีเนื้อหาไม่ครอบคลุมและไม่ได้สัดส่วนที่สมดุลกัน

2. การสร้างข้อสอบ (bias in construction) คือ การใช้ภาษาหรือข้อความบางอย่างที่เอื้อให้เกิดประโยชน์กับผู้สอบกลุ่มใดกลุ่มหนึ่ง

เทเลอร์; และออแลนโด (ซูมาลี แกลัททิงค์. 2547: 41-42; อ้างอิงจาก Taylor; & Orlando. 1987) จำแนกสาเหตุของความลำเอียงจากสิ่งที่เกิดขึ้นกับผู้สอบว่าเนื่องมาจากการสื่อสารที่ไม่ตรงกันจากแบบสอบถาม ประกอบด้วย

1. ความลำเอียงจากสถานการณ์ในข้อสอบ (situation bias) เกิดจากความเข้าใจไม่ตรงกันระหว่างผู้ออกข้อสอบและผู้เข้าสอบเนื่องจากการใช้สถานการณ์ในการออกข้อสอบและใช้ภาษาที่ไม่เหมาะสมของผู้ออกข้อสอบ

2. ความลำเอียงจากคำแนะนำหรือคำสั่งในแบบสอบ(directions bias) เกิดจากคำแนะนำหรือคำสั่งจากแบบสอบรวมไปถึงความซับซ้อน ความกำกวมของคำสั่งและความไม่คุ้นเคยจากคำแนะนำหรือคำสั่งในแบบสอบของผู้สอบ

3. ความลำเอียงจากการตีค่า (value bias) เกิดจากการตีค่าที่แตกต่างกันของผู้เข้าสอบ โดยผู้สอบจะเลือกคำตอบจากการตีค่าหรือความรู้สึกของตนเองเป็นหลัก

4. ความลำเอียงที่เกิดจากภาษา (linguistics bias) ได้แก่ ความแตกต่างจากการใช้ภาษา หรืออิทธิพลของภาษาในข้อสอบซึ่งเกิดการพูดหรือใช้ภาษาอย่างไม่เป็นมาตรฐาน ดังนั้นการเลือก ตัวเลือกของข้อสอบจะใช้มาตรฐานที่ต่างกัน

5. ความลำเอียงจากรูปแบบ(format bias) เกิดจากการปฏิบัติของข้อสอบหรือข้อกำหนด ที่ไม่สอดคล้องกับความเข้าใจหรือรูปแบบการเรียนรู้ของผู้สอบ

6. ความลำเอียงทางด้านวัฒนธรรม (cultural bias) เกิดจากผู้เข้าสอบเข้าใจเนื้อหาในแบบ สอบไขว้เขว หรือแปลเรื่องไปทางที่เกี่ยวกับวัฒนธรรมของตนจนทำให้เกิดการเข้าใจผิด

7. ความลำเอียงจากการแปลความหมายผิด (misinterpretation bias) เกิดจากการแปล ความหมายไปตามประสบการณ์และการรับรู้เดิมของตนที่มีมาก่อน

8. ความลำเอียงจากสิ่งเร้า (stimulus bias) เกิดจากผู้ออกข้อสอบแสดงสิ่งเร้าใน วัตถุประสงค์ และเนื้อหาสาระที่นำมาวิเคราะห์ไปยังผู้เข้าสอบ

จากสาเหตุที่ก่อให้เกิดการทำหน้าที่ต่างกันของข้อสอบที่กล่าวมาสรุปได้ว่า สาเหตุที่ทำให้ เกิดการทำหน้าที่ต่างกันของข้อสอบ มีหลายสาเหตุร่วมกันซึ่งอาจจะเกิดจากพื้นฐานเดิมของผู้สอบ เอง เช่น วัฒนธรรม ประเพณี เชื้อชาติ ภูมิภาค การเดา การแปลความของผู้สอบเป็นต้น หรือเกิด จากข้อสอบ เช่น ข้อสอบที่มีความกำกวม ไม่ชัดเจน หรือมีการสร้างข้อสอบที่ไม่ครอบคลุมเนื้อหา หรือเอื้อประโยชน์ให้กับฝ่ายใดฝ่ายหนึ่ง ซึ่งอาจจะทำให้เกิดการได้เปรียบเสียเปรียบกันในกลุ่มผู้สอบ ที่ต่างกัน

2.3 ลักษณะของข้อสอบที่แสดงการทำหน้าที่ต่างกัน

ลักษณะของข้อสอบโดยทั่วไปที่แสดงการทำหน้าที่ต่างกัน (สิริรัตน์ วิภาสศิลป์, 2545: 64-65)

1. เนื้อหาหรือภาษาในข้อสอบยั่วให้ผู้ตอบข้อสอบกลุ่มสนใจโกรธ เกิดการโต้แย้งหรือ เกิดอารมณ์ไม่พอใจ
2. เนื้อหาหรือภาษาในข้อสอบมีความหมายไปในทางลบ ดูถูกเหยียดหยามหรือก้าวร้าว ต่อผู้ตอบข้อสอบกลุ่มสนใจ
3. เนื้อหาหรือภาษาในข้อสอบแสดงว่าผู้ตอบข้อสอบกลุ่มสนใจมีปมด้อยเกี่ยวกับอำนาจ หรือความเป็นผู้นำ
4. เนื้อหาหรือภาษาในข้อสอบหลายๆข้อให้ความสนใจ เน้นความสำคัญและยกย่องผู้ตอบ กลุ่มอ้างอิง
5. เนื้อหาในข้อสอบมีสารสนเทศที่เป็นประโยชน์กับกลุ่มอ้างอิงมากกว่ากลุ่มสนใจ

ลักษณะข้อสอบที่แสดงการทำหน้าที่ต่างกันทางเพศ

1. รูปแบบหรือโครงสร้างของข้อสอบเป็นปัญหากับผู้ตอบข้อสอบเพศใดเพศหนึ่งมากกว่าผู้ตอบข้อสอบอีกเพศหนึ่ง
2. เนื้อหาในข้อสอบใช้สรรพนาม เป็นการให้ความสนใจความสำคัญและยกย่องเพศเฉพาะเพศใดเพศหนึ่ง
3. เนื้อหาในข้อสอบกำหนดสถานการณ์ที่เพศใดเพศหนึ่งสนใจและได้รับการปลูกฝังฝึกฝนหรือมีโอกาสพบเห็นในชีวิตประจำวันมากกว่า

ส่วนลักษณะของข้อสอบที่แสดงการทำหน้าที่ต่างกันทางเพศ ยิป;และเมลฟี (สิริรัตน์ วิชาศิลป์. 2545: 64-65; อ้างอิงจากGipps; & Murphy. 1994: 182) คือ

1. ลักษณะของโจทย์ที่มีความซับซ้อนและท้าทาย เพศชายจะชอบมากกว่าเพศหญิง
2. ลักษณะของโจทย์ที่เป็นโจทย์ปัญหาหรือการหาความสัมพันธ์และความเข้าใจเกี่ยวกับขนาดรูปทรงและมิติต่างๆ การคาดคะเนระยะทาง การมองรูปทรง(มิติสัมพันธ์) เพศชายจะชอบมากกว่าเพศหญิง
3. ลักษณะของโจทย์ที่ยากต้องใช้ความพยายามในการค้นหาคำตอบ เพศชายจะชอบมากกว่าเพศหญิง

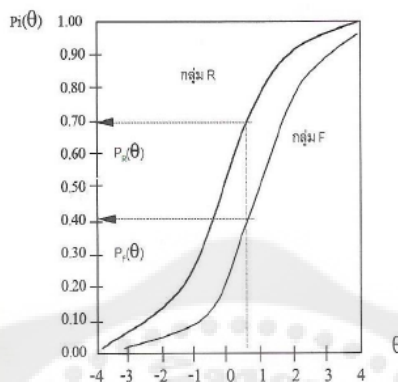
4. ลักษณะของตัวเลือกที่ต้องวิเคราะห์ซับซ้อน เพศชายจะชอบมากกว่าเพศหญิง ส่วนเพศหญิงชอบตัวเลือกที่ระบุว่า “ไม่มีคำตอบถูก” หรือ “ถูกทุกข้อ” มากกว่าเพศชาย

จากลักษณะของข้อสอบที่ทำหน้าที่ต่างกันของข้อสอบที่กล่าวมาสรุปได้ว่า ข้อสอบที่ทำหน้าที่ต่างกันนั้นจะมีเนื้อหาหรือข้อความที่เป็นการย่วยหรือให้ความสนใจกับกลุ่มใดกลุ่มหนึ่งและข้อสอบที่ทำหน้าที่ต่างกันทางเพศนั้น เพศชายจะมีความสามารถทางด้านเรขาคณิต การมองรูปภาพรูปทรงต่างๆ การวัด รวมถึงโจทย์ที่มีความยาก ซับซ้อนและท้าทายในการหาคำตอบ มากกว่าเพศหญิง

2.4 ประเภทของการทำหน้าที่ต่างกันของข้อสอบ

การทำหน้าที่ต่างกันของข้อสอบ เป็นการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มผู้สอบย่อยอย่างน้อย 2 กลุ่มขึ้นไป (ศิริชัย กาญจนวาสี. 2555: 117 -120)ปกตินิยมทำการเปรียบเทียบ 2 กลุ่ม ประกอบด้วยกลุ่มแรก เรียกว่า กลุ่มเปรียบเทียบ (Focal group หรือกลุ่ม F) เป็นกลุ่มที่สนใจศึกษาคาดว่าจะจะเป็นกลุ่มที่เสียเปรียบในการตอบข้อสอบ และกลุ่มที่สองเรียกว่ากลุ่มอ้างอิง (Reference group หรือกลุ่ม R) เป็นกลุ่มที่คาดว่าจะได้เปรียบในการตอบข้อสอบได้ถูกต้องในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ พบว่า ข้อสอบสามารถทำหน้าที่แตกต่างกันได้ 2 ประเภท เมลเลนเบอร์ก(ศิริชัย กาญจนวาสี. 2555: 117 -120; อ้างอิงจาก Mellenbergh. 1982) ได้แก่ การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป (Uniform) และแบบอนเอกรูป (Nonuniform)

1. ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF) หมายถึงข้อสอบที่ทำให้ผู้ตอบข้อสอบถูกมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างสม่ำเสมอในทุกระดับความสามารถ เมื่อพิจารณาโดยคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม จะพบว่าไม่มีปฏิสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับการเป็นสมาชิกของกลุ่ม (Group membership) ดังภาพประกอบ 2



ภาพประกอบ 2 ข้อสอบทำหน้าที่ต่างกันแบบเอกรูป (Uniform DIF)

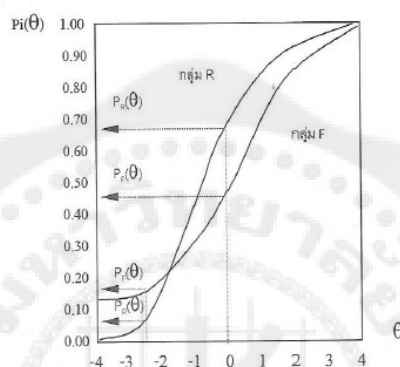
ที่มา: ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่*. หน้า 118.

2. ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป (NonUniform DIF) หมายถึงข้อสอบที่ทำให้โอกาสในการตอบถูกต้องของผู้สอบระหว่างกลุ่มแตกต่างกันอย่างไม่สม่ำเสมอในทุกระดับความสามารถ เพื่อพิจารณาโค้งคุณลักษณะข้อสอบของผู้สอบ 2 กลุ่ม พบว่ามีปฏิสัมพันธ์ร่วมกันระหว่างระดับความสามารถของผู้สอบ กับการเป็นสมาชิกของกลุ่ม เช่นที่ระดับความสามารถหนึ่งกลุ่มผู้สอบกลุ่ม R มีโอกาสในการตอบข้อสอบถูกมากกว่ากลุ่มผู้สอบกลุ่ม F แต่ที่ระดับความสามารถอีกระดับหนึ่งกลุ่มผู้สอบกลุ่ม F มีโอกาสในการตอบข้อสอบถูกมากกว่ากลุ่มผู้สอบกลุ่ม R

ตามทฤษฎีการตอบสนองข้อสอบ (Item Response Theory; IRT) สามารถพิจารณา "ปฏิสัมพันธ์" ดังกล่าวได้จากความแตกต่างระหว่างของค่าพารามิเตอร์อำนาจจำแนกของข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่ม กล่าวคือ ถ้าข้อสอบทำหน้าที่ต่างกันแบบเอกรูป แล้วโค้งลักษณะข้อสอบ (Item Characteristic Curves; ICCs) ระหว่างผู้สอบกลุ่มย่อยสองกลุ่มจะขนานกันหรือมีฟังก์ชันการตอบสนองข้อสอบ (Item Response Functions; IRFs) เหมือนกัน แต่ถ้าข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปแล้วโค้งลักษณะข้อสอบระหว่างผู้สอบกลุ่มย่อยสองกลุ่มจะไม่ขนานกันหรือมีฟังก์ชันการตอบสนองข้อสอบต่างกัน ดังนั้นความแตกต่างระหว่างโค้งลักษณะข้อสอบทั้งสองแบบจะบ่งบอกถึงขนาดและทิศทางของข้อสอบที่ทำหน้าที่ต่างกันซึ่งสามารถคำนวณได้โดยใช้สูตรการคำนวณพื้นที่ของ Raju (1990)

ข้อสอบที่ทำหน้าที่ต่างกันแบบอเนกรูป สามารถจำแนกได้เป็น 2 ลักษณะ สวามินาทาน; และโรเจอร์(ศิริชัย กาญจนวาสี. 2555: 117 -120; อ้างอิงจาก Swaminathan; & Rogers. 1990) ดังนี้

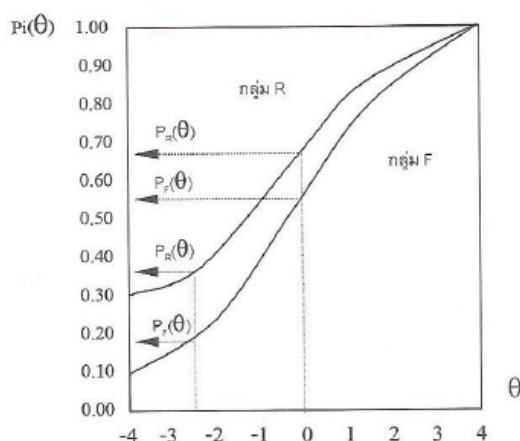
1. ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูปโดยมีปฏิสัมพันธ์ไม่เป็นลำดับ (Disordinal interaction) เป็นการทำหน้าที่ต่างกันสำหรับกลุ่มผู้สอบซึ่งเกิดขึ้น เมื่อโค้งลักษณะข้อสอบตัดกันระหว่างช่วงความสามารถของผู้สอบหรือเรียกว่าข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (Non-Unidirectional DIF) ดังภาพประกอบ 3



ภาพประกอบ 3 ข้อสอบทำหน้าที่ต่างกันแบบไม่มีทิศทาง (Non-Unidirectional DIF)

ที่มา: ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่*. หน้า 119.

2. ข้อสอบทำหน้าที่ต่างกันแบบอเนกรูป โดยมีการปฏิสัมพันธ์เป็นลำดับ (Ordinal interaction) เป็นการทำหน้าที่ต่างกัน สำหรับกลุ่มผู้สอบซึ่งเกิดขึ้น เมื่อโค้งลักษณะข้อสอบต่างกันอย่างไม่สม่ำเสมอ แต่ไม่ตัดกัน หรืออาจตัดกันนอกช่วง ความสามารถของผู้สอบตารางปลายสุดของช่วงความสามารถต่ำหรือสูง อาจเรียกข้อสอบลักษณะนี้ว่าข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียว(Unidirectional DIF) ดังภาพประกอบ 4



ภาพประกอบ 4 ข้อสอบทำหน้าที่ต่างกันแบบมีทิศทางเดียวกัน (Unidirectional DIF)

ที่มา: ศิริชัย กาญจนวาสี. (2555). *ทฤษฎีการทดสอบแนวใหม่*. หน้า 120.

จากประเภทของการทำหน้าที่ต่างกันของข้อสอบที่กล่าวมาสรุปได้ว่า ข้อสอบสามารถทำหน้าที่แตกต่างกันได้ 2 ประเภท การทำหน้าที่ต่างกันของข้อสอบแบบเอกรูป ซึ่งหมายถึงข้อสอบที่ทำให้ผู้สอบกลุ่มหนึ่งตอบได้ถูกต้องมากกว่าผู้สอบอีกกลุ่มหนึ่งอย่างสม่ำเสมอ และประเภทที่ 2 การทำหน้าที่ต่างกันแบบอเนกรูป หมายถึงข้อสอบที่ทำให้ผู้ตอบข้อสอบถูกระหว่างกลุ่มแตกต่างกันอย่างไม่สม่ำเสมอในทุกๆระดับความสามารถ

2.5 หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การตรวจการทำหน้าที่ต่างกันของข้อสอบ (DIF detection) เป็นการเปรียบเทียบผลการตอบสนองของข้อสอบเป็นรายข้อระหว่างกลุ่มผู้สอบอย่างน้อย 2 กลุ่ม ที่มีความสามารถหลัก (primary ability) ที่มุ่งวัดเท่ากัน แต่คาดว่าจะไม่มีความได้เปรียบหรือเสียเปรียบกัน โดยกลุ่มหนึ่งถือเป็น กลุ่มอ้างอิง (Reference group) ซึ่งคาดว่าจะได้เปรียบในการตอบข้อสอบข้อนั้น หรือมีโอกาสตอบข้อสอบได้ถูกต้องมากกว่า ส่วนอีกกลุ่มคือ กลุ่มเปรียบเทียบ (Focal group) ซึ่งเป็นกลุ่มสนใจศึกษาและคาดว่าจะได้เป็นกลุ่มที่เสียเปรียบในการเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบจำเป็นต้องจับคู่(matching) ผู้สอบตามความสามารถซึ่งเป็นเงื่อนไขสำคัญของการตรวจการทำหน้าที่ต่างกันของข้อสอบเกณฑ์การจับคู่(matching criteria) ที่นิยมใช้กัน มี 2 วิธี (ศิริชัย กาญจนวาสี. 2555; 120–123) ดังนี้

เกณฑ์ภายนอก (External Criterion)

การวิเคราะห์การทำหน้าที่ต่างกัน โดยใช้เกณฑ์ภายนอกนี้สามารถนำไปใช้ได้ทั้งข้อสอบรายข้อแบบสอบทั้งฉบับ โดยการใช้คะแนนจากแบบสอบอื่นเป็นเกณฑ์ภายนอกแล้วใช้เทคนิคการถดถอย (regression analysis) เพื่อทำการเปรียบเทียบเส้นกราฟความสัมพันธ์ระหว่างตัวแปรเกณฑ์กับตัวแปรทำนายระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

หลักการและจุดมุ่งหมายนี้เพื่อสร้างสมการทำนายตัวแปรเกณฑ์ ซึ่งเป็นคะแนนของแบบทดสอบอื่นจากตัวแปรทำนายซึ่งเป็นคะแนนรายข้อ หรือคะแนนแบบสอบระหว่างกลุ่มอ้างอิงและกลุ่มเปรียบเทียบในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบ จะใช้คะแนนรายข้อเป็นตัวแปรทำนาย แต่ถ้าเป็นการวิเคราะห์การทำหน้าที่ต่างกันของแบบทดสอบจะใช้คะแนนรวมของแบบสอบทั้งสองฉบับเป็นตัวแปรทำนาย สำหรับตัวแปรเกณฑ์ที่ใช้เป็นเกณฑ์ภายนอก อาจใช้คะแนนรวมทั้งฉบับ หรือเกรดเฉลี่ย หรือผลสัมฤทธิ์เกี่ยวข้องของผู้สอบ ครอนบาค(ศิริชัย กาญจนวาสี. 2555: 120–123; อ้างอิงจาก Cronbach. 1970)

สมการทำนาย สำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบแสดงผลได้ดังนี้

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

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

เมื่อ Y_i = คะแนนของตัวแปรเกณฑ์

X_i = คะแนนของตัวแปรทำนาย

A = ค่าคงที่หรือค่าตัดแกน (intercept)

B = ค่าความชัน (Slope)

การใช้เกณฑ์ภายนอกมีข้อดี คือเกณฑ์ที่ใช้มีความเป็นอิสระจากข้อสอบ และแบบสอบที่ดี ต้องมีการตรวจสอบ แต่มีจุดอ่อนตรงที่ความเหมาะสมของเกณฑ์ที่จะนำมาใช้ ในทางปฏิบัติเป็นการยากที่จะหาตัวแปรเกณฑ์ภายนอกจากแบบทดสอบฉบับอื่นที่มีความตรงเชิงทำนายและมีความยุติธรรมสำหรับกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ถ้าตัวแปรเกณฑ์ภายนอกขาดคุณสมบัติดังกล่าวจะทำให้ผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบหรือแบบทดสอบขาดความแม่นยำและความสมบูรณ์

เกณฑ์ภายใน (Internal Criterion)

การวิเคราะห์การทำหน้าที่ต่างกัน โดยใช้เกณฑ์ภายในเป็นการนำวิธีทางสถิติมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ หรือแบบทดสอบ หรือแบบสอบ โดยเน้นการพิจารณาจากโครงสร้างภายในของแบบทดสอบหลักด้วยการวิเคราะห์ผลการตอบข้อสอบและความสามารถหรือคะแนนจริงของผู้สอบจากแบบทดสอบฉบับนั้น เพื่อนำมาเปรียบเทียบระหว่างผู้สอบจากกลุ่มอ้างอิง

และกลุ่มเปรียบเทียบ ที่มีความสามารถหรือคะแนนจริงเท่ากัน จะมีผลการตอบหรือโอกาสในการตอบข้อสอบได้ถูกต้องแตกต่างกันหรือไม่ เพื่อป้องกันการทำหน้าที่ต่างกันของข้อสอบ ค่าสถิติทดสอบที่นิยมนำมาใช้พอสรุปได้ดังนี้

1. การทดสอบปฏิสัมพันธ์ (Interaction)

ในระยะเริ่มแรกของการศึกษาความลำเอียงของข้อสอบ มีการใช้สถิติทดสอบเอฟ (F-test) จากการวิเคราะห์ความแปรปรวน (ANOVA) เพื่อทดสอบปฏิสัมพันธ์ระหว่างกลุ่มผู้สอบกับข้อสอบ ถ้าการทดสอบมีนัยสำคัญเป็นสัญญาณของการทำหน้าที่ต่างกันของข้อสอบ (Cleary & Hiltion. 1968; Jensen. 1974) จากนั้นจึงทำการวิเคราะห์ต่อด้วยวิธีการ Post Hoc เพื่อระบุข้อสอบที่มีผลต่อการปฏิสัมพันธ์ ซึ่งเป็นข้อที่ทำหน้าที่ต่างกัน

วิธีการนี้มีข้อดีที่สามารถศึกษาผู้ตอบหลายๆ กลุ่มได้สะดวก แต่มีจุดอ่อนในเรื่องการควบคุมกลุ่มต่างๆ ให้มีความสามารถที่ทัดเทียมกัน ขนาดกลุ่มตัวอย่างของกลุ่มต่างๆ และอัตราการคลาดเคลื่อนประเภทที่ 1 จะสูงขึ้น ถ้าจำนวนข้อสอบเพิ่มมากขึ้น

2. การวัดความเบี่ยงเบนสัมพัทธ์ (Relative Deviation)

การคำนวณค่าความยากของข้อสอบ เช่น p, b เป็นต้น เมื่อคำนวณแยกแยะระหว่างกลุ่มและแปลงให้เป็นค่าความยากมาตรฐานสามารถนำมาพล็อตเปรียบเทียบรายข้อ ถ้าข้อใดเบี่ยงเบนไปจากแกนหลักที่คาดหมาย หรือเบี่ยงเบนเกินจากความคลาดเคลื่อนมาตรฐานของค่าความยากที่กำหนด ย่อมแสดงถึงการทำหน้าที่ต่างกันของข้อสอบ (Cleary & Hilton. 1968; Angoff & Ford. 1973) รวมทั้งสามารถคำนวณค่าสหสัมพันธ์ระหว่างค่าความยากง่ายรายข้อระหว่างกลุ่ม เพื่อแสดงถึงการทำหน้าที่ต่างกันของข้อสอบ ถ้าสหสัมพันธ์เข้าใกล้ 1.00 แสดงว่าค่าความยากสัมพัทธ์ของข้อสอบมีค่าใกล้เคียงกันระหว่างกลุ่มดังนั้นแบบสอบวัดคุณลักษณะคล้ายกันระหว่างกลุ่ม วิธีการนี้มีข้อดีและข้อเสียคล้ายการทดสอบปฏิสัมพันธ์ นอกจากนี้ค่าความยากง่ายของข้อสอบ (p) มิใช่ตัวแทนของค่าความจริงของข้อสอบ และได้รับอิทธิพลจากค่าแทรกซ้อนอื่น ได้แก่ ค่าอำนาจจำแนกและความสามารถของผู้สอบ

3. การเปรียบเทียบน้ำหนักตัวประกอบ (Factor Analysis)

เป็นเทคนิคทางสถิติที่นิยมใช้ในการตรวจสอบความตรงเชิงทฤษฎีหรือโครงสร้าง (Construct Validity) เมื่อนำมาวิเคราะห์ตัวประกอบมาใช้ในการวิเคราะห์โครงสร้างแบบทดสอบแยกตามกลุ่มผู้สอบ ความไม่สอดคล้องกันระหว่างน้ำหนักตัวประกอบบนคุณลักษณะสำคัญที่มุ่งวัด หรือความแตกต่างของค่าเฉลี่ยคะแนนตัวประกอบ (factor Score) ระหว่างกลุ่มผู้สอบ ย่อมสะท้อนการทำหน้าที่ต่างกันของข้อสอบและแบบสอบ

การใช้เทคนิคการวิเคราะห์ตัวประกอบเชิงสำรวจ (Exploratory Factor Analysis; EFA) สำหรับศึกษาการทำหน้าที่ต่างกัน จะมีจุดอ่อนในเรื่องความไม่สอดคล้องระหว่างน้ำหนักตัวประกอบ อาจเกิดจากความแตกต่างของความสามารถกลุ่มก็ได้ แนวทางที่เหมาะสมควรใช้เทคนิคการวิเคราะห์ตัวประกอบเชิงยืนยัน (Confirmatory Factor Analysis ; CFA) นอกจากนี้ยังสามารถใช้ CFA สำหรับตรวจสอบความแตกต่างระหว่างกลุ่ม ในด้านคุณลักษณะหรือความสามารถรองได้อีกด้วย (Camilli & Shepard. 1994)

การเปรียบเทียบโอกาสตอบข้อสอบถูก

การวิเคราะห์โอกาสตอบข้อสอบถูกของผู้สอบจากกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีความสามารถเท่ากัน เป็นแนวทางสำคัญที่นิยมใช้กันและเป็นที่ยอมรับกันในปัจจุบัน สำหรับบ่งชี้การทำหน้าที่ต่างกันของข้อสอบ มีการคำนวณค่าสถิติ 2 แนวทางดังนี้ เปรียบเทียบค่าสัดส่วนหรือความน่าจะเป็นในการตอบข้อสอบถูกของผู้สอบต่างกลุ่มที่มีระดับความสามารถเท่ากัน เช่นวิธี แมนเทิล-แฮนส์เซล (MH) เป็นต้น เปรียบเทียบค่าฟังก์ชันการตอบสนองข้อสอบ หรือโค้งลักษณะข้อสอบระหว่างกลุ่มที่มีระดับความสามารถเท่ากัน เป็นวิธีที่อยู่บนพื้นฐานของทฤษฎี IRT เช่นวิธีวัดความแตกต่างของพื้นที่ วิธีวัดความแตกต่างของค่าพารามิเตอร์ความยาก วิธีทดสอบไค-สแควร์ของลอร์ด (Lords' χ^2 - test) เป็นต้น วิธีการนี้มีข้อดีที่สำคัญได้แก่ การคำนวณค่าทางสถิติของข้อสอบมีความน่าเชื่อถือมีกลไกการควบคุมความสามารถเท่ากัน จึงเป็นวิธีการยอมรับกันทั่วไป แต่มีข้อจำกัดในด้านความความสลับซับซ้อนของแนวคิดพื้นฐาน และการวิเคราะห์มีความจำเป็นต้องใช้โปรแกรมคอมพิวเตอร์โดยเฉพาะ

2.6 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (DIF detection) จำแนกตามลักษณะการตรวจให้คะแนนได้เป็น 2 ประเภท คือ ข้อสอบที่มีการให้คะแนนแบบทวิภาค หรือสองค่า (Dichotomous Scoring) และข้อสอบที่มีการให้คะแนนแบบพหุวิภาค หรือหลายค่า (Polytomous Scoring) วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแต่ละประเภท

ยังสามารถจำแนกได้อีก 2 มิติ ได้แก่ มิติลักษณะของตัวแปรเกณฑ์ ซึ่งแบ่งเป็นกลุ่มวิธีที่ใช้คะแนนสังเกตได้ (Observed Score) และกลุ่มวิธีที่ใช้คะแนนสังเกตไม่ได้หรือคะแนนของตัวแปรแฝง (Latent variable) และมิติลักษณะของสถิติวิเคราะห์ ซึ่งแบ่งเป็นกลุ่มวิธีที่ใช้สถิติพาราเมตริก (Parametric approach) และกลุ่มวิธีที่ใช้สถิตินั้นพาราเมตริก (Nonparametric approach) (ศิริชัย กาญจนวาสี. 2555: 124–126) ดังแสดงในตาราง 2

ตาราง 2 วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่มีการตรวจให้คะแนนแบบทวิภาค (Dichotomous DIF) และพหุวิภาค (Polytomous DIF)

ประเภทและตัวแปรเกณฑ์	พารามетริก	นัยพารามетริก
1. DIF แบบทวิภาค		
1.1 คะแนนที่สังเกตได้ (observed score)	ANOVA Logistic Regression	TID MH STND
1.2 คุณลักษณะ/ตัวแปรแฝง (latent variable)	IRT-D2 Lord's χ^2 General IRTLR Loglinear IRTLR	SIBTEST
2. DIF แบบพหุวิภาค		
2.1 คะแนนที่สังเกตได้ (observed score)	ANOVA Polytomous Logistic Regression	Polytomous STND GMH
2.2 คุณลักษณะ/ตัวแปรแฝง (latent variable)	General IRTLR PCM	Polytomous SIBTEST GPCM

1. วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบทวิภาค กลุ่มวิธีที่ใช้คะแนนที่สังเกตได้ วิธีในกลุ่มนี้มีทวิเคราะห์ตามทฤษฎีการทดสอบแบบดั้งเดิม (CTT) หรือกลุ่มที่ไม่ใช้ทฤษฎีการตอบสนองข้อสอบ (Non-IRT approach) โดยใช้คะแนนรวมของผู้สอบเป็นเกณฑ์การจับคู่ของกลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มการวิเคราะห์ความแปรปรวน (ANOVA) วิธีการวิเคราะห์การถดถอยโลจิสติก (Logistic Regression ; LR) วิธีแปลงค่าความยากของข้อสอบ (Transformed Item Difficulty ; TID) วิธีแมนเทิล-แฮนส์เซล (Mantel-Haenzel ; MH) วิธีวัดดัชนีมาตรฐาน (Standardization ; STND) การปรับให้เป็นมาตรฐานด้วยน้ำหนัก ประกอบกลุ่มที่ใช้คุณลักษณะแฝง วิธีในกลุ่มนี้ใช้คุณลักษณะหรือตัวแปรแฝง ซึ่งวิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ (IRT) สำหรับเกณฑ์การจับคู่กลุ่มผู้สอบ วิธีการตรวจสอบที่สำคัญในกลุ่มนี้ได้แก่ วิธีวัดพื้นที่ความแตกต่างระหว่างโค้งการตอบสนอง (IRT - D2) วิธีไค-สแควร์ของลอร์ด (Lord's χ^2) วิธีอัตราส่วนไลค์ลิฮูดทั่วไป (General IRT Likelihood Ratio) วิธีซิปเทสท์ (SIBTEST)

2. วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบพหุวิภาค

2.1 กลุ่มวิธีใช้คะแนนที่สังเกตได้

วิธีการวิเคราะห์ความแปรปรวน (ANOVA)

วิธีการวิเคราะห์การถดถอยโลจิสติกพหุวิภาค (Polytomous Logistic Regression)

วิธีแมนเทิล-แฮนส์เซลทั่วไป (General Mantel-Haenzel ; GMH)

2.2 กลุ่มวิธีที่ใช้คุณลักษณะแฝง

- วิธีอัตราส่วนไลค์ลิฮูดในรูปทั่วไป (General IRT Likelihood Ratio)

- วิธีการให้คะแนนบางส่วน (Partial Credit Model ; PCM)

- วิธีชิปเทสต์พหุวิภาค (Polytomous SIBTEST)

- วิธีการให้คะแนนบางส่วนทั่วไป (Generalized partial Credit Model ;

GPCM)

2.7 เกณฑ์เปรียบเทียบคุณภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของ

ข้อสอบ (Criteria for comparison between DIF Detecting Methods)

คุณภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบนิยมพิจารณาจากดัชนีบ่งชี้สำคัญ 2 ตัว ได้แก่ อำนาจการทดสอบ หรืออัตราความถูกต้องของการตรวจพบข้อที่ทำหน้าที่ต่างกัน (Power rate) และ ความคลาดเคลื่อนของการตรวจสอบ หรืออัตราตามความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) ซึ่งเป็นโอกาสของการเกิดความคลาดเคลื่อนในลักษณะที่ตรวจพบว่า ข้อสอบทำหน้าที่ต่างกันทั้งที่ความเป็นจริงข้อสอบไม่ได้ทำหน้าที่ต่างกัน (ศิริชัย กาญจนวาสี. 2555: 151–152)

ในการคำนวณค่าสถิติตามวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ มีจุดมุ่งหมายเพื่อทดสอบนัยสำคัญของผลการตรวจสอบ โดยมีสมมติฐานศูนย์ของการทดสอบ คือ ข้อสอบไม่ได้ทำหน้าที่ต่างกัน หรือ H_0 : No DIF ผลการทดสอบสมมติฐานของวิธีการตรวจสอบ DIF วิธีต่างๆ นำสู่การตัดสินใจว่า ยอมรับสมมติฐานศูนย์ (Accept H_0) หรือปฏิเสธสมมติฐานศูนย์ (Reject H_0) ผลของการตัดสินใจมีโอกาสเกิดเหตุการณ์ 4 ลักษณะ ดังตาราง 3

ตาราง 3 คุณภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การตัดสินใจ ตามผลการตรวจสอบ	Ho: No DIF	
	ความเป็นจริง	
	H_0 ถูก	H_0 ผิด
Accept H_0	ตัดสินใจถูก (True negative) ระดับความเชื่อมั่น ($1-\alpha$)	ตัดสินใจผิด (Type II Error, β) ● False negative
Reject H_0	●ตัดสินใจผิด (Type I Error, α) ● False positive	●ตัดสินใจถูก (True positive) อำนาจการทดสอบ ($1-\beta$)

จากค่าสถิติที่คำนวณตามวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ จะนำไปสู่การตัดสินใจสรุปผลการตรวจสอบดังนี้

1. ตัดสินใจถูก มีโอกาสเกิดขึ้น 2 ลักษณะคือ การสรุปถูกว่า
 - 1.1 ข้อสอบ No DIF ตามความเป็นจริง (True negative)
 - 1.2 ข้อสอบ DIF ตามความเป็นจริง (True positive)
2. ตัดสินผิด มีโอกาสเกิดขึ้น 2 ลักษณะคือ การสรุปผิดว่า
 - 2.1 ข้อสอบ DIF ทั้งๆที่ความเป็นจริงข้อสอบ No DIF (False positive)
 - 2.2 ข้อสอบ No DIF ทั้งๆที่ความเป็นจริงข้อสอบ DIF (False negative)

เนื่องจากอำนาจการทดสอบ ($1-\beta$) กับ β เป็นค่าดัชนีมีสเกลที่ผกผันกันและความคลาดเคลื่อนประเภทที่ 1 (α) กับ $1-\alpha$ ต่างเป็นดัชนีมีสเกลที่ผกผันเช่นเดียวกัน ดังนั้น

การพิจารณาดัชนีประกันคุณภาพ 2 ตัว คือ อัตราความถูกต้องของการตรวจพบ DIF (Power rate) และอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error rate) ก็เพียงพอที่จะให้สารสนเทศครบทั้ง 4 เหตุการณ์

2.8 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ

งานวิจัยในประเทศ

พัชรีย์ ปิยภักดิ์ (2531: 64–70) ได้ทำการวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนชั้นประถมศึกษาปีที่ 6 เรื่องสมการ ที่มีตัวแปรเพศของนักเรียนและที่ตั้งของโรงเรียน (กลุ่มนักเรียนในกรุงเทพฯและจังหวัดสมุทรสาคร) ด้วยวิธีการวิเคราะห์ 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีไค-สแควร์ และวิธีโค้งลักษณะข้อสอบพบว่าวิธีโค้งลักษณะข้อสอบพบข้อสอบที่มีความลำเอียงมากที่สุด รองลงมาคือ วิธีไค-สแควร์ และวิธีการแปลงค่าความยากของข้อสอบพบข้อสอบน้อยที่สุด และพบว่ามีข้อสอบจำนวน 2 ข้อที่พบความลำเอียงซ้ำกันทั้งความลำเอียงในกลุ่มเรื่องเพศและกลุ่มผู้สอบในกรุงเทพฯและสมุทรสาคร

ทัศนีย์ พิรมนตรี (2530: 76–81) ได้ทำการวิเคราะห์ความลำเอียงของแบบสอบวิชาคณิตศาสตร์ โครงการตรวจสอบคุณภาพการศึกษา ชั้นมัธยมศึกษาปีที่ 6 ที่มีตัวแปรเป็นนักเรียนในกรุงเทพฯกับกลุ่มนักเรียนในภาคภูมิศาสตร์ ทั้ง 5 ภาค คือภาคกลาง ภาคเหนือ ภาคใต้ ภาคตะวันออกเฉียงเหนือ และภาคตะวันออกเฉียงใต้ ด้วยวิธีการวิเคราะห์ 3 วิธี คือวิธีกำหนดจุดเดลด้า วิธีทดสอบความแตกต่างระหว่างกลุ่มด้วยสถิติไค-สแควร์ ในโมเดลลอกลีเนียร์ 2 โมเดล และวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว พบว่าพบข้อสอบที่มีความลำเอียงจากวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว มากที่สุด และพบว่าการวิเคราะห์ทุกวิธีจะทำให้พบความลำเอียงซึ่งกันระหว่างกลุ่มนักเรียนกรุงเทพฯกับกลุ่มนักเรียนทุกภาค

เจน จันทร์แสง (2537: 84–92) ได้ทำการวิเคราะห์ความลำเอียงของข้อสอบของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3 เรื่องจำนวนจริง ที่มีตัว

แปรเพศและที่ตั้งของโรงเรียน (กลุ่มนักเรียนที่ศึกษาอยู่ในเมืองและนอกเมืองของจังหวัดนครสวรรค์) ด้วยวิธีการวิเคราะห์ความลำเอียง 4 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีวิเคราะห์ความแปรปรวน วิธีไค-สแควร์และวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 1 ตัว พบว่าข้อสอบในการวิเคราะห์ความลำเอียงระหว่างเพศและที่ตั้งของโรงเรียนนั้นวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 1 ตัวพบข้อสอบที่มีความลำเอียงมากที่สุด รองลงมาคือ วิธีวิเคราะห์ความแปรปรวน วิธีไค-สแควร์ และวิธีแปลงค่าความยากของข้อสอบ

จิตติมา วรรณศรี (2540: 99 - 113) ได้ศึกษาเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี Mantel-Haenszel และวิธี SIBTEST โดยใช้ข้อมูลจำลองโปรแกรม IRTDATA โดยมีเงื่อนไขของการศึกษาคือความยาวแบบทดสอบและขนาดกลุ่มตัวอย่าง ผลการศึกษาพบว่า วิธี Mantel-Haenszel และวิธี SIBTEST มีประสิทธิภาพเท่าเทียมกันในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ที่ทุกขนาดกลุ่มตัวอย่างและทุกอัตราส่วน ภายใต้ความยาวแบบทดสอบเดียวกัน โดยส่วนมาก วิธี SIBTEST มีอัตราความคลาดเคลื่อนชนิดที่ 1 มากกว่าวิธี Mantel-Haenszel เล็กน้อย และพบว่าวิธี Mantel-Haenszel และวิธี SIBTEST มีประสิทธิภาพเท่าเทียมกันทุกระดับความยาวแบบทดสอบ โดยพบว่า เมื่อใช้แบบทดสอบที่มีความยาวปานกลาง 60 ข้อ มีประสิทธิภาพในการตรวจสอบมากที่สุด

รักชนก ยี่สุนศรี (2544: 196 - 205) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบและแบบสอบสำหรับกลุ่มผู้สอบเมื่อจำแนกตามเพศและสถานที่ตั้งทางภูมิศาสตร์และโรงเรียนที่จบการศึกษา เพื่อเปรียบเทียบความเที่ยง ความตรง และฟังก์ชันสารสนเทศของแบบสอบระหว่างแบบสอบฉบับก่อนและหลังตัดข้อสอบที่พบ DIF โดยใช้ข้อมูลจากการตอบแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา วิชาภาษาอังกฤษและคณิตศาสตร์ ผลการวิจัยพบว่า แบบสอบวิชาภาษาอังกฤษทำหน้าที่ต่างกันตามเพศและสถานที่ตั้งตามภูมิศาสตร์ ส่วนแบบสอบถวามวิชาคณิตศาสตร์ทำหน้าที่ต่างกันตามเพศของผู้สอบ นอกจากนี้เมื่อเปรียบเทียบคุณภาพของแบบสอบฉบับก่อนและหลังการตัดข้อสอบที่พบว่าทำหน้าที่ต่างกัน พบว่า ในด้านความเที่ยงตรงจะไม่ต่างกัน แต่แบบสอบถวามฉบับหลังจากตัดข้อสอบที่พบว่าทำหน้าที่ต่างกันแล้ว ส่วนใหญ่จะมีค่าความเชื่อมั่นลดลง และมีค่าฟังก์ชันสารสนเทศมากขึ้น ส่วนความสัมพันธ์ระหว่างตำแหน่งของคะแนนรวมของผู้สอบก่อนและหลังจากการตัดข้อสอบที่พบว่าทำหน้าที่ต่างกันทั้งในกรณีที่ตัดทุกข้อและตัดในบางข้อ พบว่าทุกกรณีมีความสัมพันธ์ในทางบวกซึ่งกันและกันอย่างมีนัยสำคัญ

สุมาลี แก้วทนต์ (2547: 182 - 198) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบด้านเพศ ภาษาพูด และเชื้อชาติ ของแบบสอบสาระการเรียนรู้ภาษาไทยและสาระการเรียนรู้สังคมศึกษา ศาสนา และวัฒนธรรม และตรวจสอบสาเหตุของการทำหน้าที่ต่างกันของข้อสอบด้านเพศ ภาษาพูด และเชื้อชาติของแบบสอบกลุ่มสาระการเรียนรู้ภาษาไทย และกลุ่มสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม โดยใช้วิธีแมนเทิล-แฮนส์เซล และตรวจสอบสาเหตุการทำหน้าที่ต่างกันโดยผู้เชี่ยวชาญ ผลการวิจัยพบว่า การเกิดการทำหน้าที่ต่างกันของข้อสอบในด้านของเพศและภาษาพูด และเชื้อชาตินั้นเกิดจากเนื้อเรื่องที่น่าสนใจของผู้สอบมาน้อยเพียงใด

จันทนา เปรมฤดีปรีชาชาญ (2551: 83 - 89) ได้ศึกษาการทำหน้าที่ต่างกันของข้อสอบที่ทำหน้าที่ต่างกันตามเชื้อชาติของผู้สอบระหว่างวิธีการของMantel-Haenszel วิธี Lord's χ^2 และวิธี Closed Interval ผลการวิจัยพบว่า จำนวนข้อสอบที่ทำหน้าที่ต่างกันที่ได้จากการตรวจสอบสามวิธีแตกต่างกันเมื่อขนาดของกลุ่มตัวอย่างและความยาวของแบบทดสอบต่างกัน ความสอดคล้องของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเมื่อพิจารณาจากกลุ่มตัวอย่างขนาด 600 คน กับ 750 คน วิธีการตรวจสอบเหมือนกัน แบบทดสอบจำนวน 40 ข้อ พบว่าวิธีที่มีความสอดคล้องกันในการตรวจสอบสูงสุด คือ วิธี Mantel-Haenszel แบบทดสอบจำนวน 60 ข้อ พบว่าวิธีที่มีความสอดคล้องกันในการตรวจสอบสูงสุด คือ วิธี Closed-Interval Unsigned Area

งานวิจัยต่างประเทศ

แฮมเบิลตัน; และโรเจอร์ส(จันทนา เปรมฤดีปรีชา. 2551: 44;อ้างอิงจาก Hambleton; & Rogers. 1989) ได้เปรียบเทียบวิธีตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี คือ วิธีทฤษฎีการตอบข้อสอบ และวิธีแมนเทล-แฮนส์เซล (MH) ข้อมูลที่ใช้เป็นผลการสอบแบบทดสอบ New Mexico High School Proficiency Exam (NMHSPE) จำนวน150 ข้อ ของนักเรียนผิวขาว 8,000 คน และนักเรียนพื้นเมือง 2,600 คน NMHSPE เป็นแบบทดสอบชนิดเลือกตอบ 4 ตัวเลือก วัดทักษะการดำเนินชีวิต 5 ด้าน คือความรู้เกี่ยวกับแหล่งทรัพยากรของชุมชน ความรู้เกี่ยวกับผู้บริโภค ความรู้เกี่ยวกับรัฐบาลและกฎหมาย ความรู้เกี่ยวกับสุขภาพทางกายและจิต และความรู้เกี่ยวกับอาชีพ กลุ่มตัวอย่างที่ศึกษา สุ่มมากลุ่มละ 2,000 คน และแบ่งเป็นกลุ่มย่อย ๆ ละ 1,000 คน การศึกษาการทำหน้าที่ต่างกัน โดยพิจารณาร้อยละของข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันหรือไม่ต่างกันซ้ำกันในการตรวจสอบ 2 ครั้ง ผลการศึกษา พบว่าวิธีทฤษฎีการตอบข้อสอบมีความสอดคล้องของผลการตรวจสอบร้อยละ 73 ส่วนวิธีแมนเทล-แฮนส์เซล มีความสอดคล้องของผลการตรวจสอบร้อยละ 80 และพบว่าร้อยละ 47 ของข้อสอบที่ถูกระบุว่าทำหน้าที่ต่างกันในการตรวจสอบคู่ที่ 1 เป็นข้อสอบทำหน้าที่ต่างกันในการตรวจสอบคู่ที่ 2 ด้วย สำหรับวิธีแมนเทล-แฮนส์เซล และร้อยละ 61 สำหรับวิธีทฤษฎีการตอบข้อสอบในทางกลับกันพบว่า ร้อยละ 64 ของข้อสอบที่ทำหน้าที่ต่างกันในการตรวจสอบคู่ที่ 2 เป็นข้อสอบที่ทำหน้าที่ต่างกันในการตรวจสอบคู่ที่ 1 สำหรับวิธีแมนเทล-แฮนส์เซล และร้อยละ 56 สำหรับวิธีทฤษฎีการตอบข้อสอบ จะเห็นได้ว่าความคงที่ของการตรวจสอบทั้ง 2 วิธี อยู่ในระดับปานกลาง

มาซุยและมุลีนา (Mansyur; & Muliana. 2016: 437-440) การศึกษานี้มีจุดมุ่งหมายที่จะตรวจสอบลักษณะการทำหน้าที่ต่างกันของข้อสอบ และและลักษณะการทำหน้าที่ต่างกันของแบบทดสอบ สำหรับการสอบปลายภาควิชาคณิตศาสตร์ โดยวิเคราะห์บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ โดยศึกษาจากคำถามและกระดาษคำตอบของนักเรียนทั้งหมด กลุ่มตัวอย่างคือนักเรียน 286 คน เป็นนักเรียนชาย 147 คน และนักเรียนหญิง 149 คน วิเคราะห์ข้อมูลบนพื้นฐานของการตอบสนองข้อสอบ และใช้สถิติทดสอบคือไค-สแควร์ 2 พารามิเตอร์ จากการศึกษพบว่า

เมื่อวิเคราะห์ข้อคำถามทั้งหมด 40 ข้อพบว่าเกิดการทำหน้าที่ต่างกันทางเพศจำนวน 10 ข้อ และการทำหน้าที่ต่างกันทางสถานที่ตั้งจำนวน 13 ข้อ ซึ่งการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบนี้ถือเอื้อต่อเพศหญิงและนักเรียนชนบท

จากการศึกษางานวิจัยทั้งในประเทศและต่างประเทศ พบว่าการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ได้รับความนิยมจะเป็นการตรวจสอบกับข้อสอบที่ให้คะแนนแบบสองค่าเท่านั้น และวิธีที่นิยมนำมาใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบคือ วิธี Mantel-Haenszel และวิธี SIBTEST โดยวิธี MH และ SIBTEST มีประสิทธิภาพในการตรวจสอบเท่าเทียมกันในทุกขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบ



ตาราง 4 สรุปการศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบพหุวิภาคในประเทศไทย

ลำดับที่	ผู้วิจัย	กลุ่มวิธีที่ใช้คะแนนที่สังเกตได้				กลุ่มวิธีที่ใช้คุณลักษณะแฝง			การวิเคราะห์ บนพื้นฐานการ วิเคราะห์ องค์ประกอบ เชิงยืนยัน	
		ถดถอยโลจิสติก แบบจัด อันดับ	ถดถอยโลจิสติกแบบจัด อันดับหลาย มิติ	วิเคราะห์ ฟังก์ชันการ จำแนก โลจิสติก	แมนเทิล- แฮนส์เซล รูปทั่วไป	ชิปเทสต์ พหุวิภาค	การให้ คะแนน บางส่วน	เกรดเรส พอนส์ โมเดล	วิธีโครงสร้าง ความแปรปรวน ร่วมและค่าเฉลี่ย (MACS)	โมเดลเส้นตรง ระดับลดหลั่น (HLM)
1	อุทัยวรรณ สายพัฒนา				✓	✓				
2	อรินทร์ น่วมถนอม	✓	✓			✓				
3	บุษวรรษ์ แสนปลื้ม					✓				
4	วริษา ชะม้อย			✓					✓	
5	สุชาติ สิริมินนนท์	✓		✓		✓				
6	ศักดิ์สิทธิ์ ฤทธิลัน					✓				
7	ศิริรัตน์ สุคันธพฤกษ์						✓	✓		✓

3. เอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธี

Polytomous SIBTEST

3.1 แนวคิดและหลักการของวิธี Polytomous SIBTEST

เชลลีและสเตาต์ (บุษวรรณ แสนปลื้ม. 2550: 15–22; อ้างอิงจาก Shealy; & Stout.1993: 159-194) ได้เสนอวิธี SIBTEST (Simultaneous Items Bias Test) เพื่อใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Differential item functioning: DIF) การทำหน้าที่ต่างกันของแบบทดสอบ (Differential test functioning: DTF) และการทำหน้าที่ต่างกันของกลุ่มข้อสอบ (Differential Bundle Functioning: DBF) เป็นวิธีการที่พัฒนามาจากแนวคิดของการตรวจสอบการทำหน้าที่ต่างกันของแบบทดสอบบนพื้นฐานของทฤษฎีการตอบข้อสอบแบบหลายมิติ (Multidimensional IRT Modeling of DTF) มีรูปแบบเป็นนัยพารามิเตอร์แบบตัวแปรแฝง (Latent-Variable Non-Parametric) ไม่ต้องใช้ฟังก์ชันการตอบข้อสอบ หรือการประมาณค่าความสามารถแฝง สามารถวิเคราะห์ได้ทั้งในแบบทดสอบที่เป็นมิติเดียว (Unidimensional Test) และแบบทดสอบที่เป็นหลายมิติ (Multidimensional Test) จุดเด่นของวิธีการ SIBTEST คือ มีการคำนวณที่ง่าย ไม่ซับซ้อน เสียค่าใช้จ่ายไม่มาก ไม่จำเป็นต้องใช้กลุ่มตัวอย่างขนาดใหญ่ ใช้ได้ดีสำหรับการตรวจสอบการทำหน้าที่ต่างกันแบบสม่ำเสมอ (Uniform DIF) และมีทิศทางเดียว (Unidirectional DIF) นอกจากนี้ ยังสามารถใช้ได้กับการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบ 2 ค่า (Dichotomous DIF) และที่มีการให้คะแนนแบบหลายค่า (Polytomous DIF)

ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี SIBTEST ในแบบทดสอบที่เป็นมิติเดียวจะต้องมีข้อตกลงว่ามีมิติการวัด 2 มิติ $\{\theta = (\theta, \eta)\}$ คือมิติหนึ่ง เป็นความสามารถเป้าหมายหรือคุณลักษณะแฝงเพียงลักษณะเดียวที่ต้องการวัด (target ability; θ) และอีกมิติหนึ่ง เป็นคุณลักษณะแฝงแทรกซ้อนหรือความสามารถแทรกซ้อนที่ไม่ต้องการวัด (nuisance ability; η) ตัวอย่างเช่น แบบทดสอบวัดความสามารถในเรื่องคำศัพท์ภาษาอังกฤษ ข้อสอบบางข้ออาจถามความรู้พิเศษสำหรับเพศชาย เช่น ความรู้เรื่องกีฬา ส่วนข้อสอบบางข้ออาจถามความรู้พิเศษสำหรับเพศหญิง เช่น ความรู้ในเรื่องงานบ้าน ในสถานการณ์เช่นนี้ทักษะในเรื่องคำศัพท์ภาษาอังกฤษเป็นความสามารถเป้าหมายที่ต้องการวัด ซึ่งแทนด้วย θ ส่วนความรู้เรื่องกีฬาและความรู้เรื่องงานบ้านเป็นความสามารถแทรกซ้อนที่ไม่ต้องการวัด ซึ่งแทนด้วย η_1 และ η_2 ตามลำดับ ข้อสอบทุกข้อในแบบทดสอบถือว่าวัดความสามารถเป้าหมาย ส่วนข้อสอบที่ทำหน้าที่ต่างกันจะวัดทั้งความสามารถเป้าหมายและความสามารถแทรกซ้อน 1 หรือมากกว่า 1 ความสามารถ

หลักในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี SIBTEST นั้นจะเปรียบเทียบผลการตอบข้อสอบระหว่างกลุ่มอ้างอิง (R) และกลุ่มสนใจ (F) โดยแบ่งแบบทดสอบออกเป็น 2 ชุดย่อย (Subtest) คือ (1) แบบทดสอบที่มีความเที่ยงตรง (Valid Subtest) หรือแบบทดสอบที่ใช้ในการจับคู่เปรียบเทียบ (Matching Subtest) ซึ่งแบบทดสอบย่อยชุดนี้ประกอบด้วยข้อสอบที่ตั้งใจวัดความสามารถเป้าหมายเพียงอย่างเดียว (2) แบบทดสอบย่อยที่

ต้องการศึกษา (Studied Subtest) ซึ่งประกอบด้วยข้อสอบที่วัดความสามารถที่ต้องการวัดและความสามารถที่ไม่ต้องการวัด ถ้าแบบทดสอบย่อยชุดแรกมีจำนวน n ข้อแล้ว แบบทดสอบย่อยชุดที่ 2 จะมีจำนวน $N-n$ ข้อ เมื่อ N เป็นจำนวนข้อสอบทั้งหมด

กรณีที่เป็นแบบทดสอบที่มีการให้คะแนนแบบ 2 ค่า

$$\text{ให้ } x = \sum_{i=1}^n U_i$$

$$x = \sum_{i=n+1}^N U_i$$

เมื่อ x แทน คะแนนรวมจากแบบทดสอบที่มีความเที่ยงตรง

Y แทน คะแนนรวมจากแบบทดสอบที่ต้องการศึกษา

U_i แทน ผลการตอบข้อสอบข้อที่ i (ตอบถูกได้ 1 ตอบผิดได้ 0)

นำคะแนนเฉลี่ยจากการตอบข้อสอบในแบบทดสอบย่อยที่ต้องการศึกษาระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ ที่มีความสามารถระดับเดียวกันมาจับคู่เปรียบเทียบกัน โดยพิจารณาจากคะแนนรวมของแบบทดสอบย่อยที่เที่ยงตรงที่เท่ากัน ($x = k$) ได้ดังนี้

$$\bar{Y}_{RK} - \bar{Y}_{FK} ; k = 0, 1, \dots, n$$

เมื่อ $\bar{Y}_{RK}$ แทน ค่าเฉลี่ยของคะแนนรวมจากแบบทดสอบย่อยที่ต้องการศึกษาของผู้สอบกลุ่มอ้างอิง ซึ่งได้คะแนน $x = k$

$\bar{Y}_{FK}$ แทน ค่าเฉลี่ยของคะแนนรวมจากแบบทดสอบย่อยที่ต้องการศึกษาของผู้สอบกลุ่มสนใจ ซึ่งได้คะแนน $x = k$

K แทน คะแนนรวมจากแบบทดสอบย่อยที่มีความเที่ยงตรง

ค่า $\bar{Y}_{RK} - \bar{Y}_{FK}$ เป็นความแตกต่างของผลการตอบข้อสอบจากแบบทดสอบย่อยที่ต้องการศึกษา ระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ ที่มีระดับความสามารถที่ต้องการวัด (θ) เท่ากัน

ในกรณีที่ข้อสอบที่ต้องการศึกษาทำหน้าที่ไม่ต่างกัน ค่า $\bar{Y}_{RK} - \bar{Y}_{FK} = 0$ แต่ถ้า $\bar{Y}_{RK} - \bar{Y}_{FK} \neq 0$ แสดงว่าข้อสอบที่ศึกษาทำหน้าที่ต่างกัน

สมมติฐานในการทดสอบ

$$H_0 : \beta_U = 0$$

$$H_1 : \beta_U > 0$$

ซึ่ง $\hat{\beta}_U$ แทน ดัชนีบ่งชี้การทำหน้าที่ต่างกันของข้อสอบ ประเมินค่าได้จากสูตร

$$\hat{\beta}_U = \sum_{k=0}^n \hat{P}_k (\bar{Y}_{RK} - \bar{Y}_{FK})$$

เมื่อ $\hat{\beta}_U$ แทน ค่าสถิติที่ใช้ประมาณค่าปริมาณของการทำหน้าที่ต่างกันของข้อสอบ
แบบทิศทางเดียวของ $\hat{\beta}_U$
 $\hat{P}_k$ แทน สัดส่วนของผู้สอบกลุ่มสนใจที่ได้คะแนนจากแบบทดสอบที่มีความ
เที่ยงตรง $x = k$

สถิติทดสอบที่ใช้ทดสอบสมมติฐานที่เป็นกลาง (No DIF)

$$B = \frac{\hat{\beta}_U}{\hat{\sigma}(\hat{\beta}_U)}$$

โดยที่
$$\hat{\sigma}(\hat{\beta}_U) = \left[\sum_{k=0}^n \hat{P}_k^2 \left[\frac{1}{J_{RK}^{\hat{\sigma}^2}} (y | k, R) + \frac{1}{J_{FK}^{\hat{\sigma}^2}} (y | k, F) \right] \right]^{1/2}$$

เมื่อ $\hat{\sigma}(\hat{\beta}_U)$ แทน ความคลาดเคลื่อนมาตรฐานในการประมาณค่าของ $\hat{\beta}_U$
 $\hat{\sigma}^2(y | k, g)$ แทน ค่าประมาณความแปรปรวนของคะแนนจากแบบทดสอบที่
ศึกษาสำหรับกลุ่มผู้สอบ g (R หรือ F) ที่ได้คะแนนรวม $x = k$
จากแบบทดสอบที่มีความเที่ยงตรง
และ J_{RK} และ J_{FK} แทน จำนวนผู้สอบของกลุ่มอ้างอิงและกลุ่มสนใจตามลำดับที่ได้
คะแนนรวม $x = k$ จากแบบทดสอบที่มีความเที่ยงตรง

สถิติที่ใช้ในการทดสอบ B มีการแจกแจงปกติมาตรฐาน $N(0,1)$ เมื่อ $B = 0$ และจะปฏิเสธ
 H_0 ถ้าผลการทดสอบปรากฏว่า $B > Z_\alpha$ อย่างมีนัยสำคัญที่ระดับ α โดยที่ $\alpha = P[N(0,1) > Z_\alpha]$
นั่นคือข้อสอบทำหน้าที่ต่างกัน โดยจะเข้าข้างผู้สอบกลุ่มสนใจ B มีค่าเป็นบวก และจะเข้าข้างผู้สอบ
กลุ่มอ้างอิงเมื่อ B มีค่าเป็นลบ

วิธีการ SIBTEST ดังกล่าวข้างต้นเป็นวิธีการที่ใช้กับข้อสอบที่ให้คะแนนแบบ 2 ค่า
(dichotomously scored items) เชลลีและสตาต์ (Shealy and Stout, 1993) ได้พัฒนาขยายวิธีการ
SIBTEST ให้ใช้ได้กับข้อสอบที่มีการให้คะแนนแบบหลายค่า (polytomously scored items) มี
รายละเอียดดังนี้

ให้ y เป็นคะแนนจากข้อสอบที่ศึกษา ซึ่งมีลำดับขั้นที่จัดเรียงแล้ว (ordered categories) m
+ 1 ลำดับขั้น ($y = k$; $0 \leq k \leq m$)

ให้ $P_{k,g}(\theta)$ แทน ฟังก์ชันการตอบลำดับชั้นข้อสอบ (item category response function: ICRF) ซึ่งเป็นความน่าจะเป็นของการได้คะแนน k ของผู้สอบที่มีความสามารถ θ จากกลุ่ม g ($g=R$ หรือ F)

การถดถอยของคะแนนสอบบนความสามารถ สามารถนิยามเป็น ผลรวม ICRF_s ที่ให้ค่าน้ำหนัก (weighted sum of ICRF_s) ได้ดังนี้

$$E_g[y | \theta] = \sum_{k=1}^m k P_{k,g}(\theta)$$

ซึ่งการถดถอยนี้เป็น ฟังก์ชันการตอบข้อสอบ (item response function: IRF) ในกรณีที่ข้อสอบมีการให้คะแนนแบบ 2 ค่า สามารถพิจารณาเป็นข้อสอบที่มีการให้คะแนนแบบหลายค่า ที่มี $m = 1$

สำหรับโมเดลตามทฤษฎี IRT ที่เป็นแบบ 2 ค่า นั้น โครงสร้างของข้อสอบข้อหนึ่งสามารถหา IRF ตามที่กำหนดได้อย่างชัดเจน ส่วนโมเดลตามทฤษฎี IRT ที่เป็นแบบหลายขั้นนั้น โครงสร้างของข้อสอบที่มี $m + 1$ ลำดับชั้น สามารถหาฟังก์ชันการตอบลำดับชั้นข้อสอบ (ICRF) ได้ $m(\text{ICRF})$

เมื่อ $P_{kR}(\theta)$ และ $P_{kF}(\theta)$ เป็นฟังก์ชันการตอบลำดับชั้นตอน (ICRF) ของผู้สอบกลุ่มอ้างอิงและกลุ่มสนใจ ตามลำดับ นิยามของ null-DIF แบบตัวแปรแฝง เป็นดังนี้

$$P_{kR}(\theta) = P_{kF}(\theta) ; k = 1, 2, \dots, m \quad (1)$$

สมการ 1 ใช้ได้กับแบบจำลองการตอบข้อสอบที่เป็นแบบหลายค่า ซึ่งแบบจำลองที่นิยมใช้มากที่สุด 3 แบบจำลอง คือ พาเชียล เครดิต แบบจำลอง (PCM) , พาเชียล เครดิต แบบจำลองรูปทั่วไป (GPCM) และเกรด เรสพอนส์ แบบจำลอง (GRM) การพิจารณาการทำหน้าที่ต่างกันของข้อสอบโดยพิจารณาเพียงฟังก์ชันการตอบ (IRF) ที่ต่างกันระหว่างความสามารถของกลุ่มเป็นการเพียงพอ เนื่องจากการเปรียบเทียบเพียงฟังก์ชันการตอบ (IRF) ไม่มีสารสนเทศที่สูญหาย

ให้ $x_1, x_2, \dots, x_n$ แทน คะแนนจากข้อสอบที่ใช้ในการจับคู่ n ข้อ

$m_1, m_2, \dots, m_n$ แทน คะแนนที่เป็นไปได้สูงสุด (Macimum Possible Scores)

สำหรับ $x_1, x_2, \dots, x_n$ ตามลำดับ

ถ้า $m_i = 1; x_i$ คือข้อสอบที่ให้คะแนนเป็น 2 ค่า (Dichotomous)

$X = \sum_{j=1}^n X_j$ แทน คะแนนรวมที่ใช้ในการจับคู่ ; $x = 0, 1, \dots, n_H$

เมื่อ $n_H = \sum_{j=1}^n m_j$ แทน คะแนนของการจับคู่ที่เป็นไปได้สูงสุด

$\bar{Y}_{gk}$ แทน คะแนนเฉลี่ยจากข้อสอบที่ใช้ศึกษาสำหรับผู้สอบในกลุ่ม g ทั้งหมด ($g = R$ หรือ F) ที่ได้ $x = k$
 และ P_{FK} แทน สัดส่วนของผู้สอบในกลุ่มสนใจที่ได้คะแนน $x = k$ จากการสอบเพื่อจับคู่ $x_1, x_2, \dots, x_n$ นั่นคือ

$$P_{FK} = \frac{N_{FK}}{N_F}$$

เมื่อ N_F แทน จำนวนของผู้สอบทั้งหมดในกลุ่มสนใจ
 N_{FK} แทน จำนวนของผู้สอบทั้งหมดในกลุ่มสนใจที่ได้คะแนนการจับคู่ $x = k$
 P_k แทน สัดส่วนของผู้สอบทั้งหมดที่ได้คะแนน $x = k$ จากการสอบเพื่อจับคู่
 นั่นคือ

$$P_k = \frac{N_k}{N}$$

เมื่อ N แทน เป็นจำนวนผู้สอบทั้งหมด

$$N_k = N_{Rk} + N_{Fk}$$

เนื่องจากวิธีการ SIBTEST มีพื้นฐานแนวคิดของทฤษฎีการตอบข้อสอบ (IRT) การนิยามการทำหน้าที่ต่างกันของข้อสอบแบบตัวแปรแฝง (Latent-variable DIF) ก็คือ

$$E_R[Y | \theta] \neq E_F[Y | \theta]$$

และปริมาณการทำหน้าที่ต่างกันของข้อสอบ ณ ความสามารถ θ หาได้โดย

$$B_0(\theta) = E_R[Y | \theta] - E_F[Y | \theta]$$

ส่วนดัชนีของการทำหน้าที่ต่างกัน (Index of DIF) คือ

$$\beta = \int B_0(\theta) f_F(\theta) d\theta$$

แต่ถ้าพิจารณาตามทฤษฎีการทดสอบแบบมาตรฐานเดิม (CTT: $X = T + E$) แล้ว T เป็นคะแนนจริงที่ใช้ในการจับคู่ ส่วน $E = X - T$ นั้น เป็นความคลาดเคลื่อนของการวัดและถือว่ามีค่าเฉลี่ยเป็นศูนย์ทั้ง 2 กลุ่ม ให้ $f_g(t)$ เป็นความหนาแน่นของคะแนนจริงที่ใช้ในการจับคู่ในกลุ่ม g แล้ว

การถดถอยของคะแนนของข้อสอบที่ศึกษาบนคะแนนสอบของข้อสอบที่ใช้ในการจับคู่ของกลุ่ม g แทนด้วย

$$E_g[Y | t] = E[Y | T = t, G = g]$$

กรณีที่ไม่มีการทำหน้าที่ต่างกันของข้อสอบ (no DIF) แล้ว $E_R[Y | t] = E_F[Y | t]$ นั่นคือปริมาณการทำหน้าที่ต่างกันของข้อสอบ ณ ความสามารถ t หาได้โดย

$$B(t) = E_R[Y | t] - E_F[Y | t]$$

ส่วนดัชนีของการทำหน้าที่ต่างกัน (Index of DIF) คือ

$$\beta = \int B(t) f_F(t) dt$$

สถิติที่ใช้ประมาณค่าปริมาณของการทำหน้าที่ต่างกันของข้อสอบของ β คือ

$$\hat{\beta} = \sum_{k=0}^{n_H} P_k d_k$$

เมื่อ $d_k = \bar{y}_{Rk} - \bar{y}_{Fk}; k = 0, \dots, n_H$

ซึ่งเซลล์และสเตดท์ ได้ศึกษาจากข้อมูลจำลองโดยใช้ P_k แทน P_{Fk} เป็นการถ่วงน้ำหนักพบว่าสามารถควบคุมความผิดพลาดประเภท 1 ได้ดีกว่า

สถิติทดสอบ

$$B = \frac{\hat{\beta}}{\hat{\sigma}(\hat{\beta})}$$

เมื่อ $\hat{\sigma}(\hat{\beta}) = \left[\sum_{k=0}^{n_H} p_k^2 \left(\frac{\hat{\sigma}^2(y | k, R)}{N_{Rk}} + \frac{\hat{\sigma}^2(y | k, F)}{N_{Fk}} \right) \right]^{1/2}$

และ $\hat{\sigma}^2(y|k, g)$ แทน ความแปรปรวนของคะแนนจากข้อสอบที่ใช้ศึกษาของข้อสอบ
ในกลุ่ม g ($g = R$ หรือ F) ที่มีคะแนนรวมจากข้อสอบที่ใช้จับคู่
 $x = k$

สถิติที่ใช้ในการอ้างอิงการทำหน้าที่ต่างกันของข้อสอบ มักมีปัญหาในกรณีที่มีความแตกต่างของการแจกแจงความสามารถที่ต้องการวัดระหว่างกลุ่มผู้สอบ ที่เรียกว่าผลกระทบ (Impact) คือ ผู้สอบกลุ่มอ้างอิงมีความสามารถที่ต้องการวัดสูงกว่าผู้สอบกลุ่มสนใจ ก่อให้เกิดอัตราความผิดพลาดประเภทที่ 1 ที่พ้อหรือสูงเกินไป (inflate) ซึ่งในความเป็นจริงข้อสอบที่ทำหน้าที่ต่างกัน ดังนั้น จึงมีความจำเป็นต้องแก้ไขความแตกต่างของการแจกแจงความสามารถที่ต้องการวัดด้วยวิธีการปรับแก้ค่าการถดถอย (regression correction) (Shealy&Stout. 1993: 170) เพื่อแก้ไขและกำจัดอิทธิพลการเกิดความผิดพลาดประเภท 1 ที่เกิดขึ้นมากอันเนื่องมาจากผลกระทบ โดยการแปลงค่า $\bar{Y}_{RK}$ และ $\bar{Y}_{FK}$ ให้เป็น $\bar{Y}_{RK}^*$ และ $\bar{Y}_{FK}^*$ ตามลำดับที่ละคู่ ดังนี้

$$\bar{Y}_{gk}^* = \bar{Y}_{gk} + \hat{M}_{gk} (\hat{V}(k) - \hat{V}_g(k))$$

โดยที่

$$\hat{M}_{gk} \equiv \frac{\bar{Y}_{g,k+1} - \bar{Y}_{g,k-1}}{\hat{V}_g^{(k+1)} - \hat{V}_g^{(k-1)}}$$

$$\hat{V}(k) \equiv \frac{1}{2} (\hat{V}_R^{(k)} + \hat{V}_F^{(k)})$$

- เมื่อ $\bar{Y}_{RK}^*$ แทน ค่าเฉลี่ยที่ปรับแก้แล้วของคะแนนรวมจากแบบทดสอบที่ต้องการศึกษาในกลุ่มผู้สอบ g
- $\hat{M}_{gk}$ แทน ค่าประมาณคร่าวๆของค่าเบี่ยงเบนของคะแนนจริงจากแบบทดสอบที่ต้องการศึกษาที่มีต่อฟังก์ชันของคะแนนจริงจากแบบทดสอบที่มีความเที่ยงตรง
- $\hat{V}(k)$ แทน ค่าประมาณของคะแนนจากแบบทดสอบที่มีความเที่ยงตรงซึ่งได้คะแนนสังเกต $x = k$
- $\hat{V}_g^{(k)}$ แทน ค่าประมาณการถดถอยของคะแนนจริงจากแบบทดสอบที่มีความเที่ยงตรงของผู้สอบกลุ่ม g ซึ่งได้คะแนนสังเกต $x = k$

และ adjusted d_k^* ได้มาโดยการแปลงค่าเชิงเส้น (Linear Transformation) ดังนี้

$$d_k^* = \bar{Y}^* Rk - \bar{Y}^* Fk ; k = 0, 1, \dots, n_H$$

ดังนั้น สถิติที่ใช้ประมาณค่าปริมาณการทำหน้าที่ต่างกันของข้อสอบของ β จึงเปลี่ยนเป็น

$$\hat{\beta}^* = \sum_{k=0}^{n_H} P_k d_k^*$$

H_0 เป็นจริง เมื่อ B มีการแจกแจงเป็นปกติมาตรฐาน $N(0,1)$ และจะปฏิเสธ H_0 เมื่อ $|B| > 1.96$ ที่ระดับนัยสำคัญ .05

3.2 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันโดยใช้วิธี Polytomous SIBTEST งานวิจัยในประเทศ

อุทัยวรรณ สายพัฒนา(2547: 97-112) ได้เปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบหลายค่าระหว่างวิธีแมนเทิล-แฮนส์เซลทั่วไป (GMH) และวิธีโพลีโทมัสซิปเทสต์ (Polytomous SIBTEST) โดยศึกษาภายใต้ปัจจัยที่แปรเปลี่ยน 2 ปัจจัย คือ ความยาวของแบบทดสอบ และขนาดของกลุ่มตัวอย่าง เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวิชาคณิตศาสตร์ เรื่องวงกลม เป็นข้อสอบเลือกตอบแบบถูกผิด ชนิด 5 ตัวเลือก จากนั้นจะตรวจสอบการทำหน้าที่ต่างกันของข้อสอบภายใต้เงื่อนไขต่างๆโดยวิธี GMH และวิธี Polytomous SIBTEST ส่วนวิธีพหุคูณเชิงเส้น (GPCM) ใช้เป็นเกณฑ์การเปรียบเทียบกับ 2 วิธีดังกล่าว ผลการวิจัยพบว่า ประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างขนาดของกลุ่มตัวอย่างที่แตกต่างกันและความยาวของแบบทดสอบมีผลต่อความถูกต้องและความผิดพลาดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ซึ่งความแตกต่างนี้อาจจะเกิดได้จากหลายปัจจัย เช่นลักษณะของข้อสอบที่ผู้สอบสร้างขึ้นจำนวน 40 ข้อ ทำให้เกิดข้อย่อย 200 ข้อ ซึ่งอาจจะทำให้ผู้สอบเหนื่อยล้า ท้อถอย รวมถึงรูปแบบของข้อสอบที่ไม่คุ้นเคย ซึ่งจะทำให้มีผลต่อคะแนนการตอบข้อสอบด้วย

อรินทร์ นวมถนอม (2549: 225-283) ได้เปรียบเทียบอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีโพลี-ซิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ภายใต้เงื่อนไขที่แปรเปลี่ยน 4 ปัจจัย รูปแบบของข้อสอบทำหน้าที่เบี่ยงเบน สัดส่วนของข้อสอบทำหน้าที่เบี่ยงเบน ความแตกต่างของการแจกแจงความสามารถ และขนาดตัวอย่าง โดยการศึกษาในครั้งนี้ใช้การจำลองข้อมูลภายใต้โมเดลพหุคูณเชิงเส้นแบบหลายมิติ จำลองผลการตอบจากแบบทดสอบที่วัดความสามารถสองมิติจำนวน 40 ข้อ ข้อสอบทุกข้อมีรายการตอบ 5 รายการ ให้คะแนนเป็น 1, 2, 3, 4 หรือ 5 คะแนน ผลการวิจัยพบว่า วิธีการ

ถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติมีอัตราความถูกต้องใกล้เคียงกัน ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกัน และทั้งสองวิธีมีอัตราความถูกต้องสูงกว่าวิธีโพลี-ชิปเทสต์ และวิธีโพลี-ชิปเทสต์มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ภายในขอบเขตที่กำหนด ส่วนวิธีการถดถอยโลจิสติกแบบจัดอันดับ และวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าขอบเขตที่กำหนด

บุษวรรษ์ แสนปลื้ม (2550:67-73) ได้เปรียบเทียบวิธีชิปเทสต์แบบหลายค่าที่ใช้รูปแบบการคิดคะแนนต่างกันในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อคำถามวัดมโนทัศน์แห่งตน โดยใช้มาตรวัดมโนทัศน์แห่งตน ที่มีลักษณะเป็นมาตราประเมินค่า (Rating-Scale) ชนิด 5 ระดับ ผลการวิจัยสรุปได้ว่า รูปแบบการคิดคะแนนรวมทั้งฉบับมีแนวโน้มที่จะระบุข้อคำถามที่ทำหน้าที่เบี่ยงเบนได้น้อยกว่ารูปแบบการคิดคะแนนแยกองค์ประกอบ และทั้งสองรูปแบบสามารถระบุข้อคำถามที่ทำหน้าที่เบี่ยงเบนสอดคล้องสัมพันธ์กันสูง โดยที่รูปแบบการคิดคะแนนแยกองค์ประกอบมีอัตราความถูกต้องสูงกว่ารูปแบบการคิดคะแนนรวมทั้งฉบับ แต่ก็มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่ารูปแบบการคิดคะแนนรวมทั้งฉบับด้วย

งานวิจัยต่างประเทศ

ยัง ชัน ลี (Young-Sun Lee. 2009: 365-375) ได้ตรวจสอบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี SIBTEST และ Poly-SIBTEST ข้อมูลที่ได้จากกลุ่มตัวอย่างจากการทดสอบการอ่านพบว่าวิธี SIBTEST และ Poly-SIBTEST ไม่สอดคล้องกัน ผลจากการศึกษาพบว่าวิธี SIBTEST สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีที่สุด ยกเว้นในบางกรณี ในรูปแบบเดียวกันพบว่าวิธี Poly-SIBTEST มีอัตราความคลาดเคลื่อนประเภทที่ 1 น้อยกว่า วิธี Poly-SIBTEST สามารถควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 ได้ดีในเงื่อนไขข้อสอบไม่ได้ยากเกินไป ผลของวิธี SIBTEST จะไม่ได้รับผลกระทบเมื่อขนาดและเปอร์เซ็นต์ของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเพิ่มมากขึ้น แม้ว่าวิธี Poly-SIBTEST จะล้มเหลวในการควบคุมอัตราความคลาดเคลื่อนประเภทที่ 1 แต่วิธี Poly-SIBTEST ก็แสดงให้เห็นถึงอัตราความถูกต้องที่สูงกว่าวิธี SIBTEST

แคทเทอร์รีน เทเลอร์ (Catherine S. Taylor. 2012: 246-280) ได้ทำการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในด้านเพศ โดยใช้แบบทดสอบรูปแบบผสม ในวิชาด้านการอ่านและคณิตศาสตร์ โดยใช้วิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ POLY SIBTEST และ Rach กลุ่มตัวอย่างคือนักเรียนเกรด 4 เกรด 7 และเกรด 10 เครื่องมือที่ใช้ในการวิจัย ใช้แบบทดสอบอิงเกณฑ์ของการอ่านและคณิตศาสตร์ มีแบบเลือกตอบ (คะแนน 0,1) ตอบแบบสั้นๆ (คะแนน 0,1,2) ตอบแบบขยายความ (0,1,2,3,4) ผลการวิจัยพบว่าการทำหน้าที่ต่างกันของข้อสอบที่ใช้การทดสอบโดยวิธี POLY SIBTEST และ Rach นั้น วิธีการของ Rach นั้นใช้กับกลุ่มตัวอย่างขนาดใหญ่และ

สามารถวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบได้มากกว่าวิธีอื่น ในด้านการวิเคราะห์เนื้อหาเพศชายมีความสามารถด้านการระบุอธิบายหรือการให้เหตุผล และการวิเคราะห์ข้อมูล ในเพศหญิงจะมีความสามารถด้านการตีความ การวิเคราะห์ข้อความและวรรณกรรมต่างๆดีกว่าเพศชาย

4. เอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธี

GMH

4.1 แนวคิดและหลักการของวิธี MH

วิธีแมนเทล-แฮนส์เซลพัฒนาโดย Mantel และ Haenszel (ศิริชัย กาญจนวาสี. 2555: 126-131) ซึ่งแต่เดิมเป็นสถิติสำหรับเปรียบเทียบอัตราส่วนแถมต่อร่วม (Common odds ratio) และทำการทดสอบอัตราส่วนเปรียบเทียบด้วยไค-สแควร์ (Chi-square statistic) ต่อมา (Holland 1985 cited in Holland & Thayer, 1988) ได้นำวิธีแมนเทล-แฮนส์เซลมาประยุกต์ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ สำหรับหน่วยงานการบริการทดสอบทางการศึกษา (Education Testing Service ; ETS) ของประเทศสหรัฐอเมริกา วิธีแมนเทล-แฮนส์เซล เป็นที่ยอมรับจากนักวิธีวิทยาการวิจัยอย่างกว้างขวางว่าเป็นวิธีที่ใช้ง่าย สะดวก และประหยัดโดยใช้หลักการของตารางการถัว (Contingency table) แบบประเพณีนิยม (Traditional) ไม่มีการคำนวณทวนซ้ำ (Noniterative) จึงสามารถนำไปใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบได้ง่าย มีขั้นตอนการคำนวณที่ไม่สลับซับซ้อน มีการทดสอบทางสถิติแบบนันทพารามตริก (Nonparametric) ซึ่งไม่จำเป็นต้องใช้โมเดลประมาณค่า

หลักการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยวิธีแมนเทล-แฮนส์เซล จะเปรียบเทียบผลการตอบข้อสอบระหว่างผู้สอบสองกลุ่ม กลุ่มหนึ่งเรียกว่า “กลุ่มเปรียบเทียบ” (Focal group) ซึ่งเป็นกลุ่มสอบที่สนใจศึกษา ถ้าข้อสอบทำหน้าที่ต่างกันคาดว่าผู้สอบกลุ่มนี้จะเสียเปรียบในการตอบข้อสอบ ส่วนผู้สอบอีกกลุ่มหนึ่งเรียกว่า “กลุ่มอ้างอิง” (Reference group) ซึ่งเป็นกลุ่มผู้สอบที่ใช้เป็นมาตรฐานการเปรียบเทียบผลการตอบข้อสอบกับผู้สอบกลุ่มแรก ถ้าข้อสอบทำหน้าที่ต่างกัน แล้วคาดว่าผู้สอบกลุ่มหลังนี้จะได้เปรียบในการตอบข้อสอบ ในการเปรียบเทียบผลการตอบข้อสอบจะเปรียบเทียบทุกระดับความสามารถของผู้สอบกลุ่มย่อยสองกลุ่มที่มีระดับความสามารถเท่ากัน ในทางปฏิบัติมักใช้คะแนนรวมของแบบสอบเป็นเกณฑ์ในการจับคู่กลุ่มผู้สอบ

ในการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบเป็นรายข้อ เมื่อจับคู่กลุ่มผู้สอบแล้ว จะนำข้อมูลผลการตอบข้อสอบนั้น ระหว่างผู้สอบย่อย 2 กลุ่มมาจัดลงในตารางการถัวแบบ 2 X 2 (กลุ่มผู้สอบ 2 กลุ่ม X ผลการตอบ 2 แบบ) โดยที่ตารางการถัว 1 ตารางแทนคะแนนรวม 1 ระดับ ดังนั้นถ้ามีคะแนนรวมของกลุ่มผู้สอบทั้งสิ้น K ระดับ จะต้องสร้างตารางการถัวแบบ 2 X 2 ทั้งหมด K ตาราง สำหรับตารางการถัวแบบ 2 X 2 ของข้อสอบแต่ละข้อที่มีคะแนนรวมระดับ j ดังแสดงในตาราง 5

ตาราง 5 ผลการตอบข้อสอบข้อหนึ่งระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มี คะแนนรวม j

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

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

ตาราง 6 สัดส่วนของผู้สอบระหว่างผู้สอบกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ณ ระดับคะแนน j

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

- เมื่อ pR_j แทน สัดส่วนของผู้สอบกลุ่มอ้างอิงที่ระดับคะแนน j ซึ่งตอบข้อสอบถูก
 qR_j แทน สัดส่วนของผู้สอบกลุ่มอ้างอิงที่ระดับคะแนน j ซึ่งตอบข้อสอบผิด
 pF_j แทน สัดส่วนของผู้สอบกลุ่มเปรียบเทียบที่ระดับคะแนน j ซึ่งตอบข้อสอบถูก
 qF_j แทน สัดส่วนของผู้สอบกลุ่มเปรียบเทียบที่ระดับคะแนน j ซึ่งตอบข้อสอบผิด

ในการทดสอบสมมติฐานของการทำหน้าที่ต่างกันของข้อสอบจะกำหนดสมมติฐานศูนย์ (H_0) และสมมติฐานอื่น (H_1) ดังนี้

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

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

สมมติฐานศูนย์ตั้งขึ้นเป็นอิสระอย่างมีเงื่อนไขระหว่างกลุ่มผู้สอบและคะแนนผลการตอบข้อสอบที่ศึกษา ดังนั้นสัดส่วนคะแนนที่ได้จากตาราง 3 ภายใต้สมมติฐานศูนย์ที่ตั้งไว้สามารถคำนวณ เป็นค่าคาดหวัง (Expected values) ในแต่ละเซลล์ได้ดังนี้

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

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

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

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

สำหรับพารามิเตอร์ α ภายใต้สมมติฐานอื่นเรียกว่า อัตราส่วนแอดัมต่อรวม (Common odds ratio) ซึ่งคำนวณได้จาก

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

ถ้า $\alpha = 1$ แสดงว่าโอกาสของการตอบถูกระหว่างผู้สอบทั้งสองกลุ่ม มีค่าเท่ากันถ้า $\alpha > 1$ แสดงว่าผู้สอบกลุ่มอ้างอิงมีโอกาสตอบข้อสอบถูกมากกว่าผู้สอบกลุ่มเปรียบเทียบ และถ้า $\alpha < 1$ แสดงว่าผู้สอบกลุ่มเปรียบเทียบมีโอกาสตอบข้อสอบถูกมากกว่าผู้สอบกลุ่มอ้างอิง
แมนเทิลและแฮนเซล ได้เสนอวิธีประมาณค่า α จากตารางแบบ 2 X 2 จำนวน K ระดับดังนี้

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

α_{MH} เป็นค่าประมาณขนาดอิทธิพลของการทำหน้าที่ต่างกันของข้อสอบ (DIF effect size)

ซึ่งมีค่าอยู่ระหว่าง 0 ถึง ∞ ฮอลแลนด์; และเรเยอร์ (1985 cited in Holland ; & Thayer, 1988)

ได้เสนอให้แปลงค่า α_{MH} เป็นสเกลมาตรฐานเดลต้า (Delta scale ; MH_{DIF} หรือ MH_{DIF}) ดังนี้

$$MH_{DIF} = - 2.35 \ln (\alpha_{MH})$$

ค่า MH_{DIF} ดังกล่าวสามารถนำไปพิจารณาค่าความยากของข้อสอบ กล่าวคือ ถ้า MH_{DIF} มีค่าเป็นศูนย์ แสดงว่าข้อสอบนั้นมีความยากสำหรับแต่ละกลุ่มเท่ากัน ถ้า MH_{DIF} มีค่าเป็นลบ แสดงว่าข้อสอบยากสำหรับผู้สอบกลุ่มเปรียบเทียบมากกว่ากลุ่มอ้างอิง และถ้า MH_{DIF} มีค่าเป็นบวก แสดงว่าแสดงว่าข้อสอบยากสำหรับผู้สอบกลุ่มอ้างอิงมากกว่ากลุ่มเปรียบเทียบ ส่วนการประมาณค่าความคลาดเคลื่อนมาตรฐานของ MH_{DIF} สามารถคำนวณได้จากสูตร ดังนี้

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

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

$$\begin{aligned} \text{และ } U_i &= A_j D_j + \alpha_{MH} (B_j C_j) \\ V_i &= (A_j + D_j) + \alpha_{MH} (B_j + C_j) \end{aligned}$$

ในการทดสอบนัยสำคัญของสมมติฐาน จะต้องนำค่า α_{MH} หรือค่า MH_{DIF} ไปทดสอบกับสถิติแมนเทิล-แฮนส์เซล ไค-สแควร์ χ^2_{MH} ที่ระดับชั้นความเป็นอิสระเท่ากับ 1 (df = 1) สถิติ χ^2_{MH} มีสูตรในการคำนวณ ดังนี้

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

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

$$\text{Var}(A_j) = \frac{nR_j nF_j m_{1j} m_{0j}}{T_j^2 (T_j - 1)}$$

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

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

1. ค่า $\hat{\alpha}_{MH} = 1$ หรือไม่แตกต่างจาก 1 อย่างมีนัยสำคัญ แสดงว่าข้อสอบนั้นทำหน้าที่ไม่แตกต่างกันระหว่างกลุ่ม (no DIF)
2. ค่า $\hat{\alpha}_{MH} > 1$ แสดงว่าข้อสอบนั้นทำหน้าที่แตกต่างกันระหว่างกลุ่มโดยจะเข้าข้างกลุ่มอ้างอิง
3. ค่า $\hat{\alpha}_{MH} < 1$ แสดงว่าข้อสอบนั้นทำหน้าที่แตกต่างกันระหว่างกลุ่มโดยจะเข้าข้างกลุ่มเปรียบเทียบ

ส่วนค่า MH_{DIF} แปลผลการทดสอบได้ดังนี้

1. ค่า MH_{DIF} เท่ากับ 0 หรือไม่แตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติ แสดงว่าข้อสอบนั้นทำหน้าที่ไม่แตกต่างกันระหว่างกลุ่ม (no DIF)
2. ค่า MH_{DIF} แตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติ และมีค่าเป็นบวก (positive) แสดงว่าข้อสอบนั้นทำหน้าที่แตกต่างกันระหว่างกลุ่ม โดยจะเข้าข้างกลุ่มเปรียบเทียบ (F)
3. ค่า MH_{DIF} แตกต่างจาก 0 อย่างมีนัยสำคัญทางสถิติ และมีค่าเป็นลบ (negative) แสดงว่าข้อสอบนั้นทำหน้าที่แตกต่างกันระหว่างกลุ่ม โดยจะเข้าข้างกลุ่มอ้างอิง (R)

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

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบโดยวิธี MH นี้ใช้คะแนนรวมของแบบสอบเป็นเกณฑ์การจับคู่ ซึ่งมีจุดอ่อนในด้านความไม่เป็นอิสระของคะแนนรวม กับคะแนนรายข้อที่ทำการศึกษา ฮอลแลนด์; และเธเยอร์ (Holland; & Thayer, 1988) ได้เสนอวิธีแก้จุดอ่อนดังกล่าวเพื่อให้เกณฑ์การจับคู่ผู้สอบระหว่างกลุ่มมีความบริสุทธิ์ยิ่งขึ้น โดยใช้วิธีการ 2 ขั้นตอน ดังนี้

ขั้นตอนแรก นำคะแนนรวมของแบบสอบทั้งฉบับเป็นเกณฑ์การจับคู่ผู้สอบระหว่างกลุ่มย่อย 2 กลุ่ม แล้ววิเคราะห์การทำหน้าที่ต่างกันของข้อสอบเป็นรายข้อ เมื่อพบว่าข้อสอบข้อใดทำหน้าที่ต่างกันให้นำคะแนนของข้อสอบข้อนั้นออกจากคะแนนรวมของผู้สอบแต่ละคน

ขั้นตอนที่สอง ใช้คะแนนรวมของแบบสอบที่นำเอาคะแนนข้อสอบที่ทำหน้าที่ต่างกันซึ่งตรวจพบในขั้นตอนแรกออกไป เพื่อใช้เป็นเกณฑ์การจับคู่ แล้ววิเคราะห์การทำหน้าที่ต่างกันของข้อสอบซ้ำอีกครั้งหนึ่ง สำหรับนำไปใช้สรุปผลการตรวจสอบ

4.2 แนวคิดและหลักการของวิธี GMH

แมนเทิล-แฮนส์เซล (อุทัยวรรณ สายพัฒนา. 2547: 97-113; อ้างอิงจาก Mantel; & Haenszel) ได้เสนอสถิติรูปทั่วไปของ MH (Generalized MH statistic: GMH) ซึ่งเป็นวิธีการที่ขยายโดยตรงจากวิธี MH ธรรมดา ในกรณีที่ลำดับขั้นตอน (response categories) มีมากกว่า 2 ลำดับขั้น ($T > 2$) และลำดับขั้นการตอบไม่จำเป็นต้องมีการจัดเรียง ข้อมูลการตอบข้อสอบจะถูกจัดลงตารางการถักร $2 \times T \times k$ เมื่อ T เป็นจำนวนลำดับขั้นการตอบและ k เป็นจำนวนของระดับของการแบ่งชั้นตัวแปรการจับคู่ ในแต่ละระดับทั้งหมด k ระดับ ดังตาราง 7

ตาราง 7 แสดงผลการตอบข้อสอบของกลุ่มผู้สอบ ในแต่ละลำดับขั้นคะแนน

กลุ่ม	คะแนนผลของการตอบข้อสอบ					รวม
	y_1	y_2	y_3		y_T	
R	n_{R1k}	n_{R2k}	n_{R3k}	...	n_{RTk}	n_{R+k}
F	n_{F1k}	n_{F2k}	n_{F3k}	...	n_{FTk}	n_{F+k}
รวม	n_{+1k}	n_{+2k}	n_{+3k}	...	n_{+Tk}	n_{++k}

ค่า $y_1, y_2, y_3 \dots y_T$ แทน คะแนนในลำดับการตอบที่ได้จากข้อสอบข้อนั้น (เป็นจำนวนเต็ม $1, 2, \dots, T$)

n_{RTk} และ n_{FTk} แทน จำนวนของสมาชิกในกลุ่ม R และ F ตามลำดับ ณ ระดับตัวแปรจับคู่ที่ k และได้ y_T

เครื่องหมาย + แทน ผลรวมตลอดจนถึงที่บ่งชี้ไว้ เช่น n_{F+k} แทนจำนวนสมาชิกของกลุ่มสนใจทั้งหมดในระดับที่ k ซึ่งเวกเตอร์ของความถี่ของการตอบของกลุ่ม g มีการแจกแจงแบบ multivariate hypergeometric ด้วยพารามิเตอร์ n_{F+k} นั่นคือ

$$n_{Fk} = (n_{F1k}, n_{F2k}, \dots, n_{FTk})$$

และ $n_{+k} = (n_{+1k}, n_{+2k}, \dots, n_{+Tk})$

สมมติฐาน ; $H_0 : \alpha_k = \alpha = 1; k = 1, 2, \dots, k$

$H_1 : \alpha_k = \alpha \neq 1; k = 1, 2, \dots, k$

$$\begin{aligned} \text{ซึ่งแมนเทิลและแฮนส์เซล ได้กำหนด} \quad H_0 : \hat{\alpha}_{GMH} &= 1 \\ H_1 : \hat{\alpha}_{GMH} &\neq 1 \end{aligned}$$

สถิติทดสอบที่ใช้ในการทดสอบนัยสำคัญของสมมติฐาน คือ GMH_{χ^2} ซึ่งมีการแจกแจงแบบไค-สแควร์ ที่มี $df = T-1$ คำนวณได้จากสูตร

$$GMH_{\chi^2}^2 = [\sum A_k - \sum E(A_k)]' [\sum V(A_k)]^{-1} [\sum A_k - \sum E(A_k)]$$

$$\begin{aligned} \text{โดยที่} \quad A'_k &= (n_{R1k}, n_{R2k}, \dots, n_{R(T-1)k}) \\ E(A'_k) &= n_{R+k} n'_k / n_{++k} \\ n'_k &= (n_{+1k}, n_{+2k}, \dots, n_{+(T-1)k}) \end{aligned}$$

$$V(A_k) = n_{R+k} n_{F+k} \left(\frac{n_{++k} \text{diag}(n_k) - n_k n'_k}{n_{++k}^2 (n_{++k} - 1)} \right)$$

และ $\text{diag}(n_k)$ เป็นเมตริกซ์แนวทแยง (diagonal matrix) มิติ $(T-1) \times (T-1)$ ที่มีสมาชิกเป็น n_k เมื่อ A_k และ $E(A_k)$ เป็นเวกเตอร์ที่ยาว $T-1$ ที่สมนัยกับ $T-1$ ไตๆของลำดับชั้นการตอบ T และ $V(A_k)$ เป็นเมตริกซ์ความแปรปรวน-ความแปรปรวนร่วม (variance-covariance matrix) มิติ $(T-1) \times (T-1)$

งานวิจัยต่างประเทศ

เรนเดล เพ็นฟิลด์. (Randall D. Penfield. 2001: 235-259) ได้ทำการศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบและการเปรียบเทียบกระบวนการ MH 3 วิธี ประกอบด้วยวิธี MH วิธี BMH และวิธี GMH มีการตรวจสอบอัตราความคลาดเคลื่อนประเภทที่ 1 จากการทำการศึกษาพบว่า MH และ BMH จะเกิดความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธี GMH

แองเกิล ฟิดัลโก; และเดฟ บาร์ธแรม (Angel M. Fidalgo & Dave Bartram. 2010: 600-606) ได้ทำการศึกษาการสร้างความสัมพันธ์ที่มีประสิทธิภาพของการทดสอบ GMH และการทดสอบแบบ Mantel เพื่อค้นพบรูปแบบการทำหน้าที่ต่างกันของข้อสอบที่มีจำนวนมาก ในแบบทดสอบนี้จะดูอัตราความคลาดเคลื่อนประเภทที่ 1 ของการทดสอบ Mantel และใช้การจำลองข้อมูลขึ้นภายใต้ Graded Response มีการสร้างข้อสอบขึ้นประมาณ 10 หรือ 20 ข้อ มีแบบการให้คะแนนแบบหลายค่า กับข้อสอบ 5 หมวดหมู่ พารามิเตอร์ของข้อสอบเหมือนกันทุกประการ ใช้การทำซ้ำ 2 ครั้ง สำหรับการทำข้อสอบ 20 ข้อ มีรูปแบบในการดูการทำหน้าที่ต่างกันของข้อสอบ 4 รูปแบบที่มีการ

จำลองขึ้นมาเป็น มีความคงที่ สมดุล สมดุลต่ำและสมดุลสูง สถิติที่ใช้ GMH โปรแกรม GMHDIF ใช้กับการวิเคราะห์ข้อมูลที่จำลองขึ้น

ผลการวิเคราะห์ข้อมูล สิ่งที่น่าสนใจในหัวข้อแรกคือสนใจในข้อสรุปสมมติฐาน ประเภทคะแนนที่ใช้ $Q_{GMH(2)}$ คำนวณการมีอิทธิพล สามารถที่จะค้นพบรูปแบบของความแตกต่างของ DIF ดังนั้นการจำลองรูปแบบ DIF ทำให้ไม่เกิดการสมดุลต่ำ หรือเกิดการสมดุลกัน ใช้ลำดับของคะแนนเปรียบเทียบกับจำนวนเต็มปกติ ทำให้มีค่าเฉลี่ยเพิ่มขึ้น 18% (รูปแบบไม่สมดุลต่ำ) และ 42% (สมดุล) เปอร์เซ็นการค้นพบสามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดี ค้นพบการทำหน้าที่ต่างกันของข้อสอบที่มั่นคง 2 ระบบเสนอผลที่คล้ายกัน

การสนใจหัวข้อที่สอง (ความสัมพันธ์ของการทำ $Q_{GMH(1)}$ และ $Q_{GMH(2)}$) ที่ถูกค้นพบบนรูปแบบที่มีความแตกต่างกัน ผลที่แสดงให้เห็นว่าสถิติทั้งคู่สามารถควบคุมความคลาดเคลื่อนประเภทที่ 1 ได้ดี สำหรับเงื่อนไขที่เป็น 0 และไม่เป็น 0 ทำให้เปรียบเทียบได้ง่ายขึ้น ค่าเฉลี่ยจะแตกต่างกันมากระหว่าง $Q_{GMH(1)}$ และ $Q_{GMH(2)}$ โดยการใช้รูปแบบมั่นคงสม่ำเสมอ ในทางตรงกันข้าม ค่าเฉลี่ยแตกต่างกันน้อย $Q_{GMH(1)}$ มี power มากกว่า $Q_{GMH(2)}$ ผลที่ได้คือ $Q_{GMH(2)}$ มีความไวต่อค่าเฉลี่ยตรงกันข้ามกับ $Q_{GMH(1)}$ ไรต่อความรู้สึกทั้งหมด การกระจายคะแนนระหว่างกลุ่ม สถิติทั้ง 2 ตัว สามารถเข้าถึงได้ 86%

5. เอกสารและงานวิจัยที่เกี่ยวข้องกับการตรวจสอบการทำหน้าที่ต่างกันด้วยวิธี

LDFA

วิธีการถดถอยโลจิสติกสำหรับกรณีการให้คะแนนแบบสองค่าเป็นวิธีที่ใช้คะแนนสังเกตในรูปแบบพาราเมตริก (Parametric form) (สุชาติ สิริมีนนท์. 2554: 54–59) ของความสัมพันธ์ระหว่างคะแนนข้อสอบและตัวแปรที่ใช้ในการจับคู่ (Potenza; & Dorans. 1995) วิธีดังกล่าวตั้งอยู่บนโมเดลความน่าจะเป็นของการสังเกตการตอบข้อสอบที่มีคะแนนแบบสองค่า (U) ซึ่งเป็นฟังก์ชันของตัวแปรสำรวจ (Explanatory variable) สองตัว ซึ่งได้แก่ตัวแปรความสามารถที่สังเกตได้ (คะแนนแบบทดสอบรวม) (X) และตัวแปรชี้วัดกลุ่ม (G) โมเดลการถดถอยโลจิสติกจะไม่มีสัญลักษณ์แสดงข้อสอบในโมเดล สามารถเขียนแสดงได้ดังนี้ (Miller; & Spray. 1993: 108)

$$P(U | X, G) = \frac{e^{(1-U)(-\beta_0 - \beta_1 X - \beta_2 G - \beta_3 (X*G))}}{1 + e^{(1-U)(-\beta_0 - \beta_1 X - \beta_2 G - \beta_3 (X*G))}} \quad (1)$$

เราไม่สามารถขยายโมเดลการถดถอยโลจิสติกซึ่งใช้ข้อมูลการให้คะแนนแบบสองค่าไปใช้กับข้อมูลการให้คะแนนแบบหลายค่าเพียงครั้งเดียวโดยตรง ในการขยายวิธีการถดถอยโลจิสติกไปใช้กับข้อสอบที่ให้คะแนนแบบทดสอบหลายค่าข้อมูลต่างจะมีการกำหนดรหัสให้ใหม่เข้าไปในกลุ่มของเซตของข้อมูลแบบสองค่า ซึ่งแต่ละค่าก็เหมาะสำหรับการวิเคราะห์การถดถอยแบบแยกกัน มี

การขยายหลายครั้งของวิธีการถดถอยโลจิสติก เพื่อใช้กับข้อมูลที่มีการให้คะแนนแบบหลายค่าเพร็นซ์และ

มิลเลอร์ (French; & Miller. 1996) ได้แนะนำการขยายออก 3 ครั้งของโมเดลโลจิสติกสำหรับโมเดลการตอบแบบมัลติโนเมียล (Multinomial response model) โดยใช้ข้อมูลการตอบแบบจัดอันดับ (Ordinal response data) แต่ละค่าก็จะเกี่ยวข้องกับเซตของความแตกต่างของการเปรียบเทียบรายคู่ (pairwise comparisons) ระหว่างรายการคะแนน (Score categories) หรือผลรวมของรายการคะแนน โมเดลเหล่านี้ได้แก่ โมเดลโลจิทอัตราส่วนแบบต่อเนื่อง (Continuation ratio logits) โมเดลโลจิทสะสม (Cumulative logits) และโมเดลโลจิทรายการที่อยู่ติดกัน (Adjacent categories logits) ในแต่ละวิธีเหล่านี้สมมุติว่ามีรายการตอบข้อสอบ (Item response categories) จำนวน J อันดับ (Ordered) ดังนั้นทั้งสามวิธีจะเกี่ยวข้องกับการประมวลผลข้อมูลแบบจัดอันดับเข้าไปในเซตของข้อมูลที่มีการตอบแบบสองค่าจำนวน $J-1$ เซต วิธีดังกล่าวตั้งอยู่บนพื้นฐานโลจิท (Logits) หรืออัตราส่วนความน่าจะเป็นของการได้ค่าหนึ่งๆต่อความน่าจะเป็นของการได้คะแนนสูงหรือต่ำกว่าค่าหนึ่งๆ สมมุติว่าข้อสอบข้อหนึ่งมีรายการการตอบ 4 รายการ ก็จะมี 3 โลจิท และมีโค้งโลจิท 3 โค้ง ในแต่ละวิธีนั้นจะมีการคิดคำนวณการถดถอยที่แยกกันสำหรับแต่ละโลจิท เพื่อใช้ทดสอบการทำหน้าที่ต่างกันของข้อสอบ (ที่เป็นรูปแบบเดียวกัน, ไม่เป็นรูปแบบเดียวกัน) วิธีทั้งสามสามารถให้รายละเอียดเพิ่มเติมดังนี้

วิธีแรก โลจิทอัตราส่วนแบบต่อเนื่อง (Continuation ratio logits) วิธีนี้จะใช้จำนวนเล็กๆที่เพิ่มขึ้นของข้อมูลในแต่ละการถดถอยที่ต่อเนื่องกัน เพื่อที่จะทดสอบการมีอยู่ของข้อสอบที่ทำหน้าที่ต่างกัน ในการถดถอยแรกความน่าจะเป็นของการได้คะแนน 0 ถูกเปรียบเทียบกับความน่าจะเป็นของการได้คะแนนมากกว่า 0 ในการถดถอยที่สองความน่าจะเป็นของการได้คะแนน 1 ถูกเปรียบเทียบกับความน่าจะเป็นของการได้คะแนนมากกว่า 1 และในการถดถอยที่สามความน่าจะเป็นของการได้คะแนน 2 ถูกเปรียบเทียบกับความน่าจะเป็นของการได้คะแนน 3 วิธีดังกล่าวสามารถแสดงง่าย ๆ ดังนี้

เมื่อข้อมูลที่ใช้มีรายการตอบ 4 รายการในที่นี้ถ้าให้ k เป็นรายการการตอบ และ $k = 0, 1, 2, 3$

ความน่าจะเป็นของการตอบรายการที่ $Y = k$ ถูกเปรียบเทียบกับความน่าจะเป็นของการตอบรายการ $Y > k$ ดังนั้นความน่าจะเป็นที่ $Y = 0$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 1, 2, 3$ ความน่าจะเป็นที่ $Y = 1$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 2, 3$ ความน่าจะเป็นที่ $Y = 2$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 3$

วิธีที่สอง โลจิทรายการที่อยู่ติดกัน (Adjacent categories logit) วิธีนี้ยึดหลักว่า ความน่าจะเป็นการตอบรายการใดถูกเปรียบเทียบกับความน่าจะเป็นการตอบรายการที่อยู่ติดกับมัน ในการถดถอยแรก ความน่าจะเป็นของการได้คะแนน 0 ถูกเปรียบเทียบกับความน่าจะเป็นของได้คะแนน 1 ในการถดถอยที่สองความน่าจะเป็นของการได้คะแนน 1 ถูกเปรียบเทียบกับความน่าจะเป็น

เป็นของการได้คะแนน 2 และในการถดถอยที่สาม ความน่าจะเป็นของการได้คะแนน 2 ถูกเปรียบเทียบกับความน่าจะเป็นของการได้คะแนน 3 วิธีดังกล่าวสามารถแสดงง่าย ๆ ดังนี้

เมื่อข้อมูลที่ใช้มีรายการตอบ 4 รายการ ในที่นี้ถ้าให้ k เป็นรายการการตอบ และ $k = 0, 1, 2, 3$ ความน่าจะเป็นของการตอบรายการที่ $Y = k$ ถูกเปรียบเทียบกับความน่าจะเป็นของการตอบรายการที่ $Y = k + 1$ ดังนั้นความน่าจะเป็นที่ $Y = 0$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 1$ ความน่าจะเป็นที่ $Y = 1$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 2$ ความน่าจะเป็นที่ $Y = 2$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 3$ วิธีนี้พบว่า มีเพียงสองรายการถูกเปรียบเทียบในแต่ละการถดถอยทำให้สารสนเทศจำนวนมากหายไปจากการประมวลผลด้วยวิธีนี้

วิธีที่สาม โลจิสติกสะสม (Cumulative logits) วิธีนี้ยึดหลักว่า ความน่าจะเป็นของการได้คะแนนจากการตอบรายการใดหรือน้อยกว่าถูกเปรียบเทียบกับความน่าจะเป็นของการตอบรายการอื่น ๆ ในการถดถอยแรกความน่าจะเป็นของการได้คะแนน 0 ถูกเปรียบเทียบกับความน่าจะเป็นของการได้คะแนนอื่น ๆ ในการถดถอยที่สองความน่าจะเป็นของการได้คะแนน 1 หรือน้อยกว่าถูกเปรียบเทียบกับความน่าจะเป็นของการได้คะแนนอื่น ๆ และในการถดถอยที่สามความน่าจะเป็นของการได้คะแนน 2 หรือน้อยกว่าถูกเปรียบเทียบกับความน่าจะเป็นของการได้คะแนน 3 วิธีดังกล่าวสามารถแสดงง่าย ๆ ดังนี้

เมื่อข้อมูลที่ใช้มีรายการตอบ 4 รายการ ในที่นี้ถ้าให้ k เป็นรายการการตอบ และ $k = 0, 1, 2, 3$ ความน่าจะเป็นของการตอบรายการที่ $Y \leq k$ ถูกเปรียบเทียบกับความน่าจะเป็นของการตอบรายการที่ $Y > k$ ดังนั้นความน่าจะเป็นที่ $Y = 0$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 1, 2, 3$ ความน่าจะเป็นที่ $Y = 0, 1$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 2, 3$ และความน่าจะเป็นที่ $Y = 0, 1, 2$ จะถูกเปรียบเทียบกับความน่าจะเป็นที่ $Y = 3$

โมเดลโลจิสติกสะสมมีความสำคัญทั้งทางทฤษฎีและการประยุกต์ไปใช้มากกว่าสองโมเดลแรก เพราะไม่มีสารสนเทศใดสูญหายไปจากการให้รหัสและจากการศึกษาการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบหลายค่าของเฟรนช์และมิลเลอร์ (French; & Miller. 1996) โดยใช้วิธีโลจิสติกอัตราส่วนแบบต่อเนื่อง และวิธีโลจิสติกสะสม พบว่าวิธีทั้งสองเป็นวิธีที่ให้ผลค่อนข้างดี แต่การประมวลผลการถดถอยที่แยกกันสำหรับแต่ละโลจิสติกนั้น ทำให้แปรผลยากเพราะพารามิเตอร์หลายตัวต้องถูกทดสอบนัยสำคัญทางสถิติ นอกจากนี้ นักวิจัยได้ปรับขยายโมเดลโลจิสติกโดยนำโมเดลการถดถอยโลจิสติกไปประยุกต์ใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนหลายค่า (Polytomous) ด้วยวิธีอื่นๆ ดังเช่นมิลเลอร์และสเปรย์ (Miller; & Spray. 1993) ได้เสนอวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (Logistic discriminant function analysis) ซึ่งเป็นวิธีที่ง่ายกว่า และใช้สะดวกกว่าวิธีที่ได้จากการขยายแบบให้คะแนนแบบหลายค่าที่ใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ (Multinomial logistic regression: MLR) ส่วนซัมโบ (Zumbo. 1999) ได้ใช้โมเดลการถดถอยโลจิสติกแบบจัดอันดับ (Ordinal logistic regression model) โดยอาศัยวิธีโลจิสติกสะสม ทำให้ได้สารสนเทศครบถ้วน วิธีโลจิสติกทั้งหมดถูกทดสอบพร้อมกัน การตัดสินใจเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบตั้งอยู่บนการทดสอบนัยสำคัญทางสถิติ และการวัดขนาดของผล (Effect size)

5.1 แนวคิดและหลักการ

ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบหลายค่าตามวิธีการถดถอยโลจิสติกแบบมัลติโนเมียล (Multinomial logistic regression: MKR) เกิดความยุ่งยากในการแปลผล (สุชาติ สิริมินนนท์. 2554: 56–59) มิลเลอร์และสเปรย์ (Miller; & Spray. 1993) จึงได้เสนอวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (Logistic discriminant function analysis: LDFA) เพื่อใช้สำหรับตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ให้คะแนนแบบหลายค่า วิธีดังกล่าวอยู่บนโมเดลพื้นฐานที่ใช้สำหรับคัดแยกคนออกเป็นกลุ่มๆ ที่มีส่วนเหมือนกันบนพื้นฐานของการตอบข้อสอบและคะแนนแบบทดสอบ วิธีนี้ไม่ต้องการสมมติฐานของความเป็นปกติ (Normality) คือตัวแปรอิสระของวิธีนี้ไม่จำเป็นต้องมีการแจกแจงปกติ วิธีนี้ง่ายกว่าวิธีการถดถอยโลจิสติกแบบมัลติโนเมียล (MLR) เนื่องจากว่ามีเพียงแค่การถดถอยเดียวที่จำเป็นต้องใช้ในการทดสอบแต่ละโมเดล วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีความเกี่ยวข้องสัมพันธ์อย่างใกล้ชิดกับวิธีการถดถอยโลจิสติกแบบมัลติโนเมียล และการวิเคราะห์ล็อก-ลิเนียร์ (Log-linear analysis) สำหรับความแตกต่างที่เห็นได้ชัดนั้นโมเดลการถดถอยในวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก ใช้เพื่อทำนายการเป็นสมาชิกกลุ่ม ส่วนโมเดลการถดถอยในวิธีการถดถอยโลจิสติกแบบมัลติโนเมียล จะใช้เพื่อทำนายคะแนนสอบสามารถเขียนโมเดลการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกได้ดังนี้

$$P(G | X, U) = \frac{e^{(1-G)(-\alpha_0 - \alpha_1 X - \alpha_2 U - \alpha_3 (X*U))}}{1 + e^{(1-G)(-\alpha_0 - \alpha_1 X - \alpha_2 U - \alpha_3 (X*U))}} \quad (2)$$

เมื่อ	α_i (I = 0,1,2,3) แทนสัมประสิทธิ์ฟังก์ชันการจำแนก
G	แทน กลุ่มผู้สอบ
X	แทน คะแนนรวมแบบทดสอบ
U	แทน ผลการตอบข้อสอบ
P	แทน ความน่าจะเป็นของการเป็นสมาชิกกลุ่ม

โมเดลนี้โดยพื้นฐานเหมือนกับโมเดลที่ให้ไว้ในสมการที่ (1) อย่างไรก็ตามจะมีความแตกต่างบางประการกล่าวคือ ประการแรก การเป็นสมาชิกกลุ่ม (G) ในโมเดลจะแทนที่ของการตอบข้อสอบ (U) และตัวแปรกลุ่ม (G) จะเป็นตัวแปรชี้วัด (Indicator variable) กำหนดว่า G = 1 สำหรับกลุ่มอ้างอิง และ G = 0 สำหรับกลุ่มสนใจ ประการที่สอง การตอบข้อสอบ (U) ไม่จำเป็นต้องมีเพียง 2 รายการเท่านั้น แต่สามารถรับเอาไว้บนการตอบข้อสอบที่มีจำนวนของรายการใดก็ได้ ประการที่สาม ตัวแปรการตอบข้อสอบ (U) เป็นตัวทำนาย (Predictor) มากกว่าที่จะเป็นตัวแปรตาม และประการที่สี่ สัมประสิทธิ์การถดถอยที่ให้ไว้โดย α_i เมื่อ $i = 0,1,2,3$ เพื่อให้เห็นว่าแตกต่างจาก β ในโมเดลสมการ (1)

5.2 กระบวนการตรวจสอบ

ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ โดยใช้วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (LDFA) ของมิลเลอร์และสเปรย์ (Miller; & Spray. 1993) จะใช้กระบวนการตรวจสอบความเหมาะสมของข้อมูลกับโมเดล โดยจะทดสอบโมเดล 3 ขั้นตอนดังนี้

ขั้นตอนที่ 1 ทดสอบตัวแปรที่ใช้ในการจับคู่ความสามารถ (Matching variable) หรือตัวแปรสำรวจ (Explanatory variable) ในที่นี้คือคะแนนรวมแบบทดสอบ (X) ซึ่งระดับความสามารถของผู้สอบ ในขั้นตอนนี้ถือว่าอิทธิพลของตัวแปรคะแนนแบบทดสอบรวมมีค่าเท่ากันในแต่ละกลุ่ม (G) ซึ่งกำหนดรหัสตัวแปรดัมมี่ (Dummy variable) โดยให้กลุ่มอ้างอิงแทนด้วย 1 และกลุ่มสนใจแทนด้วย 0 การนำตัวแปรเข้าไปในสมการถดถอยมีลักษณะดังนี้

$$P(G | X, U) = P(G, X) = \frac{e^{(1-G)(-\alpha_0 - \alpha_1 X)}}{1 + e^{(1-G)(-\alpha_0 - \alpha_1 X)}} \quad (3)$$

สมการดังกล่าวจะทำนายถึงการเป็นสมาชิกกลุ่มจากคะแนนแบบทดสอบรวมเท่านั้น เมื่อเขียนสมการ (3) ในรูปโลจิทจะได้

$$\text{logit} = \ln \left[\frac{P(G=1)}{P(G=0)} \right] = \alpha_0 + \alpha_1 X \quad (4)$$

ขั้นตอนที่ 2 ทดสอบตัวแปรการตอบข้อสอบที่มีรายการคำตอบใดๆ (U) ซึ่งไม่จำเป็นต้องมีเพียง 2 รายการคำตอบ โดยเพิ่มตัวแปรดังกล่าวในโมเดลสมการ (3) ของขั้นตอนที่ 1 ดังนี้

$$P(G | X, U) = \frac{e^{(1-G)(-\alpha_0 - \alpha_1 X - \alpha_2 U)}}{1 + e^{(1-G)(-\alpha_0 - \alpha_1 X - \alpha_2 U)}} \quad (5)$$

สมการดังกล่าวจะทำนายการเป็นสมาชิกกลุ่มจากคะแนนแบบทดสอบรวม และคะแนนข้อสอบ เมื่อเขียนสมการ (5) ในรูปโลจิท จะได้

$$\text{logit} = \ln \left[\frac{P(G=1)}{P(G=0)} \right] = \alpha_0 + \alpha_1 X + \alpha_2 U \quad (6)$$

ขั้นตอนที่ 3 ทดสอบตัวแปรปฏิสัมพันธ์ระหว่างคะแนนแบบทดสอบรวม (ระดับความสามารถ) กับการตอบข้อสอบ ($X \times U$) โดยเพิ่มตัวแปรดังกล่าวในโมเดลสมการ (5) ของขั้นตอนที่ 2 ดังนี้

$$P(G | X, U) = \frac{e^{(1-G)(-\alpha_0 - \alpha_1 X - \alpha_2 U - \alpha_3 (X * U))}}{1 + e^{(1-G)(-\alpha_0 - \alpha_1 X - \alpha_2 U - \alpha_3 (X * U))}} \quad (7)$$

สมการดังกล่าวจะทำนายการเป็นสมาชิกกลุ่มจากคะแนนแบบทดสอบรวม , คะแนนข้อสอบและปฏิสัมพันธ์ระหว่างคะแนนแบบทดสอบรวมกับค่าคะแนนข้อสอบ เมื่อเขียนสมการ (7) ในรูปโลจิท จะได้

$$\text{logit} = \ln \left[\frac{P(G=1)}{P(G=0)} \right] = \alpha_0 + \alpha_1 X + \alpha_2 U + \alpha_3 (X * U) \quad (8)$$

5.3 การทดสอบสมมติฐาน

การทดสอบการทำหน้าที่ต่างกันที่เป็นรูปแบบเดียวกัน (Uniform DIF) และที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) สามารถทดสอบได้พร้อมกัน โดยกำหนดสมมติฐานศูนย์ (H_0) และสมมติฐานทางเลือก (H_1) ดังนี้

$$\begin{aligned} H_0 : \alpha_2 = 0 \text{ และ } \alpha_3 = 0 \\ H_1 : \alpha_2 \neq 0 \text{ หรือ } \alpha_3 \neq 0 \end{aligned} \quad (9)$$

การทดสอบนัยสำคัญทางสถิติของการทำหน้าที่ต่างกันของข้อสอบจะดำเนินการตามกระบวนการการทดสอบความเหมาะสมของข้อมูลกับโมเดล 3 ขั้นตอน โดยการนำตัวแปรเข้าทดสอบกับโมเดลตามลำดับขั้นตอนกล่าวคือ ขั้นตอนที่ 1 ใส่ตัวแปรสำรวจ (Explanatory variable) หรือตัวแปรคะแนนแบบทดสอบรวม (X) เข้าไปในสมการ ขั้นตอนที่ 2 ใส่ตัวแปรการตอบข้อสอบ (U) เข้าไปในสมการ และขั้นตอนที่ 3 ใส่ตัวแปรปฏิสัมพันธ์ระหว่างคะแนนแบบทดสอบรวมกับตัวแปรการตอบข้อสอบ ($X*U$) เข้าไปในสมการ สถิติไค-สแควร์ของอัตราส่วนความน่าจะเป็น (G^2) ถูกคำนวณออกมาในแต่ละโมเดล ความแตกต่างในค่า G^2 ระหว่างโมเดลลำดับขั้น (Hierarchical model) มีการกระจายแบบเชิงเส้นกำกับ (Asymptotically) เป็นไค-สแควร์ ที่มีระดับชั้นความเป็นอิสระเท่ากับ 1 (df = 1) สถิติการทดสอบดังกล่าวจะเป็น G^2_{diff} โดยที่

$$G^2_{diff} = -2(\ln L_{(i+1)} - \ln L_{(i)}) \quad (10)$$

เมื่อ $L_{(i+1)}$ และ $L_{(i)}$ แทนความน่าจะเป็นของโมเดลลำดับขั้นที่อยู่ติดกัน ดังนั้นในการทดสอบการทำหน้าที่ต่างกันของข้อสอบที่เป็นรูปแบบเดียวกันและที่ไม่เป็นรูปแบบเดียวกัน ถ้าค่า G^2_{diff} ที่ได้แตกต่างจากของ G^2 จากโมเดลในขั้นตอนที่ 3 และโมเดลในขั้นตอนที่ 1 สามารถเปรียบเทียบค่า G^2 ที่ระดับชั้นความเป็นอิสระเท่ากับ 2 ถ้าผลการทดสอบมีนัยสำคัญทางสถิติโดย

ปฏิเสธสมมติฐานศูนย์ นั้นแสดงว่าข้อสอบทำหน้าที่ต่างกันซึ่งเป็นผลมาจากอิทธิพลของการตอบข้อสอบ (คะแนนข้อสอบ) หรืออิทธิพลของปฏิสัมพันธ์ระหว่างคะแนนแบบทดสอบรวมกับคะแนนข้อสอบ สำหรับการทดสอบการทำหน้าที่ต่างกันของข้อสอบที่เป็นรูปแบบเดียวกัน ถ้าค่า G^2_{diff} ที่ได้จากความแตกต่างของ G^2 จากโมเดลในขั้นตอนที่ 2 และโมเดลในขั้นตอนที่ 1 สามารถเปรียบเทียบค่า G^2 ที่ระดับชั้นความเป็นอิสระเท่ากับ 1 ($df = 1$) ถ้าผลการทดสอบมีนัยสำคัญทางสถิติแสดงว่าข้อสอบข้อนั้นทำหน้าที่ต่างกันของข้อสอบที่เป็นรูปแบบเดียวกัน ในการทดสอบนัยสำคัญทางสถิติถ้าการทดสอบการทำหน้าที่ต่างกันของข้อสอบที่เป็นรูปแบบเดียวกันหรือไม่เป็นรูปแบบเดียวกัน มีนัยสำคัญทางสถิติ หมายความว่าอย่างน้อย 1 รายการคำตอบ ในการที่จะเป็นสมาชิกกลุ่ม ซึ่งได้ให้คะแนนข้อสอบและคะแนนแบบทดสอบรวม แตกต่างจากรายการซึ่งถูกทำนายจากคะแนนรวมเพียงอย่างเดียว

5.4 งานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันโดยใช้วิธี LDFA

งานวิจัยในประเทศ

วิชา ะม้อย (2550: 115-129) ได้เปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่ตรวจให้คะแนนแบบพหุวิภาค ระหว่างวิธีโครงสร้างความแปรปรวนร่วมและค่าเฉลี่ย (MACS) กับวิธีวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก (LDFA) เมื่อมีเงื่อนไขความยาวของแบบสอบและขนาดของกลุ่มตัวอย่างที่แตกต่างกัน ผลวิจัยสรุปได้ว่า วิธี LDFA มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงตั้งแต่ร้อยละ 0 ถึงร้อยละ 15.6 และวิธี MACS มีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในช่วงตั้งแต่ร้อยละ 8.3 ถึงร้อยละ 21.9 โดยวิธี LDFA เกือบทุกเงื่อนไขอยู่ในเกณฑ์ที่กำหนด ส่วนวิธี MACS พบว่าส่วนใหญ่อยู่เกินกว่าเกณฑ์ที่กำหนด เมื่อพิจารณาการเปรียบเทียบอำนาจการทดสอบพบว่า วิธี LDFA มีอำนาจการทดสอบอยู่ในช่วงตั้งแต่ร้อยละ 62.5 ถึงร้อยละ 100 สำหรับวิธี MACS พบว่ามีค่าอำนาจการทดสอบอยู่ในช่วงตั้งแต่ร้อยละ 37.5 ถึงร้อยละ 100 และพบว่าทั้งสองวิธีมีอำนาจการทดสอบในเงื่อนไขความยาวแบบสอบ 30 ข้อสูงกว่าเงื่อนไขความยาวแบบสอบ 40 ข้อทุกเงื่อนไข

สุชาติ สิริมินันท์ (2554: 137-157) ได้ศึกษาและเปรียบเทียบความถูกต้องและความคลาดเคลื่อนประเภทที่ 1 จากการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าด้วยวิธีโพลีโทมัสซิบเทสท์ วิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกและวิธีการถดถอยโลจิสติกแบบจัดอันดับ โดยใช้ข้อมูลจำลอง ข้อมูลจำลองภายใต้ปัจจัยที่แปรเปลี่ยน 4 ปัจจัย ผลวิจัยสรุปได้ว่า ปัจจัยรูปแบบของข้อสอบทำหน้าที่เบี่ยงเบนมีผลต่ออัตราความถูกต้องในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าของวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก สูงกว่าโพลีโทมัสซิบเทสท์ และวิธีการถดถอยโลจิสติกแบบจัดอันดับในทุกเงื่อนไข และในภาพรวมวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกมีประสิทธิภาพในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบสูงกว่าอีกสองวิธี

งานวิจัยต่างประเทศ

ยา-ฮุย ซู; และเว่ย-ซาง หวัง (Ya-Hui Su & Wen-Chung Wang. 2005: 313-350) ได้ทำการศึกษาประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วย 3 วิธี คือวิธีแมนเทล วิธีแมนเทล-แฮนส์เซลทั่วไป และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก โดยพิจารณาจากอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีแมนเทล และวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกให้อำนาจการตรวจสอบที่สูงกว่าวิธีแมนเทล-แฮนส์เซลทั่วไป ภายใต้ทุกเงื่อนไข ยกเว้นรูปแบบสมดุลง่าย ภายใต้รูปแบบที่สมดุลง่ายว่าค่าลักษณะข้อสอบมีลักษณะบิดเบือนไประหว่างกลุ่ม เกิดการทำหน้าที่ต่างกันของข้อสอบขนาดใหญ่ มีความคงที่ของคะแนนที่มีการบิดเบือน ในกรณีนี้จะทำให้วิธีแมนเทล-แฮนส์เซลทั่วไปได้เปรียบในการตรวจสอบทำให้มีอำนาจการทดสอบสูงกว่าวิธีแมนเทล และวิธีวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก

จากการศึกษางานวิจัยที่เกี่ยวข้องทั้งในประเทศและต่างประเทศ พบว่าการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบสองค่าเป็นที่นิยมอย่างมากแต่จะพบว่าการศึกษาภายในประเทศที่เกี่ยวกับการตรวจสอบการทำหน้าที่ต่างกันของผู้สอบที่มีการให้คะแนนแบบหลายค่า นั้นยังมีไม่มากนัก โดยข้อมูลที่น่าสนใจมีทั้งการจำลองข้อมูลและการเก็บข้อมูลจริงจากกลุ่มตัวอย่าง ซึ่งวิธีที่น่าสนใจในการนำมาตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่าคือวิธีแมนเทล-แฮนส์เซลทั่วไปและวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก รวมถึงวิธี Polytomous SIBTEST ซึ่งเป็นวิธีที่เป็นที่นิยมใช้ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบอีกด้วย

บทที่ 3

วิธีดำเนินการวิจัย

ในการวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการตามขั้นตอนดังนี้

1. การกำหนดประชากรและกลุ่มตัวอย่าง
2. การสร้างเครื่องมือที่ใช้ในการวิจัย
3. การเก็บรวบรวมข้อมูล
4. การจัดกระทำและการวิเคราะห์ข้อมูล

ประชากรและกลุ่มตัวอย่าง

ประชากร

ประชากรที่ใช้ในการวิจัยครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2559 ของโรงเรียนขนาดใหญ่ในจังหวัดกรุงเทพมหานคร สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาทั้งหมด 39 โรงเรียน จำนวน 14,299 คน เป็นนักเรียนชาย 6,603 คน และเป็นนักเรียนหญิง 7,696 คน (กลุ่มสารสนเทศ สำนักนโยบายและแผน สพฐ. 2560: ออนไลน์)

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2559 ของโรงเรียนขนาดใหญ่ในจังหวัดกรุงเทพมหานคร สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา สาเหตุที่เลือกกรุงเทพมหานครเท่านั้น เนื่องจากต้องการป้องกันการทำหน้าที่ต่างกันที่มาจากตัวแปรสถานที่ตั้งทางภูมิศาสตร์ของโรงเรียน ซึ่งได้มาจากการสุ่มแบบแบ่งกลุ่ม (Cluster Random Sampling) โดยกำหนดขนาดกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์จากตารางประมาณกลุ่มตัวอย่างด้วยความเชื่อมั่น 95% ($\alpha = .05$) เมื่อเทียบจากตารางขนาดกลุ่มตัวอย่างที่ควรเลือกจากประชากร พบว่าต้องใช้กลุ่มตัวอย่าง อย่างน้อย 385 คน (ศิริชัย กาญจนวาสี; ทวีวัฒน์ ปิตยานนท์; และดิเรก ศรีสุขโข. 2555: 150) แต่เนื่องจากงานวิจัยที่ได้ทำการศึกษาเกี่ยวกับการทำหน้าที่ต่างกันของข้อสอบนั้นพบว่า ขนาดของกลุ่มตัวอย่างไม่ควรต่ำกว่า 500 คน หรือกลุ่มตัวอย่างขนาดใหญ่อย่างน้อย 1,000 คน จะทำให้ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมีความถูกต้องในการตรวจสอบสูง ดังนั้นงานวิจัยเล่มนี้จึงกำหนดให้มีจำนวนกลุ่มตัวอย่างไม่ต่ำกว่า 1,000 คน ซึ่งมีขั้นตอนการสุ่มดังนี้

ขั้นที่ 1 แบ่งโรงเรียนในจังหวัดกรุงเทพมหานคร ตามเขตพื้นที่การศึกษาคือ สำนักงานเขตพื้นที่การศึกษามัธยมศึกษา(สพม.)เขต 1 และเขต 2 โดยเลือกกลุ่มตัวอย่างจากโรงเรียนที่มีขนาดใหญ่ เนื่องจากต้องการป้องกันการทำหน้าที่ต่างกันที่มาจากขนาดของโรงเรียน โดย สพม.1 มี

ทั้งหมด 18 โรงเรียน และ สพม.2 มีทั้งหมด 21 โรงเรียน สุ่มกลุ่มตัวอย่างด้วยวิธีการสุ่มแบบแบ่งกลุ่ม (Cluster Random Sampling) ได้สำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 1 ซึ่งมีทั้งหมด 18 โรงเรียน รวมนักเรียนทั้งหมด 6,279 คน เป็นนักเรียนชาย 2,846 คน และนักเรียนหญิง 3,433 คน

ขั้นที่ 2 สุ่มโรงเรียนในสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 1 โดยการใช้การสุ่มแบบแบ่งกลุ่ม (Cluster Random Sampling) จาก 18 โรงเรียนได้ 9 โรงเรียน

ขั้นที่ 3 สุ่มห้องเรียนจากแต่ละโรงเรียน ด้วยวิธีการสุ่มแบบแบ่งกลุ่ม (Cluster Random Sampling) ตามสัดส่วนของนักเรียน เพื่อให้ได้กลุ่มตัวอย่างตามจำนวนที่ต้องการ รายชื่อโรงเรียน , จำนวนห้องเรียนและจำนวนนักเรียน ดังแสดงในตาราง 8

ตาราง 8 รายชื่อโรงเรียน จำนวนห้องเรียน และจำนวนนักเรียน

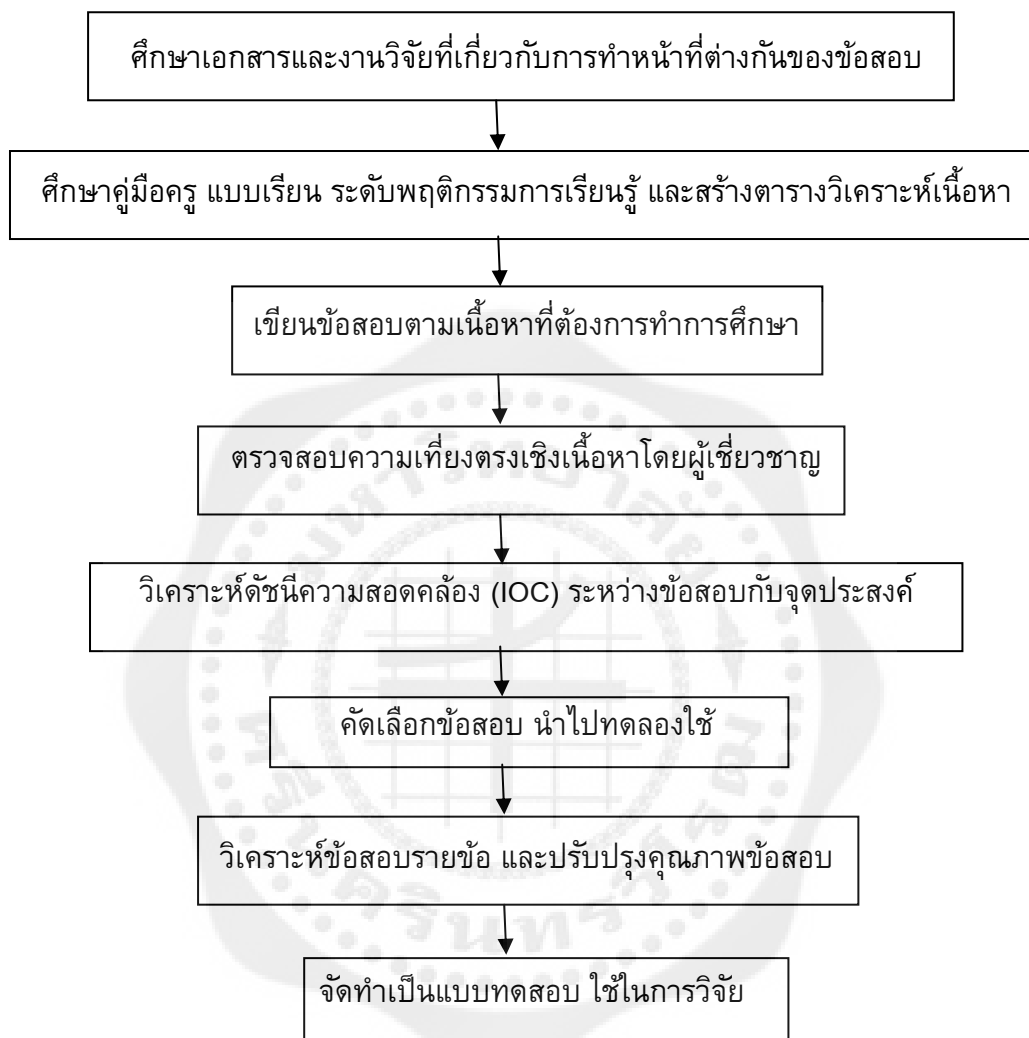
โรงเรียน	จำนวน ห้องเรียน	จำนวนนักเรียน		
		ชาย	หญิง	รวม
1. จันทรประดิษฐารามวิทยาคม	2	34	20	54
2. ปัญญาวรคุณ	2	30	32	62
3. ราชวินิตบางแคปานขำ	2	38	39	77
4. วัดอินทาราม	3	56	53	109
5. อิสลามวิทยาลัยแห่งประเทศไทย	3	55	70	125
6. วัดพุทธบูชา	3	59	75	134
7. สตรีวัดอัมพรสวรรค์	4	0	160	160
8. สุวรรณารามวิทยาคม	5	84	64	148
9. วัดราชบพิธ	5	152	0	152
รวม	29	508	513	1,021

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 เรื่องความเท่ากันทุกประการ ที่ผู้วิจัยสร้างขึ้นมีลักษณะเป็นแบบทดสอบแบบต่อเนื่องสองขั้นตอน โดยเป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก ในขั้นตอนแรกเป็นการเลือกตอบแบบถูกผิด และขั้นตอนที่สองเป็นการให้เหตุผลประกอบการเลือกตอบในขั้นตอนแรก โดยมีการให้คะแนนเป็น 0 , 1 และ 2 ตามลักษณะของการตอบ จำนวน 20 ข้อ

การสร้างเครื่องมือที่ใช้ในการวิจัย

ในการสร้างเครื่องมือสำหรับการวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการสร้างตามขั้นตอน
ดังภาพประกอบ 5



ภาพประกอบ 5 ขั้นตอนการสร้างเครื่องมือที่ใช้ในการวิจัย

จากแผนภูมิการดำเนินการสร้างเครื่องมือที่ใช้ในการวิจัย มีรายละเอียดการดำเนินการ
สร้างดังนี้

1. ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการทำหน้าที่ต่างกันของข้อสอบ เพื่อศึกษา
ลักษณะของข้อคำถามที่แสดงการทำหน้าที่ต่างกันของข้อสอบ
2. ศึกษาคู่มือครูและแบบเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 เรื่องความเท่ากัน
ทุกประการเพื่อพิจารณาโครงสร้างหลักสูตร กำหนดเนื้อหา และจำนวนแบบทดสอบ (สำนักงาน
คณะกรรมการการศึกษาขั้นพื้นฐาน กระทรวงศึกษาธิการ 2551: 41) ได้กำหนดตัวชี้วัดเรื่องความ

เท่ากันทุกประการ คือ ใช้สมบัติเกี่ยวกับความเท่ากันทุกประการของรูปสามเหลี่ยมและสมบัติของเส้นขนานในการให้เหตุผลและแก้ปัญหา โดยเนื้อหาความเท่ากันทุกประการต้องการให้นักเรียนเกิดความเข้าใจและสามารถใช้บทนิยามและสมบัติเกี่ยวกับความเท่ากันทุกประการในการให้เหตุผลเบื้องต้นทางเรขาคณิต ซึ่งมีเนื้อหาสรุปได้ดังนี้

- 2.1 ความเท่ากันทุกประการของรูปเรขาคณิต
- 2.2 ความเท่ากันทุกประการของรูปสามเหลี่ยม
- 2.3 รูปสามเหลี่ยมสองรูปที่สัมพันธ์กันแบบ ด้าน-มุม-ด้าน
- 2.4 รูปสามเหลี่ยมสองรูปที่สัมพันธ์กันแบบ มุม-ด้าน-มุม
- 2.5 รูปสามเหลี่ยมสองรูปที่สัมพันธ์กันแบบ ด้าน-ด้าน-ด้าน
- 2.6 การนำไปใช้

จากเนื้อหาในเรื่องความเท่ากันทุกประการที่ได้กล่าวมาข้างต้นสามารถสรุปได้ว่า ลักษณะของข้อสอบจะเป็นข้อสอบที่เป็นการวัดพฤติกรรมกรนำไปใช้โดยใช้สมบัติของรูปสามเหลี่ยมที่สัมพันธ์กันในลักษณะต่าง ๆ มาอธิบายและให้เหตุผลเบื้องต้นทางเรขาคณิตได้ โดยข้อสอบเป็นลักษณะข้อสอบแบบต่อเนื่องสองขั้นตอนที่จะต้องพิจารณาว่าสิ่งที่กำหนดให้เป็นจริงหรือเท็จเพราะเหตุใดพร้อมทั้งอธิบายเหตุผลประกอบ

3. สร้างตารางวิเคราะห์เนื้อหาเรื่องความเท่ากันทุกประการ โดยกำหนดให้มีจำนวนข้อสอบ 20 ข้อ ดังแสดงในตาราง 9

ตาราง 9 ตารางวิเคราะห์เนื้อหาวิชาคณิตศาสตร์ เรื่องความเท่ากันทุกประการ

เนื้อหา	จุดประสงค์การเรียนรู้	จำนวนข้อสอบ
รูปสามเหลี่ยมสองรูปที่ สัมพันธ์กันแบบ ด้าน-มุม-ด้าน	1. บอกได้ว่ารูปสามเหลี่ยมสองรูปเท่ากันทุกประการ ด้วยความสัมพันธ์แบบ ด้าน-มุม-ด้าน 2. นำสมบัติของความเท่ากันทุกประการของรูป สามเหลี่ยมสองรูปที่สัมพันธ์กันแบบ ด้าน-มุม-ด้าน ไปใช้อ้างอิงในการพิสูจน์	6
รูปสามเหลี่ยมสองรูปที่ สัมพันธ์กันแบบ มุม-ด้าน-มุม	1. บอกได้ว่ารูปสามเหลี่ยมสองรูปเท่ากันทุกประการ ด้วยความสัมพันธ์แบบ มุม-ด้าน-มุม 2. นำสมบัติของความเท่ากันทุกประการของรูป สามเหลี่ยมสองรูปที่สัมพันธ์กันแบบ มุม-ด้าน-มุม ไปใช้อ้างอิงในการพิสูจน์	6
รูปสามเหลี่ยมสองรูปที่ สัมพันธ์กันแบบ ด้าน-ด้าน-ด้าน	1. บอกได้ว่ารูปสามเหลี่ยมสองรูปเท่ากันทุกประการ ด้วยความสัมพันธ์แบบ ด้าน-ด้าน-ด้าน 2. นำสมบัติของความเท่ากันทุกประการของรูป สามเหลี่ยมสองรูปที่สัมพันธ์กันแบบ ด้าน-ด้าน-ด้าน ไปใช้อ้างอิงในการพิสูจน์	8

4. สร้างข้อสอบแบบต่อเนื่องสองขั้นตอนเรื่องความเท่ากันทุกประการจำนวน 40 ข้อ จัดพิมพ์แบบทดสอบที่สร้างแล้วนำไปให้ผู้เชี่ยวชาญซึ่งเกณฑ์ในการคัดเลือกผู้เชี่ยวชาญคือเป็นผู้มีความรู้ในเนื้อหาและการเขียนแบบทดสอบวิชาคณิตศาสตร์และมีความรู้ทางการวัดผลการศึกษาเป็นผู้ตรวจสอบความเที่ยงตรงเชิงเนื้อหาของข้อสอบเป็นรายข้อ

5. นำค่าการพิจารณาจากผู้เชี่ยวชาญมาหาดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์ (Index of Consistency: IOC) โดยใช้ดัชนีความสอดคล้องของโรวินลลีและแฮมเบิลตัน (บุญเชิด ภิญญอนันตพงษ์. 2527: 69; อ้างอิงจาก Rovinelli & Hambleton. 1972) ซึ่งมีสูตรดังนี้

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

เมื่อ IOC แทน ดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์

$\sum R$ แทน ผลรวมคะแนนความคิดเห็นของผู้เชี่ยวชาญเนื้อหาวิชาทั้งหมด

N แทน จำนวนผู้เชี่ยวชาญ

โดยผู้เชี่ยวชาญพิจารณาว่า ข้อสอบวัดได้ตรงกับจุดประสงค์ที่กำหนดตามเกณฑ์ดังนี้

- 1 แทน แน่ใจว่า ข้อสอบวัดได้ตรงกับจุดประสงค์ที่กำหนด
- 0 แทน ไม่แน่ใจว่า ข้อสอบวัดได้ตรงกับจุดประสงค์ที่กำหนด
- 1 แทน แน่ใจว่า ข้อสอบวัดได้ไม่ตรงกับจุดประสงค์ที่กำหนด

การแปลความหมายดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์ ที่คำนวณได้โดยถือว่า ถ้าค่า IOC มีค่าตั้งแต่ 0.5 ขึ้นไป แสดงว่าข้อสอบวัดได้ตรงกับจุดประสงค์ที่กำหนดผลปรากฏว่า ค่า IOC ที่ได้อยู่ในช่วง 0.6 ถึง 1

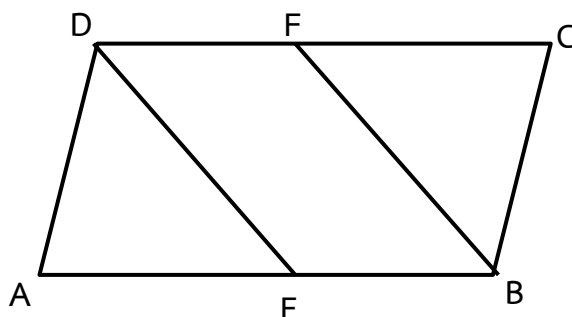
5. ปรับปรุงข้อสอบและคัดเลือกข้อสอบที่มีดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์ของความ (IOC) ตั้งแต่ 0.5 ขึ้นไปและสอดคล้องกับตารางวิเคราะห์เนื้อหาที่ผ่านการปรับปรุงและแก้ไขแล้วไปทดลองใช้ (Try out) กับนักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2559 โรงเรียนชิโนรสวิทยาลัย จำนวน 100 คน นำผลการสอบมาวิเคราะห์ค่าความยาก ค่าอำนาจจำแนก และค่าความเชื่อมั่นของแบบทดสอบ โดยคัดเลือกข้อสอบจำนวน 20 ข้อ ที่มีค่าความยาก (p) ระหว่าง 0.29–0.61 ค่าอำนาจจำแนก (r) ระหว่าง 0.16–0.45 และมีค่าความเชื่อมั่น 0.747

6. นำข้อสอบที่ผ่านการวิเคราะห์และปรับปรุงแล้วมาจัดพิมพ์เป็นแบบทดสอบฉบับจริงเพื่อนำไปใช้เป็นเครื่องมือในการเก็บรวบรวมข้อมูล

ลักษณะของเครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 เรื่องความเท่ากันทุกประการ มีลักษณะเป็นแบบทดสอบแบบต่อเนื่องสองขั้นตอน ซึ่งมีลักษณะดังนี้

ตัวอย่างข้อ (1)



1.1 จากรูปกำหนดให้ ABCD เป็นรูปสี่เหลี่ยมด้านขนาน E และ F เป็นจุดกึ่งกลางของ

AB และ

$\overline{DC}$ แล้ว $\triangle ADE \cong \triangle CBF$ ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 1.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 1.2(2)

1.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\overline{DE} = \overline{BF}$

ข. $\overline{AE} = \overline{CF}$

ค. $\overline{AD} = \overline{BC}$

ง. $\widehat{DAE} = \widehat{BCF}$

1.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

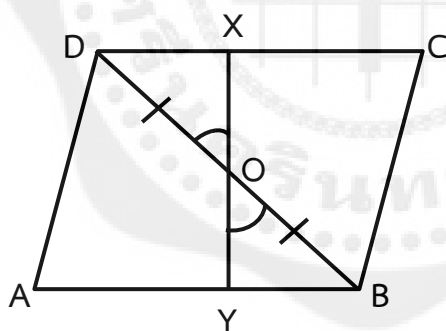
ก. $\overline{AD} = \overline{CB}$

ข. $\widehat{DAE} = \widehat{BCF}$

ค. $\widehat{ADE} = \widehat{CBF}$

ง. $\widehat{EDF} = \widehat{FBE}$

ตัวอย่างข้อ (2)



2.1 จากรูปกำหนดให้ O เป็นจุดกึ่งกลางของ $\overline{DB}$ แล้ว $\triangle DOX \cong \triangle BOY$

ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 2.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 2.2(2)

2.2(1) ถ้าข้อความเป็นจริง เพราะเหตุผลในข้อใด

ก. $\overline{OX} = \overline{OY}$

ข. $\overline{OX} = \overline{BY}$

ค. $\overline{OXD} = \overline{OYB}$

ง. $\overline{ODX} = \overline{OBY}$

2.2(2) ถ้าข้อความ เป็นเท็จ เพราะขาดสมบัติในข้อใด

ก. $\overline{OX} = \overline{OY}$

ข. $\overline{OB} = \overline{OD}$

ค. $\overline{DX} = \overline{BY}$

ง. $\overline{DOX} = \overline{BOY}$

วิธีการตรวจให้คะแนน

การตรวจให้คะแนนของแบบทดสอบเรื่องความเท่ากันทุกประการ โดยกำหนดเกณฑ์การให้คะแนนเป็น 3 ระดับ คือ 0 , 1 และ 2 ดังนี้

ตาราง 10 เกณฑ์การให้คะแนนแบบรวมของแบบทดสอบแบบต่อเนื่องสองขั้นตอน

ระดับคะแนน	เกณฑ์การให้คะแนน
2	เลือกคำตอบและบอกเหตุผลประกอบถูกต้อง
1	เลือกคำตอบถูกต้อง แต่บอกเหตุผลประกอบไม่ถูกต้อง
0	เลือกคำตอบไม่ถูกต้อง

การเก็บรวบรวมข้อมูล

ผู้วิจัยได้ดำเนินการดังนี้

1. ขอจดหมายจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ถึงผู้อำนวยการโรงเรียนสังกัดสำนักงานการศึกษาขั้นพื้นฐาน สพม. เขต 1 จังหวัดกรุงเทพฯ

2. นำหนังสือขอความร่วมมือในการเก็บรวบรวมข้อมูลจากบัณฑิตวิทยาลัยไปติดต่อประสานงานกับโรงเรียนที่เป็นกลุ่มตัวอย่างโดย

2.1 ติดต่อผู้อำนวยการหรือผู้ช่วยผู้อำนวยการฝ่ายวิชาการ เพื่อขอความอนุเคราะห์ในการเก็บรวบรวมข้อมูล

2.2 ติดต่อครูผู้สอนวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 2 เพื่อชี้แจงจุดประสงค์ของการศึกษา ขอบเขตเนื้อหาที่ต้องการวัด รูปแบบการวัดหรือลักษณะของแบบทดสอบที่ใช้วัด

รวมทั้งจุดเด่นและประโยชน์ของแบบทดสอบที่สามารถนำไปใช้ในการพัฒนาการเรียนการสอน เพื่อขอความร่วมมือในการดำเนินการสอบ

ผู้วิจัยแจกแบบทดสอบไปยังโรงเรียนที่เป็นกลุ่มตัวอย่างทั้ง 9 โรงเรียน รวม 1,350 ฉบับ และได้รับแบบทดสอบคืนกลับมา โดยนำมาคัดเลือกกระดาษคำตอบที่มีความสมบูรณ์และสามารถนำมาวิเคราะห์ได้ทั้งหมด 1,021 ฉบับ โดยแบ่งเป็นนักเรียนชาย 508 คน และนักเรียนหญิง 513 คน

3. นำผลที่ได้จากการเก็บรวบรวมข้อมูลมาหาค่าสถิติพื้นฐานและตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การวิเคราะห์ข้อมูล

ผู้วิจัยได้ดำเนินการวิเคราะห์ข้อมูลตามขั้นตอน ดังนี้

1. คำนวณค่าสถิติพื้นฐาน เพื่อบรรยายลักษณะการแจกแจงของคะแนนจากแบบทดสอบ วิชาคณิตศาสตร์ เรื่องความเท่ากันทุกประการ
2. ตรวจสอบการเป็นมิติเดียวของแบบทดสอบด้วยวิธีการวิเคราะห์องค์ประกอบ (Factor Analysis)
3. ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก วิธีแมนเทิล-แฮนส์เซลทั่วไปและวิธีโพลีโทมัสชิปเทสต์
4. คำนวณอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบระหว่างวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกกับวิธีแมนเทิล-แฮนส์เซลทั่วไป โดยวิธีโพลีโทมัสชิปเทสต์เป็นวิธีเกณฑ์

สถิติที่ใช้ในการวิจัย

ในการวิเคราะห์ข้อมูล ผู้วิจัยได้ใช้สถิติวิเคราะห์ดังต่อไปนี้

1. ค่าสถิติในการแสดงคุณภาพของแบบทดสอบ ได้แก่
 - 1.1 ค่าดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์ (Index of Consistency : IOC) (บุญเชิด ภิญโญนนตพงษ์. 2527: 69; อ้างอิงจาก Rovinelli and Hambleton. 1972) ซึ่งมีสูตรดังนี้

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

เมื่อ	IOC	แทน ดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์
	$\sum R$	แทน ผลรวมคะแนนความคิดเห็นของผู้เชี่ยวชาญเนื้อหาวิชาทั้งหมด
	N	แทน จำนวนผู้เชี่ยวชาญ

1.2 ค่าอำนาจจำแนกของแบบทดสอบ (r) คำนวณจากสูตร(ศิริชัย กาญจนวาสี. 2556: 243)

$$r = P_H - P_L$$

เมื่อ $\sum H$	แทน	คะแนนรวมรายข้อของทุกคนในกลุ่มสูง
$\sum L$	แทน	คะแนนรวมรายข้อของทุกคนในกลุ่มต่ำ
$\sum T_H$	แทน	คะแนนเต็มรวมรายข้อของทุกคนในกลุ่มสูง
$\sum T_L$	แทน	คะแนนเต็มรวมรายข้อของทุกคนในกลุ่มต่ำ
P_H	แทน	$\frac{\sum H}{\sum T_H}$
P_L	แทน	$\frac{\sum L}{\sum T_L}$

1.3 ค่าความเชื่อมั่น คำนวณได้จากสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha) (ศิริชัย กาญจนวาสี. 2556: 71) ดังนี้

$$\alpha = \left[\frac{k}{k-1} \right] \left[1 - \frac{\sum S_i^2}{S_x^2} \right]$$

เมื่อ α	แทน	สัมประสิทธิ์ความเชื่อมั่นของแบบทดสอบ
S_i^2	แทน	ความแปรปรวนของคะแนนส่วนที่ i (หรือข้อที่ i)
S_x^2	แทน	ความแปรปรวนของคะแนนรวม X
k	แทน	จำนวนส่วน (components) ที่นำมารวมเป็น X (หรือจำนวนแบบทดสอบ)

1.4 ค่าความยากง่าย (Difficulty) โดยใช้การวิเคราะห์ข้อสอบรายข้อ (Item Analysis) (ศิริชัย กาญจนวาสี. 2556: 243) ดังนี้

$$p = \frac{P_H + P_L}{2}$$

เมื่อ $\sum H$	แทน	คะแนนรวมรายข้อของทุกคนในกลุ่มสูง
$\sum L$	แทน	คะแนนรวมรายข้อของทุกคนในกลุ่มต่ำ
$\sum T_H$	แทน	คะแนนเต็มรวมรายข้อของทุกคนในกลุ่มสูง

ΣT_L	แทน	คะแนนเต็มรวมรายข้อของทุกคนในกลุ่มต่ำ
P_H	แทน	$\frac{\Sigma H}{\Sigma T_H}$
P_L	แทน	$\frac{\Sigma L}{\Sigma T_L}$

2. ค่าสถิติที่ใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

2.1 สถิติที่ใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี Polytomous

SIBTEST (บุษวรรักษ์ แสнопดิม. 2550: 56)

$$\hat{\beta}^* = \sum_{k=0}^{n_H} P_k d_k^*$$

เมื่อ $\hat{\beta}^*$ แทน สถิติที่ใช้ประมาณค่าปริมาณการทำหน้าที่ต่างกันของข้อคำถาม

$n_H = \sum_{j=1}^n m_j$ แทน คะแนนของการจับคู่ที่เป็นไปได้สูงสุด

P_k แทน สัดส่วนของผู้สอบทั้งหมดที่ได้คะแนน $x = k$ จากการสอบ
เพื่อจับคู่ นั่นคือ $P_k = \frac{N_k}{N}$

d_k^* แทน $d_k^* = \bar{Y}^* Rk - \bar{Y}^* Fk$; $k = 0, 1, \dots, n_H$ ได้มาโดยการแปลงค่าเชิงเส้น

2.2 สถิติที่ใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี GMH (เรืองเดช ศิริกิจ

. 2554: 126)

$$\alpha_{MH_i}^{\wedge} = \frac{\sum_{s=1}^s R_{1s} F_{0s} / n_s}{\sum_{s=1}^s R_{0s} F_{1s} / n_s}$$

2.3 สถิติที่ใช้ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธี LDFA ซึ่งมีสูตรในการคำนวณ ดังนี้ (สุชาติ สิริมินนนท์. 2554: 86)

$$\text{logit} = \ln \left[\frac{P(G=1)}{P(G=0)} \right] = \alpha_0 + \alpha_1 X + \alpha_2 U + \alpha_3 (X * U)$$

เมื่อ $\alpha_i (i = 0,1,2,3)$	แทน สัมประสิทธิ์ฟังก์ชันการจำแนก
G	แทน กลุ่มผู้สอบ
X	แทน ระดับความสามารถของผู้สอบ
U	แทน ผลการตอบข้อสอบ
P	แทน ความน่าจะเป็นของการเป็นสมาชิกกลุ่ม

การทดสอบข้อสอบทำหน้าที่เป็นรูปแบบเดียวกัน (Uniform DIF) และที่ไม่เป็นรูปแบบเดียวกัน (Nonuniform DIF) สามารถทดสอบได้พร้อมกัน โดยกำหนดสมมติฐานศูนย์ (H_0) และสมมติฐานทางเลือก (H_1) ดังนี้

$$H_0 : \alpha_2 = 0 \text{ และ } \alpha_3 = 0$$

$$H_1 : \alpha_2 \neq 0 \text{ และ } \alpha_3 \neq 0$$

การทดสอบนัยสำคัญทางสถิติของการทำหน้าที่ต่างกันของข้อสอบจะดำเนินการตามกระบวนการทดสอบความเหมาะสมของข้อมูลกับโมเดล 3 ขั้นตอน โดยใช้สถิติ G^2 โดยที่

$$G^2_{\text{diff}} = -2(\ln L_{i+1} - \ln L_{(i)})$$

เมื่อ $L_{(i+1)}$ และ $L_{(i)}$ แทนความน่าจะเป็นของโมเดลลำดับขั้นที่อยู่ติดกัน

สถิติ G^2 ดังกล่าวมีการแจกแจงแบบเชิงเส้นกำกับ เป็นไค-สแควร์ที่ระดับชั้นความเป็นอิสระเท่ากับ 1 (df = 1) ในการศึกษาครั้งนี้ผู้วิจัยใช้เกณฑ์ตัดสินข้อสอบที่ทำหน้าที่ต่างกันที่ระดับนัยสำคัญ .05

3. การคำนวณอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1

การคำนวณอัตราความถูกต้อง (Power rate) และอัตราความคลาดเคลื่อนประเภทที่ 1 (Type I error) ของวิธีแมนเทิล-แฮนส์เซลทั่วไป และวิธีวิเคราะห์ฟังก์ชันการจำแนกโลจิสติก (ทองอยู่ สารระ. 2543: 58 อ้างอิงจาก Narayanan and Swaminathan. 1994) ใช้สูตรการคำนวณดังนี้

$$P = \frac{n_1}{N_1} \times 100$$

$$E = \frac{n_2}{N_2} \times 100$$

เมื่อ	P	แทน ร้อยละของอัตราความถูกต้อง
	E	แทน ร้อยละของอัตราความคลาดเคลื่อนประเภทที่ 1
	n_1	แทน จำนวนข้อสอบที่ตรวจสอบได้ถูกต้องว่าทำหน้าที่ต่างกัน
	n_2	แทน จำนวนข้อสอบที่ตรวจสอบผิดพลาดว่าทำหน้าที่ต่างกัน
	N_1	แทน จำนวนข้อสอบทำหน้าที่ต่างกันแบบทดสอบ
	N_2	แทน จำนวนข้อสอบทำหน้าที่ไม่ต่างกันแบบทดสอบ



บทที่ 4

ผลการวิเคราะห์ข้อมูล

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อการศึกษาและเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกกับวิธีแมนเทิล-แฮนส์เซลทั่วไป ในข้อสอบแบบต่อเนื่องสองขั้นตอนที่มีการตรวจให้คะแนนหลายค่า ในกลุ่มผู้สอบที่มีเพศต่างกัน ซึ่งผลการวิเคราะห์ข้อมูลครั้งนี้ผู้วิจัยได้เสนอเรียงตามลำดับดังนี้

ตอนที่ 1 การวิเคราะห์ข้อมูลเบื้องต้น

ตอนที่ 2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

2.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีวิเคราะห์ฟังก์ชัน การจำแนกแบบโลจิสติก

2.2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทิล-แฮนส์เซลทั่วไป

2.3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโพลีโทมัสชิปเทสต์

ตอนที่ 3 การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี ได้แก่ วิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกกับวิธีแมนเทิล-แฮนส์เซลทั่วไป โดยมีวิธีโพลีโทมัส ชิปเทสต์เป็นเกณฑ์ โดยเรียงลำดับดังนี้

3.1 เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก วิธีแมนเทิล-แฮนส์เซลทั่วไปและวิธีโพลีโทมัสชิปเทสต์

3.2 เปรียบเทียบความถูกต้องและความคลาดเคลื่อนของการตรวจสอบระหว่าง วิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกกับวิธีแมนเทิล-แฮนส์เซลทั่วไป

สัญลักษณ์ที่ใช้ในการวิเคราะห์ข้อมูล

การนำเสนอผลการวิเคราะห์ข้อมูลของการวิจัย เรื่องการเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ในข้อสอบแบบต่อเนื่องสองขั้นตอน ระหว่างวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกกับวิธีแมนเทิล-แฮนส์เซลทั่วไป เพื่อให้นำเสนอผลการวิเคราะห์มีความสะดวกและเกิดความเข้าใจตรงกัน ผู้วิจัยจึงได้กำหนดสัญลักษณ์และความหมายที่ใช้ในการวิเคราะห์ข้อมูล ดังนี้

DIF หมายถึง การทำหน้าที่ต่างกันของข้อสอบ

LDFA หมายถึง การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก

GMH หมายถึง การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทิล-แฮนส์เซลทั่วไป

Poly-SIB	หมายถึง การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโพลีโทมัสซิปเทสท์
N	หมายถึง จำนวนคน
$\bar{X}$	หมายถึง ค่าเฉลี่ยเลขคณิต
SD	หมายถึง ค่าส่วนเบี่ยงเบนมาตรฐาน
Max	หมายถึง ค่าคะแนนสูงสุด
Min	หมายถึง ค่าคะแนนต่ำสุด
Sk	หมายถึง ค่าความเบ้
Ku	หมายถึง ค่าความโด่ง
P	หมายถึง อัตราความถูกต้อง
E	หมายถึง อัตราความคลาดเคลื่อน
*	หมายถึง $p < .05$

ตอนที่ 1 การวิเคราะห์ข้อมูลเบื้องต้น

1.1 ค่าสถิติพื้นฐานของข้อมูลผลการสอบ

การวิเคราะห์ข้อมูลส่วนนี้ผู้วิจัยได้นำคะแนนที่ได้จากการตอบแบบทดสอบวิชาคณิตศาสตร์ เรื่องความเท่ากันทุกประการ จำนวน 20 ข้อ คำนวณค่าสถิติพื้นฐาน จำแนกตามเพศปรากฏผลดังตาราง 11

ตาราง 11 ค่าสถิติพื้นฐานของแบบทดสอบวิชาคณิตศาสตร์ เรื่องความเท่ากันทุกประการ 20 ข้อ

เพศ	N	Max	Min	Mean	SD	Sk	Ku
ชาย	508	31	6	15.50	3.88	1.20	2.43
หญิง	513	36	6	15.82	3.89	1.10	2.94
รวม	1,021	36	6	15.66	3.84	1.17	2.65

จากตาราง 11 พบว่าผู้ทำแบบทดสอบทั้งหมด 1,021 คน แบบทดสอบมีความยาว 20 ข้อ มีคะแนนเต็ม 40 คะแนน ค่าคะแนนสูงสุดเท่ากับ 36 คะแนน ค่าคะแนนต่ำสุดเท่ากับ 6 คะแนน ค่าคะแนนเฉลี่ยเท่ากับ 15.66 คะแนน ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3.84 มีลักษณะการแจกแจงเป็นเบ้ทางบวก (skewness = 1.17) และค่อนข้างโด่ง (kurtosis = 2.65) เมื่อพิจารณาตามเพศของนักเรียนในการทำแบบทดสอบ พบว่านักเรียนชายจำนวน 508 คน ค่าคะแนนสูงสุดเท่ากับ 31 คะแนน ค่าคะแนนต่ำสุดเท่ากับ 6 คะแนน มีค่าคะแนนเฉลี่ยเท่ากับ 15.50 คะแนน ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3.88 ลักษณะการแจกแจงเป็นเบ้ทางบวก (skewness = 1.20) และค่อนข้างโด่ง

(kurtosis = 2.43) พบว่านักเรียนหญิงจำนวน 513 คน ค่าคะแนนสูงสุดเท่ากับ 36 คะแนน ค่าคะแนนต่ำสุดเท่ากับ 6 คะแนน มีค่าคะแนนเฉลี่ยเท่ากับ 15.82 คะแนน ค่าส่วนเบี่ยงเบนมาตรฐานเท่ากับ 3.89 ลักษณะการแจกแจงเป็นเบ้ทางบวก (skewness = 1.10) และค่อนข้างโด่ง (kurtosis = 2.94)

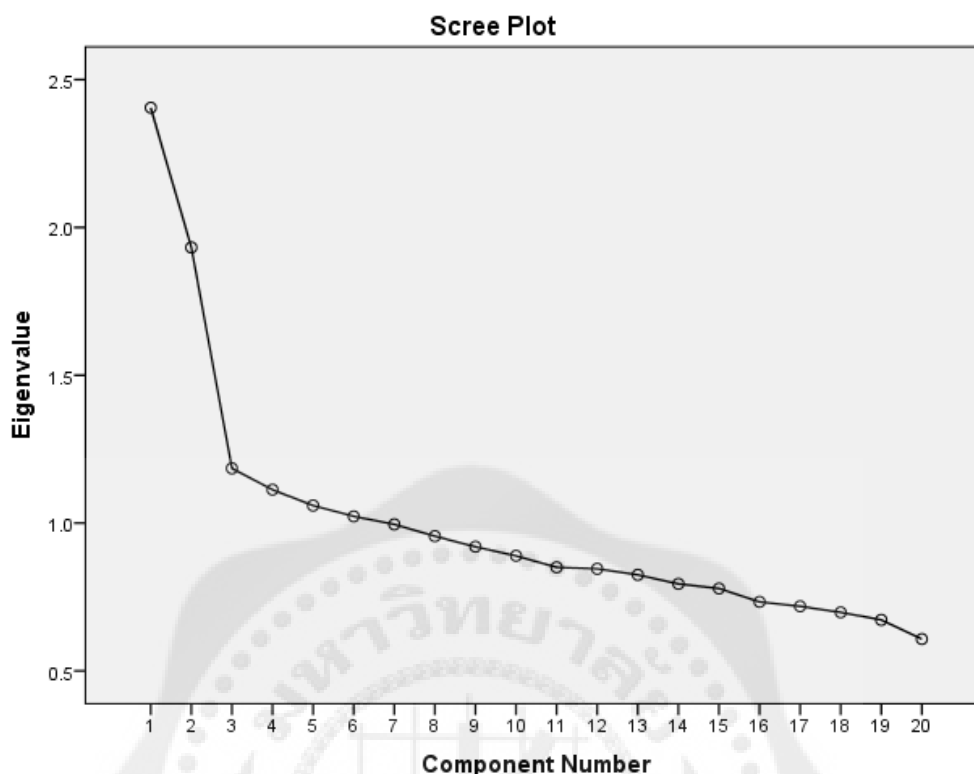
1.2 ผลการวิเคราะห์องค์ประกอบเพื่อตรวจสอบความเป็นเอกมิติของข้อมูล

การตรวจสอบความเป็นมิติเดียวของแบบทดสอบ

การตรวจสอบความเป็นมิติเดียวของแบบทดสอบเป็นข้อตกลงพื้นฐานที่สำคัญของทฤษฎีการตอบสนองข้อสอบ (IRT) เป็นการตรวจสอบว่าแบบทดสอบวัดคุณลักษณะแฝงเพียงคุณลักษณะเดียว ผู้วิจัยได้นำคะแนนที่ได้จากการตอบแบบทดสอบวิชาคณิตศาสตร์ เรื่องความเท่ากันทุกประการ จำนวน 20 ข้อ มาตรวจสอบความเป็นมิติเดียว โดยใช้การวิเคราะห์องค์ประกอบ (Factor Analysis) ด้วยวิธีการวิเคราะห์องค์ประกอบหลัก (Principle Component Analysis) หาค่าไอเกน (Eigen Value) และร้อยละของความแปรปรวน ดังตาราง 12

ตาราง 12 ค่าไอเกน และร้อยละของความแปรปรวนขององค์ประกอบ ของแบบทดสอบจำนวน 20 ข้อ

องค์ประกอบ	ค่าไอเกน	ร้อยละของความแปรปรวน
1	2.41	12.03
2	1.93	9.66
3	1.19	5.93
4	1.11	5.56
5	1.06	5.30
6	1.02	5.11



ภาพประกอบ 6 ผลการตรวจสอบความเป็นมิติเดียวกันของข้อสอบ

ผลการวิเคราะห์จากตาราง 12 และภาพประกอบ 6 พบว่าไอเกนขององค์ประกอบที่ 1 มีค่าสูงกว่าค่าไอเกนขององค์ประกอบที่ 2 ประมาณ 1.25 เท่า และตั้งแต่องค์ประกอบที่ 2 เป็นต้นไป จะมีความใกล้เคียงกัน นับว่าองค์ประกอบที่ 1 มีค่าเด่นกว่าองค์ประกอบอื่น แม้ว่าค่าร้อยละของความแปรปรวนจะไม่สูงมากนัก (12.03) หากพิจารณาตามเกณฑ์ของแรคเคส (อุทัยวรรณ สายพัฒนา. 2547: 64-65; อ้างอิงจาก Ruj. 1993: 304) ที่เสนอค่าร้อยละของความแปรปรวนควรมีค่าไม่น้อยกว่าร้อยละ 20 จึงจะถือว่าแบบทดสอบมีความเป็นเอกมิติ ซึ่งถือว่าเป็นไปได้ยากในทางปฏิบัติเนื่องจากแบบทดสอบฉบับนี้เป็นเนื้อหาเรื่องความเท่ากันทุกประการ ซึ่งประกอบด้วย ความสัมพันธ์ทางเรขาคณิต 3 ความสัมพันธ์ภายในฉบับเดียว

ดังนั้นจึงเป็นข้อจำกัดอย่างหนึ่งของการวิจัย ที่แบบทดสอบแบบต่อเนื่องสองขั้นตอนเรื่องความเท่ากันทุกประการ ความยาว 20 ข้อ ยังไม่มีความเป็นเอกมิติอย่างสมบูรณ์ แต่ผู้วิจัยได้ทำการวิเคราะห์โดยนำข้อมูลไปวิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบต่อไป

ตอนที่ 2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

2.1 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีวิเคราะห์ฟังก์ชัน การจำแนกแบบโลจิสติก

การวิเคราะห์ตอนที่ 2 เป็นการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก ซึ่งผู้วิจัยได้นำคะแนนจากการทำแบบทดสอบมาวิเคราะห์การทำหน้าที่ต่างกันของข้อคำถามตามลักษณะทางเพศของกลุ่มตัวอย่าง โดยมีกลุ่มอ้างอิง(Reference Group) คือกลุ่มนักเรียนเพศชาย และกลุ่มเปรียบเทียบ(Focal Group) คือกลุ่มนักเรียนเพศหญิง แสดงผลดังตาราง 13

ตาราง 13 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก

ข้อ	Chi - square	ข้อ	Chi - square
1	4.02	11	0.87
2	1.47	12	0.44
3	7.29*	13	1.30
4	4.64	14	3.43
5	0.61	15	1.94
6	0.52	16	6.65*
7	5.00	17	2.11
8	5.21*	18	4.03
9	3.88	19	1.26
10	3.78	20	13.29*

* $p < .05$

จากตาราง 13 พบว่าการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธี LDFA ตามแนวคิดทฤษฎีการตอบข้อสอบด้วยการทดสอบค่าไค-สแควร์ เมื่อกลุ่มตัวอย่าง 1,021 คน และแบบทดสอบยาว 20 ข้อ พบว่าข้อสอบที่ทำหน้าที่ต่างกัน (DIF) มีจำนวน 4 ข้อ ได้แก่ ข้อ 3 , 8 , 16 และ 20 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

2.2 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลทั่วไป

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลทั่วไป ผู้วิจัยได้นำคะแนนจากการทำแบบทดสอบมาวิเคราะห์การทำหน้าที่ต่างกันของข้อคำถามตามลักษณะทางเพศของกลุ่มตัวอย่าง โดยมีกลุ่มอ้างอิง(Reference Group) คือกลุ่มนักเรียนเพศชาย และกลุ่มเปรียบเทียบ(Focal Group) คือกลุ่มนักเรียนเพศหญิง แสดงผลดังตาราง 14

ตาราง 14 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลทั่วไป

ข้อ	Mantel	LOR	SE(LOR)	Z(LOR)
1	3.11	0.24	0.14	1.75
2	1.17	0.14	0.13	1.09
3	5.83	0.30	0.12	2.38*
4	2.45	-0.20	0.13	-1.58
5	0.99	0.13	0.13	0.98
6	0.05	-0.03	0.12	-0.21
7	2.88	-0.22	0.13	-1.70
8	4.34	-0.28	0.13	-2.09*
9	3.13	-0.22	0.13	-1.75
10	0.12	0.04	0.13	0.35
11	0.11	0.04	0.13	0.33
12	0.29	0.07	0.12	0.55
13	2.07	0.19	0.13	1.45
14	2.03	-0.18	0.13	-1.41
15	2.22	0.18	0.12	1.48
16	6.55	0.32	0.12	2.59*
17	1.37	-0.15	0.13	-1.16
18	2.94	-0.21	0.12	-1.71
19	0.21	-0.06	0.13	-0.46
20	1.59	-0.15	0.12	-1.24

* $Z(\text{LOR}) > 2$ หรือ $Z(\text{LOR}) < -2$

จากตาราง 14 เมื่อพิจารณาจากค่า $Z(\text{LOR}) > 2$ หรือ $Z(\text{LOR}) < -2$ พบว่ามีจำนวน 3 ข้อ คือข้อ 3 , 8 และ 16 มีค่า $Z(\text{LOR})$ เท่ากับ 2.38 , -2.09 และ 2.59 ตามลำดับ แสดงว่าข้อคำถามทั้ง 3 ข้อนั้นมีการทำหน้าที่ต่างกันระหว่างกลุ่มอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเมื่อพิจารณา ค่า LOR ที่มีค่าเป็นบวก และค่า LOR ที่มีค่าเป็นลบ พบว่าข้อ 3 และข้อ 16 มีค่า LOR เป็นบวกคือ 0.30 และ 0.32 แสดงว่าข้อคำถามนั้นเข้าข้างกลุ่มอ้างอิง (Reference group) คือกลุ่มนักเรียนเพศชาย และข้อ 8 มีค่า LOR เป็นลบคือ -0.28 แสดงว่าข้อคำถามนั้นเข้าข้างกลุ่มเปรียบเทียบ (Focal group) คือกลุ่มนักเรียนเพศหญิง

2.3 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโพลีโทมัสชิปเทสท์

การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโพลีโทมัสชิปเทสท์ ผู้วิจัยได้นำคะแนนจากการทำแบบทดสอบมาวิเคราะห์การทำหน้าที่ต่างกันของข้อคำถามตามลักษณะทางเพศของกลุ่มตัวอย่าง โดยมีกลุ่มอ้างอิง(Reference Group) คือกลุ่มนักเรียนเพศชาย และกลุ่มเปรียบเทียบ(Focal Group) คือกลุ่มนักเรียนเพศหญิง แสดงผลดังตาราง 15

ตาราง 15 ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีโพลีโทมัสชิปเทสท์

ข้อ	β	Standard error	p-value
1	0.04	0.04	0.25
2	0.07	0.04	0.12
3	0.10	0.04	0.02*
4	-0.04	0.04	0.38
5	0.08	0.04	0.03*
6	-0.01	0.05	0.78
7	0.08	0.04	0.05*
8	0.01	0.03	0.75
9	-0.10	0.04	0.02*
10	0.05	0.04	0.28
11	0.04	0.04	0.37
12	-0.03	0.04	0.42
13	0.13	0.04	0.00*
14	0.07	0.04	0.09
15	0.10	0.05	0.03*

ตาราง 15 (ต่อ)

ข้อ	β	Standard error	p-value
16	0.13	0.05	0.01*
17	0.01	0.04	0.74
18	-0.11	0.05	0.01*
19	-0.01	0.04	0.85
20	-0.11	0.05	0.01*

* p<.05

จากตาราง 15 เมื่อพิจารณาดัชนีบ่งชี้การทำหน้าที่ต่างกันของข้อสอบที่มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 พบว่ามีข้อสอบทั้งหมด 9 ข้อที่ทำหน้าที่ต่างกัน คือ ข้อ 3 , 5 , 7 , 9 , 13 , 15 , 16 , 18 และ 20 โดยข้อสอบข้อที่ 3 , 5 , 7 , 13 , 15 และ 16 เป็นข้อสอบที่เข้าข้างกลุ่มนักเรียนเพศชาย และข้อสอบข้อที่ 9 , 18 และ 20 เป็นข้อสอบที่เข้าข้างกลุ่มนักเรียนเพศหญิง

ตอนที่ 3 การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ 2 วิธี
ได้แก่วิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกกับวิธีแมนเทิล-แฮนส์เซลทั่วไป
โดยมีวิธีโพลีโตมัสชิปเทสต์เป็นเกณฑ์

3.1 เปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบระหว่างวิธี

วิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก วิธีแมนเทิล-แฮนส์เซลทั่วไปและวิธีโพลีโตมัส ชิปเทสต์

จากการวิเคราะห์ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบโดยใช้วิธีการวิเคราะห์ 3 วิธี คือวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก วิธีแมนเทิล-แฮนส์เซลทั่วไป และวิธีโพลีโตมัส ชิปเทสต์ พบว่าในแต่ละวิธีจะพบจำนวนข้อสอบที่ทำหน้าที่ต่างกันไม่เท่ากัน วิธีวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบบางวิธีสามารถพบข้อสอบที่ทำหน้าที่ต่างกันซ้ำกัน แต่ในบางวิธีก็สามารถตรวจพบข้อสอบที่ทำหน้าที่ต่างกันแตกต่างกัน ดังนั้นผู้วิจัยจึงได้สรุปจำนวนข้อสอบที่ทำหน้าที่ต่างกันจากการวิเคราะห์ทั้ง 3 วิธี ดังตาราง 16

ตาราง 16 การเปรียบเทียบผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก วิธีแมนเทิล-แฮนส์เซลทั่วไปและวิธีโพลีโตมัสชิปเทสต์

ข้อ	LDFA	GMH	Poly-SIB
1	-	-	-
2	-	-	-
3	✓	✓	✓
4	-	-	-
5	-	-	✓
6	-	-	-
7	-	-	✓
8	✓	✓	-
9	-	-	✓
10	-	-	-
11	-	-	-
12	-	-	-
13	-	-	✓
14	-	-	-
15	-	-	✓
16	✓	✓	✓
17	-	-	-
18	-	-	✓
19	-	-	-
20	✓	-	✓

จากตาราง 16 เมื่อพิจารณาผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ทั้ง 3 วิธี พบว่าวิธีโพลีโตมัสชิปเทสต์ซึ่งเป็นวิธีเกณฑ์จะพบจำนวนข้อสอบที่ทำหน้าที่ต่างกันมากที่สุดคือจำนวน 9 ข้อ รองลงมาคือวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก พบจำนวนข้อสอบที่ทำหน้าที่ต่างกันจำนวน 4 ข้อและวิธีแมนเทิล-แฮนส์เซลทั่วไปพบจำนวนข้อสอบที่ทำหน้าที่ต่างก็น้อยที่สุดคือจำนวน 3 ข้อ โดยมีข้อสอบ 2 ข้อคือข้อ 3 และข้อ 16 ที่ผลการวิเคราะห์ทั้ง 3 วิธีตรงพบตรงกัน และทุกข้อที่วิธีแมนเทิล-แฮนส์เซลทั่วไปตรงพบจะตรงกับวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกด้วยเช่นกัน และมีข้อสอบข้อที่ 8 ที่วิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกและวิธีแมนเทิล-แฮนส์เซลทั่วไปตรงพบแต่ไม่ตรงกับวิธีโพลีโตมัสชิปเทสต์ซึ่งเป็นวิธีเกณฑ์

3.2 เปรียบเทียบความถูกต้องและความคลาดเคลื่อนของการตรวจสอบระหว่าง วิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกกับวิธีแมนเทิล-แฮนส์เซลทั่วไป

เมื่อพิจารณาผลการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบทั้ง 3 วิธี โดยมีวิธี โพลีโตมัสชิปเทสต์เป็นวิธีเกณฑ์ในการตรวจสอบจะพบว่าข้อสอบบางข้อที่วิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกกับวิธีแมนเทิล-แฮนส์เซลทั่วไปตรวจพบตรงกับวิธีโพลีโตมัสชิปเทสต์ ดังนั้นผู้วิจัยจึงได้เสนออัตราความถูกต้องของการตรวจสอบและอัตราความคลาดเคลื่อนของการตรวจสอบ ดังตาราง 17

ตาราง 17 อัตราความถูกต้องและอัตราความคลาดเคลื่อนของการตรวจสอบโดยวิธีโพลีโตมัสชิปเทสต์เป็นเกณฑ์

วิธีการวิเคราะห์	อัตราความถูกต้อง (P)	อัตราความคลาดเคลื่อน (E)
LDFA	33.33	9.09
GMH	22.22	9.09

จากตาราง 17 เมื่อพิจารณาอัตราความถูกต้องพบว่าวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกเท่ากับ 33.33 % เนื่องจากวิธีการวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกสามารถตรวจพบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีโพลีโตมัสชิปเทสต์ 3 ข้อ คือข้อ 3 , 16 และข้อ 20 ทำให้มีอัตราความถูกต้องของการตรวจสอบเท่ากับ 33.33% ซึ่งสูงกว่า วิธีแมนเทิล-แฮนส์เซลทั่วไปซึ่งสามารถตรวจพบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับ วิธีโพลีโตมัสชิปเทสต์เพียง 2 ข้อคือข้อ 3 และข้อ 16 ทำให้มีอัตราความถูกต้องของการตรวจสอบเท่ากับ 22.22 % และเมื่อตรวจสอบความคลาดเคลื่อนพบว่าทั้งวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกและวิธีแมนเทิล-แฮนส์เซลทั่วไปมีการตรวจพบข้อสอบเพียงวิธีละ 1 ข้อที่ไม่ตรงกับ วิธีโพลีโตมัสชิปเทสต์ คือข้อ 8 ทำให้มีอัตราความคลาดเคลื่อนเท่ากับ 9.09 % เท่ากัน

บทที่ 5

สรุปผล อภิปรายผล และข้อเสนอแนะ

ความมุ่งหมายของการวิจัย

การวิจัยครั้งนี้ผู้วิจัยมีจุดมุ่งหมายเพื่อการศึกษาและเปรียบเทียบประสิทธิภาพของวิธีการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกกับวิธีแมนเทิล-แฮนส์เซลทั่วไปในข้อสอบแบบต่อเนื่องสองขั้นตอนที่มีการตรวจให้คะแนนหลายค่าในกลุ่มผู้สอบที่มีเพศต่างกัน

กลุ่มตัวอย่างที่ใช้ในการวิจัย

กลุ่มตัวอย่างที่ใช้ในการวิจัยครั้งนี้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ที่กำลังศึกษาอยู่ในภาคเรียนที่ 2 ปีการศึกษา 2559 ของโรงเรียนขนาดใหญ่ สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเขต 1 ซึ่งได้มาจากการสุ่มแบบแบ่งกลุ่ม (Cluster Random Sampling) จำนวน 9 โรงเรียน นักเรียนทั้งหมด 1,021 คน เป็นนักเรียนชาย 508 คน และเป็นนักเรียนหญิง 513 คน

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยเป็นแบบทดสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 เรื่องความเท่ากันทุกประการ ที่ผู้วิจัยสร้างขึ้นมีลักษณะเป็นแบบทดสอบแบบต่อเนื่องสองขั้นตอน โดยเป็นข้อสอบแบบเลือกตอบ 4 ตัวเลือก ในขั้นตอนแรกเป็นการเลือกตอบแบบถูกผิด และขั้นตอนที่สองเป็นการให้เหตุผลประกอบการเลือกตอบในขั้นตอนแรก โดยมีการให้คะแนนเป็น 0 , 1 และ 2 ตามลักษณะของการตอบ จำนวน 20 ข้อ

การเก็บรวบรวมข้อมูล

ผู้วิจัยได้ทำการเก็บรวบรวมข้อมูลโดยดำเนินการดังนี้

1. ขอจดหมายจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ถึงผู้อำนวยการโรงเรียนสังกัดสำนักงานการศึกษาขั้นพื้นฐาน สพม. เขต 1 จังหวัดกรุงเทพฯ
2. นำหนังสือขอความร่วมมือในการเก็บรวบรวมข้อมูลจากบัณฑิตวิทยาลัยไปติดต่อประสานงานกับโรงเรียนที่เป็นกลุ่มตัวอย่างโดย
 - 2.1 ติดต่อผู้อำนวยการหรือผู้ช่วยผู้อำนวยการฝ่ายวิชาการ เพื่อขอความอนุเคราะห์ในการเก็บรวบรวมข้อมูล
 - 2.2 ติดต่อครูผู้สอนวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 2 เพื่อชี้แจงจุดประสงค์ของการศึกษา ขอบเขตเนื้อหาที่ต้องการวัด รูปแบบการวัดหรือลักษณะของแบบทดสอบที่ใช้วัด

รวมทั้งจุดเด่นและประโยชน์ของแบบทดสอบที่สามารถนำไปใช้ในการพัฒนาการเรียนการสอน เพื่อขอความร่วมมือในการดำเนินการสอบ

ผู้วิจัยแจกแบบทดสอบไปยังโรงเรียนที่เป็นกลุ่มตัวอย่างทั้ง 9 โรงเรียน รวม 1,350 ฉบับ และได้รับแบบทดสอบคืนกลับมา โดยนำมาคัดเลือกกระดาษคำตอบที่มีความสมบูรณ์และสามารถนำมาวิเคราะห์ได้ทั้งหมด 1,021 ฉบับ โดยแบ่งเป็นนักเรียนชาย 508 คน และนักเรียนหญิง 513 คน

3. นำผลที่ได้จากการเก็บรวบรวมข้อมูลมาหาค่าสถิติพื้นฐานและตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ

การวิเคราะห์ข้อมูล

ผู้วิจัยได้ดำเนินการวิเคราะห์ข้อมูลตามขั้นตอน ดังนี้

1. คำนวณค่าสถิติพื้นฐาน เพื่อบรรยายลักษณะการแจกแจงของคะแนนจากแบบทดสอบ วิชาคณิตศาสตร์ เรื่องความเท่ากันทุกประการ
2. ตรวจสอบการเป็นมิติเดียวของแบบทดสอบด้วยวิธีการวิเคราะห์องค์ประกอบ (Factor Analysis)
3. ตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก วิธีแมนเทิล-แฮนส์เซลทั่วไปและวิธีโพลีโทมัสชิปเทสท์
4. คำนวณอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 ของการตรวจสอบระหว่างวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกกับวิธีแมนเทิล-แฮนส์เซลทั่วไป โดยวิธีโพลีโทมัสชิปเทสท์เป็นวิธีเกณฑ์

สรุปผลการวิเคราะห์ข้อมูล

ผลการวิจัยแบ่งเป็น 2 ส่วนคือ ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก และวิธีแมนเทิล-แฮนส์เซลทั่วไป และการเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก และวิธีแมนเทิล-แฮนส์เซลทั่วไปโดยมีวิธีโพลีโทมัสชิปเทสท์เป็นเกณฑ์

1. ผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ เมื่อใช้แบบทดสอบแบบต่อเนื่องสองขั้นตอน เรื่องความเท่ากันทุกประการจำนวน 20 ข้อ พบว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก และวิธีแมนเทิล-แฮนส์เซลทั่วไป สรุปได้ดังนี้

วิธีการวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก พบข้อสอบที่ทำหน้าที่ต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 4 ข้อได้แก่ข้อ 3 , 8 , 16 และ 20 คิดเป็นร้อยละ 20 ของจำนวนข้อสอบทั้งหมด

วิธีแมนเทิล-แฮนส์เซลทั่วไป พบข้อสอบที่ทำหน้าที่ต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 3 ข้อ ได้แก่ข้อ 3 , 8 และ 16 คิดเป็นร้อยละ 15 ของจำนวนข้อสอบทั้งหมด

2. การเปรียบเทียบผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีการวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติก และวิธีแมนเทิล-แฮนส์เซลทั่วไป โดยมีวิธีโพลีโตมัส ชิปเทสท์เป็นเกณฑ์ พบว่าเมื่อคำนวณอัตราความถูกต้องของการตรวจสอบแล้ว วิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกสามารถตรวจสอบได้ตรงกับวิธีโพลีโตมัสชิปเทสท์ 3 ข้อ คือข้อ 3 , 16 และข้อ 20 ทำให้มีอัตราความถูกต้องของการตรวจสอบเท่ากับ 33.33% ซึ่งสูงกว่า วิธีแมนเทิล-แฮนส์เซลทั่วไปที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบได้ตรงกับวิธีโพลีโตมัสชิปเทสท์เพียง 2 ข้อคือข้อ 3 และข้อ 16 ทำให้มีอัตราความถูกต้องของการตรวจสอบเท่ากับ 22.22 % และเมื่อตรวจสอบความคลาดเคลื่อนประเภทที่ 1 พบว่าทั้งวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกและวิธีแมนเทิล-แฮนส์เซลทั่วไปมีการตรวจพบข้อสอบเพียงร้อยละ 1 ข้อที่ไม่ตรงกับวิธีโพลีโตมัส ชิปเทสท์ ทำให้มีอัตราความคลาดเคลื่อนประเภทที่ 1 เท่ากับ 9.09 % เท่ากัน ดังนั้นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบทั้ง 2 วิธี ให้อัตราความถูกต้องที่ใกล้เคียงกันและมีอัตราความคลาดเคลื่อนประเภทที่ 1 เท่ากัน

อภิปรายผล

การศึกษาประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อสอบแบบต่อเนื่องสองขั้นตอนระหว่างวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกกับวิธีแมนเทิล-แฮนส์เซลทั่วไป โดยมีวิธีโพลีโตมัสชิปเทสท์เป็นวิธีเกณฑ์ สามารถอภิปรายผลได้ดังนี้

1. การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อสอบแบบต่อเนื่องสองขั้นตอนระหว่างวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกกับวิธีแมนเทิล-แฮนส์เซลทั่วไปพบว่าเมื่อพิจารณาถึงอัตราความถูกต้องและอัตราความคลาดเคลื่อนประเภทที่ 1 แล้วพบว่าวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกมีอัตราความถูกต้องใกล้เคียงกันแต่สูงกว่าวิธีแมนเทิล-แฮนส์เซลทั่วไปเพียงเล็กน้อย โดยวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกสามารถพบข้อสอบที่ทำหน้าที่ต่างกันตรงกับวิธีโพลีโตมัสชิปเทสท์ซึ่งเป็นวิธีเกณฑ์ มากกว่าวิธีแมนเทิล-แฮนส์เซลทั่วไปเพียง 1 ข้อ ทำให้มีอัตราความถูกต้องเท่ากับ 33.33% และวิธีแมนเทิล-แฮนส์เซลทั่วไปมีอัตราความถูกต้องเท่ากับ 22.22% และทั้งสองวิธีมีอัตราความคลาดเคลื่อนประเภทที่ 1 เท่ากันคือ 9.09% ซึ่งข้อสอบที่เกิดความคลาดเคลื่อนที่ทั้งสองวิธีตรวจสอบพบเป็นข้อสอบข้อเดียวกันแต่วิธีโพลีโตมัสชิปเทสท์ตรวจสอบไม่พบ ซึ่งผลที่ได้จากการศึกษาพบว่ามีความสอดคล้องกับงานวิจัยของวริษฐา ชะม้อย (2550: 115-129) ที่พบว่าวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกจะมีอำนาจการทดสอบมากขึ้นและมีอัตราความคลาดเคลื่อนประเภทที่ 1 หรือความผิดพลาดในการตรวจสอบน้อยลง เมื่อกลุ่มตัวอย่างมีขนาดเพิ่มขึ้นและวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกมีประสิทธิภาพดีกว่าวิธีโครงสร้างความแปรปรวนและค่าเฉลี่ยในทุกเงื่อนไขของการตรวจสอบการทำหน้าที่ต่างกันของ

ข้อสอบที่ให้คะแนนแบบพหุวิภาคและสอดคล้องกับการศึกษาของสุชาติ สิริมินันท์ (2554: 137-157) ที่พบว่าวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกมีประสิทธิภาพในการตรวจสอบสูงกว่าวิธีโพลีโทมัสชิปเทสต์และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ภายใต้เงื่อนไขที่แปรเปลี่ยน 4 ปัจจัย เพราะแต่ละเงื่อนไขมีอัตราความถูกต้อง 80% ขึ้นไป ซึ่งถือว่ามีความถูกต้องเพียงพอและมีอัตราความคลาดเคลื่อนประเภทที่ 1 อยู่ในขอบเขตที่กำหนดและสอดคล้องกับงานวิจัยของ อุทัยวรรณ สายพัฒนา (2547: 97-112) ที่พบว่าวิธีโพลีโทมัสชิปเทสต์มีประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบมากกว่าวิธีแมนเทิล-แฮนส์เซลทั่วไป อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในข้อสอบที่มีการให้คะแนนแบบหลายค่าที่มีจำนวน 20 ข้อและกลุ่มตัวอย่าง 1,000 คน นอกจากนี้ยังมีความเห็นว่าวิธีแมนเทิล-แฮนส์เซลทั่วไปมีประสิทธิภาพต่ำในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อสอบที่มีการให้คะแนนแบบหลายค่า และยังมีขั้นตอนในการวิเคราะห์ยุ่งยากซับซ้อนและต้องใช้กลุ่มตัวอย่างที่มีขนาดใหญ่ ซึ่งสอดคล้องกับการศึกษาของมิลเลอร์และสเปรย์ (Miller & Spray, 1993) ที่ได้ศึกษาการเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบที่มีการให้คะแนนแบบหลายค่า ด้วยวิธี LDFA กับวิธี GMH โดยใช้ข้อมูลผลการสอบการปฏิบัติวิชาคณิตศาสตร์ พบว่าวิธี GMH ขาดอำนาจในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบไม่สม่ำเสมอ และการศึกษาของเวลช์และฮูเวอร์ (Miller & Spray, 1993: 108 ; citing Welch & Hoover, 1993) ที่พบว่า แม้แต่การตรวจสอบการทำหน้าที่ต่างกันของข้อสอบแบบสม่ำเสมอ วิธี GMH ก็มีอำนาจในการตรวจสอบต่ำด้วย และเสนอแนะว่า วิธี GMH อาจจะได้เป็นวิธีการทางเลือกที่ดีที่สุดในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในการตอบข้อสอบที่เป็นแบบหลายค่า และจากการศึกษาของยา-ฮุย ซู; และเว่ย-ซาง หวัง (Ya-Hui Su & Wen-Chung Wang, 2005: 313-350) พบว่าประสิทธิภาพของการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบเมื่อพิจารณาจากอัตราความคลาดเคลื่อนประเภทที่ 1 พบว่าวิธีการวิเคราะห์ฟังก์ชันการจำแนกโลจิสติกให้อำนาจการตรวจสอบที่สูงกว่าวิธีแมนเทิล-แฮนส์เซลทั่วไป ภายใต้ทุกเงื่อนไข

2. เมื่อพิจารณาการทำหน้าที่ต่างกันของข้อสอบเมื่อผู้สอบมีความแตกต่างทางเพศที่ส่งผลต่อความสามารถทางคณิตศาสตร์แล้วพบว่า งานวิจัยที่ได้ศึกษาจะระบุว่าเพศชายมีความรู้ความสามารถมากกว่าเพศหญิงโดยเฉพาะเนื้อหาด้านรูปทรง เรขาคณิตและมีติสัมพันธ์ รวมไปถึงข้อสอบที่มีความซับซ้อนและมีความท้าทาย ซึ่งอาจกล่าวได้ว่าเพศชายจะมีความสามารถมากกว่าเพศหญิงในเรื่องดังกล่าว แต่ในการศึกษาครั้งนี้มีเครื่องมือเป็นแบบทดสอบแบบต่อเนื่องสองขั้นตอนเรื่องความเท่ากันทุกประการของนักเรียนชั้นมัธยมศึกษาปีที่ 2 ซึ่งอยู่ในสาระที่ 3 ที่เป็นการพิสูจน์เรขาคณิต พบว่านักเรียนหญิงมีคะแนนเฉลี่ยสูงกว่านักเรียนชายเล็กน้อยและเมื่อพิจารณาข้อสอบที่ตรวจพบการทำหน้าที่ต่างกันของข้อสอบพบว่าข้อสอบที่ลำเอียงเข้าข้างนักเรียนชายมากกว่านักเรียนหญิง ซึ่งผลที่ได้จากการศึกษาไม่สอดคล้องกับงานวิจัยยิปและเมลฟี (สิริรัตน์ วิชาศิลป์, 2545: 64-65; อ้างอิงจากGipps & Murphy, 1994: 182) ที่กล่าวว่าเพศชายจะมีความสามารถมากกว่าเพศหญิงในลักษณะของโจทย์ที่มีความซับซ้อนและโจทย์ที่เกี่ยวกับการหาความสัมพันธ์และความเข้าใจเกี่ยวกับรูปทรงและการมองมิติต่างๆและไม่สอดคล้องกับงานวิจัยของ โอ ลีเดีย ลุยและ

มาร์ค วิลสัน (Ou Lydia Liu and Mark Wilson. 2009: 164-184) ที่พบว่าเพศชายได้เปรียบกว่าเพศหญิงในเรื่องของปริภูมิและรูปทรงสามมิติ จากที่กล่าวมาข้างต้นอาจเกิดจากลักษณะของแบบทดสอบ เนื่องจากงานวิจัยนี้มีเครื่องมือคือแบบทดสอบแบบต่อเนื่องสองขั้นตอนที่มีการให้คะแนนแบบหลายค่า โดยแบ่งเป็นสองขั้นตอน ขั้นตอนแรกเป็นแบบเลือกตอบจริงหรือเท็จและขั้นตอนที่สองเป็นการให้เหตุผลในการตอบขั้นตอนแรก โดยข้อสอบลักษณะนี้เป็นข้อสอบที่มีการตอบคำถามอย่างต่อเนื่องเพื่อทดสอบความรู้ ความเข้าใจและกระบวนการให้เหตุผลของผู้สอบ ซึ่งเป็นแบบทดสอบที่ผู้สอบไม่คุ้นเคยเพราะในปัจจุบันการเรียนการสอนเรื่องการพิสูจน์เรขาคณิตนั้นจะเป็นการพิสูจน์ที่เป็นลำดับขั้นตอนลงมา โดยแวน ฮิลี (วินัย คำสุวรรณ. 2558: 148-152) ได้แบ่งระดับความคิดทางเรขาคณิตของผู้เรียนเป็น 5 ระดับ ดังนี้ ระดับ 0 การนึกภาพ (visualization) ผู้เรียนมองเห็นหรือนึกออกเฉพาะรูปลักษณะภายนอก ระดับ 1 การวิเคราะห์ (analysis) ผู้เรียนเริ่มต้นสร้างมโนทัศน์ทางเรขาคณิต โดยมองเห็นคุณลักษณะของรูปเรขาคณิต ระดับ 2 การนิรนัยอย่างไม่มีแบบแผน (informal deduction) ผู้เรียนสามารถบอกความสัมพันธ์ตามสมบัติของรูปเรขาคณิต ระดับ 3 การนิรนัย (deduction) ผู้เรียนสามารถเข้าใจสัจพจน์เป็นแนวทางในการศึกษา มีความสามารถในการให้เหตุผลในการพิสูจน์ โดยไม่ได้มาจากความจำ ระดับ 4 เข้มงวด (rigor) ผู้เรียนสามารถเชื่อมโยงความคิดเข้ากับความเข้าใจเรขาคณิตที่ซับซ้อนได้ ถึงแม้ว่าระดับความคิดทางเรขาคณิตจะมีถึงระดับ 4 แต่แวน ฮิลี ให้ความสนใจ 3 ระดับแรก เนื่องจากนักเรียนส่วนใหญ่จะมีระดับความคิดทางเรขาคณิตไม่เกินระดับที่ 3 ซึ่งสอดคล้องกับงานวิจัยของ อาคเนย์ จัตรบรรยงค์ (2548: 60-61) พบว่านักเรียนส่วนใหญ่จะมีความรู้ความเข้าใจในวิชาเรขาคณิตระดับ 2 และมีบางคนที่มีความรู้ความเข้าใจระดับ 3 สามารถบอกสิ่งที่โจทย์กำหนดให้และสิ่งที่ต้องการพิสูจน์ได้ แต่ในขั้นตอนพิสูจน์นั้นนักเรียนไม่สามารถให้เหตุผลได้ ซึ่งจากที่กล่าวมาจะพบว่าการพิสูจน์เรขาคณิตจะต้องมีลำดับขั้นตอนที่ชัดเจนโดยเริ่มจากการมองภาพ แล้วจึงวิเคราะห์ และบอกความสัมพันธ์ของรูปเรขาคณิต และนำไปสู่การพิสูจน์ ซึ่งข้อสอบแบบต่อเนื่องสองขั้นตอนจะเป็นการหาข้อสรุปความสัมพันธ์จากสิ่งที่โจทย์กำหนดให้แล้วให้เหตุผลจากข้อสรุปนั้น ซึ่งอาจทำให้นักเรียนเกิดความสับสนในการทำข้อสอบ เนื่องจากลักษณะของข้อสอบไม่ได้เอื้อต่อการคิดเป็นลำดับขั้นตอนทางเรขาคณิตแต่เป็นการเริ่มต้นด้วยการสรุปความสัมพันธ์และให้เหตุผลสนับสนุน ซึ่งอาจจะเป็นปัจจัยหนึ่งที่ทำให้คะแนนเฉลี่ยของนักเรียนเพศหญิงสูงกว่านักเรียนเพศชาย

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

1.1 การศึกษาครั้งนี้พบว่า การสร้างแบบทดสอบแบบต่อเนื่องสองขั้นตอน ที่มีการให้คะแนนแบบหลายค่าเรื่องความเท่ากันทุกประการ ทำให้ผู้สอบพบกับจำนวนข้อสอบที่เพิ่มมากขึ้น เนื่องจากการทำข้อสอบ 1 ข้อจะต้องมีการตอบคำถามต่อเนื่องกัน ในตอนที่ 1 และตอนที่ 2 อาจจะทำให้ผู้สอบเกิดความสับสนกับการตอบ ดังนั้นในการศึกษาข้อสอบแบบต่อเนื่องสองขั้นตอน เรื่อง

ความเท่ากันทุกประการ ควรใช้รูปแบบการเลือกตอบในขั้นตอนที่ 1 และการเขียนแสดงเหตุผลจากการตอบในขั้นตอนที่ 2 เพื่อเป็นการตรวจสอบกระบวนการให้เหตุผลและสามารถนำผลการตอบมาวินิจฉัยข้อบกพร่องได้อย่างชัดเจน

1.2 การศึกษาครั้งนี้ พบว่าเมื่อใช้แบบทดสอบแบบต่อเนื่องสองขั้นตอน จำนวน 20 ข้อ ในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบ ด้วยวิธีวิเคราะห์ฟังก์ชันการจำแนกแบบ โลจิสติก และวิธีแมนเทิล-แฮนส์เซลทั่วไป ผลการตรวจสอบพบว่าวิเคราะห์ฟังก์ชันการจำแนกแบบ โลจิสติกมีประสิทธิภาพสูงกว่าวิธีแมนเทิล-แฮนส์เซลทั่วไปเพียงเล็กน้อย ดังนั้นสามารถที่จะใช้วิธีวิเคราะห์ฟังก์ชันการจำแนกแบบโลจิสติกแทนวิธีแมนเทิล-แฮนส์เซลทั่วไป

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 ควรมีการศึกษาการทำหน้าที่ต่างกันของข้อสอบในรูปแบบของการทดสอบลักษณะอื่นๆที่แตกต่างออกไป ซึ่งอาจจะเป็นการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในข้อสอบแบบอัตนัย หรือการทดลองในรูปแบบต่างๆ

2.2 ควรมีการศึกษาเพิ่มเติมเกี่ยวกับประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบระหว่างวิธีในกลุ่ม IRT เพื่อให้ได้ผลการทดสอบที่น่าเชื่อถือมากยิ่งขึ้น

2.3 ควรมีการศึกษาเพิ่มเติมเกี่ยวกับแบบทดสอบแบบต่อเนื่องสองขั้นตอน ในเนื้อหาเรขาคณิตที่ให้ผู้เรียนได้แสดงวิธีการเป็นลำดับขั้นตอนที่ชัดเจน เพื่อนำผลที่ได้ไปพัฒนาการเรียนการสอนและกระบวนการให้เหตุผลให้มีประสิทธิภาพมากยิ่งขึ้น

บรรณานุกรม



บรรณานุกรม

- กาญจนา วัฒนสุนทร. (2538). การพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- จันทนา เปรมฤดีปรีชาชาญ. (2551). การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันตามเชื้อชาติของผู้สอบระหว่างวิธีการของ Mantel-Haenszel วิธี Lord's χ^2 และ วิธี Closed Interval Area. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- จิตติมา วรณศรี. (2540). การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีชิปเทสต์ เมื่อขนาดกลุ่มตัวอย่างและความยาวของแบบทดสอบต่างกัน. วิทยานิพนธ์ ค.ม. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- เจน จันทรแสง. (2537). การวิเคราะห์ความลำเอียงของข้อสอบวัดผลสัมฤทธิ์ทางการเรียนคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 3. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร. ถ่ายเอกสาร.
- ชวาล แพรัตกุล. (2516). เทคนิคการวัดผล. กรุงเทพฯ: โรงเรียนแพร่งนสนธิ์.
- ทัศนีย์ พิรมนตรี. (2530). การวิเคราะห์ความลำเอียงของแบบทดสอบวิชาคณิตศาสตร์โครงการตรวจสอบคุณภาพการศึกษา ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526. วิทยานิพนธ์ ค.ม. (ภาควิชาชีพการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- ทองอยู่ สาระ. (2543). การเปรียบเทียบอำนาจการตรวจสอบและการจำแนกผิดพลาดในการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันแบบสมำเสมอและแบบไม่สมำเสมอระหว่างวิธีแมนเทล-แฮนส์เซล และวิธีการถดถอยโลจิสติก โดยใช้ความยาวแบบทดสอบและขนาดกลุ่มตัวอย่างต่างกัน. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- นุชริน ไบโพธิ์. (2544) การเปรียบเทียบความสอดคล้องของผลการตรวจสอบข้อสอบที่ทำหน้าที่ต่างกันระหว่างวิธี Lord's χ^2 วิธี Mantel-Haenszel และวิธี SIBTEST. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- บุญเชิด ภิญโญนนตพงษ์. (2527). การทดสอบแบบอิงเกณฑ์:แนวคิดวิธีการ. กรุงเทพฯ: ภาควิชาพื้นฐานทางการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.

- บุษวรรษ์ แสนปลื้ม. (2550). การเปรียบเทียบวิธีชิปเทสท์แบบหลายค่าที่ใช้รูปแบบการคิดคะแนน
ต่างกันในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อคำถามวัดมโนทัศน์แห่งตน.
ปริญญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัย
ศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- พัชรี ปิยภักดิ์. (2531). การวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบวัดผลสัมฤทธิ์
ทางการเรียนชั้นประถมศึกษาปีที่ 6. ปริญญานิพนธ์ กศ.ม. (การวัดผลการศึกษา).
กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- พรณี จินตมาศ. (2540). การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบ โดยใช้ขนาด
กลุ่มผู้สอบและวิธีวิเคราะห์ต่างกัน. ปริญญานิพนธ์ กศ.ม. (การวัดผลการศึกษา).
กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- รักษนก ยี่สุนศรี. (2544). การวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบและแบบสอบด้วย
กระบวนการ ดี เอฟ ไอ ที่ สำหรับแบบสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษา
วิชาภาษาอังกฤษและวิชาคณิตศาสตร์. วิทยานิพนธ์ ค.ม. (ภาควิชาวิจัยการศึกษา).
กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- เรืองเดช ศิริกิจ. (2554). การวิเคราะห์เปรียบเทียบโมเดลการประเมินคุณภาพการจัดการศึกษา
วิชาคณิตศาสตร์: การประยุกต์ใช้โมเดลมูลค่าเพิ่มที่มีการวิเคราะห์การทำหน้าที่ต่างกัน
ของข้อสอบและการวิเคราะห์การทำหน้าที่ต่างกันของตัวลง. วิทยานิพนธ์ ครุศาสตรดุษฎี
บัณฑิต. จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- วินัย คำสุวรรณ. (2558). มโนทัศน์และการวิจัย ความเข้าใจคณิตศาสตร์. กรุงเทพฯ: แดเน็กซ์
อินเตอร์คอร์ปอเรชั่น.
- วรัชชา ชะม้อย. (2550). การเปรียบเทียบประสิทธิภาพในการตรวจสอบการทำหน้าที่ต่างกันของ
ข้อสอบที่ตรวจให้คะแนนแบบพหุภาคระหว่างวิธีโครงสร้างความแปรปรวนร่วมและ
ค่าเฉลี่ยกับวิธีการวิเคราะห์ฟังก์ชันเชิงจำแนกแบบโลจิสติก. วิทยานิพนธ์ ค.ม. (ภาควิชา
วิจัยและจิตวิทยาการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.
ถ่ายเอกสาร.
- ศิริชัย กาญจนวาสี. (2555). ทฤษฎีการทดสอบแนวใหม่. กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์
มหาวิทยาลัย.
- ศิริชัย กาญจนวาสี; ทวีวัฒน์ ปิตยานนท์; และดิเรก ศรีสุข. (2555). การเลือกใช้สถิติที่เหมาะสม
สำหรับการวิจัย. กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย
- ศิริชัย กาญจนวาสี. (2556). ทฤษฎีการทดสอบแบบดั้งเดิม. กรุงเทพฯ: สำนักพิมพ์จุฬาลงกรณ์
มหาวิทยาลัย
- สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2546). คู่มือวัดผลประเมินผลคณิตศาสตร์.
กรุงเทพฯ ศรีเมืองการพิมพ์
- (2555). คู่มือวัดผลประเมินผลคณิตศาสตร์. กรุงเทพฯ: ศรีเมืองการพิมพ์

สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี. (2557). ผลการประเมิน PISA 2012

คณิตศาสตร์ การอ่าน และวิทยาศาสตร์ นักเรียนรู้อะไร และทำอะไรได้บ้าง. กรุงเทพฯ: สสวท.

สิริรัตน์ วิภาสศิลป์. (2545). การเปรียบเทียบวิธีชิปเทสต์และดีเอฟไอทีในการตรวจสอบการทำหน้าที่เบี่ยงเบน ของข้อสอบ หมวดข้อสอบและแบบทดสอบจากข้อมูลการตอบข้อสอบที่ใช้ความสามารถหลายมิติ. วิทยานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา).

กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

สุรศักดิ์ อมรรัตนศักดิ์; และอนุสรณ์ สกุลคุ. (2549). การประเมินผลการเรียนวิชาคณิตศาสตร์.

พิมพ์ครั้งที่ 5. กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยรามคำแหง.

สุชาติ สิริมินนนท์. (2554). การเปรียบเทียบวิธีโพลี-ชิปเทสต์ วิธีการวิเคราะห์ฟังก์ชันการ

จำแนกโลจิสติก และวิธีการถดถอยโลจิสติกแบบจัดอันดับ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่า. วิทยานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

สุมาลี แก้วทนต์. (2547). สาเหตุการทำหน้าที่ต่างกันของข้อสอบสาระการเรียนรู้ภาษาไทยและสาระการเรียนรู้สังคมศึกษา ศาสนาและวัฒนธรรม. วิทยานิพนธ์ปริญญามหาบัณฑิต.

(ภาควิชาวิจัยการศึกษา). บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน กลุ่มสาระสนเทศ สำนักนโยบายและแผน. สืบค้นเมื่อ 26 เมษายน 2560, จาก <https://portal.bopp-obec.info>

อรินทร์ นวมถนอม. (2549). การเปรียบเทียบวิธีโพลี-ชิปเทสต์ วิธีการถดถอยโลจิสติกแบบจัด

อันดับและวิธีการถดถอยโลจิสติกแบบจัดอันดับหลายมิติ ในการตรวจสอบการทำหน้าที่เบี่ยงเบนของข้อสอบที่วัดความสามารถหลายมิติและให้คะแนนหลายค่า วิทยานิพนธ์ กศ.ด.(การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

อาคเนย์ ฉัตรบรรยงค์. (2548). ชุดการเรียนรู้เรื่องความเท่ากันทุกประการที่เน้นระดับขั้นการเรียนรู้เรขาคณิตของแวน ฮีลี ของนักเรียนชั้นมัธยมศึกษาปีที่ 2. วิทยานิพนธ์ กศ.ม.

(คณิตศาสตร์) กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

อ้อมใจ คงเสาร์. (2541). การศึกษาค่าพารามิเตอร์ของข้อสอบและค่าฟังก์ชันอินฟอร์เมชันจากแบบทดสอบเลือกตอบกับแบบทดสอบโคลซที่วัดมโนทัศน์ทางคณิตศาสตร์ ชั้น

ประถมศึกษาปีที่ 6: การประยุกต์ใช้ทฤษฎีการตอบสนองข้อสอบ. วิทยานิพนธ์ กศ.

ม.(การวัดผลการศึกษา). บัณฑิตวิทยาลัย มหาวิทยาลัยนเรศวร. ถ่ายเอกสาร.

- อุทัยวรรณ สายพัฒนา. (2547). การเปรียบเทียบประสิทธิภาพของผลการตรวจสอบการทำหน้าที่ต่างกันของข้อสอบในแบบทดสอบที่มีการให้คะแนนแบบหลายค่าระหว่างวิธี GMH และวิธี Polytomous SIBTES ปรินธิพนธ์ กศ.ด.(การทดสอบและวัดผลการศึกษา).
กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- Allen, M.J.&Yen, W.M. (1979). *Introduction to Measurement Theory*. Monterey: California Brooks/Cole.
- Angel M. Fidalgo & Dave Bartram. (2010). *A Comparison Between Some Generalized Mantel-Haenszel Statistics for Detecting DIF in at Simulated Under the Graded Response Model*.
- Camilli, Gregory and Lorrie A Shepard. (1994) *Methodes of Identifying Biased Test Items*. California: Sage Publications.
- Catherine S. Taylor. (2012). *Gender DIF in Reading and Mathematics Tests With Mixed Item Formats*. Hambleton, R.K. ; & H. Jane. Rogers. (1989, April) Detecting Potential Biased Test Items.Comparison of IRT Area and Mantel-Haenszel Methods. *Applied Measurement in Education*. 2(4): 246-280.
- Mansyur; & Muliana. (2016). *Detecting Differential Item Functioning and Differential Test Functioning on Math School Final-exam*. Advanced Science Engineering Information Technology. 437-440.
- Miller, T. R.; & Spray , J.A. (1993). Logistic Discriminant Function Analysis for DIF Identification of Polytomous Scored Items. *Journal of Educational Measurement*. 107-122.
- Narayanan, P.; & Swaminathan H. (1994). Performance of the Mantel-Haenszel and Simultaneous Item Bias Procedures for Detecting Differential Item Functioning. *Applied Psychological Measurement*. 18(4): 315-328
- Ou Lydia Liu; & Mark Wilson. (2009). Gender Differences in Large-Scale Math Assessment: PISA Trend 2000 and 2003. *Applied Measurement in Education*. pp.164-184.
- Popham, W.J. (1981). *Modern educational Measurment*. Englewood Cliffs, New Jersey: Prentice-Hall.
- Randall D. Penfield. (2001). *Assessing Differential Item Functioning Among Multiple Group: A Comparison of Three Mantel-Haenszel Procedures*.

- Ya-Hui Su; & Wen-Chung Wang. (2005). Efficiency of the Mantel, Generalized Mantel - Haenszel, and Logistic Discriminant Function Analysis Methods in Detecting Differential Item Functioning for Polytomous Items. *Applied Measurement in Education*. 313-350.
- Young-Sun Lee. (2009). Examining type I error and power for detection of differential item and testlet functioning, *The Asia Pacific Education Review*. 365–375



ภาคผนวก



ภาคผนวก ก

รายนามผู้เชี่ยวชาญในการตรวจเครื่องมือที่ใช้ในการวิจัย



รายนามผู้เชี่ยวชาญในการตรวจเครื่องมือที่ใช้ในการวิจัย

ดร. มนตา ตุลย์เมธการ	อาจารย์ประจำภาควิชาการวัดผลและการวิจัยการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ดร. นวรินทร์ ตาก้อนทอง	อาจารย์ประจำสำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ
อาจารย์ปานวสน์ มหาลเลิศ	อาจารย์ประจำสำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ
อาจารย์ชัยพร นาควิเชียร	ตำแหน่ง ครู วิทยฐานะ ครูชำนาญการพิเศษ โรงเรียนวัดราชโอรส จังหวัดกรุงเทพฯ
อาจารย์พรวิทย์ ศรียะพันธุ์	ตำแหน่ง ครู วิทยฐานะ ครูชำนาญการพิเศษ โรงเรียนเทพศิรินทร์ร่มเกล้า จังหวัดกรุงเทพฯ

ภาคผนวก ข

ค่าดัชนีความสอดคล้อง (Index of item – objective congruence :IOC)
ของแบบทดสอบแบบต่อเนื่องสองขั้นตอนเรื่องความเท่ากันทุกประการ

ตาราง 18 ค่าดัชนีความสอดคล้องระหว่างข้อสอบกับจุดประสงค์ (Index of item – objective congruence :IOC) ของแบบทดสอบแบบต่อเนื่องสองขั้นตอนเรื่องความเท่ากันทุกประการ

ข้อที่	ความคิดเห็นของผู้เชี่ยวชาญ					ค่าดัชนีความสอดคล้อง ของเนื้อหา (IOC)	ผลการพิจารณา
	1	2	3	4	5		
1	1	0	1	1	1	0.8	ใช้ได้
2	1	1	1	1	1	1	ใช้ได้
3	1	1	1	0	1	0.8	ใช้ได้
4	1	1	1	1	1	1	ใช้ได้
5	1	1	1	1	1	1	ใช้ได้
6	1	1	1	1	1	1	ใช้ได้
7	1	1	1	1	1	1	ใช้ได้
8	1	1	1	1	1	1	ใช้ได้
9	1	1	1	1	1	1	ใช้ได้
10	1	1	1	1	1	1	ใช้ได้
11	1	1	1	1	1	1	ใช้ได้
12	1	1	1	1	1	1	ใช้ได้
13	1	1	1	1	1	1	ใช้ได้
14	1	1	1	1	1	1	ใช้ได้
15	1	1	1	1	1	1	ใช้ได้
16	1	1	1	1	1	1	ใช้ได้
17	1	1	1	1	1	1	ใช้ได้
18	1	1	1	1	1	1	ใช้ได้
19	1	1	1	1	1	1	ใช้ได้
20	1	1	1	1	1	1	ใช้ได้
21	1	1	1	1	1	1	ใช้ได้
22	1	1	1	1	1	1	ใช้ได้
23	0	1	1	1	1	0.8	ใช้ได้
24	1	1	1	0	1	0.8	ใช้ได้
25	1	1	1	1	1	1	ใช้ได้

ตาราง 18 (ต่อ)

ข้อที่	ความคิดเห็นของผู้เชี่ยวชาญ					ค่าดัชนีความสอดคล้อง ของเนื้อหา (IOC)	ผลการพิจารณา
	1	2	3	4	5		
26	1	1	1	1	1	1	ใช้ได้
27	1	1	1	1	1	1	ใช้ได้
28	0	1	1	1	1	0.8	ใช้ได้
29	1	1	1	1	1	1	ใช้ได้
30	1	1	1	1	1	1	ใช้ได้
31	1	1	0	1	1	0.8	ใช้ได้
32	1	1	1	1	1	1	ใช้ได้
33	1	1	1	1	1	1	ใช้ได้
34	-1	1	1	1	1	0.6	ใช้ได้
35	1	1	1	1	1	1	ใช้ได้
36	1	1	1	1	1	1	ใช้ได้
37	1	1	1	1	1	1	ใช้ได้
38	1	1	1	1	1	1	ใช้ได้
39	1	1	1	1	1	1	ใช้ได้
40	1	1	1	1	1	1	ใช้ได้

ภาคผนวก ค

ค่าความยากง่าย (p) ค่าอำนาจจำแนก (r) และค่าความเชื่อมั่น
ของแบบทดสอบแบบต่อเนื่องสองขั้นตอน (Cronbach's Alpha)

ตาราง 19 ผลการวิเคราะห์ค่าความยากง่าย (p) ค่าอำนาจจำแนก (r) และค่าความเชื่อมั่นของ
แบบทดสอบแบบต่อเนื่องสองขั้นตอน (Cronbach's Alpha)

ข้อที่	p	ผลการ วิเคราะห์	r	ผลการ วิเคราะห์	ผลการ พิจารณา
1	0.51	ผ่านเกณฑ์	0.09	ไม่ผ่านเกณฑ์	คัดออก
2	0.43	ผ่านเกณฑ์	0.20	ผ่านเกณฑ์	คัดเลือก
3	0.33	ผ่านเกณฑ์	0.19	ไม่ผ่านเกณฑ์	คัดออก
4	0.34	ผ่านเกณฑ์	0.24	ผ่านเกณฑ์	คัดเลือก
5	0.30	ผ่านเกณฑ์	0.21	ผ่านเกณฑ์	คัดเลือก
6	0.52	ผ่านเกณฑ์	0.23	ผ่านเกณฑ์	คัดเลือก
7	0.54	ผ่านเกณฑ์	0.16	ไม่ผ่านเกณฑ์	คัดเลือก
8	0.26	ผ่านเกณฑ์	0.09	ไม่ผ่านเกณฑ์	คัดออก
9	0.43	ผ่านเกณฑ์	0.36	ผ่านเกณฑ์	คัดเลือก
10	0.59	ผ่านเกณฑ์	0.33	ผ่านเกณฑ์	คัดเลือก
11	0.35	ผ่านเกณฑ์	0.20	ผ่านเกณฑ์	คัดเลือก
12	0.53	ผ่านเกณฑ์	0.17	ไม่ผ่านเกณฑ์	คัดเลือก
13	0.33	ผ่านเกณฑ์	0.18	ไม่ผ่านเกณฑ์	คัดออก
14	0.37	ผ่านเกณฑ์	0.20	ผ่านเกณฑ์	คัดเลือก
15	0.29	ผ่านเกณฑ์	0.24	ผ่านเกณฑ์	คัดเลือก
16	0.43	ผ่านเกณฑ์	0.24	ผ่านเกณฑ์	คัดเลือก
17	0.43	ผ่านเกณฑ์	0.45	ผ่านเกณฑ์	คัดเลือก
18	0.61	ผ่านเกณฑ์	0.30	ผ่านเกณฑ์	คัดเลือก
19	0.49	ผ่านเกณฑ์	0.15	ไม่ผ่านเกณฑ์	คัดออก
20	0.31	ผ่านเกณฑ์	0.06	ไม่ผ่านเกณฑ์	คัดออก
21	0.34	ผ่านเกณฑ์	0.22	ผ่านเกณฑ์	คัดเลือก
22	0.32	ผ่านเกณฑ์	0.24	ผ่านเกณฑ์	คัดเลือก
23	0.48	ผ่านเกณฑ์	0.26	ผ่านเกณฑ์	คัดเลือก
24	0.34	ผ่านเกณฑ์	0.17	ไม่ผ่านเกณฑ์	คัดออก
25	0.52	ผ่านเกณฑ์	0.23	ผ่านเกณฑ์	คัดเลือก
26	0.50	ผ่านเกณฑ์	0.21	ผ่านเกณฑ์	คัดเลือก
27	0.49	ผ่านเกณฑ์	0.18	ไม่ผ่านเกณฑ์	คัดออก
28	0.47	ผ่านเกณฑ์	0.23	ผ่านเกณฑ์	คัดเลือก

ตาราง 19 (ต่อ)

ข้อที่	p	ผลการวิเคราะห์	r	ผลการวิเคราะห์	ผลการพิจารณา
29	0.46	ผ่านเกณฑ์	0.10	ไม่ผ่านเกณฑ์	คัดออก
30	0.51	ผ่านเกณฑ์	-0.10	ไม่ผ่านเกณฑ์	คัดออก

ข้อสอบที่ผ่านเกณฑ์จำนวน 18 ข้อ ทำการคัดเลือกเพิ่มอีก 2 ข้อเพื่อให้ครบ 20 ข้อ โดยได้คัดเลือกข้อ 7 และข้อ 12 ที่มีความยากง่ายผ่านเกณฑ์ และมีค่าอำนาจจำแนกใกล้เคียงกับ 0.2 โดยข้อสอบที่คัดเลือกมีความยากง่ายอยู่ระหว่าง 0.29 – 0.61 และมีค่าอำนาจจำแนกอยู่ระหว่าง 0.16 – 0.45 และค่าความเชื่อมั่นของแบบทดสอบแบบต่อเนื่องสองขั้นตอน โดยใช้สูตร Cronbach's Alpha มีค่าเท่ากับ 0.747



ภาคผนวก ง

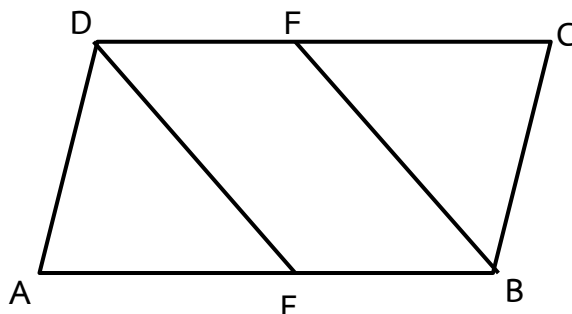
แบบทดสอบแบบต่อเนื่องสองขั้นตอน เรื่องความเท่ากันทุกประการ
สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2

แบบทดสอบแบบต่อเนื่องสองขั้นตอน เรื่องความเท่ากันทุกประการ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2

คำอธิบายวิธีทำแบบทดสอบ

1. แบบทดสอบฉบับนี้เป็นแบบทดสอบแบบต่อเนื่องสองขั้นตอน มีคำถามทั้งหมด **20 ข้อ**
ให้เวลาทำ **60 นาที**
2. คำถามในแบบทดสอบนี้เป็นแบบเลือกตอบ ซึ่งมีคำตอบที่ถูกต้องเพียงคำตอบเดียว
เมื่อเลือกได้คำตอบแล้วให้ตอบในกระดาษคำตอบ
3. คำถามในแบบทดสอบนี้ ใน **1 ข้อ** จะประกอบด้วย **2 ตอน**
 - ตอนที่ 1** เป็นการเลือกตอบในลักษณะของข้อความที่กำหนดให้เป็นจริงหรือ
เป็นเท็จ
 - ตอนที่ 2** เป็นการให้เหตุผลจากการตอบคำถามที่ต่อเนื่องมาจากการตอบใน
ตอนที่ 1
 - ในกรณีตอนที่ 1 นักเรียนตอบ “จริง” ให้เลือกทำตอนที่ 2 ที่ระบุเหตุผลที่ทำให้
ข้อความเป็นจริง
 - ในกรณีตอนที่ 1 นักเรียนตอบ “เท็จ” ให้เลือกทำตอนที่ 2 ที่ระบุสมบัติที่ขาด
หายไป
4. นักเรียนต้องทำข้อสอบทั้ง 2 ตอนในแต่ละข้อ
5. ควรระวังในการกากบาทคำตอบให้ตรงกับข้อความคำถามในแต่ละตอน
6. นักเรียนสามารถทดลองในกระดาษข้อสอบได้
7. ควรทำข้อสอบด้วยความตั้งใจและทำให้ครบทุกข้อคำถาม

ตัวอย่างข้อ (1) ดังนี้



1.1 จากรูปกำหนดให้ ABCD เป็นรูปสี่เหลี่ยมด้านขนาน E และ F เป็นจุดกึ่งกลางของ $\overline{AB}$ และ $\overline{DC}$

แล้ว $\triangle ADE \cong \triangle CBF$ ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 1.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 1.2(2)

1.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\overline{DE} = \overline{BF}$

ข. $\overline{AE} = \overline{CF}$

ค. $\overline{AD} = \overline{BC}$

ง. $\angle DAE = \angle BCF$

1.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

ก. $\overline{AD} = \overline{CB}$

ข. $\angle DAE = \angle BCF$

ค. $\angle ADE = \angle CBF$

ง. $\angle EDF = \angle FEB$

วิธีคิด

จากข้อความที่กำหนดให้จะได้ว่า $\triangle ADE \cong \triangle CBF$ ข้อความนี้เป็นจริง

ดังนั้นจึงเลือกตอบข้อ 1.2(1) ข. $\overline{AE} = \overline{CF}$

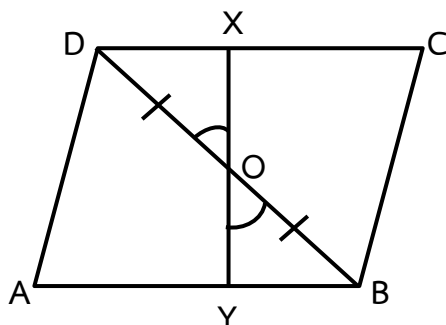
เนื่องจาก $\overline{AD} = \overline{BC}$ โจทย์กำหนดให้ (สมบัติของรูปสี่เหลี่ยมด้านขนาน)

$\angle DAE = \angle BCF$ โจทย์กำหนดให้ (สมบัติของรูปสี่เหลี่ยมด้านขนาน)

$\overline{AE} = \overline{CF}$ เนื่องจาก E และ F เป็นจุดกึ่งกลางของ $\overline{AB}$ และ $\overline{DC}$

ดังนั้น $\triangle ADE \cong \triangle CBF$ (ความสัมพันธ์แบบ ด้าน – มุม – ด้าน)

ตัวอย่างข้อ (2) ดังนี้



2.1. จากรูปกำหนดให้ O เป็นจุดกึ่งกลางของ $\overline{DB}$ แล้ว $\triangle DOX \cong \triangle YOY$
ข้อความเป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 2.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 2.2(2)

2.2(1) ถ้าข้อความเป็นจริง เพราะเหตุผลในข้อใด

ก. $\overline{OX} = \overline{OY}$

ข. $\overline{OX} = \overline{BY}$

ค. $\widehat{OXD} = \widehat{OYB}$

ง. $\widehat{DXO} = \widehat{BYO}$

2.2(2) ถ้าข้อความเป็นเท็จ เพราะขาดสมบัติในข้อใด

ก. $\overline{OX} = \overline{OY}$

ข. $\overline{OB} = \overline{OD}$

ค. $\overline{DX} = \overline{BY}$

ง. $\widehat{DOX} = \widehat{BOY}$

วิธีคิด

จากข้อความที่กำหนดให้จะได้ว่า $\triangle DOX \cong \triangle YOY$ ข้อความเป็นเท็จ

เนื่องจากข้อความที่กำหนดให้ไม่เพียงพอที่จะตอบว่า $\triangle DOX \cong \triangle YOY$

ดังนั้นจึงเลือกตอบข้อ 2.2(2) ก. $\overline{OX} = \overline{OY}$

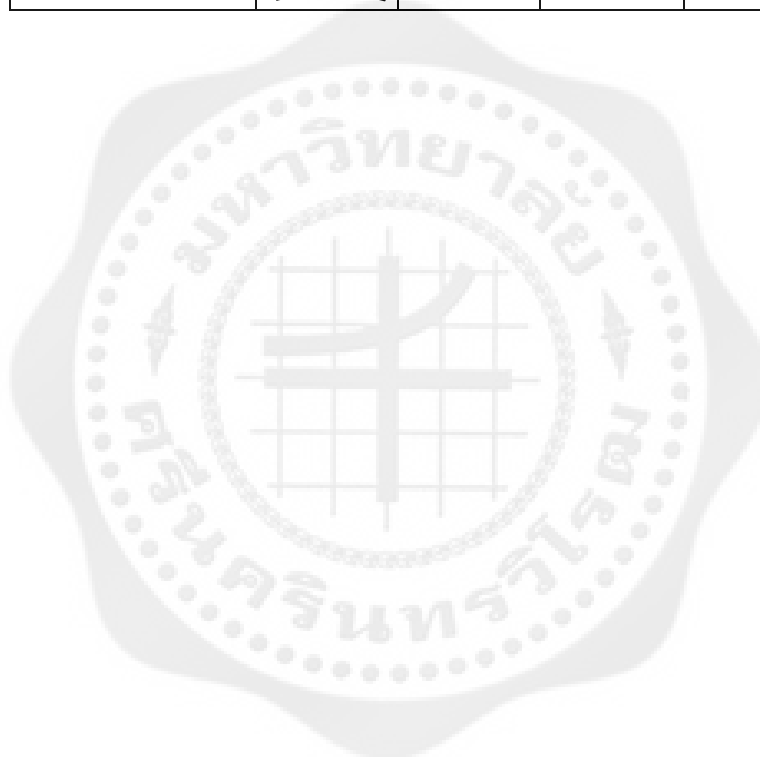
เนื่องจาก ถ้ามีสมบัติ $\overline{OX} = \overline{OY}$

จะทำให้ $\triangle DOX \cong \triangle YOY$ (ความสัมพันธ์แบบ ด้าน – มุม – ด้าน)

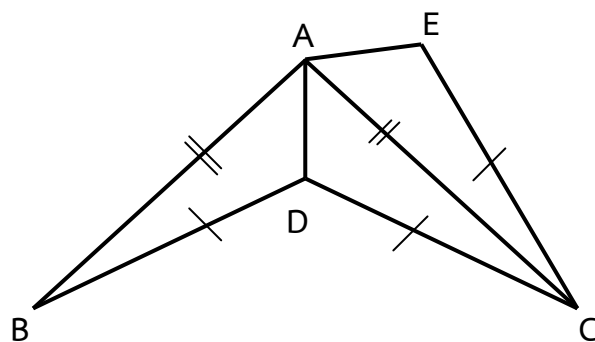
ตัวอย่างกระดาษคำตอบ

ชื่อ นามสกุล.....ชั้น.....เลขที่.....
 โรงเรียน เพศ ☐ ชาย ☐ หญิง

ข้อ 1	ก	ข	ค	ง
1.1	<del> </del>			
1.2		<del> </del>		
ข้อ 2	ก	ข	ค	ง
2.1		<del> </del>		
2.2	<del> </del>			



1.



1.1 จากรูปกำหนดให้สี่เหลี่ยม ADCE เป็นรูปสี่เหลี่ยมรูปว่าว แล้ว $\triangle ABD \cong \triangle AEC$
ข้อความเป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 1.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 1.2(2)

1.2(1) ถ้าข้อความเป็นจริง เพราะเหตุผลในข้อใด

ก. $\overline{AD} = \overline{AE}$

ข. $\widehat{ADC} = \widehat{AEC}$

ค. $\widehat{BAD} = \widehat{CAE}$

ง. $\overline{DC} = \overline{EC}$

1.2(2) ถ้าข้อความเป็นเท็จ เพราะขาดสมบัติในข้อใด

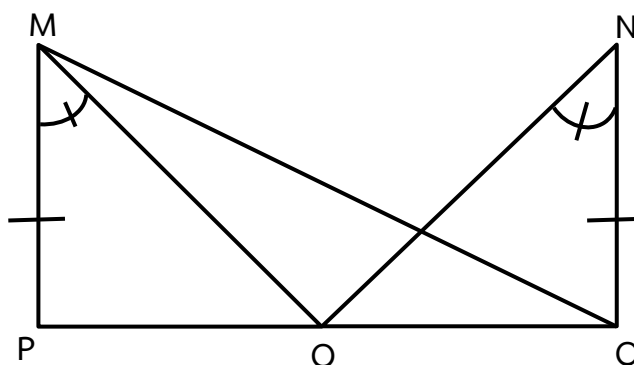
ก. $\overline{AC} = \overline{AC}$

ข. $\widehat{BAD} = \widehat{CAE}$

ค. $\overline{DC} = \overline{EC}$

ง. $\widehat{ADC} = \widehat{AEC}$

2.



2.1 จากรูปกำหนดให้ $\triangle PMO \cong \triangle QNO$ แล้วข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 2.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 2.2(2)

2.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\overline{NO} = \overline{MO}$

ข. $\widehat{MPO} = \widehat{NQO}$

ค. $\overline{PO} = \overline{QO}$

ง. $\widehat{PCM} = \widehat{QON}$

2.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

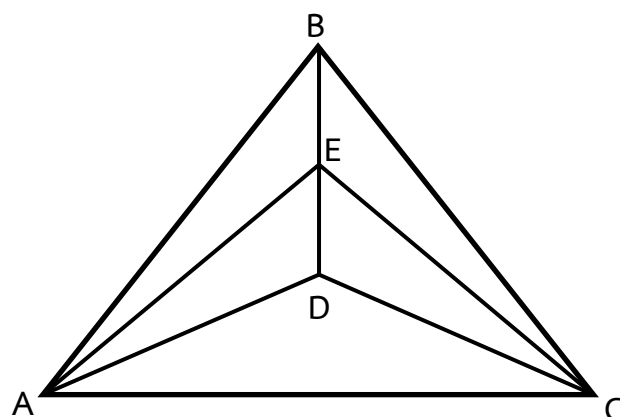
ก. $\widehat{MQP} = \widehat{QON}$

ข. $\widehat{MOP} = \widehat{NOQ}$

ค. $\overline{PO} = \overline{QO}$

ง. $\widehat{OFM} = \widehat{OQN}$

3.



3.1 จากรูปกำหนดให้สามเหลี่ยม ABC และสามเหลี่ยม ADC เป็นรูปสามเหลี่ยมหน้าจั่ว แล้ว $\triangle AED \cong \triangle CED$ ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 3.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 3.2(2)

3.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\overline{AE} = \overline{CE}$

ข. $\widehat{DAC} = \widehat{DCA}$

ค. $\widehat{BAC} = \widehat{BCA}$

ง. $\overline{AD} = \overline{CD}$

3.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

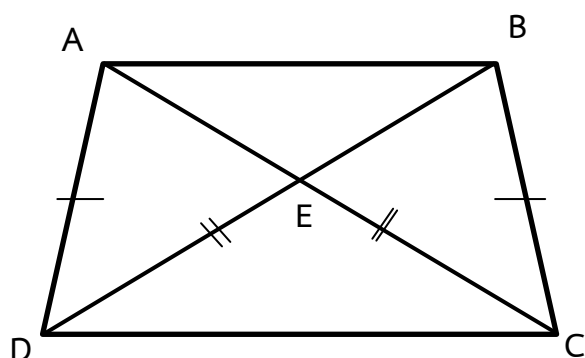
ก. $\widehat{AED} = \widehat{CED}$

ข. $\overline{AE} = \overline{CE}$

ค. $\overline{DE} = \overline{DE}$

ง. $\widehat{EAD} = \widehat{ECD}$

4.



4.1 จากรูปกำหนดให้ $\overline{DB} = \overline{CA}$ และ $\overline{AD} = \overline{BC}$ แล้ว $\triangle DAE \cong \triangle CBE$
ข้อความเป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 4.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 4.2(2)

4.2(1) ถ้าข้อความเป็นจริง เพราะเหตุผลในข้อใด

ก. $\angle DEA = \angle CEB$

ข. $\overline{AB} = \overline{AB}$

ค. $\overline{DC} = \overline{DC}$

ง. $\angle AEB = \angle DEC$

4.2(2) ถ้าข้อความเป็นเท็จ เพราะขาดสมบัติในข้อใด

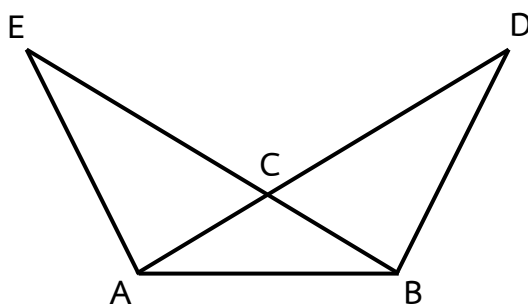
ก. $\angle AEB = \angle DEC$

ข. $\overline{EB} = \overline{EA}$

ค. $\overline{DC} = \overline{DC}$

ง. $\angle AED = \angle BEC$

5.



5.1 จากรูปกำหนดให้สามเหลี่ยม ABC เป็นสามเหลี่ยมหน้าจั่ว ให้ $\overline{AD} = \overline{BE}$
แล้ว $\triangle AEC \cong \triangle BDC$ ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 5.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 5.2(2)

5.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\overline{EC} = \overline{DC}$

ข. $\overline{AE} = \overline{BD}$

ค. $\widehat{EAB} = \widehat{DBA}$

ง. $\widehat{CAB} = \widehat{CBA}$

5.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

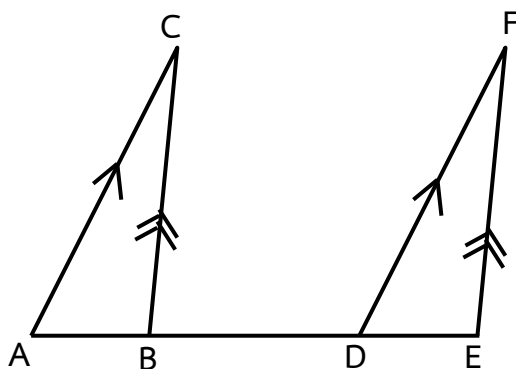
ก. $\overline{AE} = \overline{BD}$

ข. $\overline{AE} = \overline{AD}$

ค. $\widehat{EAB} = \widehat{DBA}$

ง. $\widehat{CAB} = \widehat{CBA}$

6.



6.1 จากรูปกำหนดให้ $AC \parallel DF$ และ $BC \parallel EF$ แล้ว $\triangle ABC \cong \triangle DEF$
ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 6.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 6.2(2)

6.2(1) ถ้าข้อความเป็นจริง เพราะเหตุผลในข้อใด

ก. $\angle CEA = \angle FED$

ข. $\overline{BC} = \overline{EF}$

ค. $\angle CAB = \angle FDE$

ง. $\overline{AC} = \overline{DF}$

6.2(2) ถ้าข้อความเป็นเท็จ เพราะขาดสมบัติในข้อใด

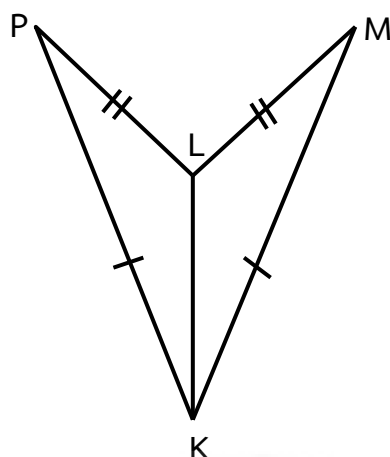
ก. $\overline{AB} = \overline{DE}$

ข. $\angle ACB = \angle DFE$

ค. $\overline{AC} = \overline{DF}$

ง. $\angle CEA = \angle FED$

7.



7.1 จากรูปกำหนดให้ $\triangle PLK \cong \triangle MLK$ แล้วข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 7.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 7.2(2)

7.2(1) ถ้าข้อความเป็นจริง เพราะเหตุผลในข้อใด

ก. $\widehat{PLK} = \widehat{MLK}$

ข. $\widehat{KPL} = \widehat{KML}$

ค. $\overline{LK} = \overline{LK}$

ง. $\widehat{PKL} = \widehat{MKL}$

7.2(2) ถ้าข้อความเป็นเท็จ เพราะขาดสมบัติในข้อใด

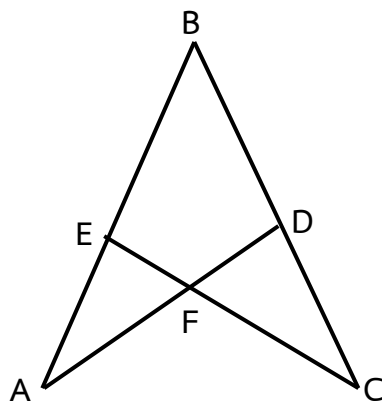
ก. $\widehat{PKL} = \widehat{MKL}$

ข. $\widehat{KML} = \widehat{KPL}$

ค. $\overline{LK} = \overline{LK}$

ง. $\widehat{PLK} = \widehat{MLK}$

8.



8.1 จากรูปกำหนดให้ $\overline{AE} = \overline{CE}$ แล้ว $\triangle ABD \cong \triangle CBE$ ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 8.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 8.2(2)

8.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\overline{BE} = \overline{BD}$

ข. $\overline{AEF} = \overline{CDF}$

ค. $\overline{AF} = \overline{CF}$

ง. $\overline{AE} = \overline{CD}$

8.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

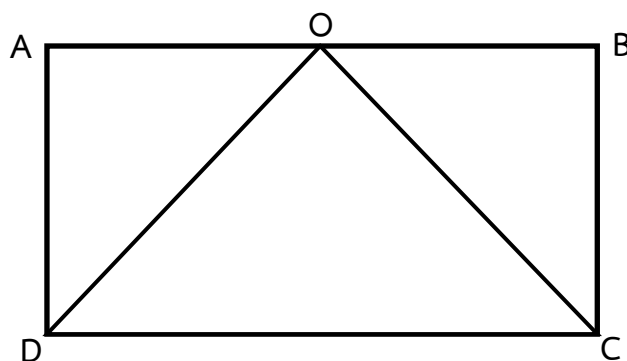
ก. $\widehat{BAD} = \widehat{BCE}$

ข. $\overline{AD} = \overline{CE}$

ค. $\widehat{ADB} = \widehat{CEB}$

ง. $\overline{BD} = \overline{BE}$

9.



9.1 จากรูปกำหนดให้รูปสี่เหลี่ยม ABCD เป็นรูปสี่เหลี่ยมมุมฉาก ที่มี O เป็นจุดกึ่งกลางของ $\overline{AB}$ แล้ว $\triangle DAO \cong \triangle CBO$ ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 9.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 9.2(2)

9.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\overline{AD} = \overline{CO}$

ข. $\overline{DO} = \overline{CO}$

ค. $\overline{AO} = \overline{BC}$

ง. $\overline{AD} = \overline{BC}$

9.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

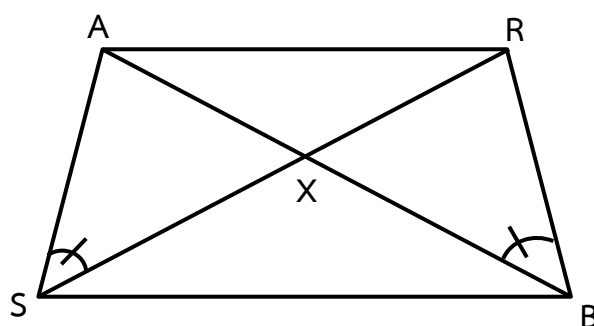
ก. $\overline{AO} = \overline{BO}$

ข. $\overline{DO} = \overline{CO}$

ค. $\overline{AB} = \overline{DC}$

ง. $\widehat{DAO} = \widehat{CBO}$

10.



10.1 จากรูปกำหนดให้สามเหลี่ยม AXR เป็นสามเหลี่ยมหน้าจั่ว แล้ว $\triangle SAR \cong \triangle BRA$ ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 10.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 10.2(2)

10.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\overline{AR} = \overline{AR}$

ข. $\overline{SA} = \overline{BR}$

ค. $\angle ASR = \angle RBA$

ง. $\angle XSR = \angle XRB$

10.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

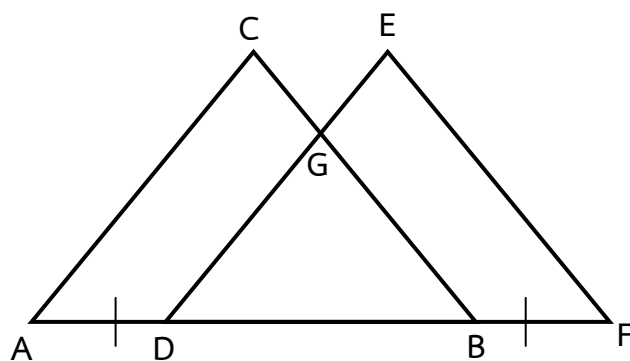
ก. $\overline{BA} = \overline{SR}$

ข. $\overline{AR} = \overline{AR}$

ค. $\overline{SA} = \overline{BR}$

ง. $\angle SAR = \angle BRA$

11.



11.1 จากรูปกำหนดให้สามเหลี่ยม DGB เป็นสามเหลี่ยมด้านเท่า และ $\overline{AD} = \overline{FB}$
แล้ว $\triangle ACB \cong \triangle FED$ ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 11.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 11.2(2)

11.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\widehat{GDB} = \widehat{GBD}$

ข. $\overline{AC} = \overline{FE}$

ค. $\widehat{CAD} = \widehat{EFB}$

ง. $\overline{DG} = \overline{BG}$

11.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

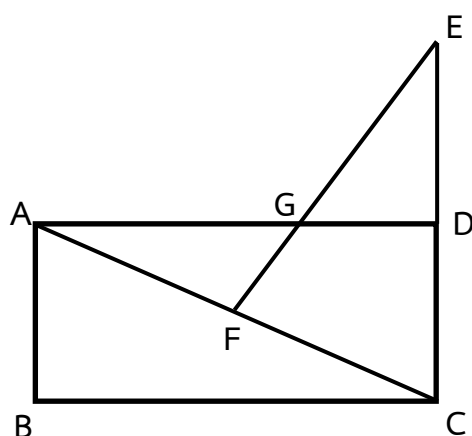
ก. $\widehat{CAD} = \widehat{EFB}$

ข. $\overline{DB} = \overline{DB}$

ค. $\widehat{GDB} = \widehat{GBD}$

ง. $\overline{CB} = \overline{ED}$

12.



12.1 จากรูปกำหนดให้สี่เหลี่ยม ABCD เป็นรูปสี่เหลี่ยมผืนผ้า โดย $\overline{AC} = 2\overline{AB}$ โดยมีจุด F เป็นจุดกึ่งกลางของ $\overline{AC}$ และ $\overline{CD} = \overline{DE}$ และ $\overline{AD} = \overline{EF}$ แล้ว $\triangle ACD \cong \triangle ECF$ ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 12.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 12.2(2)

12.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\overline{CE} = \overline{AC}$

ข. $\widehat{FEC} = \widehat{DAC}$

ค. $\widehat{EFC} = \widehat{ADC}$

ง. $\overline{BC} = \overline{AD}$

12.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

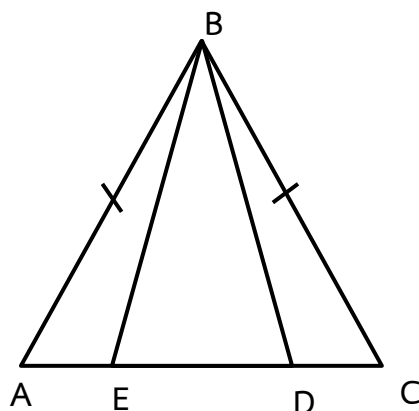
ก. $\overline{CD} = \overline{DE}$

ข. $\widehat{FEC} = \widehat{DAC}$

ค. $\widehat{EFC} = \widehat{ADC}$

ง. $\overline{AD} = \overline{EF}$

13.



13.1 จากรูปกำหนดให้ สามเหลี่ยม ABC เป็นรูปสามเหลี่ยมหน้าจั่ว แล้ว $\triangle ABE \cong \triangle CBD$ ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 13.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 13.2(2)

13.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\widehat{BAE} = \widehat{BCD}$

ข. $\overline{BE} = \overline{BD}$

ค. $\overline{AE} = \overline{DB}$

ง. $\widehat{BED} = \widehat{BDB}$

13.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

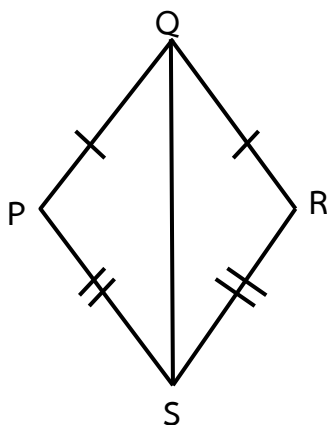
ก. $\overline{BE} = \overline{BD}$

ข. $\overline{AE} = \overline{CD}$

ค. $\widehat{AEB} = \widehat{CDB}$

ง. $\widehat{EDB} = \widehat{BCD}$

14.



14.1 จากรูปกำหนดให้ $\triangle PQS \cong \triangle RQS$ แล้วข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 14.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 14.2(2)

14.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\overline{QS} = \overline{QS}$

ข. $\angle PQS = \angle RQS$

ค. $\angle PQS = \angle RQS$

ง. $\angle PSQ = \angle RSQ$

14.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

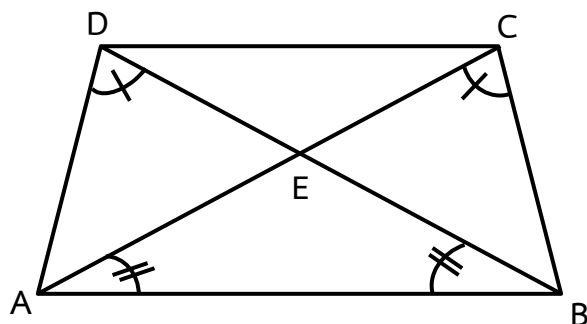
ก. $\overline{QS} = \overline{QS}$

ข. $\angle PQS = \angle RQS$

ค. $\angle PQS = \angle RQS$

ง. $\angle PSQ = \angle RSQ$

15.



15.1 จากรูปกำหนดให้ $\triangle ABD \cong \triangle BAC$ แล้วข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 15.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 15.2(2)

15.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\overline{AE} = \overline{AE}$

ข. $\widehat{ADB} = \widehat{BCA}$

ค. $\overline{AD} = \overline{BC}$

ง. $\widehat{AED} = \widehat{CEB}$

15.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

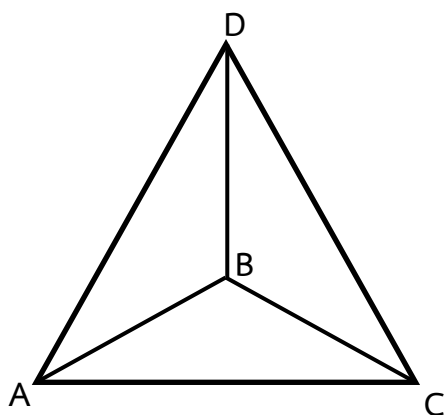
ก. $\overline{DB} = \overline{CA}$

ข. $\widehat{AED} = \widehat{BEC}$

ค. $\overline{AB} = \overline{AB}$

ง. $\widehat{DAE} = \widehat{CBE}$

16.



16.1 จากรูปกำหนดให้สามเหลี่ยม ABC เป็นรูปสามเหลี่ยมหน้าจั่ว แล้ว $\triangle ABD \cong \triangle CBD$ ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 16.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 16.2(2)

16.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\widehat{DAB} = \widehat{DCB}$

ข. $\overline{BD} = \overline{BD}$

ค. $\overline{AB} = \overline{CB}$

ง. $\widehat{ADB} = \widehat{CDB}$

16.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

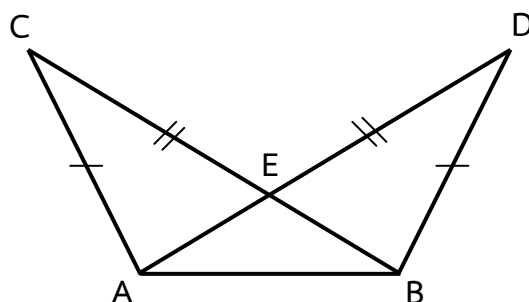
ก. $\overline{AD} = \overline{CD}$

ข. $\widehat{BAC} = \widehat{BCA}$

ค. $\widehat{BDA} = \widehat{CDB}$

ง. $\overline{AB} = \overline{CB}$

17.



17.1 จากรูปกำหนดให้สามเหลี่ยม ABC และสามเหลี่ยม ABD เป็นรูปสามเหลี่ยมสองรูปที่มี $\overline{AC} = \overline{BD}$ และ $\overline{BC} = \overline{AD}$ แล้ว $\triangle ABC \cong \triangle ABD$

ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 17.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 17.2(2)

17.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\widehat{CAB} = \widehat{DBA}$

ข. $\overline{AE} = \overline{BE}$

ค. $\widehat{AEC} = \widehat{BED}$

ง. $\overline{AB} = \overline{AB}$

17.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

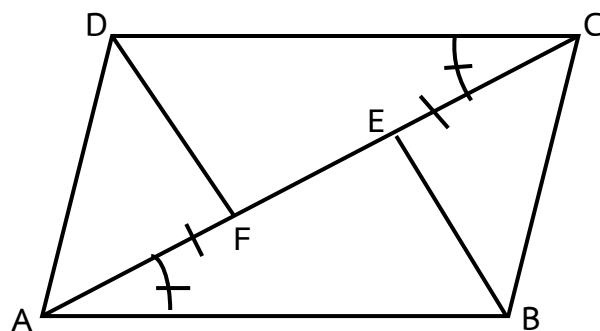
ก. $\widehat{EAB} = \widehat{EBA}$

ข. $\overline{CE} = \overline{DE}$

ค. $\widehat{AEC} = \widehat{BED}$

ง. $\overline{AE} = \overline{BE}$

18.



18.1 จากรูปกำหนดให้ $\widehat{BAE} = \widehat{DCF}$, $\overline{AF} = \overline{EC}$ และ $\widehat{AEB} = \widehat{CFD} = 90^\circ$
แล้ว $\triangle ABE \cong \triangle CDF$ ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 18.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 18.2(2)

18.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\overline{AB} = \overline{CD}$

ข. $\overline{BE} = \overline{DF}$

ค. $\overline{AE} = \overline{CF}$

ง. $\overline{AD} = \overline{BC}$

18.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

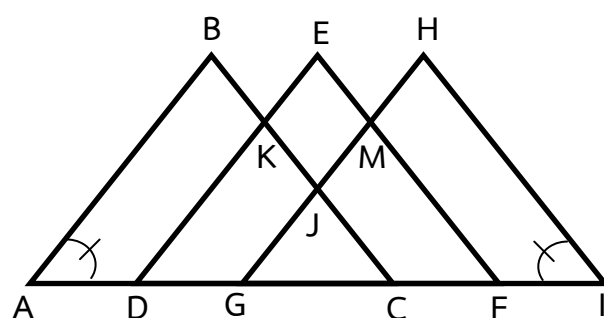
ก. $\overline{AC} = \overline{AC}$

ข. $\overline{AE} = \overline{DC}$

ค. $\widehat{FDC} = \widehat{ABE}$

ง. $\widehat{DAC} - \widehat{ACB}$

19.



19.1 จากรูปกำหนดให้สามเหลี่ยม ABC สามเหลี่ยม DEF และสามเหลี่ยม GHI
 สร้างอยู่บน $\overline{AI}$ ให้ $\angle BAC = \angle HIG$ และ สามเหลี่ยม GJC เป็นสามเหลี่ยมหน้าจั่ว
 และ $\overline{AG} = \overline{IC}$ แล้ว $\triangle ABC \cong \triangle IGH$ ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 19.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 19.2(2)

19.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\angle BAC = \angle HIG$

ข. $\overline{AB} = \overline{GH}$

ค. $\overline{BC} = \overline{HI}$

ง. $\overline{AC} = \overline{IG}$

19.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

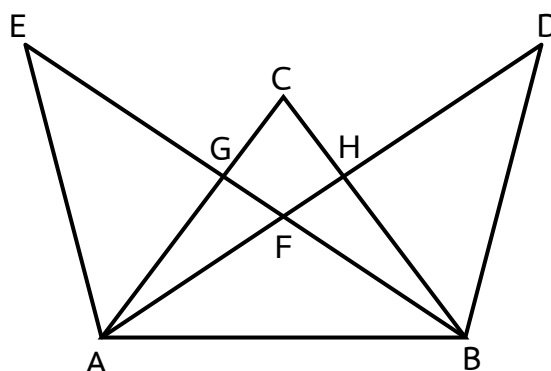
ก. $\angle BAC = \angle HIG$

ข. $\overline{GJ} = \overline{CJ}$

ค. $\angle BAC = \angle HIG$

ง. $\overline{BC} = \overline{HI}$

20.



20.1 จากรูปกำหนดให้สามเหลี่ยม AFB เป็นสามเหลี่ยมหน้าจั่ว และ $\overline{BE} = \overline{AD}$

แล้ว $\triangle AEF \cong \triangle BDF$ ข้อความนี้เป็นจริงหรือเท็จ

ก. เป็นจริง

ข. เป็นเท็จ

ถ้าเลือกตอบข้อ ก. ให้ตอบคำถามข้อที่ 20.2(1)

ถ้าเลือกตอบข้อ ข. ให้ตอบคำถามข้อที่ 20.2(2)

20.2(1) ถ้าข้อความนี้เป็นจริง เพราะเหตุผลในข้อใด

ก. $\overline{AG} = \overline{BH}$

ข. $\overline{EF} = \overline{DF}$

ค. $\angle AEG = \angle BDH$

ง. $\angle AEG = \angle BDH$

20.2(2) ถ้าข้อความนี้เป็นเท็จ เพราะขาดสมบัติในข้อใด

ก. $\overline{AE} = \overline{BD}$

ข. $\overline{AG} = \overline{BH}$

ค. $\angle AFB = \angle BFA$

ง. $\angle AEG = \angle BDH$

ภาคผนวก จ

ผลที่ได้จากการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วย
วิธีวิเคราะห์การจำแนกแบบโลจิสติก

Case Processing Summary

	N	Marginal Percentage
item1 0	241	23.6%
1	691	67.7%
2	89	8.7%
1 1	508	49.8%
2	513	50.2%
Valid	1021	100.0%
Missing	0	
Total	1021	
Subpopulation	2	

Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	23.721 ^a	.000	0	.
1 1	27.737	4.017	2	.134

Case Processing Summary

	N	Marginal Percentage
item2 0	519	50.8%
1	401	39.3%
2	101	9.9%
1 1	508	49.8%
2	513	50.2%
Valid	1021	100.0%
Missing	0	
Total	1021	
Subpopulation	2	

Likelihood Ratio Tests

	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	24.440 ^a	.000	0	.
item3	25.913	1.472	2	.479

Case Processing Summary

		N	Marginal Percentage
item3	0	472	46.2%
	1	431	42.2%
	2	118	11.6%
	1	508	49.8%
	2	513	50.2%
Valid		1021	100.0%
Missing		0	
Total		1021	
Subpopulation		2	

Likelihood Ratio Tests

	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	24.701 ^a	.000	0	.
item3	31.986	7.285	2	.026

Case Processing Summary

		N	Marginal Percentage
item4	0	260	25.5%
	1	584	57.2%
	2	177	17.3%
	1	508	49.8%
	2	513	50.2%
Valid		1021	100.0%
Missing		0	
Total		1021	
Subpopulation		2	

Likelihood Ratio Tests

	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	24.924 ^a	.000	0	.
	29.561	4.637	2	.098

Case Processing Summary

		N	Marginal Percentage
item5	0	191	18.7%
	1	685	67.1%
	2	145	14.2%
	1	508	49.8%
	2	513	50.2%
Valid		1021	100.0%
Missing		0	
Total		1021	
Subpopulation		2	

Likelihood Ratio Tests

Effect	Model Fitting	Likelihood Ratio Tests		
	Criteria			
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	24.243 ^a	.000	0	.
item6	24.850	.607	2	.738

Case Processing Summary

		N	Marginal Percentage
item6	0	459	45.0%
	1	387	37.9%
	2	175	17.1%
	1	508	49.8%
	2	513	50.2%
Valid		1021	100.0%
Missing		0	
Total		1021	
Subpopulation		2	

Likelihood Ratio Tests

Effect	Model Fitting	Likelihood Ratio Tests		
	Criteria			
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	25.231 ^a	.000	0	.
item6	25.747	.516	2	.773

Case Processing Summary

	N	Marginal Percentage
item7 0	167	16.4%
1	570	55.8%
2	284	27.8%
Valid 1	508	49.8%
2	513	50.2%
Valid	1021	100.0%
Missing	0	
Total	1021	
Subpopulation	2	

Likelihood Ratio Tests

Effect	Model Fitting	Likelihood Ratio Tests		
	Criteria			
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	24.932 ^a	.000	0	.
item7	29.928	4.996	2	.082

Case Processing Summary

	N	Marginal Percentage
item8 0	454	44.5%
1	567	55.5%
Valid 1	508	49.8%
2	513	50.2%
Valid	1021	100.0%
Missing	0	
Total	1021	
Subpopulation	2	

Likelihood Ratio Tests

	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	13.341 ^a	.000	0	.
1.1111	18.550	5.209	1	.022

Case Processing Summary

	N	Marginal Percentage
item9 0	271	26.5%
1	561	54.9%
2	189	18.5%
1.1111 1	508	49.8%
2	513	50.2%
Valid	1021	100.0%
Missing	0	
Total	1021	
Subpopulation	2	

Likelihood Ratio Tests

	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	25.056 ^a	.000	0	.
1.1111	28.939	3.883	2	.143

Case Processing Summary

		N	Marginal Percentage
item10	0	510	50.0%
	1	394	38.6%
	2	117	11.5%
	1	508	49.8%
	2	513	50.2%
Valid		1021	100.0%
Missing		0	
Total		1021	
Subpopulation		2	

Likelihood Ratio Tests

Effect	Model Fitting	Likelihood Ratio Tests		
	Criteria			
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	24.658 ^a	.000	0	.
	28.433	3.775	2	.151

Case Processing Summary

		N	Marginal Percentage
item11	0	529	51.8%
	1	366	35.8%
	2	126	12.3%
	1	508	49.8%
	2	513	50.2%
Valid		1021	100.0%
Missing		0	
Total		1021	
Subpopulation		2	

Likelihood Ratio Tests

	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	24.743 ^a	.000	0	.
item12	25.613	.870	2	.647

Case Processing Summary

		N	Marginal Percentage
item12	0	401	39.3%
	1	504	49.4%
	2	116	11.4%
	1	508	49.8%
	2	513	50.2%
Valid		1021	100.0%
Missing		0	
Total		1021	
Subpopulation		2	

Likelihood Ratio Tests

	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	24.668 ^a	.000	0	.
item12	25.111	.442	2	.802

Case Processing Summary

		N	Marginal Percentage
item13	0	411	40.3%
	1	517	50.6%
	2	93	9.1%
	1	508	49.8%
	2	513	50.2%
Valid		1021	100.0%
Missing		0	
Total		1021	
Subpopulation		2	

Likelihood Ratio Tests

Effect	Model Fitting	Likelihood Ratio Tests		
	Criteria			
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	24.327 ^a	.000	0	.
	25.627	1.300	2	.522

Case Processing Summary

		N	Marginal Percentage
item14	0	226	22.1%
	1	545	53.4%
	2	250	24.5%
	1	508	49.8%
	2	513	50.2%
Valid		1021	100.0%
Missing		0	
Total		1021	
Subpopulation		2	

Likelihood Ratio Tests

	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	25.199 ^a	.000	0	.
item15	28.633	3.433	2	.180

Case Processing Summary

		N	Marginal Percentage
item15	0	487	47.7%
	1	370	36.2%
	2	164	16.1%
	1	508	49.8%
	2	513	50.2%
Valid		1021	100.0%
Missing		0	
Total		1021	
Subpopulation		2	

Likelihood Ratio Tests

	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	25.127 ^a	.000	0	.
item15	27.065	1.938	2	.379

Case Processing Summary

		N	Marginal Percentage
item16	0	476	46.6%
	1	352	34.5%
	2	193	18.9%
	1	508	49.8%
	2	513	50.2%
Valid		1021	100.0%
Missing		0	
Total		1021	
Subpopulation		2	

Likelihood Ratio Tests

Effect	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	25.281 ^a	.000	0	.
	31.932	6.651	2	.036

Case Processing Summary

		N	Marginal Percentage
item17	0	321	31.4%
	1	564	55.2%
	2	136	13.3%
	1	508	49.8%
	2	513	50.2%
Valid		1021	100.0%
Missing		0	
Total		1021	
Subpopulation		2	

Likelihood Ratio Tests

	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	24.757 ^a	.000	0	.
item18	26.871	2.114	2	.348

Case Processing Summary

		N	Marginal Percentage
item18	0	330	32.3%
	1	498	48.8%
	2	193	18.9%
	1	508	49.8%
	2	513	50.2%
Valid		1021	100.0%
Missing		0	
Total		1021	
Subpopulation		2	

Likelihood Ratio Tests

	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	25.259 ^a	.000	0	.
item18	29.293	4.034	2	.133

Case Processing Summary

		N	Marginal Percentage
item19	0	307	30.1%
	1	571	55.9%
	2	143	14.0%
	1	508	49.8%
	2	513	50.2%
Valid		1021	100.0%
Missing		0	
Total		1021	
Subpopulation		2	

Likelihood Ratio Tests

Effect	Model Fitting	Likelihood Ratio Tests		
	Criteria			
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Intercept	24.796 ^a	.000	0	.
	26.057	1.261	2	.532

Case Processing Summary

		N	Marginal Percentage
item20	0	401	39.3%
	1	429	42.0%
	2	191	18.7%
	1	508	49.8%
	2	513	50.2%
Valid		1021	100.0%
Missing		0	
Total		1021	
Subpopulation		2	

Likelihood Ratio Tests

	Model Fitting Criteria	Likelihood Ratio Tests		
	-2 Log Likelihood of Reduced Model	Chi-Square	df	Sig.
Effect				
Intercept	25.286 ^a	.000	0	.
เพศ	38.571	13.285	2	.001



ภาคผนวก จ

ผลที่ได้จากการวิเคราะห์การทำหน้าที่ต่างกันของข้อสอบด้วย
วิธีแมนเทิล-แฮนส์เซลทั่วไป

Opened the text file: d:\score_MH.txt

Number of Cases: 1021

Number of Variables: 21

DIF analysis: Nonparametric tests for polytomous items

Stratifying variable: Sum of item responses

Stratum size: 1

Number of strata: 31

Number of reference group members: 508

Number of focal group members: 513

Grouping variable: Var1

Reference Value = 1, Focal Value = 2

DIF STATISTICS: POLYTOMOUS ITEMS

Name	Mantel	L-A LOR	LOR SE	LOR Z	COX'S B	COX SE	COX Z
Var 2	3.114	.239	.137	1.745	.211	.1197	1.763
Var 3	1.172	.138	.127	1.087	.112	.1036	1.081
Var 4	5.826	.295	.124	2.379	.241	.0997	2.417
Var 5	2.446	-.2	.127	-1.575	-.165	.1052	-1.568
Var 6	.985	.131	.134	.978	.116	.1169	.992
Var 7	.045	-.026	.122	-.213	-.02	.0923	-.217
Var 8	2.879	-.222	.131	-1.695	-.182	.1074	-1.695
Var 9	4.337	-.276	.132	-2.091	-.276	.1327	-2.08
Var 10	3.134	-.22	.126	-1.746	-.175	.0991	-1.766
Var 11	.124	.044	.126	.349	.034	.0952	.357
Var 12	.108	.042	.127	.331	.033	.099	.333
Var 13	.292	.068	.124	.548	.054	.1001	.539
Var 14	2.072	.185	.128	1.445	.157	.1093	1.436
Var 15	2.034	-.179	.127	-1.409	-.145	.102	-1.422
Var 16	2.219	.181	.122	1.484	.136	.091	1.495
Var 17	6.554	.316	.122	2.59	.225	.0878	2.563
Var 18	1.368	-.148	.128	-1.156	-.122	.1045	-1.167
Var 19	2.939	-.208	.122	-1.705	-.158	.0921	-1.716
Var 20	.216	-.058	.126	-.46	-.047	.1018	-.462
Var 21	1.594	-.151	.122	-1.238	-.113	.0893	-1.265

Reference Value = 1, Focal Value = 2

CONDITIONAL DIFFERENCES: Intervals of size 3

Lower	6	9	12	15	18	21	24	27	30	33
Upper	9	12	15	18	21	24	27	30	33	36.1
Var 2	.14	.08	.02	.07	-.03	.31	.11	1.5	.5	.
Var 3	-.48	.05	.04	.11	-.11	.09	.33	.8	.5	.
Var 4	.2	.04	.07	.07	.32	-.1	.16	.1	0	.
Var 5	-.3	.08	-.04	-.13	-.11	-.07	.03	.3	-2	.
Var 6	.3	.01	.06	-.11	.23	.32	.02	.3	0	.
Var 7	-.15	.01	0	.11	-.31	-.43	.35	.8	.5	.
Var 8	-.4	.1	-.03	-.15	-.18	.02	.06	0	0	.
Var 9	-.33	-.1	-.1	.04	-.05	-.2	.14	0	0	.
Var 10	.03	-.24	-.08	-.03	.02	.23	-.24	.1	1	.
Var 11	-.12	.03	-.01	.03	.25	-.16	.17	-.7	0	.
Var 12	-.16	-.2	.05	.19	-.09	-.22	-.14	-.6	-1	.
Var 13	.02	-.04	.04	.08	-.01	-.4	-.48	.9	2	.
Var 14	.22	.03	.04	.08	.11	-.21	.02	.2	0	.
Var 15	.2	-.01	-.03	-.11	-.21	.16	-.11	-.1	0	.
Var 16	.2	-.09	.08	.13	-.05	.72	-.1	-.9	0	.
Var 17	.13	0	.1	.17	.18	.18	.13	.5	0	.
Var 18	.21	.05	-.01	-.12	-.07	-.19	-.19	-.4	-1	.
Var 19	-.14	.02	-.06	-.14	.01	-.31	-.16	-.1	-1	.
Var 20	-.22	.13	.02	-.09	-.14	.38	-.19	-1.5	0	.
Var 21	.48	-.04	-.09	-.18	.24	-.04	.32	-1.4	0	.

ประวัติย่อผู้วิจัย



ประวัติย่อผู้วิจัย

ชื่อ ชื่อสกุล	นางสาวอัมมาภรณ์ ตันติดำรงกุล
วันเดือนปีเกิด	23 มีนาคม 2532
สถานที่เกิด	กรุงเทพฯ
สถานที่อยู่ปัจจุบัน	8/14 ถนนเทอดไท ตำบลบางยี่เรือ อำเภอธนบุรี จังหวัดกรุงเทพฯ 10600
ตำแหน่งหน้าที่การงาน	ครู คศ.1
สถานที่ทำงานปัจจุบัน	โรงเรียนวัดราชโอรส อำเภอจอมทอง จังหวัดกรุงเทพฯ

ประวัติการศึกษา

พ.ศ. 2550	ชั้นมัธยมศึกษาปีที่ 6 จากโรงเรียนสตรีวัดอัมพรสวรรค์
พ.ศ. 2555	ปริญญาการศึกษาบัณฑิต (กศ.บ.) สาขาวิชาคณิตศาสตร์ จาก มหาวิทยาลัยศรีนครินทรวิโรฒ
พ.ศ. 2560	ปริญญาการศึกษามหาบัณฑิต (กศ.ม.) สาขาวิชาการวิจัยและพัฒนาศักยภาพมนุษย์ (แขนงการทดสอบและวัดผลการศึกษา) จาก มหาวิทยาลัยศรีนครินทรวิโรฒ