

๑-๕ ๖

การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบ
วิธีปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา

ปริชญานิพนธ์

ของ

จิตสุดา ธารพร

14 ส.ค. 2540

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา

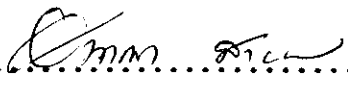
สิงหาคม 2539

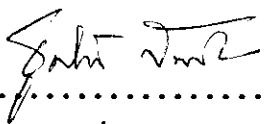
ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

๐๖๐ ๕๖๖๔

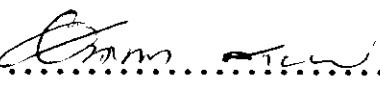
คณะกรรมการควบคุมและคณะกรรมการสอบได้พิจารณาปฏิญานพนธ์ฉบับนี้แล้ว
เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาการศึกษามหาบัณฑิต
วิชาเอกการวัดผลการศึกษา ของมหาวิทยาลัยศรีนครินทรวิโรฒได้

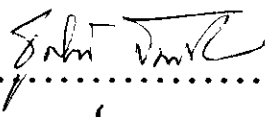
คณะกรรมการควบคุม

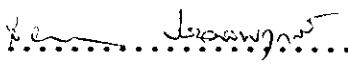
..........ประธาน
(รศ.อังคณา สายยศ)

..........กรรมการ
(ดร.ชูศักดิ์ ชัมภลชิต)

คณะกรรมการสอบ

..........ประธาน
(รศ.อังคณา สายยศ)

..........กรรมการ
(ดร.ชูศักดิ์ ชัมภลชิต)

..........กรรมการที่แต่งตั้งเพิ่มเติม
(รศ.ดร.ส. วาสนา ประवालพฤษ์)

บัณฑิตวิทยาลัยอนุมัติให้รับปฏิญานพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา ของมหาวิทยาลัยศรีนครินทรวิโรฒ

..........คณบดีบัณฑิตวิทยาลัย
(ดร.ศิริรุภา พูลสุวรรณ)

วันที่ 16 ..เดือน สิงหาคม พ.ศ. 2539

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้ สำเร็จลุล่วงได้ด้วยดี เพราะได้รับความอนุเคราะห์อย่างยิ่งจาก รองศาสตราจารย์อังคณา สาขยศ และ ดร.ชูศักดิ์ ชัมภลจิตต์ ประธานและกรรมการ ที่ปรึกษา ปริญญานิพนธ์ ที่ให้คำปรึกษาแนะนำและข้อคิดเห็นต่างๆ ตลอดเวลาที่ศึกษาและทำปริญญานิพนธ์ มาโดยตลอดจนสำเร็จ ผู้วิจัยกราบขอบพระคุณเป็นอย่างสูง

กราบขอบพระคุณ รองศาสตราจารย์ ดร.ส.วาสนา ประवालพฤกษ์ กรรมการสอบ ปริญญานิพนธ์ที่กรุณาให้ความรู้ และคำแนะนำเพิ่มเติม อันเป็นประโยชน์อย่างยิ่งสำหรับการทำ ปริญญานิพนธ์ในครั้งนี้ และทำให้ปริญญานิพนธ์ฉบับนี้สมบูรณ์ยิ่งขึ้น

กราบขอบพระคุณคณาจารย์ ภาควิชาวัดผลและวิจัยการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร ที่ได้ให้ความรู้ และการอบรมสั่งสอน

ขอขอบพระคุณ อาจารย์สมกิจ กิจพูนวงศ์ อาจารย์รุ่งฤดี ทรัพย์นิธิ อาจารย์ บุญเรือง ศรีเจริญ และ คุณเกศเกล้า หนูมีง ที่กรุณาให้ความอนุเคราะห์ ในการวิเคราะห์ข้อมูลด้วยเครื่อง คอมพิวเตอร์

ขอขอบพระคุณ ผู้เชี่ยวชาญทุกท่าน ที่ได้กรุณาตรวจสอบเครื่องมือในการวิจัย ตลอดจน ให้คำแนะนำและข้อคิดเห็นต่างๆ ในการแก้ไขข้อบกพร่องของเครื่องมือที่ใช้ในการศึกษาครั้งนี้

ขอขอบคุณ ผู้บริหารโรงเรียน ผู้ช่วยฝ่ายวิชาการคณะครู-อาจารย์ ของโรงเรียนที่เป็น กลุ่มตัวอย่าง และ ขอขอบคุณนักเรียนทุกคนที่ให้ความร่วมมือในการเก็บรวบรวมข้อมูลเป็นอย่างดี

ขอขอบพระคุณ อาจารย์ฉัณษา และ อาจารย์นพพร ๗ นคร ที่กรุณาช่วยอำนวยความสะดวกและให้ความช่วยเหลือในการเก็บรวบรวมข้อมูล และให้กำลังใจตลอดเวลาที่ศึกษา และใน การทำวิจัย และขอขอบพระคุณ ดร.เรวดี อินทสระระ อาจารย์ณลินี ๗ นคร ที่ให้ความรู้ความ ช่วยเหลือในการวิเคราะห์ข้อมูลตลอดจนให้คำแนะนำ และข้อคิดเห็นในด้านต่างๆ

ขอขอบคุณ นักวัดผลรุ่นพี่ๆ เพื่อนารุ่น 30 และ น้องๆ ผู้ซึ่งให้ความช่วยเหลือและให้กำลังใจ ด้วยความปรารถนาดี

กราบขอบพระคุณพ่อ แม่ ญาติพี่น้อง ที่ให้กำลังใจ กำลังทรัพย์ สนับสนุนทุนการศึกษาแก่ ผู้วิจัย ขอกราบขอบพระคุณ อาจารย์ระวีวรรณ พันธุ์พานิช ครู-อาจารย์ ที่ให้คำแนะนำ ให้การ อบรมสั่งสอน ตลอดจนให้ความช่วยเหลือสนับสนุนทางการศึกษาและให้กำลังใจแก่ผู้วิจัยตลอดมา

จิตส์ดา ธารพร

สารบัญ

บทที่	หน้า
1 บทนำ.....	1
ภูมิหลัง.....	1
ความมุ่งหมายของการศึกษาค้นคว้า.....	6
ความสำคัญของการศึกษาค้นคว้า.....	6
ขอบเขตของการศึกษาค้นคว้า.....	6
นิยามศัพท์เฉพาะ.....	7
2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	10
ความเป็นมาและวิธีการตรวจสอบความล่าเอียงของข้อสอบ.....	10
ความหมายของความล่าเอียงของข้อสอบ.....	15
สาเหตุที่ก่อให้เกิดความล่าเอียงของข้อสอบ.....	16
หลักการวิเคราะห์ความล่าเอียงของข้อสอบ.....	17
การวิเคราะห์ความล่าเอียงของข้อสอบโดยวิธีทฤษฎีการวัดแบบคลาสสิกอล.....	20
การวิเคราะห์ความล่าเอียงของข้อสอบโดยวิธีของแมนเทล-แฮนส์เซล.....	30
ทฤษฎีการตอบข้อคำถาม.....	36
การประยุกต์ทฤษฎีการตอบข้อคำถามเพื่อศึกษาความล่าเอียงของข้อสอบ.....	46
งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความล่าเอียงของข้อสอบ.....	61
สมมติฐานของการศึกษาค้นคว้า.....	79
3 วิธีดำเนินการศึกษาค้นคว้า.....	80
ประชากร.....	80
กลุ่มตัวอย่าง.....	80
เครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	82
วิธีดำเนินการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	83
ลักษณะของเครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	87

วิธีดำเนินการเก็บรวบรวมข้อมูล.....	89
สถิติที่ใช้ในการวิเคราะห์ข้อมูล.....	90
4 ผลการวิเคราะห์ข้อมูล.....	100
สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล.....	100
การเสนอผลการวิเคราะห์ข้อมูล.....	101
ผลการวิเคราะห์ข้อมูล.....	102
5 สรุปผล อภิปรายผล และข้อเสนอแนะ.....	129
ความมุ่งหมายของการศึกษาค้นคว้า.....	129
กลุ่มตัวอย่าง.....	129
เครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	129
วิธีดำเนินการเก็บรวบรวมข้อมูล.....	130
การเสนอผลการวิเคราะห์ข้อมูล.....	130
สรุปผลการวิเคราะห์ข้อมูล.....	131
อภิปรายผล.....	134
ข้อเสนอแนะในการวิจัย.....	137
บรรณานุกรม.....	139
ภาคผนวก.....	152
ประวัติย่อของผู้วิจัย.....	189

บัญชีตาราง

ตาราง		หน้า
1	ผลการสอบของกลุ่มตัวอย่างกลุ่ม R และ F ที่มีคะแนนรวมอยู่ในช่วงคะแนน j....	31
2	สัดส่วนการตอบข้อสอบของกลุ่มประชากร 2 กลุ่ม.....	32
3	จำนวนนักเรียนในกลุ่มตัวอย่างที่ใช้ในการวิจัย จำแนกตามโรงเรียน.....	81
4	ค่าสถิติพื้นฐานของค่าพารามิเตอร์ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษาของกลุ่มนักเรียนที่นับถือศาสนาพุทธ และศาสนาอิสลาม.....	102
5	ค่าดัชนีความลำเอียงของข้อสอบด้วยวิธีการแปลงค่าความยากของข้อสอบจากแบบ ทดสอบโจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษาระหว่าง กลุ่มนักเรียนที่นับถือศาสนาพุทธและศาสนาอิสลาม.....	106
6	ค่าดัชนีความลำเอียงของข้อสอบด้วยวิธีโค้งลักษณะข้อสอบจากแบบทดสอบโจทย์ปัญหา ทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษาระหว่างกลุ่มนักเรียนที่นับถือ ศาสนาพุทธและศาสนาอิสลาม.....	109
7	ค่าดัชนีความลำเอียงของข้อสอบด้วยวิธีของแมนเทล-แฮนส์เชล (α_{MH} ; χ^2_{MH}) จากแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทาง ภาษาระหว่างกลุ่มนักเรียนที่นับถือศาสนาพุทธและศาสนาอิสลาม.....	113
8	สรุปจำนวนข้อสอบที่มีความลำเอียงในแต่ละวิธีการวิเคราะห์ความลำเอียง ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจ ทางภาษา.....	117
9	ผลการทดสอบความแตกต่างของจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธี วิเคราะห์ความลำเอียงต่างกันของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์.....	119
10	ผลการทดสอบความแตกต่างของจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธี วิเคราะห์ความลำเอียงต่างกันของแบบทดสอบความเข้าใจทางภาษา.....	120

11	ค่าความเชื่อมั่นของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบ ความเข้าใจทางภาษา หลังจากคัดข้อสอบที่มีความลำเอียงออกด้วย วิธีแปลงค่าความยากของข้อสอบ(TID)วิธีโค้งลักษณะข้อสอบ(ICC) และวิธีของแมนเทล-แฮนส์เชล(MH)ซึ่งคำนวณโดยใช้สูตรแบบKR-20.....	121
12	ค่าความเชื่อมั่นของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบ ความเข้าใจทางภาษา หลังจากคัดข้อสอบที่มีความลำเอียงออกด้วย วิธีแปลงค่าความยากของข้อสอบ(TID)วิธีโค้งลักษณะข้อสอบ(ICC) และ วิธีของแมนเทล-แฮนส์เชล(MH)โดยใช้สูตรการคำนวณแบบแบ่งครั้งฉบับ.....	123
13	ผลการทดสอบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบโจทย์ปัญหา ทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา หลังจากคัดเลือก ข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโค้ง ลักษณะข้อสอบ และวิธีของแมนเทล-แฮนส์เชล ที่คำนวณโดยใช้สูตรแบบ KR-20.....	125
14	ผลการทดสอบความแตกต่างของค่าความเชื่อมั่นรายข้อของแบบทดสอบ โจทย์ปัญหาทางคณิตศาสตร์ หลังจากคัดข้อสอบที่มีความลำเอียงออก ด้วยวิธีวิเคราะห์ความลำเอียงที่แตกต่างกัน 3 วิธี.....	126
15	ผลการทดสอบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบโจทย์ปัญหา ทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา หลังจากคัดข้อสอบ ที่มีความลำเอียงออก ด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะ ข้อสอบ และวิธีของแมนเทล-แฮนส์เชล ที่คำนวณจากสูตรแบบ แบ่งครั้งฉบับ.....	127
16	ค่าความยาก(P)ค่าอำนาจจำแนก(r)เฉพาะตัวถูกของแบบทดสอบโจทย์ปัญหา ทางคณิตศาสตร์ จากการทดสอบครั้งที่ 1.....	156
17	ค่าความยาก(P)ค่าอำนาจจำแนก(r)เฉพาะตัวถูกของแบบทดสอบความ เข้าใจทางภาษา จากการทดสอบครั้งที่ 1.....	157

18	ค่าไอเกน เพอร์เซ็นต์ของความแปรปรวน และ เพอร์เซ็นต์ของความแปรปรวนสะสมจากการนำแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์มาวิเคราะห์องค์ประกอบ.....	159
19	ค่าไอเกน เพอร์เซ็นต์ของความแปรปรวน และ เพอร์เซ็นต์ของความแปรปรวนสะสมจากการนำแบบทดสอบความเข้าใจทางภาษามาวิเคราะห์องค์ประกอบ.....	160
20	ค่าพารามิเตอร์ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์.....	162
21	ค่าพารามิเตอร์ของแบบทดสอบความเข้าใจทางภาษา.....	164
22	ค่าเฉลี่ยของข้อสอบแต่ละข้อระหว่างกลุ่มนักเรียนที่นับถือศาสนาพุทธกับกลุ่มนักเรียนที่นับถือศาสนาอิสลามของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา.....	167

บัญชีภาพประกอบ

ภาพประกอบ	หน้า
1 แสดงกราฟ ของค่าเฉลี่ยของคนผิวดำกลุ่ม 1 และคนผิวดำกลุ่ม 2 จาก แบบทดสอบวัดความถนัดทางการเรียนด้านภาษา.....	25
2 แสดงกราฟ ของค่าเฉลี่ยของคนผิวขาวกลุ่ม 5 และคนผิวขาวกลุ่ม 6 จาก แบบทดสอบวัดความถนัดทางการเรียนด้านภาษา.....	26
3 แสดงกราฟ ของค่าเฉลี่ยของคนผิวขาวกลุ่ม 5 และคนผิวดำกลุ่ม 1 จาก แบบทดสอบวัดความถนัดทางการเรียนด้านภาษา.....	27
4 แสดงเส้นโค้งลักษณะข้อสอบของนอร์มัล ออร์โรว์ โมเดล (Normal Ogive Model).....	40
5 แสดงโค้งลักษณะข้อสอบ ซึ่งมีลักษณะไม่เปลี่ยนแปลงตามจำนวนและ ระดับความสามารถของผู้สอบ.....	48
6 แสดงโค้งลักษณะข้อสอบ ข้อ 2 ระหว่างกลุ่มนักเรียนผิวขาวและผิวดำ.....	49
7 แสดงโค้งลักษณะข้อสอบ ข้อ 24 ระหว่างกลุ่มนักเรียนผิวขาวและผิวดำ.....	50
8 แสดงโค้งลักษณะข้อสอบ ข้อ 71 ระหว่างกลุ่มนักเรียนผิวขาวและผิวดำ.....	51
9 แสดงโค้งลักษณะข้อสอบระหว่างกลุ่ม R และกลุ่ม F.....	52
10 แสดงโค้งลักษณะข้อสอบระหว่างกลุ่ม R และกลุ่ม F.....	53
11 แสดงลำดับขั้นในการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	83
12 แสดงการกระจายของคู่อันดับค่าเฉลยระหว่างกลุ่มนักเรียนที่นับถือศาสนาพุทธ กับกลุ่มนักเรียนที่นับถือศาสนาอิสลามของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ จำนวน 40 ข้อ.....	169
13 แสดงการกระจายของคู่อันดับค่าเฉลยระหว่างกลุ่มนักเรียนที่นับถือศาสนาพุทธ กับกลุ่มนักเรียนที่นับถือศาสนาอิสลามของแบบทดสอบความเข้าใจทางภาษา จำนวน 40 ข้อ.....	170

บทที่ 1

บทนำ

ภูมิหลัง

การวัดผลการศึกษาเป็นองค์ประกอบที่สำคัญอย่างหนึ่ง ของการจัดการศึกษาที่เกิดขึ้นมาพร้อมๆ กับการอบรมสั่งสอนในทุกสถานที่และทุกสมัยในทุกระดับ เพราะผลจากการวัดทำให้เราทราบว่าจัดการศึกษาประสบความสำเร็จมากน้อยเพียงใด และยังเป็นพื้นฐานในการตัดสินใจของครูและนักการศึกษา ที่จะใช้ปรับปรุงการเรียนการสอน การแนะแนว การประเมินผลหลักสูตรแบบเรียน ตลอดจนการจัดระบบการบริหารทั่วไปของโรงเรียน (อนันต์ ศรีโสภณ. 2524 : 1) นอกจากนี้ ความรู้เกี่ยวกับการวัดผลและประเมินผลยังได้นำมาใช้ภายในโรงเรียนอยู่ตลอดเวลา เพื่อช่วยในการตัดสินใจเกี่ยวกับการให้การศึกษาแก่นักเรียนเป็นไปอย่างถูกต้องและเชื่อถือได้ (สมเกียรติ บัณฑิตพร. 2525 : 1)

การวัดผลทางการศึกษาเป็นการวัดความสามารถหรือคุณลักษณะต่างๆ ที่เป็นคุณลักษณะภายใน (traits) ของบุคคลที่เป็นนามธรรมและมีองค์ประกอบที่สลับซับซ้อนไม่สามารถสังเกตและวัดโดยตรงได้ ต้องใช้วิธีการวัดทางอ้อม ซึ่งจะวัดสิ่งที่มีความสัมพันธ์เชิงปริมาณกับสิ่งที่ต้องการจะวัดแล้วนำผลที่ได้จากการวัดอ้างอิง (infer) หรืออธิบายกลับไปสู่สิ่งที่ต้องการวัดจริง หรือทำนายความสามารถที่แท้จริง (True ability or, True score) อันเป็นคุณลักษณะภายใน (Latent Traits) ของบุคคลที่ถูกวัด (Lord. 1980 : 3) การวัดผลการศึกษาจะมีคุณภาพหรือไม่ หรือสามารถวัดได้ครอบคลุมเนื้อหาและความเป็นจริงเพียงใด สัมพันธ์อยู่กับวิธีการวัดและเครื่องมือที่ใช้ในการวัดเป็นสำคัญ ถ้าเครื่องมือที่ใช้มีคุณภาพดีก็จะทำให้ได้ข้อมูลที่ถูกต้องน่าเชื่อถือ ซึ่งจะส่งผลให้การลงสรุปและการตัดสินใจมีความถูกต้องและมีประสิทธิภาพ การวัดผลการศึกษานั้นสามารถกระทำได้หลายวิธี แต่ที่นิยมใช้กันมากที่สุดในโรงเรียน คือการทดสอบ (ชวาล แพรัตกุล. 2516 : 87-88) ซึ่งเครื่องมือที่นิยมใช้ในการทดสอบคือแบบทดสอบ ดังนั้น แบบทดสอบที่นำมาใช้ในการวัดจะต้องมีคุณภาพที่เชื่อถือได้ มีความคลาดเคลื่อนในการวัดน้อยที่สุดและวัดได้ตรงความสามารถที่แท้จริงกับสิ่งที่ต้องการวัด (อวยพร วิบูลย์กาญจน์. 2526 : 7) แบบทดสอบที่ใช้กันมากที่สุดในการวัดผลสัมฤทธิ์ทางการเรียนนั้นเป็นแบบปรนัยชนิดเลือกตอบ (Multiple-Choice items) เนื่องจากเป็นแบบทดสอบที่ออกได้ครอบคลุมเนื้อหา สามารถวัดสมรรถภาพสมองด้านต่างๆ ได้อย่าง

เหมาะสม สะดวกในการดำเนินการสอบและการตรวจให้คะแนน ครูและนักการศึกษาส่วนใหญ่มีความเข้าใจและคุ้นเคยกับเครื่องมือประเภทนี้ และแบบทดสอบชนิดนี้จะให้ความยุติธรรมแก่นักเรียนและอาจให้ผลการวัดได้ดีกว่าแบบทดสอบชนิดอื่น ๆ (ชวาล แพริตกุล. 2516 : 150) จากความสำคัญของแบบทดสอบดังกล่าว จึงมีการสร้างและพัฒนาแบบทดสอบขึ้นเพื่อใช้ในโอกาสต่าง ๆ กัน โดยจำแนกชนิดของแบบทดสอบออกตามเนื้อหาที่จะวัด และเน้นหนักคุณภาพของแบบทดสอบทางด้านความยุติธรรม (Fairness) ความเชื่อมั่น (Reliability) ความเที่ยงตรง (Validity) ความยาก และอำนาจจำแนก นอกจากนี้จะต้องระมัดระวังไม่ให้เกิดความลำเอียงหรือการได้เปรียบเสียเปรียบของกลุ่มผู้สอบบางกลุ่ม ไม่ว่าจะเป็น นักเรียนเก่ง ปานกลาง อ่อน หรือเชื้อชาติ สังคม วัฒนธรรม และพื้นที่ใดๆ ก็ตาม (Intasuwat. 1979 : 1 ; Citing Green. 1972) ซึ่งคุณลักษณะของข้อสอบดังกล่าวนี้ คือ ข้อสอบจะต้องปราศจากความลำเอียง เพราะความลำเอียงของข้อสอบ เป็นคุณลักษณะที่สำคัญที่มีผลต่อคุณภาพของแบบทดสอบ ดังนั้น การพัฒนาแบบทดสอบให้มีคุณภาพดีที่น่าเชื่อถือ และมีความยุติธรรมกับกลุ่มผู้สอบที่มีจำนวนมากและมีความแตกต่างกันดังกล่าว ควรจะมีการวิเคราะห์ความลำเอียงของข้อสอบก่อนที่จะนำไปใช้ในการสอบแต่ละครั้ง

การวิเคราะห์ความลำเอียงของข้อสอบเท่าที่ผ่านมาได้รับความสนใจจากนักการศึกษากันมากขึ้น ซึ่งในปัจจุบันการวิเคราะห์ความลำเอียงของข้อสอบระหว่างผู้สอบแบบกลุ่มย่อยนี้ เป็นสิ่งที่ปฏิบัติกันมา ถือเป็นเสมือนขั้นตอนของการพัฒนาแบบทดสอบโดยทั่วไป แบบทดสอบทุกฉบับไม่ว่าจะเป็นแบบทดสอบวัดผลสัมฤทธิ์ แบบทดสอบวัดความถนัด รวมทั้งแบบทดสอบวัดทักษะปฏิบัติ ควรได้รับการวิเคราะห์ความลำเอียงก่อนที่จะนำไปใช้ในการสอบ (Berk. 1982 : 3) โดยเฉพาะแบบทดสอบที่นำไปใช้สอบกับกลุ่มผู้สอบกลุ่มย่อย ๆ จำนวนมากที่มีภูมิหลังทางวัฒนธรรม สภาพเศรษฐกิจ เพศ สังคม ภาษา และเชื้อชาติ แตกต่างกัน จากการศึกษาของสำนักงานคณะกรรมการการศึกษาแห่งชาติ พบว่า ความแตกต่างทางวัฒนธรรมของครอบครัวอันเป็นผลมาจากฐานะทางเศรษฐกิจและสังคม มีผลทำให้ผลสัมฤทธิ์ทางการเรียนของนักเรียนต่างกัน ทั้งนี้อาจจะเป็นเพราะนักเรียนมีเจตคติ และความถนัดแตกต่างกันตามพื้นฐานทางวัฒนธรรมของตนเอง (สำนักงานคณะกรรมการการศึกษาแห่งชาติ. 2524 : 78) ซึ่งสอดคล้องกับงานวิจัยของ พิชรี ปิยภัณท์ (2531 : 65-67) ที่ได้ทำการวิเคราะห์ความลำเอียงของแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ พบว่าข้อสอบบางข้อมีความลำเอียงต่อกลุ่มเพศใดเพศหนึ่งและกลุ่มที่อยู่ในสภาพแวดล้อมที่แตกต่างกัน และยังสอดคล้องกับงานวิจัยของ ชัยชัย เผ่าพงษ์ (2527 : 73-76) ที่ได้ทำการศึกษาความลำเอียง

ของแบบทดสอบวัดความถนัดทางคณิตศาสตร์และภาษา พบว่า ข้อสอบบางข้อมีความลำเอียงต่อกลุ่มนักเรียนเพศใดเพศหนึ่งโดยเฉพาะเช่นเดียวกัน

สำหรับงานวิจัยในต่างประเทศมีการศึกษากันหลายท่านเช่นกัน เช่น งานวิจัยของ กรีน (Intasuwan. 1979 :1 ; Citing Green.1972) พบว่า กลุ่มของข้อสอบซึ่งถูกเลือกมาจากคลังข้อสอบเดียวกันของแบบทดสอบวัดผลสัมฤทธิ์แคลิฟอร์เนีย (California Achievement Test) ปี 1970 เมื่อนำไปทดสอบกับกลุ่มตัวอย่างที่มีความแตกต่างกันทางด้าน เชื้อชาติ ฐานะทางเศรษฐกิจ สังคม และสภาพภูมิศาสตร์ จะให้ผลการทดสอบที่แตกต่างกัน ซึ่งผลการทดสอบที่แตกต่างกันระหว่างกลุ่มผู้สอบที่มีความแตกต่างกันทางด้านเชื้อชาติ วัฒนธรรมนั้น มีอิทธิพลมาจากสิ่งแวดล้อม ความเคยชินที่แตกต่างกัน ข้อสอบประเภทนี้เป็นข้อสอบที่มีความลำเอียง (Item Bias) และจากผลงานวิจัยของ เชปาร์ด (Shepard. 1982 : 11) พบว่า เมื่อนำแบบทดสอบการคิดคำนวณทางคณิตศาสตร์ไปทดสอบกับนักเรียนผิวขาว และนักเรียนผิวดำในระดับมัธยมศึกษา มีข้อความของ โจทย์ปัญหาบางข้อความที่ผู้สอบทั้งสองกลุ่มเข้าใจไม่ตรงกัน คือ ลำเอียงต่อกลุ่มใดกลุ่มหนึ่งแต่เมื่อเปลี่ยนจากการถามด้วยแบบทดสอบมาเป็นการถามโดยผู้ดำเนินการสอบเองผู้สอบมีความเข้าใจในเนื้อหาหรือข้อความถามได้ตรงกัน นอกจากนี้ งานวิจัยของ สกบ ลักษณะ (Burrill. 1982 : 168 ; Citing Laksana. 1979) ได้ศึกษาถึงความลำเอียงทางเพศและผิว ของแบบทดสอบทางด้านภาษาและคณิตศาสตร์ ก็พบ ความลำเอียงของข้อสอบจากแบบทดสอบทางด้านภาษาและคณิตศาสตร์จำนวนหนึ่ง ระหว่างกลุ่มนักเรียนชายกับนักเรียนหญิง และนักเรียนผิวดำกับนักเรียนผิวขาว ทั้งนี้เนื่องจากข้อสอบที่ใช้ เช่น ข้อสอบวิชาคณิตศาสตร์จำเป็นต้องอาศัยความสามารถด้านการอ่านโจทย์ปัญหา ถ้ากลุ่มเพศชายหรือเพศหญิง และกลุ่มผิวดำหรือผิวขาวมีความสามารถทางการอ่านไม่เท่ากัน ข้อสอบข้อนั้นย่อมมีความลำเอียงทางการทดสอบ หรือข้อสอบที่ใช้วัดความสามารถทางภาษา ถ้าเพศชายหรือหญิง และนักเรียนผิวดำหรือผิวขาว มีการได้เปรียบเสียเปรียบเกี่ยวกับภาษาที่ใช้แตกต่างกันย่อมทำให้มีความลำเอียงทางการทดสอบได้เช่นกัน

จากผลการวิจัยดังกล่าว แสดงให้เห็นว่า แบบทดสอบวัดความสามารถด้านต่าง ๆ มีข้อสอบบางข้อที่มีความลำเอียง เมื่อนำไปทดสอบกับกลุ่มที่มีความแตกต่างกันทางด้าน เพศ เชื้อชาติ ฐานะทางเศรษฐกิจ สังคม วัฒนธรรม ศาสนา และสภาพภูมิศาสตร์ จะทำให้มีผลการทดสอบที่แตกต่างกัน ทั้งนี้เพราะระหว่างกลุ่มผู้สอบนั้นมีอิทธิพลมาจากสิ่งแวดล้อม วัฒนธรรม และความเคยชินที่ต่างกัน ฉะนั้นในการสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน ข้อสอบบางข้อจึงอาจมีความ

ค่าเฉลี่ยกับกลุ่มบางกลุ่มที่มีลักษณะแตกต่างกันดังกล่าวข้างต้น จึงอาจจะเป็นเหตุให้มีการได้เปรียบเสียเปรียบเกี่ยวกับเนื้อหาวิชาของข้อสอบ (Popham. 1981 : 181)

จากการศึกษาเปรียบเทียบการวิเคราะห์ความลำเอียงของข้อสอบที่ผ่านมาจนถึงปัจจุบัน มีวิธีวิเคราะห์หลายวิธีแต่วิธีที่นิยมใช้กันมากในการเปรียบเทียบ ได้แก่ วิธีแปลงค่าความยาก (Transformed Item Difficulty) การวิเคราะห์ความแปรปรวน (Analysis of Variance) การวิเคราะห์ด้วย ไคสแควร์ (Chi-Square) วิธีวิเคราะห์องค์ประกอบ (Factor Analysis) วิธีวิเคราะห์การถดถอย (Regression Analysis) วิธีล็อกลิเนียร์ (Log-Linear) วิธีถดถอยโลจิสติก (Logistic Regression) วิธีค่าอำนาจจำแนกของข้อสอบ (Item Discrimination Indices) วิธีของแมนเทิล-แฮนส์เซล (Mantel-Haenszel) ซึ่งเป็นวิธีที่ใช้หลักการของทฤษฎีการวัดแบบคลาสสิกอล (Classical Test Theory) นอกจากวิธีวิเคราะห์ความลำเอียงดังกล่าวแล้ว ยังมีวิธีวิเคราะห์ความลำเอียงโดยใช้หลักการของทฤษฎีการตอบข้อคำถาม (Item Response Theory) ได้แก่ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ (Item Characteristic Curve-3 Parameter) วิธีโค้งลักษณะข้อสอบ 2 พารามิเตอร์ (Item Characteristic Curve-2 Parameter) และวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ (Item Characteristic Curve-1 Parameter) ซึ่งวิธีการวิเคราะห์ความลำเอียงของข้อสอบแต่ละวิธีที่กล่าวมามีจุดเด่นและจุดด้อยแตกต่างกัน มีนักวิจัยหลายท่านได้ทำการศึกษาเปรียบเทียบวิธีการวิเคราะห์ความลำเอียงของข้อสอบแต่ละวิธี เช่น รัตเนอร์และคนอื่นๆ (Subkoviak 1984 ; 5 ; Citing Rudner and Others. 1978) ได้ศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบ 4 วิธี คือ วิธีแปลงค่าความยาก วิธีไคสแควร์ วิธีโค้งลักษณะข้อสอบ และวิธีวิเคราะห์องค์ประกอบ พบว่า วิธีโค้งลักษณะข้อสอบ และวิธีแปลงค่าความยาก เป็นวิธีที่เหมาะสมที่สุดสอง ลักษณะ (Burrill. 1982 : 168 ; Citing Laksana. 1979) ศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียง 2 วิธี คือ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และวิธีวิเคราะห์ความแปรปรวน พบว่า วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ เป็นวิธีที่ให้ผลครอบคลุม แต่วิธีวิเคราะห์ความแปรปรวนใช้ได้ดีในทางปฏิบัติ เชปาร์ดและคณะ (Berk. 1982 : 169 ; Citing Shepard, Camilli and Averrill. 1980) ได้ทำการศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียง 5 วิธี คือ วิธีไคสแควร์ วิธีแปลงค่าความยาก วิธีพอยท์ไบสรีเรียล วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ พบว่า วิธีโค้งลักษณะข้อสอบ

3 พารามิเตอร์ เป็นวิธีที่เหมาะสมในทางทฤษฎี ถูกต้องและตรงที่สุด สำหรับวิธีไคสแควร์ให้ผลสอดคล้องกับวิธีไคงลักษณะข้อสอบ 3 พารามิเตอร์ ส่วนสวามินาธานและโรเจอร์ส (Swaminathan and Rogers.1990) ศึกษาความลำเอียงของข้อสอบด้วยวิธีการถดถอยแบบโลจิสต์และวิธีของแมนเทล-แฮนส์เซล พบว่า ทั้งสองวิธีให้ผลใกล้เคียงกันและสามารถใช้แทนกันได้

ในประเทศไทยก็ได้มีการศึกษาเปรียบเทียบวิธีการวิเคราะห์ความลำเอียงของข้อสอบหลายท่านเช่นกัน เช่น งานวิจัยของ ทศนีย์ พิรมนตรี (2530 : 76-78) ได้ศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบ 3 วิธี คือวิธีกำหนดจุดเฉดตา วิธีไคสแควร์ และวิธีไคงลักษณะข้อสอบ 3 พารามิเตอร์ พบว่า วิธีไคงลักษณะข้อสอบ 3 พารามิเตอร์ พบข้อสอบมีความลำเอียงมากที่สุด สุรศักดิ์ อมรรัตนศักดิ์ (2531 : 200-207) ได้ทำการศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบ 4 วิธี คือ วิธีวิเคราะห์ความแปรปรวน วิธีแปลงค่าความยาก วิธีไคงลักษณะข้อสอบ 3 พารามิเตอร์ และ วิธีไคงลักษณะข้อสอบ 1 พารามิเตอร์ พบว่า วิธีไคงลักษณะข้อสอบ 3 พารามิเตอร์ พบข้อสอบที่มีความลำเอียงมากที่สุด รองลงมา เป็นวิธีวิเคราะห์ความแปรปรวน และวิธีไคสแควร์ พบข้อสอบที่มีความลำเอียงน้อยที่สุด

ผลการวิจัยดังกล่าว จะเห็นว่าเป็นการศึกษาเปรียบเทียบวิธีการวิเคราะห์ความลำเอียงของข้อสอบโดยยึดทฤษฎีการวัดแบบคลาสสิกอล และทฤษฎีการตอบข้อคำถาม แต่ผลการวิจัยส่วนมากพบว่า วิธีไคงลักษณะข้อสอบ 3 พารามิเตอร์ เป็นวิธีที่ให้ผลการวิเคราะห์ถูกต้องและตรงที่สุด เพราะสามารถวิเคราะห์และอธิบายโอกาสในการตอบข้อสอบถูกของผู้สอบที่มีความสามารถต่ำมากๆ ได้ รองลงมาเป็นวิธีไคงลักษณะข้อสอบ 1 พารามิเตอร์ สำหรับวิธีแปลงค่าความยาก วิธีไคสแควร์ วิธีวิเคราะห์ความแปรปรวน และวิธีของแมนเทล-แฮนส์เซลนั้น แม้จะไม่เหมาะสมในเชิงทฤษฎีมากนัก แต่เป็นวิธีการที่ใช้ได้ดีในทางปฏิบัติ ง่ายต่อการคำนวณและการแปลผล และจากการศึกษารายงานผลการวิจัยที่ผ่านมา ก็ยังไม่สามารถระบุได้ว่าวิธีวิเคราะห์ความลำเอียงของข้อสอบวิธีใดดีที่สุด และให้ผลการวิเคราะห์ถูกต้องที่สุด อีกทั้งยังไม่มีการศึกษาว่า ศาสนา วัฒนธรรม และภาษา จะส่งผลทำให้เกิดความลำเอียงของข้อสอบขึ้นหรือไม่ และจากการศึกษารายงานการวิจัยก็ยังไม่ปรากฏว่ามีการศึกษาความลำเอียงของข้อสอบวัดผลสัมฤทธิ์ในกลุ่มผู้สอบย่อยที่นับถือศาสนาต่างกันว่าจะให้ผลอย่างไร

ด้วยเหตุผลดังกล่าวข้างต้น ผู้วิจัยจึงสนใจที่จะศึกษาเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบ โดยใช้วิธีวิเคราะห์ 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีไคง

ลักษณะข้อสอบ 3 พารามิเตอร์ และวิธีของแมนเทิล-แฮนส์เฟล ระหว่างกลุ่มนักเรียนที่นับถือศาสนาต่างกัน คือ กลุ่มนักเรียนที่นับถือศาสนาพุทธกับกลุ่มนักเรียนที่นับถือศาสนาอิสลาม ในด้านจำนวนข้อสอบที่มีความลำเอียง และค่าความเชื่อมั่นของแบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกแล้ว จะให้ผลการวิเคราะห์ความลำเอียงแตกต่างกันหรือสอดคล้องกันเพียงไร อันจะเป็นแนวทางในการพัฒนาการนำเอาวิธีการวิเคราะห์ความลำเอียงมาใช้ในทางปฏิบัติ ตลอดจนการปรับปรุงแก้ไขข้อสอบที่มีความลำเอียงให้สามารถนำไปใช้ได้อย่างยุติธรรมต่อไป

๑. ความมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียงต่างกัน ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา
2. เพื่อเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังการคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วระหว่างวิธีวิเคราะห์ความลำเอียงต่างกัน ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา

๒. ความสำคัญของการศึกษาค้นคว้า

ผลจากการศึกษาค้นคว้าครั้งนี้ ทำให้ทราบว่าวิธีวิเคราะห์ความลำเอียงของข้อสอบวิธีใดสามารถค้นพบจำนวนข้อสอบที่มีความลำเอียงได้มากที่สุด และทราบถึงค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว สำหรับใช้เป็นแนวทางในการตัดสินใจเลือกวิธีวิเคราะห์ความลำเอียงที่เหมาะสมมาใช้ในการพัฒนาแบบทดสอบ ซึ่งจะทำได้แบบทดสอบที่มีคุณภาพสามารถนำไปทดสอบกับนักเรียนในกลุ่มต่าง ๆ ได้อย่างมีความยุติธรรมและมีประสิทธิภาพต่อไป

ขอบเขตของการศึกษาค้นคว้า

1. ประชากรที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นนักเรียนชายและนักเรียนหญิงชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษา ในจังหวัดกระบี่ จำนวน 16 โรงเรียน 94 ห้องเรียน มีจำนวนนักเรียนทั้งหมด 4,028 คน
2. กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนชายและนักเรียนหญิง ที่นับถือศาสนาพุทธและศาสนาอิสลาม ชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2538 ของ

โรงเรียนสังกัดกรมสามัญศึกษา ในจังหวัดกระบี่ ซึ่งเลือกมาโดยการสุ่มอย่างง่าย (Simple Random Sampling) โดยมีห้องเรียนเป็นหน่วยการสุ่ม (Sampling Unit) จำนวน 58 ห้องเรียน มีจำนวนนักเรียน 2,689 คน กลุ่มตัวอย่างแบ่งเป็น 2 กลุ่ม คือ

2.1 กลุ่มตัวอย่างที่ใช้ในการศึกษาเปรียบเทียบ ผลการวิเคราะห์ความลำเอียงด้านจำนวนข้อสอบที่มีความลำเอียง จำนวน 1,936 คน เป็นนักเรียนที่นับถือศาสนาพุทธ จำนวน 1,026 คน และเป็นนักเรียนที่นับถือศาสนาอิสลาม จำนวน 910 คน

2.2 กลุ่มตัวอย่างที่ใช้ในการศึกษาเปรียบเทียบ ผลการวิเคราะห์ความลำเอียงด้านความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว จำนวน 753 คน

3. ตัวแปรที่ศึกษา

3.1 ตัวแปรอิสระ ได้แก่ วิธีวิเคราะห์ความลำเอียงที่แตกต่างกัน 3 วิธี

3.1.1 วิธีแปลงค่าความยากของข้อสอบ

3.1.2 วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์

3.1.3 วิธีของแมนเทล-แฮนส์เชล

3.2 ตัวแปรตาม ได้แก่

3.2.1 จำนวนข้อสอบที่มีความลำเอียง

3.2.2 ความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว

4. เนื้อหาวิชาที่สอบ

เนื้อหาวิชาคณิตศาสตร์ (ค 204) ชั้นมัธยมศึกษาปีที่ 2 เรื่อง สมการและอสมการ อัตราส่วนและร้อยละ ปริมาตรและพื้นที่ผิว ตามหลักสูตรมัธยมศึกษาตอนต้น พุทธศักราช 2521 (ฉบับปรับปรุง พ.ศ. 2533) ของสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) และแบบทดสอบความเข้าใจทางภาษา ชั้นมัธยมศึกษาปีที่ 2

นิยามศัพท์เฉพาะ

1. แบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน หมายถึง แบบทดสอบที่วัดความรู้ ทักษะ และสมรรถภาพทางสมองด้านต่าง ๆ ที่ผู้เรียนได้เรียนไปแล้วตามที่กำหนด

2. แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ หมายถึง ข้อสอบที่วัดความสามารถทางด้านการแก้ปัญหา ซึ่งแบบทดสอบประเภทนี้ประกอบด้วยคำถามประเภทโจทย์ต่าง ๆ ที่มีภาษาเข้ามา

เกี่ยวข้องกับงานองเดียวกับโจทย์คณิตศาสตร์ทั่วไป แต่ลักษณะของคำถามมุ่งที่จะวัดความสามารถในการแปลความหมาย และการหาความสัมพันธ์ของตัวเลข จำนวน และการใช้เทคนิค หรือวิธีการแก้ปัญหาต่าง ๆ

3. แบบทดสอบความเข้าใจทางภาษา หมายถึง ข้อสอบที่วัดความสามารถในการแปลความ ตีความ ขยายความ และการจับใจความสำคัญหรือสรุปสาระสำคัญ ตลอดจนใช้เหตุผลพิจารณาตัดสินใจเพื่อลงสรุปในสถานการณ์ที่กำหนดให้ได้

4. ความลำเอียงของข้อสอบ หมายถึง ข้อสอบที่ทำให้ผลการสอบของบุคคลที่มีลักษณะต่างกัน แต่มีความสามารถเท่ากันได้ผลการสอบต่างกัน ในการวิจัยครั้งนี้ศึกษาความลำเอียงของกลุ่มตัวอย่างที่นับถือศาสนาต่างกัน

5. การวิเคราะห์ความลำเอียง หมายถึง การหาค่าดัชนีความลำเอียงของข้อสอบในแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา สำหรับการวิจัยครั้งนี้ใช้ตัวแปรที่วัดความลำเอียง คือ ศาสนา ซึ่งจำแนกเป็นศาสนาพุทธ และศาสนาอิสลาม

6. วิเคราะห์ความลำเอียงของข้อสอบ หมายถึง การหาค่าดัชนีความลำเอียงของข้อสอบ ซึ่งในที่นี้ใช้วิเคราะห์ 3 วิธี ดังนี้

6.1 วิธีแปลงค่าความยากของข้อสอบ (Transformed Item Difficulties Approach) หมายถึง การหาค่าดัชนีความลำเอียงของข้อสอบโดยพิจารณาค่าระยะห่างตั้งฉากจากคู่อันดับค่าเฉลยของข้อสอบแต่ละข้อไปยังเส้นแกนหลัก ข้อสอบที่มีความลำเอียงจะต้องมีค่าระยะห่างตั้งฉากมากกว่า $3S_d$ หรือน้อยกว่า $-3S_d$

6.2 วิธีโค้งลักษณะข้อสอบ (Item Characteristic Curve Method) หมายถึง การวิเคราะห์หาค่าดัชนีความลำเอียงของข้อสอบโดยหาค่าพารามิเตอร์ 3 ตัว ซึ่งคำนวณโดยใช้โปรแกรม BILOG3 แล้วสร้างเป็นโค้งลักษณะข้อสอบ แล้วดูความแตกต่างของพื้นที่ ที่คำนวณได้จากโค้งลักษณะข้อสอบของกลุ่มผู้สอบที่นับถือศาสนาพุทธกับกลุ่มผู้สอบที่นับถือศาสนาอิสลาม ข้อสอบที่มีความลำเอียงจะต้องมีพื้นที่ใต้โค้งต่างกันตั้งแต่ .40 ขึ้นไป

6.2.1 โค้งลักษณะข้อสอบ หมายถึง เส้นกราฟของข้อสอบซึ่งแสดงความสัมพันธ์ระหว่างระดับความสามารถของผู้สอบกับค่าความน่าจะเป็นของการตอบข้อสอบถูกต้องหนึ่งข้อ โค้งลักษณะข้อสอบนี้จะถูกกำหนดด้วยพารามิเตอร์ของข้อสอบ

6.2.2 ค่าพารามิเตอร์ของข้อสอบ หมายถึง ค่าความยากของข้อสอบ ค่าอำนาจจำแนกของข้อสอบ และค่าสัมประสิทธิ์การเดา ซึ่งคำนวณโดยใช้โปรแกรม BILOG3

6.3 วิธีของแมนเทล-แฮนส์เชล หมายถึง การวิเคราะห์หาดัชนีความล่าเอียงของข้อสอบ โดยพิจารณาเปรียบเทียบผลการตอบข้อสอบของกลุ่มผู้สอบ 2 กลุ่ม ซึ่งข้อสอบที่มีความล่าเอียง จะต้องมียค่า α_{HH} มากกว่า 1 หรือ χ^2_{HH} ที่คำนวณได้มีนัยสำคัญทางสถิติที่ระดับ .05

7. กลุ่มอ้างอิง (Reference Group : R) หมายถึง กลุ่มผู้สอบที่คาดว่าจะจะเป็นกลุ่มที่ได้ประโยชน์จากข้อสอบที่ล่าเอียง คือ เป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบที่ล่าเอียงได้ถูกต้องมากกว่าผู้สอบอีกกลุ่มหนึ่งที่ได้คะแนนจากแบบทดสอบเท่ากัน สำหรับกลุ่มอ้างอิงในการวิจัยครั้งนี้ คือ กลุ่มผู้สอบที่นับถือศาสนาพุทธ

8. กลุ่มเปรียบเทียบ (Focal Group : F) หมายถึง กลุ่มผู้สอบที่คาดว่าจะจะเป็นกลุ่มที่เสียประโยชน์จากข้อสอบที่ล่าเอียง คือ เป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบที่ล่าเอียงได้ถูกต้องต่ำกว่าผู้สอบกลุ่มอ้างอิงที่มีคะแนนจากแบบทดสอบเท่ากัน สำหรับกลุ่มเปรียบเทียบในการวิจัยครั้งนี้ คือ กลุ่มผู้สอบที่นับถือศาสนาอิสลาม

9. ผลการวิเคราะห์ความล่าเอียงของข้อสอบ หมายถึง ผลที่ได้จากการคำนวณด้วยวิธีวิเคราะห์ต่างกัน ซึ่งได้แก่ วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และวิธีของแมนเทล-แฮนส์เชล ในการวิจัยครั้งนี้ หมายถึงผลดังนี้

9.1 จำนวนข้อสอบที่มีความล่าเอียง หมายถึง จำนวนข้อสอบในแบบทดสอบแต่ละฉบับที่ให้ผลการสอบแตกต่างกันระหว่างกลุ่มผู้สอบที่นับถือศาสนาพุทธและกลุ่มผู้สอบที่นับถือศาสนาอิสลาม ซึ่งวัดได้จากแบบทดสอบที่ผู้วิจัยสร้างขึ้นทั้ง 2 ฉบับ และทำการวิเคราะห์ด้วยวิธีต่างกัน ดังข้อ 6

9.2 ค่าความเชื่อมั่น (Reliability) ของแบบทดสอบ หมายถึง คุณสมบัติของแบบทดสอบที่วัดความรู้ความสามารถของนักเรียนได้คงที่แน่นอน ซึ่งในการวิจัยครั้งนี้จะศึกษาหาความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อที่มีความล่าเอียงออกแล้ว ซึ่งจะคำนวณค่าความเชื่อมั่น โดยใช้สูตรคูเดอร์ริชาร์ดสัน 20 (Kuder-Richardson 20) และสูตรแบบแบ่งครึ่งฉบับ (Split-half) ของแบบทดสอบโดยใช้สูตรของ Spearman-Brown

10. ผู้เชี่ยวชาญ หมายถึง ครูผู้สอนวิชาภาษาไทยและครูผู้สอนวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 2 ที่มีประสบการณ์ด้านการสอนและการเขียนข้อสอบในวิชาดังกล่าวอย่างน้อย 5 ปี และนักวัดผลการศึกษา

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องกับการศึกษาค้นคว้าครั้งนี้ ผู้วิจัยได้นำเสนอเนื้อหาต่างๆ ตามลำดับ ดังนี้

1. ความเป็นมาของการตรวจสอบความลำเอียงของข้อสอบ
2. ความหมายของความลำเอียงของข้อสอบ
3. สาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบ
4. หลักการวิเคราะห์ความลำเอียงของข้อสอบ
5. การวิเคราะห์ความลำเอียงของข้อสอบโดยยึดหลักการวัดแบบคลาสสิกอล
6. การวิเคราะห์ความลำเอียงของข้อสอบโดยวิธีของ แมนเทล-แฮนส์เซล
7. การวิเคราะห์ความลำเอียงของข้อสอบโดยยึดหลักการตอบข้อคำถาม
8. งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบ
 - 8.1 งานวิจัยภายในประเทศ
 - 8.2 งานวิจัยในต่างประเทศ

ความเป็นมาของการตรวจสอบความลำเอียงของข้อสอบ

การศึกษความลำเอียงของข้อสอบหรือแบบทดสอบนั้น เริ่มมาจากการสังเกตปรากฏการณ์ของผลการสอบคัดเลือก เช่น การสอบคัดเลือกเพื่อเข้าศึกษาต่อเพื่อบรรจุเข้าทำงาน หรือการสอบเพื่อเลื่อนตำแหน่ง หรือในการสอบอื่นๆ ซึ่งพบว่าไม่เป็นไปตามสัดส่วนระดับสติปัญญาของผู้เข้าสอบ หรือโครงสร้างของประชากร จึงเป็นเหตุให้เกิดความลำเอียงหรือความไม่ยุติธรรมในการสอบ และส่งผลให้เกิดการศึกษาความลำเอียงของแบบทดสอบและข้อสอบรายข้อในแบบทดสอบที่ใช้ (Test item bias) ในระยะต่อ ๆ มา/แนวคิดในการวิเคราะห์ความลำเอียงของข้อสอบและการตรวจสอบความยุติธรรมของการใช้ข้อสอบได้เริ่มปรากฏและได้รับความสนใจตั้งแต่ปี 1910 โดยอัลเฟรด บินเน็ต และ ซีมอน (Alfred Binet and Simon, 1961/1973) ได้ทดสอบเด็กที่มีฐานะทางเศรษฐกิจและสังคมที่แตกต่างกัน พบว่า มีข้อสอบบางข้อที่น่าจะวัดในสิ่งที่เป็นผลกระทบของการฝึกฝนทางด้านวัฒนธรรม (เช่น ที่บ้าน หรือ ที่โรงเรียน) มากกว่าที่จะวัดความสามารถทางสมองหลัง

จากการศึกษาครั้งนี้ บีเนต และไซมอน ได้มีการพิจารณาอย่างละเอียดและได้เอาข้อสอบข้อเหล่านั้นออกจากแบบทดสอบ เช่น ถ้าเป็นเด็กอายุ 7 ปี ข้อคำถามที่เกี่ยวกับนิ้วมือ การคัดลอกประโยค และการเรีอกชื่อเหรียญ จะถูกคัดออกจากแบบทดสอบ โดยทั่วไป ในการทดสอบจะพิจารณาเฉพาะข้อคำถามที่ไม่ขึ้นกับการฝึกฝนของทางบ้าน (home training) ความตั้งใจ (attention) ภาษา (Language) นิสัยในการดูภาพ (habit of looking at picture) และแบบฝึกหัดเชิงวิชาการ (scholastic exercise) ในปี 1912 วิลเลียม สเตร์น (William Stern) ได้ศึกษาความแตกต่างระหว่างชนชั้นต่างๆ (Class difference) ในประเทศเยอรมัน ผลที่ได้มีความคล้ายคลึงกับที่บีเนตพบ คือ ความแตกต่างระหว่างชนชั้นนี้มีนัยสำคัญ สเตร์น พยายามที่จะค้นหา และทำความเข้าใจเกี่ยวกับแหล่งที่ทำให้เกิดความแตกต่าง เพื่อที่จะเสนอแนะว่า แบบทดสอบต่าง ๆ ที่ได้พิสูจน์มาแล้วนั้น เข้าข้างชนชั้นหนึ่งมากกว่าชนชั้นอื่น ๆ อย่างชัดเจน

ต่อมาในปี 1951 มิงงานวิจัยของ เอลส์ (Ells) เดวิส (Davis) ฮาวิเกอร์สท (Havighurst) แฮร์ริค (Herrick) และไทเลอร์ (Tyler) ซึ่งถือเป็นการเริ่มต้นของการศึกษาเกี่ยวกับความลำเอียงของข้อสอบแนวใหม่ โดยเอลส์และคณะสนใจความแตกต่างของการวัดบางอย่าง ซึ่งไม่มีความแม่นยำพอที่จะสะท้อนให้เห็นความสามารถได้ เช่น ความแตกต่างทางด้านสติปัญญาของนักเรียน อาจขึ้นอยู่กับเนื้อหาเฉพาะของข้อสอบแต่ละข้อในแบบทดสอบ และไม่ได้สะท้อนความสามารถที่แท้จริงของนักเรียน จุดมุ่งหมายหลักเพื่อที่จะอธิบายความแตกต่างระหว่างกลุ่มนักเรียน โดยใช้แบบทดสอบวัดสติปัญญา โดยมุ่งไปที่ข้อคำถามแต่ละข้อเช่นเดียวกับบีเนต และสเตร์น คือดูว่าข้อคำถามใดมีความลำเอียงหรือไม่ แล้ววิเคราะห์ด้วยวิธีที่ค่าความยากของข้อสอบระหว่างกลุ่มที่แตกต่างกัน ผลการศึกษาพบว่า ยังมีความแตกต่างระหว่างกลุ่มนักเรียนที่มีระดับชั้นทางสังคม และภูมิหลังของเชื้อชาติ และวัฒนธรรมต่างกันในการตอบแบบทดสอบ เอลส์ และคณะได้พยายามพัฒนาวิธีการปฏิบัติและแก้ไขอิทธิพลของระดับชั้นทางสังคมและภูมิหลังของเชื้อชาติที่มีผลต่อการตอบข้อสอบ เพื่อใช้จัดข้อสอบที่มีความลำเอียงในแบบทดสอบต่างๆ โดยการเปรียบเทียบกลุ่มของเชื้อชาติที่อยู่ในชั้นทางสังคม

ผลจากการศึกษาของเอลส์ได้นำไปสู่การศึกษาอย่างกว้างขวางในช่วง 20 ปี ที่ผ่านมา เมื่อรัฐมีความมุ่งหวังที่จะทำให้เกิดความยุติธรรมในการใช้ผลการสอบในการเข้าศึกษาและโอกาสในการมีงานทำ ดังนั้นในช่วงปลายปี 1960 และ 1970 ได้มีความตื่นตัวเกี่ยวกับความลำเอียงของข้อสอบโดยมุ่งไปที่แบบทดสอบวัดสติปัญญา เพราะเป็นแบบทดสอบที่ถูกนำไปใช้ในทางด้านการศึกษา

การคัดเลือก (Selection) และการจัดบุคคลเข้าสู่ตำแหน่ง (Placement) อย่างแพร่หลายจนกระทั่งปี 1971 บริษัท Griggs V. Duke Power ได้วินิจฉัยแบบทดสอบต่าง ๆ แล้วชี้ให้เห็นว่าผลการทดสอบเหล่านั้นใช้ในการคัดเลือกอาจไม่เหมาะสม เมื่อพิจารณาตามเชื้อชาติของผู้สมัครงาน ยกเว้นกรณีที่ผลของการปฏิบัติงานมีความสอดคล้องกับการทดสอบในช่วงเวลาเดียวกัน มีการโต้แย้งในเรื่องความลำเอียงของข้อสอบมากขึ้น เมื่อ เจนเซน (Jensen. 1969) ได้เสนอบทความลงในสารทางการศึกษาของฮาร์วาร์ด (Harvard Educational Review. 1969) ฉบับที่ 39 ปี 1969 โดยกล่าวว่าไอคิว เป็นผลมาจากการถ่ายทอดทางพันธุกรรมมากกว่าสิ่งแวดล้อม ความแตกต่างของไอคิวระหว่างคนผิวขาวกับคนผิวดำนั้นไม่ได้มาจากสิ่งแวดล้อมเพียงอย่างเดียว และพันธุกรรมอธิบายความแปรปรวนผลการปฏิบัติงานของแต่ละคนได้ประมาณร้อยละ 80 จากบทความของเจนเซนนี้เองทำให้สมาคมนักจิตวิทยาคนผิวดำเสนอว่า การทดสอบคนผิวดำทุกชนิดควรทำให้เท่าเทียมกันมากที่สุด เท่าที่จะเป็นไปได้ อีกทั้งยังมีบทความและวิธีการทางสถิติในการวิเคราะห์ความลำเอียงของข้อสอบเกิดขึ้นมากมาย ซึ่งแสดงให้เห็นว่าปัญหาเรื่องความลำเอียงของข้อสอบมีความสำคัญ เพื่อจะสร้างแบบทดสอบที่มีความยุติธรรมทางวัฒนธรรม และให้ผลการวิเคราะห์ที่มีความถูกต้องที่สุด (Camilli and Shepard. 1994 : 4-7) นอกจากนี้ ยังมีเหตุการณ์ที่ทำให้มีการตื่นตัวเป็นอันมากในเรื่องความลำเอียงของข้อสอบ คือ กรณีที่มาร์โค เดอฟูนิส (Marco Defunis) และคณะซึ่งถูกปฏิเสธจากโรงเรียนกฎหมายของมหาวิทยาลัยแห่งวอชิงตันได้ฟ้องร้องว่าเขาได้คะแนนการสอบสูงกว่าผู้ได้รับการคัดเลือกบางคนที่มีมหาวิทยาลัยรับเข้าศึกษา และได้ยื่นฎีกาฟ้อง ชาร์ลส์ โอเดการ์ด (Charles Odegard) อธิการบดีมหาวิทยาลัยดังกล่าวและคณะ เพื่อให้พิจารณาทบทวนการคัดเลือกนักศึกษาใหม่ หลังจากนั้นการพิจารณาตรวจสอบความลำเอียงของข้อสอบระหว่างผู้สอบกลุ่มย่อยที่มีลักษณะแตกต่างกันในเรื่อง เพศ เชื้อชาติ ศาสนา หรือ วัฒนธรรมได้ปฏิบัติกันมาจนปัจจุบัน (Brelan & Ironson. 1976 : 89) ซึ่งเป็นเสมือนส่วนหนึ่งของขั้นตอนการพัฒนาแบบทดสอบโดยทั่ว ๆ ไป โดยเฉพาะการพัฒนาแบบทดสอบมาตรฐาน เช่น ศูนย์บริการการทดสอบทางการศึกษา (Educational Testing Service) ของประเทศสหรัฐอเมริกา ได้มีการตรวจสอบความลำเอียงทั้งก่อนและหลังการนำแบบทดสอบไปทดลองใช้ การตรวจสอบความลำเอียงของข้อสอบก่อนนำไปทดลองใช้ จะใช้การตรวจสอบโดยผู้เชี่ยวชาญ ซึ่งประกอบด้วยบุคคลหลายฝ่ายตรวจสอบความไว (Sensitivity) ของแบบทดสอบโดยพิจารณาถึงรูปแบบของข้อสอบ เนื้อหา ภาพประกอบ คำที่ใช้และอื่น ๆ เพื่อไม่ให้มีความลำเอียงหรือเกิดการ

ได้เปรียบเทียบเปรียบเทียบระหว่างผู้สอบกลุ่มย่อยที่มีลักษณะแตกต่างกัน ส่วนการตรวจสอบความลำเอียงของข้อสอบหลังจากการทดลองใช้นั้นจะใช้วิธีการทางสถิติ

อย่างไรก็ตามเท่าที่ผ่านมาได้มีผู้เสนอวิธีการวิเคราะห์ความลำเอียงของข้อสอบ หรือแบบทดสอบไว้หลายวิธีด้วยกัน วิธีที่รู้จักและนำมาใช้มากในอดีตจนถึงปัจจุบัน โดยพิจารณาจากค่าสถิติซึ่งมีรายละเอียด ดังนี้

1. การพิจารณาความแตกต่างในค่าเฉลี่ยระหว่างกลุ่ม ซึ่งกรณีเช่นนี้อาจไม่ได้แสดงถึงความลำเอียงของข้อสอบหรือแบบทดสอบ เพราะกลุ่มย่อยทั้งสองกลุ่มนั้นอาจจะมี ความแตกต่างกันอยู่แล้วในสิ่งที่ต้องการวัด

2. การวิเคราะห์ตัวประกอบ เป็นการดูความสัมพันธ์ระหว่างคะแนนข้อๆ ในแบบทดสอบ สำหรับกลุ่มย่อย คำนวณความลำเอียง ได้แก่ ความแตกต่างของน้ำหนักตัวประกอบ หรือความแตกต่างอย่างมีนัยสำคัญระหว่างค่าเฉลี่ยของคะแนนตัวประกอบระหว่างกลุ่มย่อย ซึ่งโดยมากจะใช้เทคนิคการวิเคราะห์แบบ maximum likelihood

3. การเปรียบเทียบค่าความยาก เป็นการพิจารณาลำดับที่สัมพันธ์ของค่าความยากที่ได้จากการวิเคราะห์แยกกลุ่ม ซึ่งพิจารณาปฏิสัมพันธ์ระหว่างข้อสอบและกลุ่มผู้สอบ ข้อสอบข้อใดที่มีลักษณะของความสัมพันธ์สำหรับบางกลุ่มแตกต่างไปจากข้ออื่นๆ ในแบบทดสอบเดิยกันถือว่าข้อสอบข้อนั้นมีความลำเอียง วิธีการนี้ทำได้ทั้งการวิเคราะห์ความแปรปรวนและการใช้แผนภูมิเส้น

ผู้ที่พัฒนาวิธีการนี้ได้สมบูรณ์ที่สุด คือ แองกอฟฟ์ (Angoff) ซึ่งแปลงค่าความยากให้เป็นคะแนนมาตรฐาน Z เพื่อจัดความสัมพันธ์เชิงเส้นโค้งระหว่างค่าความยากของข้อสอบ 2 ชุด แล้วแปลงค่าคะแนนมาตรฐาน Z เป็นเดลตา (Δ) มีค่าเฉลี่ยเท่ากับ 13 และค่าความส่วนเบี่ยงเบนมาตรฐานเท่ากับ 4 เพื่อกำจัดค่าที่เป็นลบ แล้วพิจารณาค่าเดลตาของข้อสอบข้อเดียวกันจากกลุ่มที่ต่างกันโดยใช้นาเวเส้นทะแยงที่เป็นเส้นแกนหลัก ข้อสอบจะลำเอียงหรือไม่พิจารณาได้จากระยะทางจากจุดศูนย์กลางค่าเดลตาที่ลากไปตั้งฉากกับเส้นแกนหลัก

4. การใช้ทฤษฎีการตอบข้อคำถามเป็นการเปรียบเทียบความน่าจะเป็นในการตอบข้อสอบถูกของกลุ่มผู้สอบย่อย โดยพิจารณาถึงค่าพารามิเตอร์ของข้อสอบ และผู้สอบตามแบบจำลองที่เลือกใช้ซึ่งมีทั้งชนิด 1, 2 และ 3 พารามิเตอร์

5. การวิเคราะห์ด้วยค่าโคสแควร์ มีผู้เสนอหลายคน เช่น วิธีของชูเนแมน (Scheuneman, 1979) เป็นวิธีการที่แปลงค่าความสามารถของผู้สอบในกลุ่มเป็นช่วงๆ อาจเป็น 3 หรือ 5 ช่วง

แล้วใส่ข้อมูลในตาราง 2 ทาง (จำนวนกลุ่ม x การตอบถูก-ผิด) จากนั้นคำนวณค่าไคสแควร์แต่ละตารางและเมื่อรวมค่าทุกตารางจะได้ค่าไคสแควร์ ที่มีขึ้นแห่งความเป็นอิสระเท่ากับจำนวนกลุ่ม

วิธีการวิเคราะห์ด้วยไคสแควร์นี้จะถือว่าข้อสอบมีความลำเอียงก็ต่อเมื่อ โอกาสในการตอบถูกจากผู้สอบ 2 กลุ่มหรือมากกว่าแตกต่างกัน โดยจะประมาณความสามารถของผู้สอบจากคะแนนดิบ แล้วคำนวณหาสัดส่วนการตอบถูกและความถี่ที่คาดหวัง แล้วจึงคำนวณค่าไคสแควร์ ถ้าค่าไคสแควร์ที่คำนวณได้มากกว่าค่าไคสแควร์ที่เปิดในตารางแสดงว่าข้อสอบข้อนั้นลำเอียง

6. เทคนิค ลอกลินีียร์ (loglinear) หรือ แบบจำลองโลจิสติกที่เบี่ยงเบนจากไคสแควร์และได้มีการนำมาขยายโดยการทำซ้ำ หลังจากการจัดข้อสอบที่ลำเอียงออกไปแล้วในแต่ละรอบของการวิเคราะห์ (Mellenbergh. 1982)

7. วิธีของแมนเทล-แฮนส์เซล (Mantel-Haenszel) เป็นวิธีที่แตกแยกออกมาจาก Loglinear เป็นการศึกษาตัวแปรประเภทแบ่งสองในกลุ่มที่แตกต่างกันด้วยตัวแปรบล็อก และนำมาใช้เป็นตัวบ่งชี้ความลำเอียงของข้อสอบ

8. วิธีทฤษฎีการตอบข้อคำถามแบบเทย์ม ที่เสนอโดย ลินน์ และฮาร์นิช (Linn and Harnisch. 1981) เป็นวิธีวิเคราะห์ในลักษณะทฤษฎีการตอบข้อคำถามแบบ 3 พารามิเตอร์ โดยรวมกลุ่มข้อเข้าด้วยกัน แล้วแบ่งความสามารถเป็นช่วงดูความแตกต่างระหว่างความน่าจะเป็นของการตอบถูกตามที่คาดหวัง ซึ่งประมาณได้ด้วยทฤษฎีการตอบข้อคำถามกับความน่าจะเป็นของการตอบถูกจากข้อมูลเชิงประจักษ์พิจารณาเป็นช่วงความสามารถแล้วนำมารวมกันเป็นดัชนีรวมของความลำเอียงของแต่ละข้อ และใช้ได้ในการที่มีกลุ่มตัวอย่างน้อยทำให้ใช้ทฤษฎีการตอบข้อคำถามแบบเต็มรูปแบบไม่ได้ (Linn and Harnisch. 1981)

9. วิธี SIBTEST หรือ การทดสอบ SIB เป็นวิธีล่าสุดที่เสนอโดยซีลีและสเตทท์ (Shealey and Stout) เป็นวิธีที่ดัดแปลงจากวิธีทฤษฎีการตอบข้อคำถาม สามารถใช้ได้ ทั้งเพื่อ การตรวจสอบความลำเอียงของข้อสอบรายข้อและการตรวจสอบความลำเอียงของกลุ่มข้อสอบ หรือแบบทดสอบทั้งฉบับ

สำหรับการวิจัยครั้งนี้วิธีการวิเคราะห์ความลำเอียงของข้อสอบที่ผู้วิจัยนำมาใช้ในการศึกษามี 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีการใช้ทฤษฎีการตอบข้อคำถาม และวิธีของแมนเทล-แฮนส์เซล ซึ่งจะกล่าวถึงรายละเอียดต่อไป

ความหมายของความลำเอียงของข้อสอบ

ความลำเอียงของข้อสอบได้มีผู้ให้ความหมายไว้ต่าง ๆ ดังนี้ คือ

เคลียร์และฮิลตัน (Cleary and Hilton. 1968 : 61) ให้ความหมายของข้อสอบที่มีความลำเอียงว่า เป็นข้อสอบที่เมื่อนำไปทดสอบกับบุคคลกลุ่มใดกลุ่มหนึ่งที่มีความสามารถเท่ากันแล้ว ปรากฏว่ามีคะแนนเฉลี่ยของข้อสอบข้อเดียวกันของแต่ละกลุ่มไม่เท่ากัน

แอลเลน และ เยน (Allen and Yen. 1979 : 129) ให้ความหมายของข้อสอบที่มีความลำเอียงว่าเป็นข้อสอบที่วัดความสามารถ หรือคุณลักษณะทางจิตวิทยาของผู้สอบแต่ละกลุ่มได้ไม่ตรงกัน เช่น เมื่อนำข้อสอบวัดความสามารถด้านคณิตศาสตร์ที่ใช้ภาษาสื่อความหมายของโจทย์ปัญหาไปทดสอบกับผู้สอบ 2 กลุ่มที่มีความสามารถทางการอ่านไม่เท่ากัน ข้อสอบนั้นมีความลำเอียงต่อผู้สอบที่มีความสามารถทางการอ่านสูง และข้อสอบที่มีความลำเอียงนี้จะมีโด่งลักษณะของข้อสอบของผู้สอบแต่ละกลุ่มแตกต่างกัน

ลอร์ด (Lord. 1980 : 212) สับโคเวียค และคนอื่นๆ (Subkoviak and Others. 1984 : 49) แมคคอลลีย์ และเมนโดซา (McCanley and Mendoza. 1985 : 389) แฮมเบิลตัน และคณะ (Hambleton et. al. 1991 : 110) ฮอลล์แลนด์ และ ไวเนอร์ (Holland and Wainer. 1993 : 4 ; Citing Angoff. 1993) ให้ความหมายของข้อสอบที่มีความลำเอียงว่า เป็นข้อสอบที่ผู้สอบที่มีความสามารถเท่ากัน แต่อยู่ต่างกลุ่มกัน มีโอกาสตอบข้อสอบข้อนั้นได้ถูกไม่เท่ากัน

ปอปแฮม (Popham. 1981 : 181) ให้ความหมายของความลำเอียงของข้อสอบว่า หมายถึง ความชอบหรือความโน้มเอียงที่ทำให้เกิดการพิจารณาตัดสินเป็นไปอย่างไม่ยุติธรรม

ออสเตอร์ลันด์ (Osterlind. 1983 : 55) ให้ความหมายว่า ความลำเอียงเป็นความคลาดเคลื่อนอย่างมีระบบทางการวัด

คามิลลี และเชฟเพิร์ด (Camilli and Shepard. 1994 : 8) ได้ให้ความหมายของข้อสอบที่มีความลำเอียงว่า เป็นความคลาดเคลื่อนอย่างเป็นระบบหรือความไม่เที่ยงตรงของการวัด ทำให้สมาชิกบางกลุ่มได้ประโยชน์หรือเสียประโยชน์ เช่น แบบทดสอบหรือข้อสอบที่เข้าข้างกลุ่มนักเรียนชายมากกว่ากลุ่มนักเรียนหญิง หรือคนผิวขาวกับคนผิวดำ หรือระหว่างคนที่อยู่ในภูมิภาคต่างกัน

นอกจากนี้ยังมีผู้ให้ความหมายของความลำเอียงของข้อสอบ เฉพาะด้าน ดังนี้

เมห์เรนส์ และเลห์แมน (Mehrens and Lehmann. 1978 : 435-544) ให้ความหมาย ของข้อสอบที่มีความลำเอียงด้านวัฒนธรรม (Cultural Bias) ว่าหมายถึง ข้อสอบที่ทำให้กลุ่มผู้สอบสองกลุ่มที่มีความสามารถเท่ากัน แต่ที่มีความแตกต่างทางวัฒนธรรมมีคะแนนเฉลี่ยแตกต่างกัน หรือ เป็นข้อสอบที่วัดคุณลักษณะทางจิตวิทยา หรือผลสัมฤทธิ์ในเนื้อหาเดียวกันไม่ตรงกัน และ ให้ความหมายข้อสอบที่มีความลำเอียงระหว่างเพศ (Sex Bias) ว่าหมายถึง ข้อสอบที่มีเนื้อหา ลำเอียงต่อเพศใดเพศหนึ่งโดยเฉพาะ และมีผลต่อการทำข้อสอบแต่ละข้อ

เอลส์ และ คนอื่นๆ (Jensen. 1980 : 369 ; Citing Eells and Others. 1951) ให้ความหมายของข้อสอบที่มีความลำเอียงด้านวัฒนธรรมว่า หมายถึง ข้อสอบที่ทำให้ผู้สอบ ซึ่งคุ้นเคยกับวิชาเฉพาะหรือกระบวนการเฉพาะมีโอกาสตอบถูกมากกว่าผู้สอบคนอื่น ที่มีลักษณะของ กลุ่มแตกต่างกัน เป็นต้น

จากความหมายดังกล่าวข้างต้นสามารถสรุปได้ว่า ข้อสอบที่มีความลำเอียง หมายถึง ข้อสอบที่เมื่อนำไปทดสอบกับกลุ่มผู้สอบที่มีความสามารถเท่ากันภายใต้ประชากรเดียวกัน แต่มีลักษณะของกลุ่มย่อยแตกต่างกันทางด้านใดด้านหนึ่ง เช่น เพศ เชื้อชาติ ศาสนา วัฒนธรรม สภาพภูมิศาสตร์ หรือเศรษฐกิจ แล้วปรากฏว่าโอกาสในการที่จะตอบข้อสอบข้อนั้นถูกไม่เท่ากัน หรือ ทำให้ในแต่ละกลุ่มย่อยมีการได้เปรียบเสียเปรียบจากการสอบข้อสอบเดียวกันไม่เท่ากัน ซึ่งอาจกล่าวได้ว่าเป็นข้อสอบที่ขาดความเที่ยงตรงในคุณลักษณะที่ต้องการวัด นั่นคือ เป็นข้อสอบที่วัดความสามารถหรือคุณลักษณะทางจิตวิทยาของผู้สอบแต่ละกลุ่มได้ไม่ตรงกัน เมื่อทำการวิเคราะห์คุณสมบัติของข้อสอบตามกลุ่มย่อยของผู้สอบแล้ว จะได้คุณสมบัตินั้นไม่เท่ากัน

สาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบ

จากการนำข้อสอบที่มีความลำเอียงมาศึกษาภายหลังจากการวิเคราะห์ความลำเอียงด้วยค่าสถิติแบบต่าง ๆ ในแบบทดสอบหลาย ๆ ฉบับ เพื่อศึกษาถึงสาเหตุหรือแหล่งที่ก่อให้เกิดความลำเอียงในตัวคำถามเหล่านั้น พบว่า สาเหตุที่น่าจะก่อให้เกิดความลำเอียงมีหลายสาเหตุด้วยกัน ซึ่งสามารถสรุปสาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบมากที่สุด ได้ดังต่อไปนี้ (Scheuneman. 1982 : 180-195)

1. การเดา (guessing) อาจเกิดจากข้อสอบนั้นยากเกินไปหรือเวลาไม่เพียงพอ จะก่อให้เกิดความไม่เท่าเทียมกันในโอกาสการตอบข้อสอบของผู้สอบแต่ละคน

2. ความเร็ว (Speed) หรือเวลาในการตอบ จะทำให้เกิดการเดาหรือในกรณีเวลาน้อยอาจจะทำข้อสอบไม่ทัน ซึ่งจะมีผลกับข้อสอบข้อหลัง ๆ โดยเฉพาะในการศึกษาความลำเอียงของข้อสอบวัดความถนัด

3. ความกำกวมหรือความไม่ชัดเจน (Unclear) ของข้อคำถาม นั่นคือ ข้อคำถามขาดความเป็นปรนัย การใช้ภาษาถิ่นหรือใช้คำที่ไม่เป็นภาษากลางในการสื่อความหมาย ซึ่งจะก่อให้เกิดความลำเอียงกับกลุ่มภาษาใดภาษาหนึ่งขึ้นได้

4. ลำดับขึ้นของคำถาม (Series) อาจจะเป็นสิ่งที่ก่อให้เกิดความสับสน หรือขึ้นหาคำตอบของข้อสอบบางข้อได้

5. สถานภาพทางสังคมหรือความเกี่ยวข้องทางสังคม (Social implication) ก็ เป็นสิ่งที่ก่อให้เกิดความแตกต่างระหว่างกลุ่มผู้สอบแต่ละกลุ่มได้

6. ประสบการณ์หรือการฝึกฝนของแต่ละกลุ่มย่อย เป็นสิ่งที่ก่อให้เกิดการได้เปรียบเสียเปรียบของแต่ละกลุ่มค่อนข้างจะชัดเจน

7. องค์ประกอบทางวัฒนธรรม ความเป็นอยู่ ขนบธรรมเนียมประเพณี เชื้อชาติ ศาสนา ก็จะเอื้อให้กับบางกลุ่มย่อย จึงทำให้เกิดการได้เปรียบเสียเปรียบในบางเนื้อหาวิชาได้

นอกจากนี้ความลำเอียงของข้อสอบอาจเกิดจากสาเหตุหรือแหล่งที่สำคัญ 2 แหล่ง คือ

1. การเลือกเนื้อหา (bias in selection) คือ ผู้สร้างข้อสอบเลือกเนื้อหาเฉพาะส่วนใดส่วนหนึ่งมาสร้างข้อสอบ ทำให้ได้ข้อสอบที่มีเนื้อหาไม่ครอบคลุมและไม่ได้สัดส่วนที่สมดุลกัน

2. การสร้างข้อสอบ (bias in construction) คือ การใช้ภาษาหรือข้อความบางอย่างที่เอื้อให้เกิดประโยชน์กับผู้สอบกลุ่มใดกลุ่มหนึ่ง

หลักการวิเคราะห์ความลำเอียงของข้อสอบ

การวิเคราะห์ความลำเอียงของข้อสอบ สามารถทำได้ 2 ขั้นตอน คือ ขั้นตอนแรกเป็นขั้นตอนของการสร้างแบบทดสอบโดยให้ผู้เชี่ยวชาญพิจารณาว่าข้อสอบมีความลำเอียงหรือไม่ วัดในคุณลักษณะที่ต้องการวัดเดียวกันหรือไม่ ส่วนขั้นตอนที่สองใช้เทคนิควิธีการทางสถิติเข้ามาช่วยหลังจากที่นำแบบทดสอบไปใช้แล้ว นำผลการสอบมาวิเคราะห์คุณลักษณะการตอบข้อสอบของกลุ่มผู้สอบที่

แตกต่างกัน ซึ่งอาจจะเป็น เพศ เชื้อชาติ ศาสนา วัฒนธรรม หรือสภาพภูมิศาสตร์ ระดับสติปัญญา เป็นต้น ว่าข้อสอบนั้นยังคงทำหน้าที่เบี่ยงเบนไปหรือเอื้อประโยชน์ต่อกลุ่มใดกลุ่มหนึ่งหรือไม่ ซึ่งการศึกษาความลำเอียงของข้อสอบที่มีพื้นฐานจากการใช้หลักสถิตินี้ สามารถแบ่งศึกษาได้เป็น 2 วิธี ใหญ่ ๆ คือ การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้เกณฑ์ภายนอกและเกณฑ์ภายใน (Rudner, Getson and Knight. 1980) ซึ่งแต่ละวิธีมีรายละเอียด ดังนี้

การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้เกณฑ์ภายนอก

การวิเคราะห์ความลำเอียงของข้อสอบ โดยวิธีนี้สามารถวิเคราะห์ได้ทั้งรายข้อและรายฉบับ ซึ่งวิธีการวิเคราะห์จะเขียนกราฟแสดงความสัมพันธ์ระหว่างคะแนนที่ใช้เป็นตัวแปรเกณฑ์ภายนอกกับตัวแปรทำนาย วิธีนี้มีจุดมุ่งหมายเพื่อวิเคราะห์ความถดถอยของตัวแปรทั้งสอง แล้วเปรียบเทียบค่าความชัน (Slope) และค่าการตัดแกน (Intercept) ของเส้นกราฟระหว่างกลุ่มผู้สอบ 2 กลุ่ม ในการวิเคราะห์ความลำเอียงของข้อสอบทั้งฉบับจะใช้คะแนนรวมของข้อสอบฉบับนั้น จากผู้สอบแต่ละคนเป็นตัวแปรทำนาย แต่ถ้าเป็นการวิเคราะห์ความลำเอียงของข้อสอบเป็นรายข้อ จะใช้ค่าความชุก (p) ของข้อสอบแต่ละข้อเป็นตัวแปรทำนาย ส่วนตัวแปรที่ใช้เป็นเกณฑ์ภายนอกของการวิเคราะห์ความลำเอียงของข้อสอบรายข้อหรือรายฉบับ จะใช้คะแนนรวมหรือเกรดเฉลี่ยหรือผลสัมฤทธิ์ของงานบางอย่างที่ทำให้ทำ (Cronbach. 1970 ; Anastasi. 1976)

ดังนั้นการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีนี้ จึงเป็นการเปรียบเทียบเส้นกราฟที่แสดงความสัมพันธ์ระหว่างตัวแปรเกณฑ์กับตัวแปรทำนายของกลุ่มผู้สอบที่ต้องการวิเคราะห์ความลำเอียงของข้อสอบ ถ้าเส้นกราฟดังกล่าวมีค่าความชันและค่าการตัดแกนแตกต่างกันในแต่ละกลุ่ม แล้วข้อสอบข้อนั้น หรือแบบทดสอบฉบับนั้นก็จะมีความลำเอียงต่อกลุ่มผู้สอบที่มีค่าการตัดแกน หรือค่าความชันมากกว่า (Haebara. 1979)

อย่างไรก็ตามการวิเคราะห์ความลำเอียงของข้อสอบตามวิธีการที่กล่าวมามีจุดอ่อนตรงที่ในทางปฏิบัติแล้วเป็นการยากมากที่จะหาตัวแปรเกณฑ์ภายนอกที่มาจากแบบทดสอบที่มีความเที่ยงตรงเชิงพยากรณ์และมีความยุติธรรมสำหรับกลุ่มที่นำมาวิเคราะห์ความลำเอียง ถ้าหากตัวแปรเกณฑ์ภายนอกขาดคุณสมบัติดังกล่าว จะทำให้ผลการวิเคราะห์ความลำเอียงของข้อสอบตามวิธีนี้ขาดความสมบูรณ์และถูกต้องเท่าที่ควร การวิเคราะห์ความลำเอียงของข้อสอบตามวิธีนี้ส่วนใหญ่่นำมาใช้กับข้อสอบคัดเลือก จากความยุ่งยากในการสร้างเกณฑ์ภายนอกที่ไม่ลำเอียงนี้นักวิจัยทั้งหลายจึงหัน

มาตรฐานข้อสอบโครงสร้างภายในของแบบทดสอบ โดยอาศัยวิธีการทางสถิติเป็นเครื่องตัดสินว่าข้อค้นพบนั้นมีความลำเอียงทางการทดสอบกับกลุ่มผู้สอบหรือไม่ ซึ่งวิธีการดังกล่าวนี้จะใช้ข้อมูลจากผลการตอบคำถามในแบบทดสอบเท่านั้น จึงเรียกวิธีนี้ว่า การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้เกณฑ์ภายใน

การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้เกณฑ์ภายใน

การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้เกณฑ์ภายในนี้ยึดหลักการที่ว่า ข้อสอบจะมีความยุติธรรมไม่ลำเอียงต่อกลุ่มผู้สอบก็ต่อเมื่อมีผู้สอบที่อยู่ต่างกลุ่มกัน มีความสามารถเท่ากันแล้วจะมีคะแนนจริงของผลการสอบเท่ากัน ถ้าข้อสอบที่มีผู้สอบที่ระดับความสามารถเท่ากัน แต่อยู่ต่างกลุ่มกันมีคะแนนจริงของผลการสอบแตกต่างกัน เป็นข้อสอบที่มีความลำเอียง การวิเคราะห์ความลำเอียงโดยวิธีการนี้อาจแบ่งออกได้เป็น 3 วิธีใหญ่ๆ (Laksana and Coffman. 1980) ดังนี้

1. ใช้หลักการวัดความเบี่ยงเบนสัมพัทธ์ (Relative Deviation) ของข้อสอบแต่ละข้อ จากแนวโน้มเข้าสู่ส่วนกลาง วิธีนี้มีข้อตกลงเบื้องต้นว่าข้อสอบส่วนใหญ่ในแบบทดสอบมีความเป็นเอกพันธ์ ในการวัดความสามารถใดความสามารถหนึ่งในผู้สอบต่างกลุ่มกันข้อสอบข้อใดที่เบี่ยงเบนไปจากส่วนกลางมากกว่าที่คาดหวังไว้ ก็อาจตั้งข้อสงสัยได้ว่าเป็นข้อสอบที่มีความลำเอียง วิธีการหนึ่งที่ใช้กันบ่อยๆ ก็คือการวิเคราะห์ความแปรปรวน (ANOVA) ข้อสอบที่ลำเอียง คือข้อสอบที่มีปฏิสัมพันธ์ระหว่างข้อกับกลุ่มผู้สอบมีนัยสำคัญทางสถิติ ส่วนอีกวิธีหนึ่ง ได้แก่ การเขียนกราฟ วิธีนี้ค่าพารามิเตอร์เกี่ยวกับค่าความยาก เช่น ค่า P ค่าเดลตา หรือค่า b จากกลุ่มหนึ่งจะลงจุดคู่อันดับ คู่กับค่าพารามิเตอร์จากอีกกลุ่มหนึ่ง ข้อสอบข้อใดที่เบี่ยงเบนไปจากเส้นแกนหลักมาก ถือว่าเป็นข้อสอบที่มีความลำเอียง

2. ใช้หลักการประเมินความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) ของแบบทดสอบวิธีการหนึ่งที่ใช้กันมาก คือ การวิเคราะห์องค์ประกอบ ดัชนีความลำเอียงของข้อสอบตามวิธีนี้ก็คือความแตกต่าง (Discrepancy) ของน้ำหนักองค์ประกอบ (Factor Loading) จากผู้สอบต่างกลุ่มกัน หรือความแตกต่างอย่างมีนัยสำคัญ ระหว่างค่าเฉลี่ยของคะแนนองค์ประกอบ (Factor Scores) จากแต่ละกลุ่มที่นำมาเปรียบเทียบกัน ดัชนีความลำเอียงทั้งสองที่กล่าวมาเป็นตัวชี้ว่า ข้อสอบไม่ได้วัดสิ่งเดียวกันในผู้สอบต่างกลุ่มกัน

3. ใช้หลักการประมาณค่าโอกาสในการตอบข้อสอบแต่ละข้อถูก วิธีการที่ใช้กันทั่ว ๆ ไป คือ วิธีไคสแควร์ (Chi-Square) และวิธีโค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC) วิธีการทั้งสองนี้มีความคล้ายคลึงกันในเรื่องการใช้โอกาสในการตอบถูกที่แตกต่างกัน จาก ผู้สอบ 2 กลุ่ม หรือมากกว่า เป็นต้นนี้ความลำเอียงของข้อสอบ สำหรับข้อแตกต่างระหว่างวิธีการ ทั้งสองก็คือ วิธีไคสแควร์จะประมาณความสามารถของผู้สอบ โดยใช้คะแนนดิบ ส่วนวิธีโค้งลักษณะ ข้อสอบจะประมาณความสามารถของผู้สอบจากคุณลักษณะแฝง

ในการวิจัยครั้งนี้ ผู้วิจัยเลือกใช้วิธีการวิเคราะห์ความลำเอียงของข้อสอบโดยไม่ใช้เกณฑ์ ภายนอก ทั้งนี้เพราะในการหาตัวแปรเกณฑ์ภายนอกให้มีความเที่ยงตรงเชิงพยากรณ์และยุติธรรมนั้น เป็นสิ่งที่ทำได้ยากมาก หากตัวแปรเกณฑ์ภายนอกขาดคุณสมบัติดังกล่าวจะทำให้ผลการวิเคราะห์ความ ลำเอียงมีความคลาดเคลื่อนเกิดขึ้นได้ ด้วยเหตุนี้จึงเลือกใช้วิธีวิเคราะห์ความลำเอียงโดยใช้เกณฑ์ ภายใน 3 วิธี ได้แก่ วิธีการแปลงค่าความยากของข้อสอบ (TID) วิธีการวิเคราะห์ด้วยโค้งลักษณะ ข้อสอบที่มีพารามิเตอร์ 3 ตัว (ICC3) และการวิเคราะห์ด้วยวิธีของ แมนเทล-แฮนส์เซล (MH) เป็นกรอบทฤษฎีในการวิเคราะห์ความลำเอียงของข้อสอบซึ่งมีรายละเอียด ดังนี้

การวิเคราะห์ความลำเอียงของข้อสอบโดยยึดทฤษฎีการวัดแบบคลาสสิกอล

(Classical Test Theory)

การวิเคราะห์ความลำเอียงของข้อสอบโดยยึดทฤษฎีการวัดแบบคลาสสิกอล มีนักการศึกษา และนักวิจัยได้สร้างวิธีการวิเคราะห์ความลำเอียงของข้อสอบไว้หลายวิธี เช่น วิธีการแปลงค่า ความยากของข้อสอบ (Transformed Item Difficulty Approach) วิธีไคสแควร์ (Chi-square Methods) วิธีล็อกลิเนียร์ (Log-Linear Methods) วิธีของแมนเทล-แฮนส์เซล (Mantel-Haenzel) วิธีการบรรเทาการมาตรฐาน (Standardization Procedure) วิธีการ ถดถอยโลจิสติก (Logistic Regression Method) การวิเคราะห์ตัวลวง (Distractor Analysis) วิธี Golden Rule วิธีการวิเคราะห์องค์ประกอบ (Factor Analysis) วิธีการ วิเคราะห์การถดถอย (Regression Analysis) กระบวนการค่าอำนาจจำแนกของข้อสอบ (Item Discrimination Procedures) เป็นต้น แต่วิธีการวิเคราะห์ความลำเอียงของข้อสอบ ที่นักศึกษานิยมใช้ส่วนมากมีดังนี้ คือ

1. การวิเคราะห์ความแปรปรวน (Analysis of Variance)
2. วิธีไคสแควร์ (Chi-Square ; χ^2)
3. วิธีแปลงค่าความยากของข้อสอบ (Transformed Item Difficulty Approach)
4. กระบวนการแมนเทล-แฮนสเซล (Mantel-Haenszel Procedure)

การวิเคราะห์ความแปรปรวน (ANOVA)

การวิเคราะห์ความแปรปรวน เป็นวิธีหนึ่งที่สามารถนำมาประยุกต์ใช้ในการหาความลำเอียงของข้อสอบได้ โดยถือว่าข้อสอบที่ไม่ได้วัดความสามารถเดียวกันจากผู้สอบต่างกลุ่มกันเป็นข้อสอบที่มีความลำเอียง ซึ่งหมายความว่าถ้ามีข้อสอบที่ง่ายสำหรับกลุ่มหนึ่ง แต่ยากสำหรับอีกกลุ่มหนึ่ง ก็อาจพิจารณาได้ว่าข้อสอบนั้นเป็นข้อสอบที่มีความลำเอียง และยังเป็นการแสดงให้เห็นอีกว่าข้อสอบนั้นไม่ได้วัดความสามารถเดียวกันในผู้สอบแต่ละกลุ่ม วิธีการวิเคราะห์นี้จะใช้วิธีวิเคราะห์ความแปรปรวนแบบ 2 ตัวประกอบ คือการจัดข้อสอบและกลุ่มผู้สอบลงใน Two-factor Design แล้วทำการทดสอบปฏิสัมพันธ์ระหว่างข้อสอบกับกลุ่มผู้สอบ ซึ่งมีรูปแบบการวิเคราะห์ ดังนี้

$$Y_{ij} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ij}$$

เมื่อ	μ	แทน	ค่าเฉลี่ยของทั้งหมด
	α_i	แทน	อิทธิพลจากข้อสอบ ข้อที่ i ($i = 1, 2, \dots, I$)
	β_j	แทน	อิทธิพลจากกลุ่มผู้สอบ กลุ่มที่ j ($j = 1, 2, \dots, J$)
	$(\alpha\beta)_{ij}$	แทน	ปฏิสัมพันธ์ระหว่างข้อสอบและกลุ่มผู้สอบ
	ϵ_{ij}	แทน	ความคลาดเคลื่อนแบบสุ่มภายในเซลล์ i, j

การวิเคราะห์ความลำเอียงโดยวิธีนี้มีข้อตกลงว่า แบบทดสอบจะต้องมีลักษณะเป็นเอกพันธ์ นั่นคือ แบบทดสอบนั้นต้องวัดความสามารถหรือคุณลักษณะเพียงลักษณะเดียว และยังมีข้อตกลงเบื้องต้นอีก 2 ประการ คือ

1. ความแปรปรวนของข้อสอบแต่ละข้อต้องเท่ากัน

2. ความแปรปรวนร่วมของข้อสอบแต่ละคู่ต้องเท่ากัน

รัดเนอร์ (Rudner. 1977 ; Cited in Laksana. 1977) ได้กล่าวว่า ในการศึกษาเกี่ยวกับการตอบข้อสอบนั้นมีแหล่งความแปรปรวนที่ควรพิจารณาหลายประการ ดังนี้

1. ข้อสอบ (Items) มีข้อสอบบางข้อที่ยากกว่าข้ออื่น
 2. กลุ่มผู้สอบ (Groups) กลุ่มผู้สอบหนึ่งอาจจะถูกวัดคุณลักษณะที่สูงกว่าอีกกลุ่มหนึ่ง
 3. ความแปรปรวนในกลุ่ม (Subject Within Groups) ผู้สอบในแต่ละกลุ่มมีความสามารถแตกต่างกัน
 4. ผลของปฏิสัมพันธ์ระหว่างข้อสอบกับกลุ่มผู้สอบ
- วิธีการหาความลำเอียงของข้อสอบโดยวิธีการวิเคราะห์ความแปรปรวนนั้นมีขั้นตอน ดังนี้
1. จัดข้อสอบ และกลุ่มผู้สอบ ลงในตัวแปรผลการตอบข้อสอบและกลุ่มผู้สอบ
 2. คำนวณสัดส่วนในการตอบข้อสอบแต่ละข้อระหว่างกลุ่มผู้สอบ 2 กลุ่ม
 3. ทดสอบปฏิสัมพันธ์ระหว่างข้อกับกลุ่มผู้สอบ หากพบว่ามีปฏิสัมพันธ์อย่างมีนัยสำคัญทางสถิติ แสดงว่าแบบทดสอบฉบับนั้นมีความลำเอียง
 4. หากการทดสอบในข้อ 3 มีนัยสำคัญทางสถิติให้ทำการทดสอบต่อเป็นรายข้อว่ามีข้อใดบ้างที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ ข้อที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติก็คือ ข้อที่มีความลำเอียงนั่นเอง

การวิเคราะห์ความลำเอียงด้วยวิธีไคสแควร์ (Chi-Square : χ^2)

ผู้ที่เสนอการวิเคราะห์ความลำเอียงวิธีนี้คือ ชูเนแมน (Berk. 1982 :120 ; Citing Scheuneman. 1975, 1979) โดยมีหลักการว่า ข้อสอบที่ไม่ลำเอียง หมายถึง ข้อสอบที่มีความน่าจะเป็นของการตอบถูกในกลุ่มผู้สอบที่มีความสามารถเหมือนกันจะเท่ากัน โดยไม่คำนึงถึงความแตกต่างของวัฒนธรรม เชื้อชาติ เพศ หรืออื่นๆ ความสามารถของผู้สอบวัดได้โดยคะแนนรวมของข้อสอบที่มีความเป็นเอกพันธ์หรือตัวแทนของข้อสอบที่เราสนใจ วิธีไคสแควร์ (Chi-Square) มีหลายวิธีด้วยกัน แต่ที่นิยมใช้ในการวิเคราะห์ความลำเอียงมีอยู่ 2 วิธี ดังนี้

1. วิธีไคสแควร์ของ ชูเนแมน (Scheuneman. 1979) ใช้ค่าของไคสแควร์ที่ใช้กันทั่วไปเป็นดัชนีความลำเอียงของข้อสอบโดยแบ่งคะแนนรวมของผู้สอบที่ใช้แทนความสามารถออกเป็นช่วงๆ (ประมาณ 3 ถึง 5 ช่วง) ก็จะสามารถคำนวณค่าไคสแควร์ที่ใช้เป็นค่าดัชนีความลำเอียงได้

หรือจะตรวจสอบความมีนัยสำคัญทางสถิติ ซึ่งข้อสอบที่มีความลำเอียง คือ ข้อที่มีนัยสำคัญทางสถิติ
วิธีนี้ ชูเนแมนได้กำหนดให้ $C2$ เป็นค่าดัชนีความลำเอียง ที่ขึ้นอยู่กับเฉพาะความถี่ของการตอบ
ข้อสอบทุกเท่านั้น ค่า $C2$ คำนวณได้จากสูตร

$$C2 = \sum_{j=1}^J \frac{(E_{1j} - O_{1j})^2}{E_{1j}} + \sum_{j=1}^J \frac{(E_{2j} - O_{2j})^2}{E_{2j}} ; \quad df = (I-1)(J-1)$$

เมื่อ $C2$	แทน	Scheuneman's Chi-Square
O_{1j}, O_{2j}	แทน	ความถี่ที่สังเกตได้ของการตอบข้อสอบทุกของกลุ่มที่ 1 และกลุ่มที่ 2 ในแต่ละช่วงคะแนนที่ j
$E_{1j}, E_{2j} = P_j N_{ik}$	แทน	ความถี่ที่คาดหวังไว้ของการตอบข้อสอบทุกของกลุ่มที่ 1 และกลุ่มที่ 2 ในแต่ละช่วงคะแนนที่ j
I	แทน	จำนวนกลุ่มที่ทำการวิเคราะห์
J	แทน	จำนวนช่วงคะแนน
df	แทน	degrees of freedom

2. วิธีโคสแควร์แบบเติมรูป วิธีนี้ ชูเนแมน (Scheuneman, 1981) สร้างขึ้นเพื่อแก้ไข
ข้อบกพร่อง ของค่าดัชนี $C2$ ที่ได้จากวิธีที่ 1 ในด้านการแจกแจงที่ไม่เป็นการแจกแจงอย่างสม่ำเสมอ
และการทดสอบนัยสำคัญทางสถิติที่ไม่เหมาะสม โดยการนำเอาสัดส่วนของการตอบผิดในแต่ละช่วง
รวมเข้าไปด้วยกัน ทำให้ได้สูตรใหม่ ดังนี้

$$\chi^2_{c_{u11}} = \sum_{j=1}^J \frac{(E_{1j} - O_{1j})^2}{E_{1j}(1-P_{0j})} + \sum_{j=1}^J \frac{(E_{2j} - O_{2j})^2}{E_{2j}(1-P_{0j})} ; \quad df = (I-1)(J-1)$$

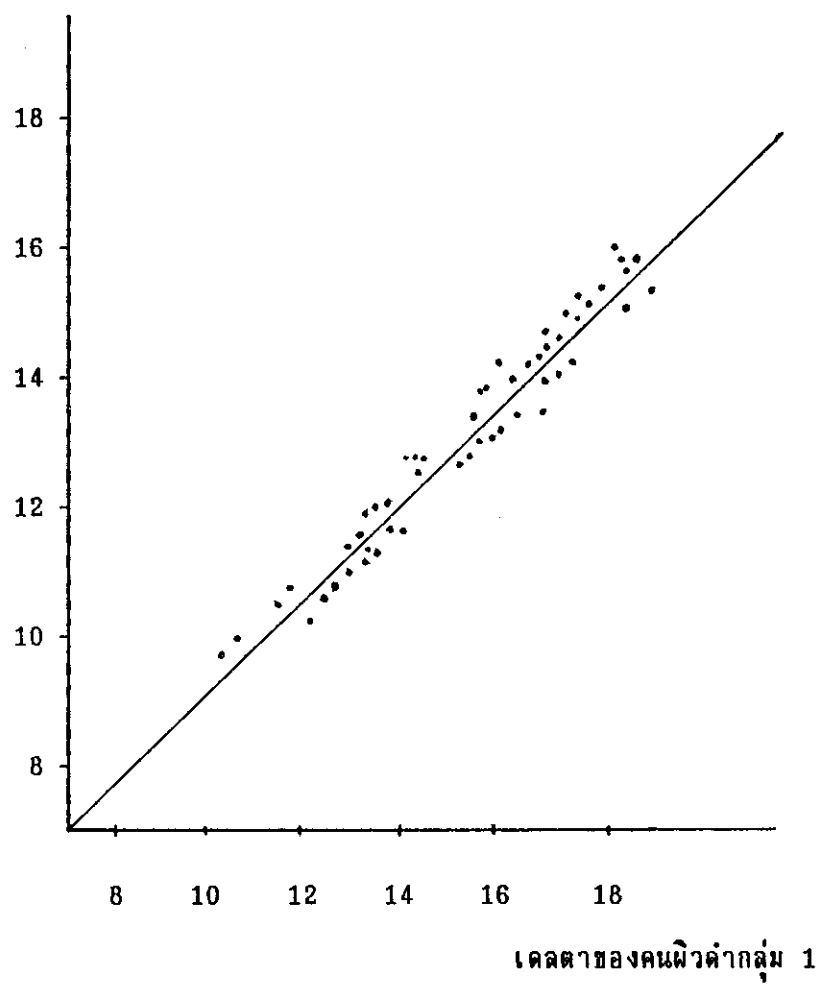
เมื่อ P_j แทน สัดส่วนของผู้สอบที่ตอบข้อสอบถูกในช่วงคะแนนที่ j
 N_{jk} แทน จำนวนของผู้สอบในช่วงคะแนนที่ j ของกลุ่มที่ k

ข้อได้เปรียบของวิธีไคสแควร์ คือ เป็นวิธีที่ใช้กลุ่มตัวอย่างขนาดเล็กได้และคำนวณได้ง่าย ลื่นเปลืองน้อย และที่สำคัญก็คือใช้การจับคู่ระหว่างกลุ่มต่าง ๆ กับระดับความสามารถ แต่จากการศึกษา พบว่า ค่าดัชนี $C2$ และไคสแควร์แบบเต็มรูป มีแนวโน้มให้ผลที่เหมือนกันมาก (Scheuneman, 1986) อาจจะต่างกันที่จุดศัณยฆาตเท่านั้น แต่เมื่อกลุ่มตัวอย่างมีขนาดใหญ่หรือมีจำนวนไม่น้อยกว่า 300 คน ไคสแควร์แบบเต็มรูป มีแนวโน้มที่ให้ค่าบางอย่างที่ดีกว่า และใช้ได้ดีกว่า ในการวิเคราะห์ความลำเอียงของข้อสอบที่ยาก ๆ ส่วนดัชนี $C2$ ใช้ได้กับกลุ่มตัวอย่างตั้งแต่ 100 คนขึ้นไปและจะใช้ได้ดี สำหรับการวิเคราะห์ความลำเอียงของข้อสอบที่ง่าย ๆ แต่ทั้งสองวิธีก็ถูกวิจารณ์ว่า กลุ่มที่นำมาเปรียบเทียบแต่ละกลุ่ม อาจจะมีความสามารถไม่เท่ากัน ภาย ในความกว้างของแต่ละช่วงคะแนน

การแปลงค่าความยากของข้อสอบ (Transformed Item Difficulty)

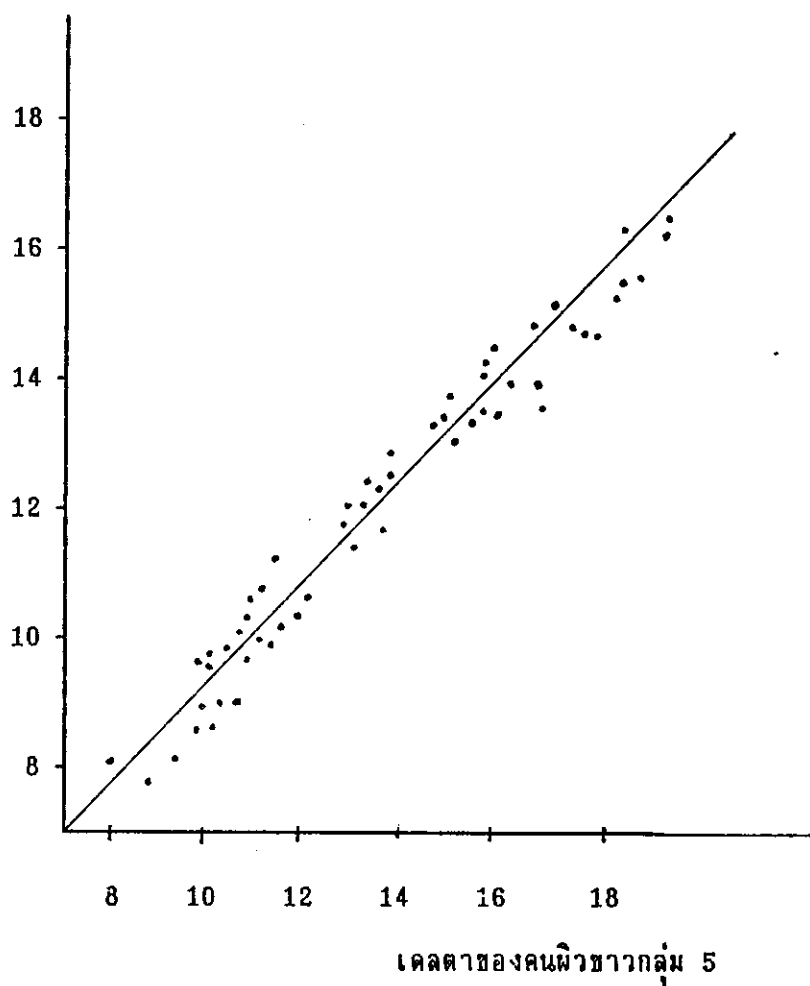
ผู้ที่เสนอการวิเคราะห์ความลำเอียงของข้อสอบวิธีนี้ คือ แองกอฟฟ์ (Berk, 1982 : 97 ; Citing Angoff, 1972 ; also Angoff & Ford, 1973) การวิเคราะห์ความลำเอียงของข้อสอบตามวิธีการของ แองกอฟฟ์ ก็คือจะต้องคำนวณหาค่าความยาก (P-Values) ของข้อสอบแต่ละข้อจากผู้สอบแต่ละกลุ่ม (คือกลุ่ม 1 กลุ่ม 2 ดังนั้นค่าความยากของข้อสอบแต่ละข้อจะมี 2 ค่า ตามจำนวนกลุ่มที่แบ่ง) ในการเปรียบเทียบค่าความยาก (P-Values) นี้จะใช้วิธีการแปลงค่าความยาก (P-Values) เป็นคะแนนมาตรฐาน Z จากคะแนนมาตรฐาน Z แปลงให้เป็นค่าความยากมาตรฐาน เดลตา (Δ) (Delta) ด้วยสมการ $\Delta = 4Z + 13$ แล้วนำค่าเดลตา (Delta) แต่ละค่าลงจุดคู่อันดับบนกราฟ จุดต่าง ๆ บนกราฟ ซึ่งแทนค่าเดลตาจะอยู่ในรูปวงรี แล้วคำนวณหาระยะทางที่จุดต่างๆ เหล่านี้ห่างจากเส้นแกนหลัก (Major Axis) ข้อสอบข้อใดที่ห่างจากเส้นแกนหลักน้อยกว่า $-3Sd$ หรือมากกว่า $3Sd$ จะสรุปว่าข้อสอบข้อนั้นเป็นข้อสอบที่มีความลำเอียง ดังตัวอย่างในภาพประกอบ 1, 2 และ 3

เดลตาของคณพืวค้ากลุ่ม 2



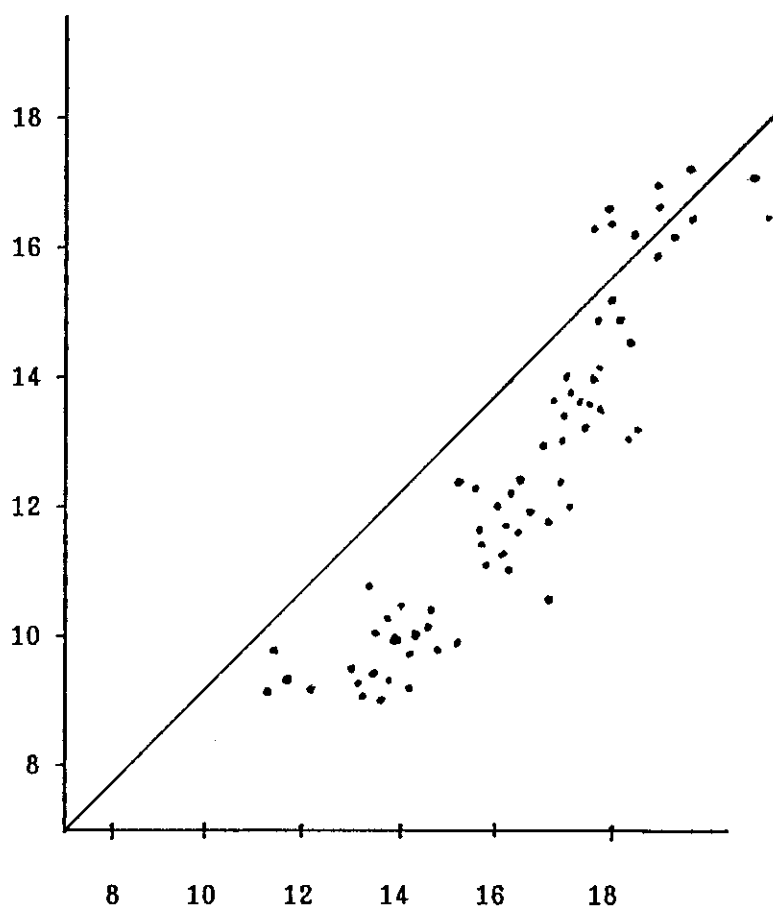
ภาพประกอบ 1 แสดงกราฟของค่าเดลตา ของคณพืวค้ากลุ่ม 1 และคณพืวค้ากลุ่ม 2 จากแบบทดสอบวัดความถนัดทางการเรียนด้านภาษา (Preliminary Scholastic Aptitude (PSAT)-Verbal) (Berk. 1982 : 99 ; Citing Angoff & Ford. 1973)

เดคลาของคณพืชาวกลุ่ล 6



ภาพประกอบ 2 แสดงกราฟของค่าเดคลา ของคณพืชาวกลุ่ล 5 และคณพืชาว กลุ่ล 6 จากแบบทดสอบวัดความถนัดทางการเรียนด้านภาษา (Preliminary Scholastic Aptitude (PSAT)-Verbal) (Berk. 1982 : 100 ; Citing Angoff & Ford. 1973)

เดลตาของคนผิวขาวกลุ่ม 5



เดลตาของคนผิวดำกลุ่ม 1

ภาพประกอบ 3 แสดงกราฟ ของค่าเดลตา ของคนผิวขาวกลุ่ม 5 และคนผิวดำ กลุ่ม 1 จากแบบทดสอบวัดความถนัดทางการเรียนด้านภาษา (Preliminary Scholastic Aptitude (PSAT)-Verbal) (Berk. 1982 : 101 ; Citing Angoff & Ford. 1973)

เมื่อลงจุดอันดับค่าเฉลี่ยของข้อสอบแต่ละข้อของผู้เข้าสอบแต่ละกลุ่มแล้ว จะเห็นว่า ข้อสอบที่ไม่มีความลำเอียงจะได้กราฟ ดังภาพประกอบ 1 และภาพประกอบ 2 ซึ่งภาพประกอบ 1 เป็นกราฟของค่าเฉลี่ยจากกลุ่มตัวอย่างคนผิวดำ 2 กลุ่ม โดยใช้แบบทดสอบวัดความถนัดทางการเรียนด้านภาษา (Berk. 1982 : 99 ; Citing Angoff & Ford. 1973) ภาพประกอบ 2 เป็นกราฟของค่าเฉลี่ยที่ได้จากการทดสอบกลุ่มตัวอย่างคนผิวขาว 2 กลุ่ม โดยใช้แบบทดสอบชุดเดียวกันจะเห็นว่าจุดต่าง ๆ ของค่าเฉลี่ยของกลุ่มผู้เข้าสอบ 2 กลุ่ม จะกระจายอยู่ในแนวเส้นแกนหลักคือ ไม่มีการเบี่ยงเบนไปจากเส้นแกนหลัก

ส่วนภาพประกอบ 3 เป็นกราฟของค่าเฉลี่ยที่ได้จากการทดสอบกลุ่มตัวอย่างคนผิวดำกับคนผิวขาว โดยใช้แบบทดสอบชุดเดียวกันจะเห็นว่าจุดต่าง ๆ ของค่าเฉลี่ยของผู้เข้าสอบ 2 กลุ่มนี้ จะกระจายห่างจากเส้นแกนหลัก คือ มีการเบี่ยงเบนไปจากเส้นแกนหลัก ซึ่งข้อสอบเหล่านี้ยากต่อกลุ่มผู้สอบคนผิวดำ ดังนั้น แบบทดสอบชุดนี้มีปฏิสัมพันธ์ระหว่างข้อสอบกับเชื้อชาติ (Item Race Interaction) และเรียกได้ว่าเป็นข้อสอบที่มีความลำเอียงต่อเชื้อชาติ ซึ่งคำนวณได้ดังนี้

$$\text{จากสมการเส้นแกนหลัก} \quad y = ax + b$$

เมื่อ a แทน ความชันของเส้นแกนหลัก

b แทน ค่าคงที่ของการตัดแกน Y

โดยที่ค่า a และ b คำนวณได้จากสูตร

$$a = \frac{(S_y^2 - S_x^2) \pm \sqrt{(S_y^2 - S_x^2)^2 + 4r_{xy}^2 S_x^2 S_y^2}}{2r_{xy} S_x S_y}$$

$$b = M_y - aM_x$$

เมื่อ	x	แทน	ค่าเฉลี่ยของกลุ่มที่ 1
	y	แทน	ค่าเฉลี่ยของกลุ่มที่ 2
	M_x	แทน	ค่าเฉลี่ยของค่าเฉลี่ยของกลุ่มที่ 1
	M_y	แทน	ค่าเฉลี่ยของค่าเฉลี่ยของกลุ่มที่ 2
	S_x^2	แทน	ค่าความแปรปรวนของค่าเฉลี่ยกลุ่มที่ 1
	S_y^2	แทน	ค่าความแปรปรวนของค่าเฉลี่ยกลุ่มที่ 2
	S_x	แทน	ค่าความเบี่ยงเบนมาตรฐานของค่าเฉลี่ยกลุ่มที่ 1
	S_y	แทน	ค่าความเบี่ยงเบนมาตรฐานของค่าเฉลี่ยกลุ่มที่ 2
	r_{xy}	แทน	สหสัมพันธ์ของค่าเฉลี่ย กลุ่มที่ 1 และกลุ่มที่ 2

แองกอฟฟ์ (Angoff) เสนอการคำนวณระยะทางตั้งฉากจากจุดคู่อันดับค่าเฉลี่ยของ
ข้อสอบแต่ละข้อไปยังเส้นแกนหลักจากสูตร

$$d_i = \frac{ax_i - y_i + b}{\sqrt{a^2 + 1}}$$

$$\text{และ } S_d = \sqrt{1 - r_{xy}}$$

เมื่อ	d_i	แทน	ระยะทางตั้งฉากจากจุดคู่อันดับค่าเฉลี่ยของ ข้อสอบแต่ละข้อไปยังเส้นแกนหลัก
	S_d	แทน	ค่าเบี่ยงเบนมาตรฐานของระยะ d

การวิเคราะห์ความลำเอียงของข้อสอบโดยวิธีของแมนเทล-แฮนส์เซล (Mantel-Haenszel)

เทคนิคแมนเทล-แฮนส์เซล (MH) เป็นเทคนิคที่แมนเทลและแฮนส์เซลได้เสนอขึ้นใช้ตั้งแต่ปี 1959 แต่ฮอลแลนด์และเทเซอร์ (Holland and Thayer) ได้นำเสนอเพื่อใช้ในการวิเคราะห์ความลำเอียงของข้อสอบในปี 1986 และในปีนั้นก็ได้มีการวิเคราะห์ความลำเอียงของข้อสอบด้วยเทคนิคแมนเทล-แฮนส์เซล (MH) โดย โดแรนส์ (Dorans) แฮมเบิลตัน โรเจอร์ส และอะราสมิท (Hambleton, Rogers and Arrasmith) ฮอลแลนด์ เทเซอร์ และไรท์ (Wright) (Raju, Bode and Larsen. 1989 : 2)

เทคนิค แมนเทล-แฮนส์เซล (MH) เป็นเทคนิคที่มีความคล้ายคลึงกับวิธีไคสแควร์ที่เสนอโดยชูเนแมน (1979) มาราสคูโลและสโลเตอร์ (Marascuilo and Slayghter. 1981) และเมลเลนเบิร์ก (Mellenberg. 1982) โดยเป็นการเปรียบเทียบผลการตอบข้อสอบของกลุ่มผู้สอบ 2 กลุ่ม หรือที่ฮอลแลนด์เรียกว่า กลุ่มอ้างอิงและกลุ่มเปรียบเทียบกับ (reference and focal group) โดยที่กลุ่มแรกจะใช้อ้างอิงถึงกลุ่มที่คาดว่าจะได้ประโยชน์จากข้อสอบและกลุ่มหลังเป็นกลุ่มเสียประโยชน์จากข้อสอบในกรณีที่ข้อสอบมีความลำเอียง โดยมีการตรวจสอบทุก ๆ ระดับคะแนนรวมจากการสอบ ข้อสอบข้อใดที่ผู้สอบทั้ง 2 กลุ่ม ทำได้เท่าๆ กัน จะเป็นข้อสอบที่ถือได้ว่าไม่มีความลำเอียงต่อกลุ่มใดกลุ่มหนึ่ง วิธีการในการวิเคราะห์มี ดังนี้

ถ้าให้ $N_{r,j}$ และ $N_{f,j}$ แทน จำนวนผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบกับที่มีคะแนนรวมในช่วงคะแนน j และให้ T_j ซึ่งเท่ากับ $N_{r,j} + N_{f,j}$ เป็นจำนวนผู้สอบรวมทั้ง 2 กลุ่ม ที่ได้คะแนนรวมจากการสอบอยู่ในช่วงคะแนน j แล้วนำมาเขียนในรูปตาราง 2 ทาง แสดงผลการตอบถูก (1) และผิด (0) ดังตารางที่ 1

ตาราง 1 ข้อมูลผลการสอบของกลุ่มตัวอย่างกลุ่ม R และ F ที่มีคะแนนรวมอยู่ใน ช่วงคะแนน j

กลุ่มผู้สอบ	คะแนนที่ได้จากข้อสอบที่ต้องการวิเคราะห์ความลำเอียง			
		1	0	รวม
	R	A_j	B_j	$N_{R,j}$
	F	C_j	D_j	$N_{F,j}$
	รวม	$M_{1,j}$	$M_{0,j}$	T_j

เมื่อ A_j , B_j , C_j และ D_j เป็นความถี่ที่สังเกตได้ของการตอบถูก(1)และตอบผิด(0) ในช่วงคะแนน j และแสดงในรูปสัดส่วนได้ ดังตารางที่ 2

ตาราง 2 สัดส่วนการตอบข้อสอบของกลุ่มประชากร 2 กลุ่ม

กลุ่ม ผลการตอบ	1	0	รวม
R	$P_{R,j}$	$Q_{R,j}$	1
F	$P_{F,j}$	$Q_{F,j}$	1

โดยที่ $P_{R,j}$ แทน สัดส่วนของกลุ่มอ้างอิงที่อยู่ในช่วงความสามารถ j ที่ตอบข้อสอบถูก

$P_{F,j}$ แทน สัดส่วนของกลุ่มสนใจที่อยู่ในช่วงความสามารถ j ที่ตอบข้อสอบถูก

$Q_{R,j}$ แทน $1 - P_{R,j}$

$Q_{F,j}$ แทน $1 - P_{F,j}$

จากตาราง 1 จะได้ว่า ในช่วงระดับคะแนน j A_j และ B_j คือ จำนวนผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบข้อที่ถูกต้อง และไม่ถูกต้อง ตามลำดับ ในขณะที่ C_j และ D_j คือ จำนวนผู้สอบกลุ่มเปรียบเทียบที่ตอบข้อสอบข้อที่ถูกต้องและไม่ถูกต้อง ตามลำดับ เช่นกัน

ตามเทคนิควิธี แมนเทิล-แฮนส์เฟลด์ (MH) ที่เสนอโดยฮอลล์แลนด์และเชสเตอร์ นั้น ข้อมูลการตอบข้อสอบแต่ละข้อจะถูกนำมาแสดงในรูปตาราง 2 ทาง เป็นจำนวนตารางเท่ากับคะแนนรวมที่มีผู้สอบทำได้ บวกด้วย 1 จากนั้นจะคำนวณค่าความเป็นไปได้ในรูปของค่าสัดส่วนการตอบข้อสอบถูก-ผิด ระหว่างกลุ่มโดยกำหนดให้เป็นค่า α ดังนี้

$$\alpha_{MH} = \frac{\sum_{j=1}^S A_j D_j / T_j}{\sum_{j=1}^S B_j C_j / T_j}$$

ค่า α_{MH} จะมีค่าระหว่าง 0 และ ∞ โดยที่ $\alpha_{MH} = 1$ จะแสดงถึงสมมติฐานทางสถิติที่ว่าไม่มีความแตกต่างหรือไม่มีความลำเอียง ซึ่งแสดงว่ากลุ่มตัวอย่างทั้ง 2 กลุ่ม ทำข้อสอบได้ผลเช่นเดียวกันในแต่ละระดับคะแนนรวม ค่า $A_j D_j / T_j$ และ $B_j C_j / T_j$ มีค่าเท่ากันทั้ง 2 กลุ่ม ทำให้ α_{MH} มีค่าเท่ากับ 1 ถ้าค่า α_{MH} มีค่ามากกว่า 1.00 นั่นคือ สมาชิกของกลุ่มที่ 1 จะตอบข้อสอบข้อนั้นได้ถูกต้องมากกว่าสมาชิกในกลุ่มที่ 2 แต่ถ้าหากว่าค่า α_{MH} น้อยกว่า 1.00 แสดงว่าสมาชิกกลุ่มที่ 2 ตอบข้อสอบข้อนั้นได้ถูกต้องมากกว่าสมาชิกในกลุ่มที่ 1 ซึ่งแสดงถึงความแตกต่างหรือความลำเอียงที่เกิดขึ้นระหว่างกลุ่มตัวอย่าง 2 กลุ่ม

อย่างไรก็ตามในทางปฏิบัติแล้ว แม้ว่าจะมีการควบคุมให้ผู้สอบ 2 กลุ่ม มีความสามารถในการที่ต้องการวัดเท่ากัน แต่เป็นไปได้ที่ผลการตอบข้อสอบของผู้สอบทั้ง 2 กลุ่ม จะเป็นเช่นเดียวกัน ดังกล่าว แมนเทิล-ฮันส์เชล (1959) จึงเสนอค่าสถิติไคสแควร์เพื่อทดสอบค่าที่ได้ว่าจะมีความแตกต่างจาก 1.00 อย่างมีนัยสำคัญทางสถิติหรือไม่ ที่ระดับขึ้นแห่งความเป็นอิสระเท่ากับ 1 ดังนี้

$$MH-CHISQ = \frac{\left[\sum_{j=1}^S \left| \frac{A_j D_j}{T_j} - E(A_j D_j) \right| - 0.5 \right]^2}{\sum_{j=1}^S \text{Var}(A_j D_j)}$$

โดยที่ค่าผลรวมเป็นค่าที่ได้จากการรวมทุกชั้นของคะแนนรวม และค่า $E(A_j)$ มีค่าดังนี้

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

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

และสมมติฐานที่แสดงว่าข้อสอบไม่ลำเอียง ได้แก่ สัดส่วน จำนวนคนตอบถูกในกลุ่มอ้างอิง เท่ากับสัดส่วนจำนวนคนตอบถูกในกลุ่มเปรียบเทียบกับ สำหรับทุกชั้นของคะแนน j

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

สมมติฐานทางสถิตินี้เป็นสมมติฐานของความเป็นอิสระอย่างมีเงื่อนไขของสมาชิกกลุ่ม และคะแนนที่ได้จากการตอบข้อสอบที่ต้องการตรวจสอบความลำเอียง และภายใต้สมมติฐานดังกล่าว จะได้ค่าคาดหวังของการตอบข้อสอบ ดังนี้

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

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

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

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

สมมติฐานอื่นของ แมนเทิล-แฮนส์เชล ได้แก่

$$H_1 : \frac{P_{R,j}}{Q_{R,j}} \neq \alpha \frac{P_{F,j}}{Q_{F,j}} ; j = 1, 2, \dots, k \text{ เมื่อ } \alpha \neq 1$$

เมื่อ $\alpha = 1$ ซึ่งสอดคล้องกับสมมติฐานทางสถิติ จะได้ว่า

$$H_0 : \frac{P_{R,j}}{Q_{R,j}} = \frac{P_{F,j}}{Q_{F,j}}$$

$$\text{และค่าประมาณของ } \alpha_{MH} = \frac{P_{R,j}}{Q_{R,j}} \cdot \frac{P_{F,j}}{Q_{F,j}} = \frac{P_{R,j} Q_{F,j}}{P_{F,j} Q_{R,j}} \text{ สำหรับทุก ๆ } j = 1, \dots, k$$

นอกจากนั้น ซอลแลนด์และเกเซอร์ ได้เสนอแนะให้แปลงค่า α_{MH} ให้เป็นเดลตา (Δ) เพื่อให้เป็นค่าที่มีค่าเฉลี่ยเป็น 0 ดังนี้

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

ค่า Δ_{MH} มีค่าอยู่ระหว่าง -2.6 ถึง 2.6 และใช้เป็นดัชนีแสดงความลำเอียงของข้อสอบ
ถ้า α_{MH} มีค่าเท่ากับ 1 หรือ Δ_{MH} มีค่าเท่ากับ 0 แล้ว แสดงว่า ข้อสอบข้อนั้นไม่ลำเอียง

จากที่ผ่านมามีผู้นำวิธีการวิเคราะห์ความลำเอียงของแมนเทล-แฮนส์เชล (MH) มาศึกษาหลายท่าน เช่น คลอส์เชอร์และคนอื่นๆ (Clauser and Others. 1991) ใช้วิธีแมนเทล-แฮนส์เชลวิเคราะห์ความลำเอียงอย่างสม่ำเสมอ พบว่า วิธีแมนเทล-แฮนส์เชล จะใช้ได้ดีในกรณีที่ข้อสอบมีค่าอำนาจจำแนกสูง แต่จะไม่สามารถวิเคราะห์ข้อสอบที่มีความยากสูงมากได้ ซึ่งสอดคล้องกับข้อค้นพบของ เมเซอร์ และ คนอื่นๆ (Mazor and Others. 1991) ซูดวีคส์ และทอลแมน (Sudweeks and Tolman. 1990) ที่พบว่า ข้อสอบที่ลำเอียงมักเป็นข้อสอบที่ยาก บาภิและเฟอรารา (Baghi and Ferrara. 1990) พบว่า เมื่อใช้ขนาดกลุ่มตัวอย่าง 750 คนขึ้นไป วิธีแมนเทล-แฮนส์เชลให้แนววิธีใช้ทฤษฎีการตอบข้อคำถาม 3 พารามิเตอร์ได้ สวามินาธาน และ โรเจอร์ส (Swaminathan and Rogers. 1990) ซึ่งศึกษาจากข้อมูลจำลอง พบว่า วิธีแมนเทล-แฮนส์เชลวิเคราะห์ได้ดีกว่าการถดถอยแบบโลจิสต์เล็กน้อย โดยตรวจค้นได้ถูกต้อง ร้อยละ 75 ในกรณีใช้กลุ่มตัวอย่าง 250 คน และตรวจค้นได้ถูกต้องร้อยละ 100 กรณีกลุ่มตัวอย่าง 500 คน กรณีที่ลำเอียงอย่างสม่ำเสมอและกรณีไม่สม่ำเสมอที่ติดกันที่ปลายข้างใดข้างหนึ่ง แต่วิธีแมนเทล-แฮนส์เชลมีค่าใช้จ่ายน้อยกว่าโลจิสต์ประมาณ 3-4 เท่า แฮมเบิลตันและคนอื่นๆ (Hambleton and Others. 1986) พบว่า วิธีแมนเทล-แฮนส์เชล ให้ค่าใกล้เคียงกับวิธีของทฤษฎีการตอบข้อคำถาม ทั้งที่ใช้ค่าความแตกต่างของค่าเฉลี่ยกำลังสองและการตรวจสอบความแตกต่างของพื้นที่รวมได้คง แต่วิธีแมนเทล-แฮนส์เชล มีค่าใช้จ่ายต่ำกว่าและใช้เวลาน้อยกว่า ลินเคอร์ (Linacre. 1988) พบว่า วิธีแมนเทล-แฮนส์เชล ใช้ได้ดีในสถานการณ์จำลองทุกสถานการณ์พอ ๆ กับโปรแกรม PROX ของ ราสช์ (Rasch) แต่ควรใช้เกณฑ์ในการคำนวณความลำเอียงและความคลาดเคลื่อนหลายๆ เกณฑ์ และใช้ตัวประมาณค่าความคลาดเคลื่อนมาตรฐานมากกว่า 1 ตัว

ทฤษฎีการตอบข้อคำถาม (Item Response Theory)

ทฤษฎีการตอบข้อคำถาม (Item Response Theory) บางครั้ง เรียกว่าทฤษฎีคุณลักษณะแฝง (Latent trait Theory) หรือทฤษฎีโค้งลักษณะข้อสอบ (Item Characteristic Curve Theory) หรือทฤษฎีการทดสอบแบบใหม่ (Modern Test Theory) เป็นต้น (พวงจิต อินทสุวรรณ. 2525 : 54) เป็นทฤษฎีที่ศึกษาเกี่ยวกับความสัมพันธ์ของการตอบข้อสอบของผู้สอบและคุณลักษณะภายใน (Latent Trait) ของผู้สอบ กล่าวคือ การตอบข้อสอบของผู้สอบสามารถได้รับการทำนายได้ด้วยคุณลักษณะเฉพาะของผู้สอบ หรือกล่าวอีกนัยหนึ่งก็คือเราสามารถนำระดับความสามารถหรือคุณลักษณะ

ที่ประมาณได้ไปทำนาย หรืออธิบายการตอบข้อสอบถูกของผู้สอบได้ (Hambleton and Cook. 1997 : 75 ; Citing Lord and Novick. 1968 ; Hambleton and Swaminathan. 1985 : 9) เป้าหมายหลักของทฤษฎีการตอบข้อคำถาม คือ ต้องการหาคุณภาพรายข้อของแบบทดสอบและการประมาณค่าความสามารถของผู้สอบ ซึ่งต้องอาศัยความสัมพันธ์กันระหว่างโมเดลและข้อมูล ทฤษฎีนี้ได้เข้ามามีบทบาทในวงการวัดผลการศึกษาและจิตวิทยาตั้งแต่ปี 1950 โดยในระบอบการพัฒนาเป็นไปอย่างเชื่องช้ามาก เนื่องจากความยุ่งยากในการคำนวณ และขาดแคลนโปรแกรมคอมพิวเตอร์และเครื่องคอมพิวเตอร์ที่ใช้ในการวิเคราะห์ข้อมูลตามทฤษฎี จึงทำให้ชะงักการพัฒนาอยู่ระยะหนึ่ง แต่ในปัจจุบันได้มีการพัฒนาโปรแกรมคอมพิวเตอร์ขึ้นมาเรื่อยๆ จนทำให้เป็นที่นิยมนำใช้กันอย่างแพร่หลายและเป็นที่สนใจของนักวัดผลการศึกษามากยิ่งขึ้น

หลักการของทฤษฎีการตอบข้อคำถาม

ทฤษฎีการตอบข้อคำถามเป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคลกับพฤติกรรมการตอบสนองต่อข้อสอบของบุคคลนั้น ๆ โดยทฤษฎีนี้มีความเชื่อว่าพฤติกรรมการตอบสนองต่อข้อสอบของบุคคลจะถูกกำหนด โดยคุณลักษณะภายในหรือความสามารถที่มีอยู่ในตัวบุคคล (Lord and Novick. 1968 : 358) ซึ่งไม่สามารถวัดหรือสังเกตได้ โดยตรง แต่สามารถที่จะอธิบายหรือพยากรณ์ได้ จากพฤติกรรมที่บุคคลนั้นตอบสนองต่อข้อสอบหรือจากคะแนนการตอบข้อสอบ ซึ่งเป็นสิ่งที่สามารถวัดหรือสังเกตได้ในรูปแบบจำลองทางคณิตศาสตร์ที่เรียกว่า ฟังก์ชัน การตอบสนองข้อสอบ หรือ ฟังก์ชันลักษณะข้อสอบ หรือโค้งลักษณะข้อสอบ (Hambleton and Swaminathan. 1985 : 13) ฟังก์ชันทางคณิตศาสตร์ที่แสดงความสัมพันธ์ระหว่างระดับความสามารถกับพฤติกรรมการตอบสนองต่อข้อสอบ สามารถเขียนให้อยู่ในรูปสมการทางคณิตศาสตร์ได้ดังนี้ (Lord and Novick. 1968 : 360)

$$P_i(\theta) = \text{Prob}(U_i = 1 | \theta) \quad \text{เมื่อ } U_i = 0, 1$$

จากสมการข้างต้นนี้ หมายถึง โอกาสที่ผู้สอบซึ่งมีระดับความสามารถ θ จะตอบคำถามข้อ i ได้ถูกต้อง

ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อคำถาม

ทฤษฎีการตอบข้อคำถาม มีข้อตกลงเบื้องต้นดังนี้

1. ความเป็นมิติเดียว (Unidimensionality) เป็นแบบทดสอบที่ประกอบด้วยข้อคำถามที่เป็นเอกพันธ์ นั่นคือ แบบทดสอบนั้นจะต้องมุ่งวัดความสามารถอย่างใดอย่างหนึ่งที่มีลักษณะเด่น ๆ เพียงความสามารถเดียว หากไม่กำหนดข้อตกลงเบื้องต้นเช่นนี้จะทำให้โมเดลที่ใช้มีความสลับซับซ้อนมาก (Hambleton and Swaminathan. 1985 : 38) สำหรับการตรวจสอบว่าแบบทดสอบนั้นวัดในมิติเดียวหรือไม่ สามารถทำได้โดยการวิเคราะห์องค์ประกอบ (Factor Analysis) แล้วสังเกต ค่าไอเกน (Eigen Value) ขององค์ประกอบสูงสุดว่าแตกต่างจากค่าไอเกนขององค์ประกอบที่เหลืออย่างชัดเจนหรือไม่ ในขณะที่ค่าไอเกนขององค์ประกอบที่เหลือทั้งหมดมีค่าใกล้เคียงกัน (Warm. 1978 : 104 ; Lord. 1980 : 21)

2. ความเป็นอิสระในการตอบข้อสอบ (Local Independence) เป็นการกำหนดข้อตกลงเบื้องต้นเกี่ยวกับความเป็นอิสระในการตอบข้อสอบ กล่าวคือ การตอบข้อสอบข้อใดข้อหนึ่งจะไม่มีผลต่อการตอบข้อสอบข้ออื่นๆ หรือจะกล่าวในเชิงคณิตศาสตร์ได้ว่า ความเป็นอิสระในการตอบข้อสอบ หมายถึง โอกาสที่จะตอบข้อสอบถูกต้องทั้งหมดจะเท่ากับผลคูณของโอกาสการตอบข้อสอบถูกต้องเป็นรายข้อ (Hambleton and Swaminathan. 1985 : 23) ในการตรวจสอบว่าข้อสอบแต่ละข้อเป็นไปตามข้อตกลงเกี่ยวกับความเป็นอิสระในการตอบข้อสอบหรือไม่ ให้สังเกตความเป็นมิติเดียวของแบบทดสอบถ้าแบบทดสอบมีความเป็นมิติเดียวแล้ว แบบทดสอบนั้นจะมีความเป็นอิสระในการตอบข้อสอบด้วย (Lord. 1980 : 19 ; Hambleton and Swaminathan. 1985 : 24)

3. โค้งลักษณะข้อสอบ (Item Characteristic Curve) เป็นฟังก์ชันทางคณิตศาสตร์ที่แสดงถึงความสัมพันธ์ระหว่าง ความน่าจะเป็นในการตอบข้อสอบข้อนั้นได้ถูกต้องตรงกับระดับความสามารถที่วัดได้ โดยใช้ชุดของข้อสอบหรือแบบทดสอบฉบับนั้น และความน่าจะเป็นที่ผู้สอบจะตอบข้อสอบได้ถูกต้องขึ้นอยู่กับโค้งลักษณะข้อสอบของแต่ละโมเดลที่ใช้ แต่จะไม่ขึ้นอยู่กับการแจกแจงความสามารถของกลุ่มตัวอย่าง (Hambleton and Swaminathan. 1985 : 25) กล่าวคือ คุณสมบัติของโค้งลักษณะของข้อสอบจะไม่แปรเปลี่ยนไปตามกลุ่มตัวอย่าง ฉะนั้น จึงทำให้ความน่าจะเป็นในการตอบข้อสอบถูกต้องแต่ละข้อ ไม่แปรเปลี่ยนด้วย โดยผู้สอบที่มีความสามารถในระดับสูง จะมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมากกว่าผู้สอบที่มีความสามารถในระดับต่ำ

เนื่องจากมีความเชื่อเกี่ยวกับความสัมพันธ์ระหว่างความน่าจะเป็นในการตอบข้อสอบได้ ถูกต้องกับระดับความสามารถที่วัดได้หลายความเชื่อ จึงทำให้เกิด โมเดล ของโด่งลักษณะข้อสอบ ขึ้นมาหลายโมเดลด้วยกัน ดังจะกล่าวต่อไปนี้

โมเดลของทฤษฎีการตอบข้อคำถาม

ทฤษฎีการตอบข้อคำถาม ได้มีการพัฒนาในรูปแบบต่าง ๆ ทำให้เกิดโมเดลเฉพาะขึ้น ซึ่งแต่ละโมเดลจะแตกต่างกันไปตามฟังก์ชันทางคณิตศาสตร์ และจำนวนพารามิเตอร์ที่ใช้อธิบาย โด่งลักษณะข้อสอบของแต่ละโมเดล ซึ่งถ้าจะจัดแบ่งโมเดลตามลักษณะของข้อมูล หรือการตรวจ ให้คะแนน สามารถแบ่งได้ 3 ประเภท ดังนี้ (Hambleton and Swaminathan. 1985 : 35)

1. โมเดลที่มีการตรวจให้คะแนนแบบสองค่า (Dichotomous) โดยตอบถูกได้ 1 คะแนน และถ้าตอบผิดให้ 0 คะแนน แบ่งออกเป็นโมเดลต่าง ๆ ดังนี้

1.1 กัตต์แมน เฟอร์เฟ็คท์ โมเดล (Guttman Perfect Model)

1.2 ลาเทนท์ ดิสแท้นซ์ โมเดล (Latent Distance Model)

1.3 ลาเทนท์ ลินีเยร์ โมเดล (Latent Linear Model)

1.4 นอร์มัล ออร์โงว์ โมเดล ที่มีพารามิเตอร์ 1, 2, 3 ตัว (One -, Two -, Three - Parameter Normal Ogive Model)

1.5 โลจิสติก โมเดล ที่มีพารามิเตอร์ 1, 2, 3, 4 ตัว (One -, Two -, Three -, Four - Parameter Logistic Model)

1.6 ราส์ช โมเดล (Rasch Model)

2. โมเดลที่มีการตรวจให้คะแนนแบบหลายค่า (Multi-Dichotomous) โดยใช้กับ ข้อสอบที่ให้คะแนนแบบลำดับขั้น เช่น แบบทดสอบวัดเจตคติ แบบทดสอบวัดภาคปฏิบัติ และแบบ ทดสอบวัดบุคลิกภาพ เป็นต้น ซึ่งแบ่งออกเป็นโมเดลต่าง ๆ ดังนี้

2.1 นอมินอล เรสพอน โมเดล (Nominal Response Model)

2.2 เกรด เรสพอน โมเดล (Graded Response Model)

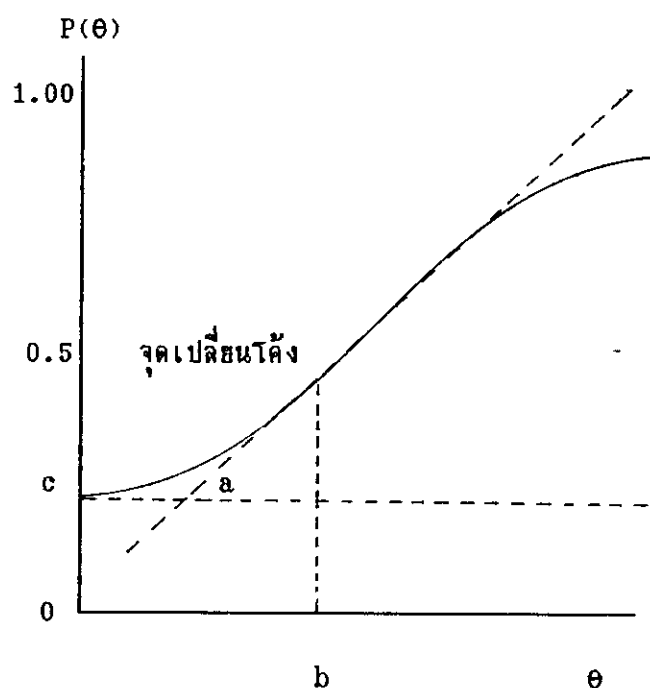
2.3 พาเชียล เครดิต โมเดล (Partial Credit Model)

3. โมเดลที่มีการตรวจให้คะแนนแบบต่อเนื่อง (Continuous) ได้แก่ คอนทินิวอัล เรสพอน โมเดล (Continuous Response Model) โมเดลนี้พัฒนาขึ้นโดย ซามีโงมา

(Samejima) (Hambleton and Swaminathan. 1985 : 52) สามารถนำไปใช้เป็นเครื่องมือของการทดสอบทางจิตวิทยา หรือการศึกษาทางด้านเจตคติโดยผู้สอบต้องตอบคำถามลงในมาตราวัดแบบต่อเนื่อง (Continuous Scale)

สำหรับในการศึกษาครั้งนี้ จะนำเสนอเฉพาะโมเดลที่มีการตรวจให้คะแนนแบบสองค่า ของ นอร์มัล ออร์ไวว์ โมเดล และโลจิสติก โมเดล ซึ่งเป็นโมเดลที่กล่าวถึงกันมาก และใช้กันแพร่หลาย ดังนี้

1. นอร์มัล ออร์ไวว์ โมเดล (Normal Ogive Model) เส้นโค้งลักษณะข้อสอบตามโมเดลนี้จะมีลักษณะเป็นรูปตัว S และแต่ละข้อจะมีโค้งลักษณะข้อสอบแตกต่างกันไปทั้งนี้จะขึ้นอยู่กับค่าพารามิเตอร์ของข้อสอบ ดังภาพประกอบ 4



ภาพประกอบ 4 แสดงเส้นโค้งลักษณะข้อสอบของนอร์มัล ออร์ไวว์ โมเดล (Normal Ogive Model)

จากภาพประกอบ 4 แสดงให้เห็นว่า โค้งลักษณะข้อสอบเป็นกราฟที่แสดงความสัมพันธ์ระหว่างโอกาสที่จะตอบข้อสอบข้อใดข้อหนึ่งถูกต้องกับระดับความสามารถของผู้สอบ โดยผู้สอบที่มีระดับความสามารถสูงมีโอกาสดังกล่าวจะตอบข้อสอบถูกมากกว่าผู้สอบที่มีระดับความสามารถต่ำ ค่าความ

สามารถและค่าความยากของข้อสอบจะอยู่ในสเกลเดียวกัน โดยมีค่าอยู่ระหว่าง $-\alpha$ ถึง $+\alpha$ แต่ในทางปฏิบัติแล้ว ระดับความสามารถจะมีค่าอยู่ระหว่าง -3 ถึง $+3$ ค่า -3 หมายถึง ผู้สอบมีระดับความสามารถต่ำมาก ค่า $+3$ หมายถึง ผู้สอบมีระดับความสามารถสูงมาก ค่าความยากจะมีค่าอยู่ระหว่าง -3 ถึง $+3$ เช่นเดียวกับค่าอำนาจจำแนกของข้อสอบ (Hambleton and Swaminathan, 1985 : 50) ส่วนค่าการเดานั้น หมายถึง ค่าปลายต่ำสุด ของโค้งลักษณะข้อสอบ ซึ่งมีค่าอยู่ระหว่าง 0 ถึง 1

ความสัมพันธ์ดังกล่าวนี้จะอยู่ในรูปฟังก์ชันการแจกแจงแบบปกติ หรือฟังก์ชันปกติ ออร์ไวจ (Cumulative Normal Ogive) ดังนี้

$$P_i(\theta) = c_i + (1 - c_i) \int_{-\alpha}^{a_i(\theta - b_i)} (1/\sqrt{2\pi}) e^{-z^2/2} dz \quad (1)$$

เมื่อ $P_i(\theta)$ แทน โอกาสที่ผู้สอบซึ่งมีความสามารถ θ จะตอบคำถามข้อที่ i ได้ถูกต้อง

a_i แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ i

b_i แทน ค่าความยากของข้อสอบข้อที่ i

c_i แทน ค่าการเดา

θ แทน ระดับความสามารถของผู้สอบ

Z แทน ความเบี่ยงเบนปกติ (Normal Deviation) จากการแจกแจง ซึ่งมีค่าเฉลี่ยเท่ากับ b_i และค่าความเบี่ยงเบนมาตรฐานเท่ากับ $1/a_i$

e แทน ค่าคงที่มีค่าประมาณ 2.71828...

ค่าพารามิเตอร์ของข้อสอบ (a_i , b_i และ c_i) จะเป็นตัวที่ทำให้รูปร่างของโค้งลักษณะของข้อสอบแต่ละข้อแตกต่างกันไป สมการ (1) เป็นสมการที่มีพารามิเตอร์ 3 ตัว คือ พารามิเตอร์

ค่าความยาก (b_i) พารามิเตอร์ค่าอำนาจจำแนก (a_i) และพารามิเตอร์ค่าการเดา (c_i) จึงอาจเรียกว่า นอร์มัล ออร์โรว์ โมเดล แบบ 3 พารามิเตอร์ (Three-parameter Normal Ogive Model) แต่หากในสถานการณ์การสอบที่ผู้สอบมีโอกาสตอบถูกโดยการเดา มีน้อย สมการ (1) ก็จะเหลือค่าพารามิเตอร์เพียง 2 ตัว คือ พารามิเตอร์ค่าความยาก (b_i) และพารามิเตอร์ค่าอำนาจจำแนก (a_i)

$$P_i(\theta) = \int_{-\infty}^{a_i(\theta - b_i)} (1/\sqrt{2\pi}) e^{-z^2/2} dz$$

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

$$-z^2/2$$

$$\frac{1}{\sqrt{2\pi}} \quad \text{แทน Normal density function}$$

b_i แทน ระดับความยากของข้อสอบข้อที่ i

a_i แทน ค่าอำนาจจำแนกของข้อสอบข้อที่ i

θ แทน ระดับความสามารถของผู้สอบ

ในกรณีที่กำหนดให้ข้อสอบทุกข้อมีค่าอำนาจจำแนกเท่ากันหมด และค่าการเดาเป็นศูนย์ โมเดล ก็จะเหลือ พารามิเตอร์ค่าความยาก (b_i) เพียงตัวเดียว ซึ่งจะเรียกว่า นอร์มัล ออร์โรว์ โมเดล ที่มีพารามิเตอร์ 1 ตัว โดยมีลักษณะของโมเดล ดังนี้ (Hambleton and Swaminathan, 1985 : 50)

$$P_i(\theta) = \int_{-\infty}^{(\theta - b_i)} (1/\sqrt{2\pi}) e^{-z^2/2} dz$$

นอร์มัล ออร์โรว์ นั้นมีปัญหามากประการทั้งในด้านทฤษฎีและคณิตศาสตร์ จึงได้มีผู้พยายามคิดหาแบบจำลองที่มีรูปร่างคล้ายคลึงกันกับ นอร์มัล ออร์โรว์ โมเดล แต่มีความสะดวกกว่าในทางคณิตศาสตร์ โมเดล ดังกล่าวนี้ก็คือ โลจิสติก โมเดล (Logistic Model)

2. โลจิสติก โมเดล (Logistic Model) โลจิสติก โมเดล เป็นโมเดลที่มีลักษณะใกล้เคียงกับนอร์มัล ออร์โรว์ โมเดล แต่สะดวกในการประมาณค่ามากกว่า นอกจากนี้ยังเป็นโมเดลที่ได้พัฒนาเพื่อให้เหมาะสมกับสถานการณ์การทดสอบ เช่น ในกรณีที่ตอบข้อสอบได้ด้วยการเดาปลายค่าสุดของโดเมนลักษณะข้อสอบไม่เป็นศูนย์ แสดงว่าการทำข้อสอบย่อมมีการเดาและในกรณีที่ผู้สอบทำข้อสอบด้วยความสะเพร่า โมเดลนี้จะเพิ่มค่าความรอบคอบขึ้นอีกหนึ่งค่า ซึ่งจะเห็นได้จากปลายสูงสุดของโดเมนลักษณะข้อสอบมีค่าน้อยกว่า 1 โลจิสติก โมเดล แบ่งเป็น 4 โมเดลย่อย ๆ ตามพารามิเตอร์ ดังนี้ (Hambleton and Swaminathan. 1985 : 34-49)

2.1 โลจิสติก โมเดล ที่มีพารามิเตอร์สองตัว (Two-Parameter Logistic Model)

เบิร์ตบอมได้เสนอโมเดลโลจิสติกที่มีค่าพารามิเตอร์ 2 ตัว คือ พารามิเตอร์ค่าความยาก (b_i) และพารามิเตอร์ค่าอำนาจจำแนก (a_i) ในปี ค.ศ. 1968 มีสมการดังนี้

$$P_i(\theta) = \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}} ; \quad i = 1, 2, \dots, n$$

เมื่อ θ แทน ระดับความสามารถที่แท้จริงที่คำนวณจากคะแนนจริง แต่ปรับหน่วยให้เป็นมาตรฐาน

$P_i(\theta)$ แทน ค่าความน่าจะเป็นที่ผู้สอบที่มีระดับความสามารถ θ จะทำข้อสอบข้อที่ i ได้ถูกต้อง

a_i แทน ค่าอำนาจจำแนกซึ่งมีค่าเป็นสัดส่วนโดยตรงกับค่าความชันของโค้ง ณ จุดเปลี่ยนโค้ง หรือจุดที่ชันที่สุด

- b_i แทน ค่าความยากที่แสดงระดับความสามารถที่แท้จริง ที่จุดโค้งชันที่สุด หรือในกรณีที่ไม่มีการเดา ค่า b_i คือ 0 ณ จุดความน่าจะเป็น .50
- D แทน Scaling Factor ซึ่งมีค่า 1.7 เป็นค่าที่มากที่สุดในการปรับ Logistic Model และ Ogive Model ให้สอดคล้องกัน
- e แทน ค่าคงที่ มีค่าประมาณ 2.71828...

2.2 โลจิสติก โมเดล ที่มีพารามิเตอร์สามตัว (Three-Parameter Model)

โมเดลที่มีพารามิเตอร์สามตัว ได้ดัดแปลงมาจากโมเดลที่ใช้พารามิเตอร์สองตัว ค่าพารามิเตอร์ที่เบร์นบอมได้เพิ่มขึ้นอีก 1 ตัว คือ ค่าสัมประสิทธิ์การเดา (c_i) ซึ่งแสดงได้ด้วยสมการทางคณิตศาสตร์ ดังนี้

$$P_i(\theta) = c_i + (1-c_i) \frac{e^{Da_i(\theta-b_i)}}{1 + e^{Da_i(\theta-b_i)}} ; i = 1, 2, \dots, n$$

เมื่อ D แทน ค่าคงที่มีค่าประมาณ 1.7...

c_i แทน ค่าความน่าจะเป็นที่ผู้สอบที่มีความสามารถต่ำมากมีโอกาสจะทำข้อสอบข้อนี้ได้ถูกต้อง

2.3 โลจิสติก โมเดล ที่มีพารามิเตอร์ตัวเดียว (One-Parameter Logistic Model)

โมเดลนี้จะกำหนดให้ค่าพารามิเตอร์การเดาของข้อสอบมีค่าเท่ากับศูนย์ ($c_i = 0$) และถือว่าข้อสอบมีค่าอำนาจจำแนกเท่ากัน ($a_i = 1$) โมเดลก็จะเหลือค่าพารามิเตอร์ค่าความยาก (b_i) เพียงตัวเดียว ซึ่งตรงกับ โมเดลของทฤษฎีการตอบข้อคำถามที่พัฒนาขึ้นโดย ราล์ฟ โดยที่ฟังก์ชันของโมเดลนั้นสามารถอธิบายได้ด้วยพารามิเตอร์ของข้อสอบเพียงตัวเดียว คือ พารามิเตอร์ค่าความยาก (b_i) หรือที่นิยม เรียกว่า ราล์ฟ โมเดล ซึ่งมีลักษณะของ โมเดล ดังนี้ (Hambleton and Swaminathan. 1985 : 47)

$$P_i(\theta) = \frac{e^{Da(\theta - b_i)}}{1 + e^{Da(\theta - b_i)}} ; i = 1, 2, \dots, n$$

เมื่อ $\bar{a}$ แทน ค่าอำนาจจำแนกที่เป็นกลาง

หรือ

$$P_i(\theta) = \frac{e^{\theta - b_i}}{1 + e^{\theta - b_i}} ; i = 1, 2, \dots, n$$

4. โลจิสติก โมเดล ที่มีพารามิเตอร์ 4 ตัว (Four-Parameter Logistic Model)
โมเดลนี้เสนอโดย แม็คโดแนลด์ (McDonald) ในปี ค.ศ. 1967 พัฒนาโดยลอร์ด (Lord) และ บาร์ตัน (Barton ปี ค.ศ. 1981) โมเดลนี้เพิ่มค่าความรอบคอบ (Careless) หรือพารามิเตอร์ อีก 1 ตัว ในโมเดลพารามิเตอร์ 3 ตัว เพื่อแก้ปัญหาที่ผู้สอบที่มีความสามารถในระดับสูงมักตอบ ข้อสอบผิด อาจจะเป็นเพราะความไม่รอบคอบหรือผู้สอบมีความสามารถอยู่เหนือกว่าแบบทดสอบ จึง ทำให้ไม่สามารถเลือกคำตอบที่ถูกต้องได้ มีลักษณะของโมเดล ดังนี้ (Hambleton and Swaminathan. 1985 : 34-49)

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

$$i = 1, 2, \dots, n$$

เมื่อ $\hat{\theta}_i$ แทน ค่าที่ผู้สอบซึ่งมีความสามารถในระดับสูงแต่มีความสะเพร่า
ตอบข้อสอบไม่ถูกต้อง โดยมีค่าน้อยกว่า 1 เล็กน้อย

การประยุกต์ทฤษฎีการตอบข้อคำถามเพื่อศึกษาความลำเอียงของข้อสอบ

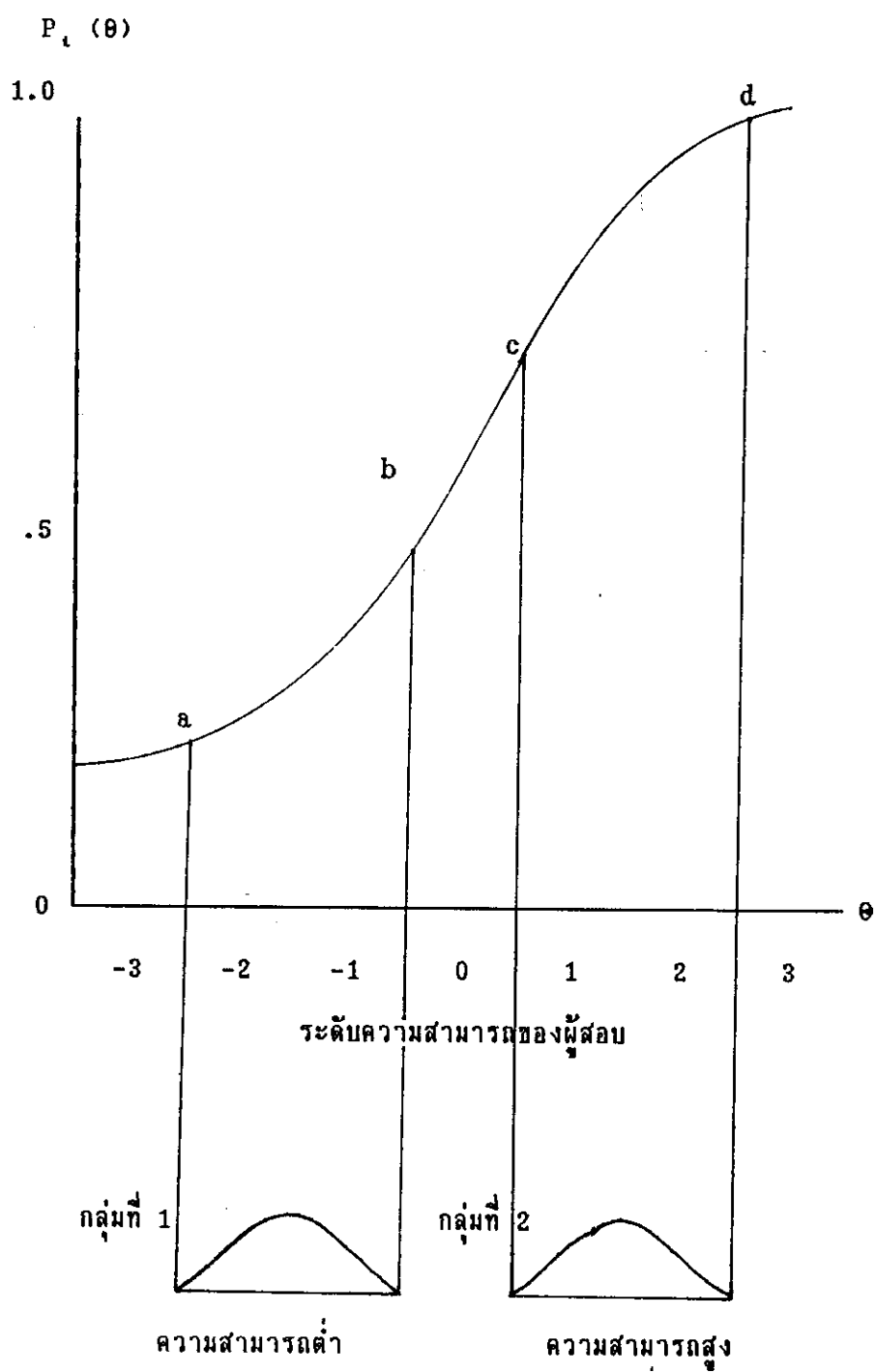
จากข้อตกลงเบื้องต้น 3 ประการของทฤษฎีการตอบข้อคำถาม คือ ความเป็นมิติเดียวของข้อสอบ ความเป็นอิสระในการตอบข้อสอบ และโด่งลักษณะข้อสอบ สามารถนำมาประยุกต์ใช้เป็นแนวทางในการวิเคราะห์ความลำเอียงของข้อสอบได้ดังนี้ คือ ถ้ามีแบบทดสอบฉบับหนึ่งวัดความสามารถของผู้สอบมากกว่า 1 กลุ่ม แล้วปรากฏว่า ข้อสอบข้อเดียวกันวัดความสามารถของผู้สอบได้ไม่ตรงกัน เช่น กลุ่มหนึ่ง อาจจะเผชิญกับเนื้อหาข้อสอบข้อนั้นมาก่อน หรือเมื่อตอบข้อสอบข้ออื่นถูกแล้ว เป็นเหตุทำให้ความน่าจะเป็นของการตอบข้อสอบข้อนั้นถูกมีค่ามากกว่าอีกกลุ่มหนึ่ง ถึงแม้ว่าภาษาในแต่ละฉบับจะมีความเป็นเอกพันธ์กันทุกข้อ แต่เมื่อพิจารณาเปรียบเทียบข้อสอบเป็นรายข้อระหว่างกลุ่มผู้สอบทุกกลุ่ม โดยนำผลการสอบของแต่ละกลุ่มมาสร้างโด่งลักษณะข้อสอบบนกราฟเดียวกัน เพื่ออธิบายลักษณะของข้อสอบจะพบว่าโด่งลักษณะข้อสอบจากผู้สอบแต่ละกลุ่มแตกต่างกัน ความแตกต่างของโด่งลักษณะข้อสอบของกลุ่มต่าง ๆ จะเป็นตัวบ่งชี้ว่าผู้สอบที่มีระดับความสามารถเท่ากันแต่อยู่ต่างกลุ่มกันมีโอกาสจะตอบข้อสอบได้ถูกต้องไม่เท่ากัน เพราะระดับความสามารถที่เท่ากันของผู้สอบนั้นอาจมาจากความสามารถต่างชนิดกัน ข้อสอบที่มีลักษณะเช่นนี้ถือว่าเป็นข้อสอบที่มีความลำเอียงต่อกลุ่มที่มีความน่าจะเป็นการตอบข้อสอบถูกมากกว่ากลุ่มที่มีระดับความสามารถต่ำ ซึ่งความลำเอียงของข้อสอบนี้แสดงให้เห็นได้ด้วยโด่งลักษณะข้อสอบ กล่าวคือ ถ้าผู้สอบต่างกลุ่มกันมีโด่งลักษณะข้อสอบข้อเดียวกันแตกต่างกัน ข้อสอบข้อนั้นถือว่าเป็นข้อสอบที่มีความลำเอียงแต่ในทำนองกลับกันหากโด่งลักษณะข้อสอบของผู้สอบต่างกลุ่มกันมี

ลักษณะโค้งเหมือนกันก็ย่อมแสดงว่าข้อสอบข้อนั้นไม่ลำเอียง สำหรับวิธีการวิเคราะห์ความลำเอียงด้วยโค้งลักษณะข้อสอบนั้นจะใช้คอมพิวเตอร์โปรแกรม LOGISTIC หรือโปรแกรม BILOG 3 ในการประมาณค่าโค้งลักษณะข้อสอบที่มี 2 พารามิเตอร์ หรือ 3 พารามิเตอร์และใช้โปรแกรม BICAL ในการประมาณค่าโค้งลักษณะข้อสอบที่มี 1 พารามิเตอร์

จากการศึกษารายงานการวิจัยเกี่ยวกับการวิเคราะห์ความลำเอียงของข้อสอบด้วยโค้งลักษณะข้อสอบ พบว่า วิธีวิเคราะห์ที่ใช้กันส่วนมากมี 2 แบบ คือ แบบ 3 พารามิเตอร์ เป็นแบบที่คำนวณพารามิเตอร์ของข้อสอบ 3 ตัว คือ ค่าความยาก (b_i) ค่าอำนาจจำแนก (a_i) และค่าสัมประสิทธิ์การเดา (c_i) โดยใช้แบบจำลองโลจิสติก (Three-Parameter Logistic Model) และแบบ 1 พารามิเตอร์ ที่คำนวณค่าพารามิเตอร์ของข้อสอบ 1 ตัว คือ ค่าความยาก (b_i) เพราะกำหนดให้ค่าอำนาจจำแนก (a_i) ของข้อสอบทุกข้อเท่ากัน หรือให้ค่าอำนาจจำแนกมีค่า เท่ากับ 1 และค่าสัมประสิทธิ์การเดา (c_i) เท่ากับ 0 โดยใช้แบบจำลองโลจิสติก หรือแบบจำลองของราสช์ (One-Parameter Logistic Model or Rasch Model) ซึ่งแต่ละแบบต่างก็มีข้อดีข้อเสียแตกต่างกัน กล่าวคือ การวิเคราะห์ด้วยโค้งลักษณะข้อสอบแบบคำนวณค่าพารามิเตอร์ของข้อสอบ 3 ตัว มีความสะดวก ง่าย และรวดเร็ว ในการวิเคราะห์ไม่จำเป็นต้องประมาณค่าพารามิเตอร์ของข้อสอบทุกตัว และสามารถหาข้อสอบที่เหมาะสม (Fit) กับวิธีการวิเคราะห์ได้ ถึงแม้ว่าจะมีข้อจำกัดในเรื่องเกี่ยวกับการเดาและอำนาจจำแนกที่ไม่เป็นจริงในทางปฏิบัติ สำหรับการวิเคราะห์ด้วยโค้งลักษณะข้อสอบแบบคำนวณค่าพารามิเตอร์ของข้อสอบ 3 ตัวนั้น แม้ว่าความถูกต้องของการประมาณค่าอำนาจจำแนก (a_i) และสัมประสิทธิ์ในการตอบถูกโดยการเดายังค่อนข้างต่ำและต้องวิเคราะห์ด้วยคอมพิวเตอร์ที่ค่อนข้างจะซับซ้อนและเสียค่าใช้จ่ายแพง (Berk. 1982 : 154) แต่ก็ให้ผลการวิเคราะห์ที่มีความถูกต้องและสมบูรณ์กว่า เพราะสามารถวิเคราะห์และอธิบายถึงโอกาสในการตอบข้อสอบถูกของผู้สอบในกลุ่มที่มีความสามารถต่ำมาก ๆ ได้

คุณสมบัติของโค้งลักษณะข้อสอบ

ในการศึกษาความไม่เปลี่ยนแปลงเกี่ยวกับโค้งลักษณะของข้อสอบนี้ ลอร์ด และโนวิก (Lord and Novick. 1968) ได้ทำการศึกษาเกี่ยวกับค่าพารามิเตอร์ของข้อสอบข้อเดียวกันระหว่างกลุ่มผู้สอบที่มีความสามารถต่างกัน ซึ่ง นำมาเขียนสรุปได้ ดังภาพประกอบ 5

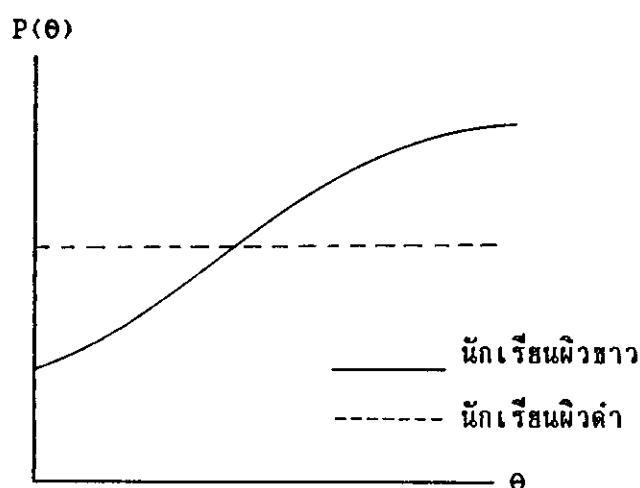


ภาพประกอบ 5 แสดงโค้งลักษณะข้อสอบ ซึ่งมีลักษณะไม่เปลี่ยนแปลงตามจำนวน และระดับความสามารถของผู้สอบ

จากภาพประกอบ 5 แสดงว่า ในกรณีที่กลุ่มผู้สอบมีระดับความสามารถต่ำค่าความน่าจะเป็นของการตอบข้อสอบถูกของผู้สอบสามารถจะอธิบายได้ด้วยโค้งลักษณะของข้อสอบในช่วงระยะ a-b และในทำนองเดียวกันเมื่อผู้สอบมีระดับความสามารถสูงค่าความน่าจะเป็นของการตอบข้อสอบถูกก็ย่อมสามารถอธิบายได้ด้วยโค้งลักษณะของข้อสอบในช่วง c-d นั่นคือ ถ้าค่าพารามิเตอร์ของข้อสอบไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบแล้ว ข้อสอบที่นำไปทดสอบกับผู้ที่มีความสามารถแตกต่างกัน เมื่อนำข้อมูลที่ได้ในแต่ละกลุ่มมาวิเคราะห์จะมีค่าพารามิเตอร์ของข้อสอบเท่ากัน ถ้าข้อสอบข้อนี้วัดความสามารถอย่างเดียวกันของกลุ่มตัวอย่างทั้งสองกลุ่มแล้ว เส้นโค้งลักษณะข้อสอบก็ย่อมมีเส้นเดียว เส้นโค้งในช่วง a-b และ ช่วง c-d ย่อมเป็นส่วนหนึ่งของโค้งลักษณะข้อสอบข้อนี้ เมื่อเส้นโค้งลักษณะข้อสอบมีได้เส้นเดียว ตัวพารามิเตอร์ซึ่งเป็นตัวกำหนดรูปร่างลักษณะของเส้นโค้งย่อมมีได้ค่าเดียวเช่นกัน หรืออาจกล่าวได้ว่า ถ้าค่าพารามิเตอร์ของข้อสอบไม่แปรเปลี่ยนไปตามกลุ่มผู้สอบแล้ว ข้อสอบที่นำไปสอบกับผู้สอบที่มีความสามารถแตกต่างกัน เมื่อนำข้อมูลที่ได้จากแต่ละกลุ่มมาวิเคราะห์จะมีค่าพารามิเตอร์ของข้อสอบเป็นค่าเดียวกัน

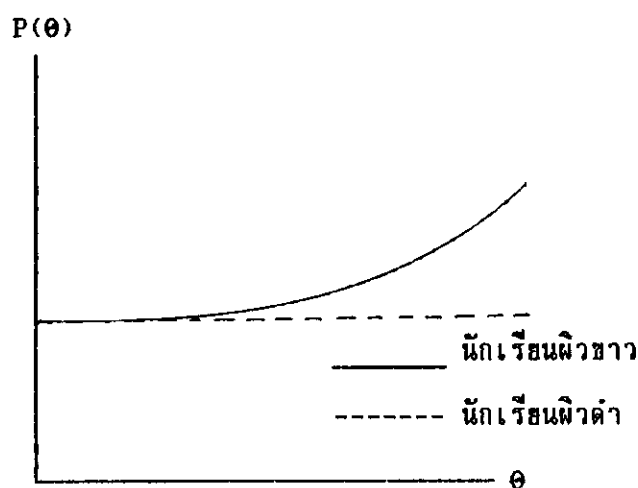
เพื่อให้เข้าใจง่ายขึ้นในที่นี้จะขอเสนอตัวอย่างผลการวิเคราะห์ความลำเอียงของข้อสอบวัดความถนัดทางการเรียนด้านภาษา (Verbal Scholastic Aptitude) ที่ใช้ทดสอบกับนักเรียนผิวขาวและผิวดำในสหรัฐอเมริกา (Lord, 1977 : 136-137) จำนวน 3 ข้อ

ดังภาพประกอบ 6-8



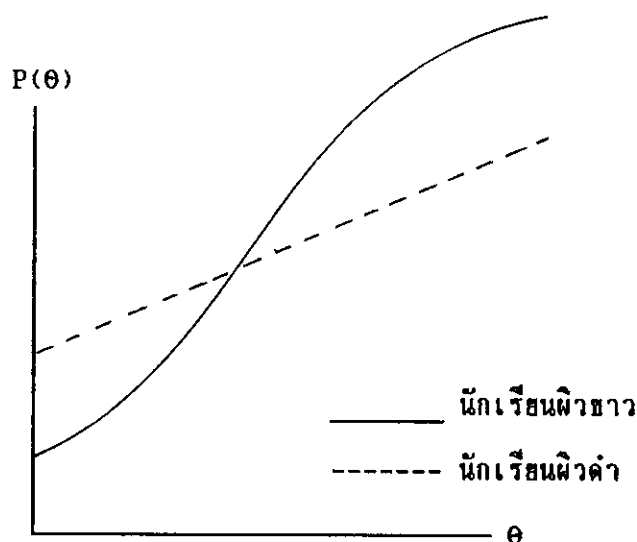
ภาพประกอบ 6 แสดงโค้งลักษณะข้อสอบ ข้อ 2 ระหว่างนักเรียนผิวขาวและผิวดำ

จากภาพประกอบ 6 แสดงให้เห็นว่า ข้อสอบข้อนี้ไม่สามารถจำแนกนักเรียนผิวดำได้ เพราะไม่ว่าที่ระดับความสามารถใดก็ตามนักเรียนผิวดำ มีโอกาสที่จะตอบข้อสอบข้อนี้ได้ถูกต้องเท่า ๆ กัน ถ้าแบ่งตามช่วงระดับความสามารถเฉพาะในกลุ่มผู้สอบที่มีระดับความสามารถสูง นักเรียนผิวขาวมีโอกาที่จะตอบข้อสอบข้อนี้ได้ถูกมากกว่านักเรียนผิวดำ ซึ่งตรงกันข้ามกับในกลุ่มผู้สอบที่มีระดับความสามารถต่ำ นักเรียนผิวดำมีโอกาสดตอบข้อสอบข้อนี้ได้ถูกมากกว่านักเรียนผิวขาว



ภาพประกอบ 7 แสดงโค้งลักษณะข้อสอบ ข้อ 24 ระหว่างนักเรียนผิวขาวและผิวดำ

จากภาพประกอบ 7 แสดงให้เห็นว่า ข้อสอบข้อนี้ยากเพราะโอกาสในการตอบถูกต่ำมากสำหรับในกลุ่มผู้สอบที่เป็นนักเรียนผิวดำในทุก ๆ ระดับความสามารถ และยากสำหรับนักเรียนผิวขาวที่มีความสามารถต่ำด้วย ส่วนนักเรียนผิวขาวที่มีความสามารถสูงมีแนวโน้มที่จะตอบข้อสอบถูกมากขึ้น

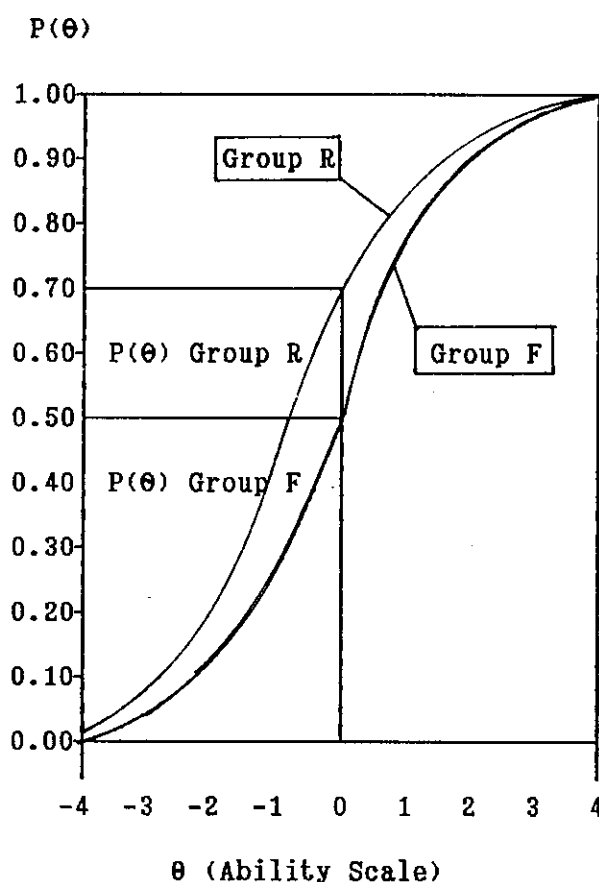


ภาพประกอบ 8 แสดงโค้งลักษณะข้อสอบ ข้อ 71 ระหว่างนักเรียนผิวขาวและผิวดำ

จากภาพประกอบ 8 แสดงให้เห็นว่า ในกลุ่มผู้สอบที่มีระดับความสามารถสูงนักเรียนผิวขาวมีโอกาสที่จะตอบข้อสอบ ข้อนี้ ได้ถูกต้องมากกว่านักเรียนผิวดำที่มีระดับความสามารถเดียวกัน แต่สำหรับในกลุ่มผู้สอบที่มีระดับความสามารถต่ำแล้วนักเรียนผิวดำมีโอกาสที่จะตอบถูกมากกว่านักเรียนผิวขาว

จะเห็นว่าการวิเคราะห์ความลำเอียงของข้อสอบด้วยโค้งลักษณะข้อสอบจะพิจารณาจากผลการตอบข้อสอบกับโอกาสในการตอบข้อสอบข้อนั้นๆ ถูก ระหว่างกลุ่มผู้สอบตั้งแต่ 2 กลุ่ม หากโค้งลักษณะข้อสอบมีความแตกต่างกัน แสดงว่าข้อสอบข้อนั้นจะทำหน้าที่เบี่ยงเบนไปจากปกติ หรือทำให้เกิดการได้เปรียบเสียเปรียบระหว่างกลุ่มผู้สอบที่มีลักษณะแตกต่างกัน และค่าพารามิเตอร์ที่เป็นตัวกำหนดโค้งลักษณะข้อสอบตามทฤษฎีการตอบข้อคำถามนั้น มี 3 ค่า คือ ค่าความยาก (b_i) ค่าอำนาจจำแนก (a_i) และค่าสัมประสิทธิ์การเดา (c_i) ในการทดสอบว่าข้อสอบข้อใดมีความลำเอียงหรือไม่และจะใช้วิธีใดในการวิเคราะห์ความลำเอียงของข้อสอบให้มีความถูกต้องเหมาะสมเพียงใดนั้น ขึ้นอยู่กับลักษณะความแตกต่างของโค้งลักษณะข้อสอบที่มีความลำเอียง ซึ่งสามารถแบ่งออกเป็น 2 แบบ คือ

1. แบบลำเอียงอย่างสม่ำเสมอหรือแบบลำเอียงคงที่ คือ โด่งลักษณะข้อสอบของกลุ่มต่างๆ จะมีค่าพารามิเตอร์ของข้อสอบเท่ากัน 2 ค่า คือ ค่าอำนาจจำแนกของข้อสอบ (a_i : ความชัน) ของแต่ละกลุ่มเท่ากัน และมีค่าสัมประสิทธิ์การเดา (c_i) ที่ใกล้เคียงกันหรือเท่ากัน วิธีนี้จะมีเฉพาะค่าความยากของข้อสอบ (b_i) ตัวเดียวเท่านั้นที่แตกต่างกันและโด่งลักษณะข้อสอบของ กลุ่มต่างๆ จะไม่ตัดกันตลอดต่อเนื่องของทุกระดับความสามารถ ดังภาพประกอบ 9

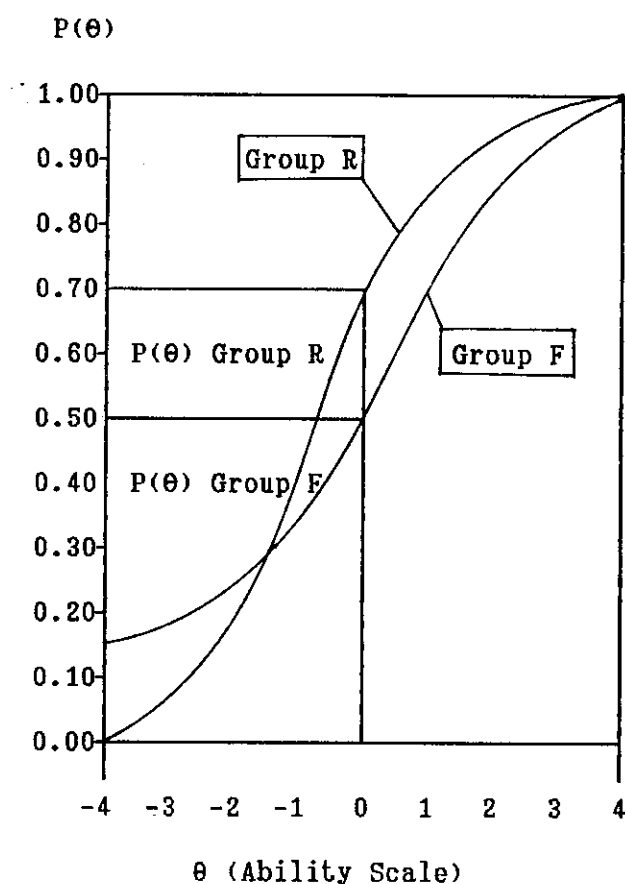


ภาพประกอบ 9 แสดงโด่งลักษณะข้อสอบระหว่างกลุ่ม R และกลุ่ม F

จากภาพประกอบ 9 แสดงให้เห็นว่า ในกลุ่มผู้สอบที่มีระดับความสามารถ (θ) เดียวกัน ผู้สอบที่อยู่ในกลุ่ม R มีโอกาสในการตอบข้อสอบข้อนี้ได้ถูกต้องมากกว่าผู้สอบที่อยู่ในกลุ่ม F ซึ่งในการวิเคราะห์ความลำเอียงของข้อสอบ จะใช้ค่าพารามิเตอร์เพียงตัวเดียว คือ ค่าความยากของข้อสอบ (b_i) มาคำนวณค่าดัชนีความลำเอียงของข้อสอบ นั่นคือ เป็นการวิเคราะห์ความลำเอียง

ของข้อสอบตามแบบจำลอง โลจิสติกที่มี 1 พารามิเตอร์ หรือ แบบจำลองของราสช์ (Rasch Model) นั้นเอง การคำนวณพื้นที่ระหว่างโค้งลักษณะข้อสอบวิธีนี้สามารถคำนวณได้ง่ายเพราะพื้นที่ที่คำนวณได้มีเฉพาะเครื่องหมายบวกเพียงอย่างเดียวเท่านั้น

2. แบบจำลองเชิงอย่างไม่สม่ำเสมอหรือแบบจำลองเชิงไม่คงที่ คือ โค้งลักษณะข้อสอบของกลุ่มต่าง ๆ จะมีค่าพารามิเตอร์ของข้อสอบแตกต่างกัน 3 ค่า คือ ค่าความยาก (b_i) ค่าอำนาจจำแนก (a_i) และค่าสัมประสิทธิ์การเดา (c_i) ทำให้โค้งลักษณะข้อสอบของกลุ่มต่างๆ มีลักษณะตัดกัน ดังภาพประกอบ 10



ภาพประกอบ 10 แสดงโค้งลักษณะของข้อสอบระหว่างกลุ่ม R และกลุ่ม F

จากภาพประกอบ 10 แสดงให้เห็นว่า ที่ระดับความสามารถสูงผู้สอบในกลุ่ม R มีโอกาสที่จะตอบข้อสอบข้อนี้ได้ถูกต้องมากกว่าผู้สอบในกลุ่ม F ที่มีระดับความสามารถเดียวกัน แต่สำหรับ

ในกลุ่มผู้สอบที่มีระดับความสามารถต่ำแล้ว ผู้สอบในกลุ่ม F มีโอกาสที่จะตอบข้อสอบถูกมากกว่าผู้สอบในกลุ่ม R ในการวิเคราะห์หาความลำเอียงของข้อสอบจะนำค่าพารามิเตอร์ ทั้ง 3 ค่า มาคำนวณค่าดัชนีความลำเอียงของข้อสอบ นั่นคือ เป็นการวิเคราะห์ความลำเอียงของข้อสอบตามแบบจำลอง 3 พารามิเตอร์ การคำนวณค่าดัชนีความลำเอียงตามวิธีนี้จะมีความยุ่งยากมาก และซับซ้อน เพราะว่าค่าพารามิเตอร์ของโดเมนลักษณะข้อสอบแต่ละข้อจะแตกต่างกันทุกค่า ส่วนในกรณีที่ใช้พื้นที่ระหว่างโดเมนลักษณะข้อสอบเป็นตัวดัชนีความลำเอียงของข้อสอบนั้นการคำนวณพื้นที่จะมีความยุ่งยากมากขึ้น เพราะพื้นที่ ที่คำนวณได้จะต้องมีทั้งทางบวกและทางลบ เนื่องจากโดเมนลักษณะข้อสอบของผู้สอบทั้งสองกลุ่มมีลักษณะตัดกัน

วิธีการวิเคราะห์ความลำเอียงของข้อสอบโดยใช้ทฤษฎีการตอบข้อคำถาม

นิยามความลำเอียงของข้อสอบ ในแง่ของทฤษฎีการตอบข้อคำถามนั้นข้อสอบจะลำเอียงถ้าฟังก์ชันการตอบข้อสอบข้อนั้น ๆ ของประชากรกลุ่มย่อย ๆ 2 กลุ่ม ไม่เป็นเส้นเดียวกัน ซึ่งจากนิยามดังกล่าวข้างต้น การวิเคราะห์ความลำเอียงของข้อสอบตามแนวทฤษฎีการตอบข้อคำถามสามารถทำได้โดยการเปรียบเทียบฟังก์ชันการตอบข้อสอบของผู้สอบกลุ่มย่อย ที่ต้องการศึกษาและการเปรียบเทียบ ในทฤษฎีการตอบข้อคำถามมีวิธีการวิเคราะห์หลายวิธี ซึ่งพอสรุปได้ดังนี้ คือ

1. แบบจำลองโลจิสติกชนิด 1 พารามิเตอร์ ซึ่งกำหนดให้ค่าการเดา (c_i) เท่ากับ 0 และค่าอำนาจจำแนก (a_i) ของข้อสอบทุกข้อ เท่ากับ 1.00 มีค่าความยาก (b_i) เพียงตัวเดียวที่แปรเปลี่ยนไปตามกลุ่มผู้สอบ วิธีนี้เป็นการวิเคราะห์ความแตกต่างของค่าพารามิเตอร์ความยาก (b_i) ของข้อสอบจากโดเมนลักษณะข้อสอบของแต่ละกลุ่ม

เริ่มด้วยการทดสอบความเหมาะสมของข้อมูลทั้ง 2 กลุ่ม กับแบบจำลอง โดยทดสอบรายข้อด้วย $H_1(0)$ ซึ่งมีการกระจายแบบไคสแควร์ (χ^2) มี $df = N - 1$ เพื่อตัดข้อที่ไม่เหมาะสมออก แล้วจึงนำไปวิเคราะห์ความลำเอียงรายข้อด้วย Z_i

$$H_i(\theta) = \sum_{j=1}^N \frac{[U_{ij} - P_i]^2}{[\hat{P}_i \cdot Q_i]}$$

เมื่อ U_{ij} แทน คะแนนจากการตอบข้อสอบแบบ 0 - 1 ของคนที่ j ข้อที่ i

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

$$Z_i = \frac{b_{iA} - b_{iB}}{\sqrt{SE_{b_{iA}}^2 + SE_{b_{iB}}^2}}$$

$$SE_{b_i}^2 = I_{a_i} / (I_{a_i} I_{b_i} I_{a_i b_i}^2)$$

ถ้าค่า Z_i ที่คำนวณได้มีค่าสูง แสดงว่า ข้อสอบนั้นยากสำหรับกลุ่มหนึ่งมากกว่าอีกกลุ่มหนึ่ง หรืออาจจะวิเคราะห์ความลำเอียงเป็นรายข้อด้วย t ซึ่งคำนวณได้ จากสูตร

$$t = \frac{b_{iA} - b_{iB}}{\sqrt{SE_{b_{iA}}^2 + SE_{b_{iB}}^2}}$$

เมื่อ b_{iA} , b_{iB} แทน ค่าความยากของข้อสอบที่ได้จากการวิเคราะห์ในกลุ่มที่ 1 และกลุ่มที่ 2 ตามลำดับ

SE^2_{b1A} , SE^2_{b1B} แทน ค่าความคลาดเคลื่อนมาตรฐานของข้อสอบที่ได้จากการ
วิเคราะห์จากกลุ่มที่ 1 และกลุ่มที่ 2 ตามลำดับ

ค่า t ที่คำนวณได้คือ ดัชนีความลำเอียงของข้อสอบ ซึ่ง ดรบา (Berk. 1982 :
138 ; Citing Draba. 1977) เสนอให้ใช้ค่า $t \geq 2.4$ เป็นเกณฑ์ในการตัดสินใจว่าข้อสอบมี
ความลำเอียง สำหรับวิธีการนี้เรียกอีกอย่างหนึ่งว่า เป็นการวิเคราะห์ความลำเอียงตามแบบ
จำลองของราส์ (Rasch Model)

2. แบบจำลองโลจิสติกชนิด 2 พารามิเตอร์ กำหนดให้ค่าการเดา (c_i) เท่ากับ 0
ส่วนค่าความยาก (b_i) และ ค่าอำนาจจำแนก (a_i) แปรเปลี่ยนไปตามกลุ่มผู้สอบ วิธีนี้เริ่มด้วย
การประมาณค่าพารามิเตอร์ของข้อสอบแยกตามกลุ่มผู้สอบ กลุ่มย่อย A และ B ด้วยวิธี Maximum
Likelihood แล้วเปรียบเทียบถึงลักษณะการตอบข้อสอบด้วยค่า F ซึ่งคำนวณได้จากสูตร

$$F = \frac{SSE(A) - [SSE(A) - SSE(B)]}{SSE(A) + SSE(B)} \cdot \frac{J_A + J_B - 4}{2}$$

เมื่อ $SEE(A)$ แทน ความคลาดเคลื่อนมาตรฐานของข้อสอบที่ได้จากการวิเคราะห์
กลุ่มที่ 1

$SEE(B)$ แทน ความคลาดเคลื่อนมาตรฐานของข้อสอบที่ได้จากการวิเคราะห์
กลุ่มที่ 2

$SSE(Pooled)$ แทน ความคลาดเคลื่อนมาตรฐานของข้อสอบที่ได้จากการวิเคราะห์
กลุ่มที่ 1 กับ กลุ่มที่ 2

J_A , J_B แทน จำนวนช่วงความสามารถ

และมีค่าขึ้นแห่งความเป็นอิสระ ($J_A + J_B - 4$) คือ 2 ซึ่งการทดสอบด้วยค่า F จะบอกว่า
เส้นถดถอยรวม สามารถอธิบายความแปรปรวนของสัดส่วนการตอบของเส้นถดถอยจากแต่ละกลุ่ม
ได้เท่ากันหรือไม่ ถ้าเท่ากัน หรือไม่มีนัยสำคัญทางสถิติก็ไม่จำเป็นต้องใช้เส้นถดถอยแยกกัน

เพราะความสัมพันธ์ระหว่างความสามารถและการตอบข้อสอบของผู้สอบทั้งสองกลุ่มเป็นเช่นเดียวกัน และข้อสอบก็ทำหน้าที่ในการวัดความสามารถของคน 2 กลุ่มนี้ได้เหมือนกัน

3. แบบจำลองโลจิสติกชนิด 3 พารามิเตอร์ เป็นแบบที่ให้ค่าความยาก (b_i) ค่าอำนาจจำแนก (a_i) และค่าสัมประสิทธิ์การเดา (c_i) ของข้อสอบ แปรเปลี่ยนไปตามกลุ่มย่อยของผู้สอบ การคำนวณหาดัชนีความลำเอียงของข้อสอบวิธีนี้ เริ่มจากการประมาณค่าความสามารถของผู้สอบและค่าพารามิเตอร์ของข้อสอบเป็นรายข้อรวมของทุกคนทุกกลุ่มแล้วแปลงค่าพารามิเตอร์ที่ได้จากผู้สอบ 2 กลุ่ม ให้อยู่บนมาตรฐานเดียวกันบนค่าประมาณความยาก จากนั้นแยกประมาณค่าความยากและค่าอำนาจจำแนกตามกลุ่มย่อย โดยใช้ค่าการเดาที่ประมาณได้จากการรวมทุกกลุ่ม การทำค่าความยากให้เป็นมาตรฐานนั้นจะต้องตัดข้อสอบที่มีค่าความยากสูงมากหรือต่ำมากออก

การวิเคราะห์ความลำเอียงของข้อสอบตามแบบจำลอง โลจิสติก 3 พารามิเตอร์ มีหลายวิธี เช่น ลินน์ ลีวีน ฮาสติงส์ และอาร์ดรอพ (Hulin, Drasgows and Parson. 1983 : 176 ; Citing Linn, Levine, Hastings and Wardrop. 1981) ได้พัฒนาการทดสอบความลำเอียงของข้อสอบจากความแตกต่างระหว่างโค้งลักษณะของข้อสอบโดยตรง โดยพิจารณาความแตกต่างของพื้นที่ใต้โค้งลักษณะข้อสอบของผู้สอบ 2 กลุ่มย่อย เป็นเกณฑ์การตัดสินความลำเอียงของข้อสอบ ผู้ที่ศึกษาความลำเอียงโดยการเปรียบเทียบความแตกต่างของโค้งลักษณะข้อสอบ เช่น รัดเนอร์ (Rudner. 1977 : Citing in Berk. 1982 : 118) รัดเนอร์ เกทสัน และ ไนท์ (Rudner, Getson and Knight. 1980) ไอรอนสัน และ ซับโคเวียค (Ironson and Subkoviak. 1979) ซึ่งเกณฑ์ที่ใช้พิจารณาความลำเอียง คือ มีพื้นที่ใต้โค้งต่างกันตั้งแต่ .20 ขึ้นไป ในขณะที่ ทศนิษฐ์ นิรมนตรี ใช้ค่ามากกว่า .40 การใช้เกณฑ์ความลำเอียง .20 หรือ .40 แทนที่จะใช้ 0.00 นั้น เนื่องจากยอมให้มีความคลาดเคลื่อนที่อาจจะเกิดจากการสุ่ม ผู้ที่นำดัชนีนี้ไปใช้และใช้เกณฑ์ที่ต่างกัน ได้แก่ คิม และ โคเฮน (Kim and Cohen. 1991) โดยมีข้อตกลงเบื้องต้นว่า การกระจายของความแตกต่างของพื้นที่ใต้โค้งที่วัดได้มีลักษณะเป็นโค้งปกติ เกณฑ์การตัดสินความลำเอียงจะพิจารณาค่าที่มากกว่าค่าวิกฤตชนิดหางเดียวที่ระดับ .05 และ .01

นอกจากนี้ ลินน์ และคนอื่นๆ (Linn and Others. 1981) ได้อธิบายเกี่ยวกับการหาดัชนีความลำเอียงของข้อสอบ โดยใช้รากที่สองของค่าเฉลี่ยความแตกต่างกำลังสองของโค้งลักษณะข้อสอบ ของกลุ่มผู้สอบ 2 กลุ่ม เป็นดัชนีความลำเอียงของข้อสอบในช่วงความสามารถ -3 ถึง 3 ที่ได้จากการคำนวณ จากสูตร ดังนี้

$$RMSD = \frac{1}{600} \left[\sum_{j=1}^J [P_{iA}(\theta_j) - P_{iB}(\theta_j)]^2 \right]^{1/2}$$

โดยได้จากการแบ่งช่วงความสามารถจาก -3 ถึง 3 เป็น 600 ช่วง ค่าสัมบูรณ์ของความแตกต่างของโค้งทั้ง 2 ที่จุดกึ่งกลางของ 600 ช่วง จะเป็น $P_{iA}(\theta_j) - P_{iB}(\theta_j)$ คูณด้วยความยาวของช่วงแต่ละช่วงคือ .01 การบวกค่าพื้นที่ทั้ง 600 ช่วง จะได้ค่าประมาณของพื้นที่ทั้งหมดที่อยู่ระหว่างโค้งทั้งสอง

ดัชนี 2 ตัวแรก แสดงค่าความลำเอียงต่อกลุ่มใดกลุ่มหนึ่งโดยตรง แต่ 2 ตัวหลังเป็นการให้ภาพรวมของความลำเอียงของข้อสอบ ค่ายิ่งเข้าใกล้ 0 แสดงว่า โค้งลักษณะข้อสอบทั้งสองมีความใกล้เคียงมาก การคำนวณค่าดัชนีดังกล่าวจะแสดงความลำเอียงได้ดีกว่าการพิจารณาเฉพาะความแตกต่างของค่าพารามิเตอร์ ซึ่งอาจเกิดขึ้นได้โดยพื้นที่โค้งลักษณะข้อสอบทั้งสองไม่ต่างกันเลขก็เป็นได้ ดังที่ ลินน์ และคณะ พบว่า ข้อสอบข้อที่ i สำหรับกลุ่มหนึ่งมีค่า $a = 1.8$ $b = 3.5$ และ $c = .2$ อีกกลุ่มหนึ่งมีค่า $a = .5$ $b = 5.0$ และ $c = .2$ เมื่อพิจารณาแล้วจะเห็นว่าค่า a และ b ต่างกันมาก แต่ค่าความแตกต่างระหว่างโค้งลักษณะข้อสอบทั้ง 2 โค้ง มีค่าเท่ากับ .05 ซึ่งเป็นค่าที่ยอมรับได้ในการสรุปว่าข้อสอบนั้นไม่ลำเอียง

เชปาร์ต และคนอื่นๆ (Shepard and Others. 1985 : 81) กล่าวว่า เมื่อพบว่า โค้งลักษณะข้อสอบของกลุ่มอ้างอิง และกลุ่มเปรียบเทียบบแตกต่างกันมากพอที่จะเป็นหลักฐานได้ว่า ข้อสอบที่ศึกษาอยู่ไม่ได้วัดลักษณะแฝงเดียวกันสำหรับกลุ่มผู้สอบทั้งสองแล้ว จะใช้ดัชนีที่บ่งชี้ถึงระดับความลำเอียง ดังนี้

1. พื้นที่ชนิดไม่มีเครื่องหมาย (Unsigned Area) เป็นค่าสัมบูรณ์ของพื้นที่ระหว่างโค้งลักษณะข้อสอบทั้งสอง

2. ผลบวกกำลังสองแบบที่ 1 (SOS 1) เป็นค่าผลบวกของความแตกต่างระหว่างกำลังสองทุกค่าของความสามารถที่เป็นไปได้

3. ผลบวกกำลังสองแบบที่ 2 (SOS 2) คล้าย SOS 1 แต่ให้น้ำหนักความแตกต่างกำลังสองด้วยส่วนกลับของความคลาดเคลื่อนของความแปรปรวนในการประมาณค่าความเป็นไปได้ ดังนั้นถ้ามีการประมาณโค้งอย่างไม่ดีในส่วนใดส่วนหนึ่งแล้ว จะทำให้ลดประโยชน์ต่อความแตกต่างระหว่างสองกลุ่ม

4. ไคสแควร์ ($IRT - \chi^2$) เป็นการทดสอบนัยสำคัญทางสถิติ โดยการเปรียบเทียบความแตกต่างของค่าความชุกและค่าอำนาจจำแนกระหว่างกลุ่มพร้อม ๆ กัน โดยลอร์ด (Lord, 1980)

5. พื้นที่ชนิดมีเครื่องหมาย เหมือนพื้นที่ชนิดไม่มีเครื่องหมายแต่มีการให้เครื่องหมายไว้เพื่อแสดงให้เห็นว่ากลุ่มใดเสียประโยชน์ และกลุ่มใดได้ประโยชน์ ถ้าไคสแควร์ลักษณะการตอบข้อสอบตัดกัน แสดงว่า มีความลำเอียงต่างกันสำหรับผู้มีความสามารถต่างกัน และค่าสัมบูรณ์ของพื้นที่ชนิดมีเครื่องหมายจะน้อยกว่าชนิดไม่มีเครื่องหมาย

6. ผลบวกกำลังสองแบบที่ 3 (SOS 3) เป็นผลบวกกำลังสองของความแตกต่างชนิดมีเครื่องหมายของ SOS 1 โดยการใช้เครื่องหมายของความแตกต่างแต่ละตัวที่ยังไม่ยกกำลังสอง

7. ผลบวกกำลังสองแบบที่ 4 (SOS 4) เป็นดัชนีคู่ขนานกับ SOS 2 แต่เป็นผลบวกที่มีการให้น้ำหนักและให้เครื่องหมาย

การวัดพื้นที่ของความแตกต่างระหว่างไคสแควร์ข้อสอบของกลุ่มตัวอย่าง เมื่อพิจารณาถึงขนาดของความลำเอียงของข้อสอบนั้นมีอยู่ 2 ลักษณะ คือ การวัดแบบช่วงเปิดและแบบช่วงปิด (Open and Closed Intervals) กรณีของการวัดแบบช่วงปิดจะกำหนดขอบเขตของการวัดให้อยู่ในช่วงความสามารถที่กำหนด ในขณะที่การวัดแบบช่วงเปิดจะวัดภายในช่วงความสามารถทั้งหมดเพื่อให้ได้พื้นที่ ที่แน่นอน (Exact Area) ระหว่างไคสแควร์ข้อสอบ 2 ไคส ภายใต้การวัดแบบช่วงปิดนั้นสามารถวัดพื้นที่ได้ 2 ลักษณะ คือ พื้นที่ชนิดมีเครื่องหมายและไม่มีเครื่องหมาย (Signed and Unsigned Areas) โดยมีสูตรการคำนวณทั่ว ๆ ไป ดังนี้

จากฟังก์ชันการตอบข้อสอบแบบ 3 พารามิเตอร์

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

$$\text{โดยที่ } P^*(\theta) = (1 + e^{-D(a(\theta - b))})^{-1}$$

เมื่อ a , b และ c เป็นพารามิเตอร์ที่แสดงลักษณะของข้อสอบ และ D เป็นค่าคงที่ กำหนดให้ มีค่าเท่ากับ 1.7

พื้นที่ระหว่างความสามารถที่กำหนดช่วงไว้ 2 จุดบนสเกลความสามารถคำนวณได้จาก

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

ในการศึกษาความลำเอียงของข้อสอบนั้น จะมีพารามิเตอร์ของ 2 กลุ่ม คือ กลุ่มอ้างอิง ได้แก่ a_R , b_R และ c_R และกลุ่มเปรียบเทียบ คือ a_F , b_F และ c_F พื้นที่ชนิดที่มีเครื่องหมาย (Closed - interval Signed Area - CSA) สามารถคำนวณได้จาก

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

และพื้นที่ชนิดไม่มีเครื่องหมาย (Closed - interval Unsigned Area - CUA) หรือ ϕ สามารถคำนวณได้จาก

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

สงบ ลักษณะ และ คอฟแมน (สุรศักดิ์ อมรรัตนศักดิ์. 2531 : 54 ; อ้างอิงมาจาก สงบ ลักษณะ และ คอฟแมน.1980 : 11) เสนอให้ใช้ค่า $CUA > .20$ เป็นเกณฑ์ของความลำเอียงของข้อสอบ

การวิเคราะห์ความลำเอียงของข้อสอบด้วยทฤษฎีการตอบข้อคำถามนั้น ชัยโคเวียค แม็คไฮรอนสัน และเครก (Subkoviak, Mack, Ironson and Craig. 1984) พบว่า จากการเปรียบเทียบกับวิธีแปลงค่าความยาก และไคสแควร์แบบ 5 ช่วง พบว่า การวิเคราะห์ด้วยทฤษฎีการตอบข้อคำถามแบบ 3 พารามิเตอร์ ให้ความถูกต้องมากที่สุด ซึ่งเชปาร์ด คามิลลี และ

วิลเลียมส์ (Shepard, Camilli and Williams. 1984 : 93-128) ค้นพบสอดคล้องกัน
ในแง่ที่ว่า เป็นวิธีที่ดีที่สุด ดีกว่าวิธีไคสแควร์และการแปลงค่าความชาก บากิและเฟอร์รารา (Baghi
and Ferrara. 1990) พบว่า ในขนาดกลุ่มตัวอย่าง 750 คน สามารถใช้วิธีแมนเทิล-แฮนส์เชล
แทนการใช้ทฤษฎีการตอบข้อคำถามชนิด 3 พารามิเตอร์ได้ คิมและโคเฮน (Kim and Kohen.
1991) พบว่า การเปรียบเทียบความแตกต่างของพื้นที่ใต้โค้งโดยไม่มีเครื่องหมาย จะดีกว่าแบบ
มีเครื่องหมาย รวมทั้งมีความคลาดเคลื่อนต่ำกว่าแบบมีเครื่องหมาย และระบุว่าค่าเฉลี่ยโดยที่
ไม่ค่าเฉลี่ยจริง น้อยกว่าวิธีอื่น ๆ คิดเป็นร้อยละ 5 หือ ลอเตนสเลเจอร์และพาร์ค
(Lautenschlager and Park. 1988) พบว่า วิธีของลินน์ (Linn. 1980) ดีกว่าวิธีของวอร์ม
(Warm. 1978) ส่วนรัดเนอร์ เกทสัน และไนท์ (Rudner, Getson and Knight. 1980)
พบว่า วิธีการตรวจสอบความลำเอียงด้วยทฤษฎีการตอบข้อคำถามแบบ 3 พารามิเตอร์ ไม่มีผล
กระทบจากความยาวของข้อสอบ

การใช้โปรแกรมการถดถอยแบบโลจิสต์ ของสวามินาธาน และโรเจอร์ส (Swaminathan
and Rogers. 1990 : 361-370) โดยใช้ข้อมูลจำลอง พบว่า สามารถตรวจค้นความลำเอียง
ของข้อสอบได้ร้อยละ 50-75 แล้วแต่ความยาวของข้อสอบ และขนาดของกลุ่มตัวอย่าง ในกรณี
ที่เป็น การลำเอียงอย่างไม่มีสม่าเสมอ และสามารถตรวจค้นความลำเอียงอย่างสม่าเสมอได้ใกล้เคียง
กับวิธีของแมนเทิล-แฮนส์เชล (MH) แต่การถดถอยแบบโลจิสต์ใช้ค่าจำมากกว่าประมาณ
3-4 เท่า นอกจากนี้โรเจอร์ส (Rogers. 1989) ได้ใช้การถดถอยแบบโลจิสต์เพื่อตรวจค้นความ
ลำเอียงของข้อสอบแบบไม่มีสม่าเสมอ พบว่า สามารถใช้ได้ดีกับกลุ่มตัวอย่างขนาด 250 คนขึ้นไป
และจะตรวจค้นความลำเอียงอย่างสม่าเสมอได้ดีเท่ากับวิธีการของแมนเทิล-แฮนส์เชล (MH) แต่
การถดถอยแบบโลจิสต์จะตรวจค้นความลำเอียงได้ดีกว่าวิธีของแมนเทิล-แฮนส์เชล (MH) ในกรณี
ของความลำเอียงแบบไม่มีสม่าเสมอ

งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบ

งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบนั้นแบ่งเป็นงานวิจัยภายใน
ประเทศและงานวิจัยต่างประเทศ ดังนี้

งานวิจัยภายในประเทศ

งานวิจัยในประเทศที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบเป็นการเปรียบเทียบผลการวิเคราะห์ความลำเอียง ที่ใช้วิธีวิเคราะห์ต่างกัน โดยพิจารณาความลำเอียงที่เกิดขึ้นระหว่างประชากรกลุ่มย่อยที่มีเพศต่างกัน สภาพภูมิศาสตร์ต่างกัน และการพัฒนาเกณฑ์การตัดสินความลำเอียงทางเพศ ผู้วิจัยรวบรวมและสรุปได้ ดังนี้

ชัยชัย เผ่าพงศ์ (2527 : 73-76) ได้ศึกษาความลำเอียงของข้อสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์และภาษา ระหว่างกลุ่มเพศชายและเพศหญิง ด้วยวิธีโดคิงลักษณะข้อสอบแบบจำลองโลจิสติก 3 พารามิเตอร์ โดยใช้แบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์และภาษา ในระดับมัธยมศึกษา ซึ่งพัฒนาโดยสำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร กลุ่มตัวอย่างที่ใช้ในการศึกษา เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2524 ซึ่งประกอบด้วยทุกภาคภูมิศาสตร์ทั่วประเทศ 5 ภาค คือ ภาคเหนือ ภาคกลาง ภาคใต้ ภาคตะวันออก และ ภาคตะวันออกเฉียงเหนือ จำนวน 5,248 คน แบ่งเป็นกลุ่มที่ใช้วิเคราะห์ความลำเอียงของแบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์ จำนวน 2,947 คน เป็นนักเรียนชาย 1,610 คน และนักเรียนหญิง 1,337 คน อีกกลุ่มหนึ่งใช้วิเคราะห์ความลำเอียง ของแบบทดสอบด้านภาษา จำนวน 2,301 คน เป็นนักเรียนชาย 1,316 คน นักเรียนหญิง 985 คน

ผลการศึกษาพบว่า แบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์ จำนวน 30 ข้อ มีข้อสอบที่มีความลำเอียงต่อกลุ่มใดกลุ่มหนึ่งโดยเฉพาะ 9 ข้อ คือ ลำเอียงต่อกลุ่มนักเรียนชาย 7 ข้อ และลำเอียงต่อกลุ่มนักเรียนหญิง 2 ข้อ เมื่อพิจารณาแยกตามระดับความสามารถมีข้อสอบที่ลำเอียงต่อกลุ่มนักเรียนชายและหญิงในระดับความสามารถปานกลางขึ้นไปมี จำนวน 5 ข้อ ซึ่งวัดในเรื่องร้อยละ การหาปริมาตรและการหาความยาวของเส้นในด้านของรูปสามเหลี่ยม จำนวนเรื่องละ 1 ข้อ อีกจำนวน 2 ข้อ เป็นเรื่องเกี่ยวกับโจทย์ปัญหา จากข้อสอบจำนวน 5 ข้อ ดังกล่าวนี้ เป็นข้อสอบที่มีความลำเอียงต่อกลุ่มนักเรียนหญิงในระดับความสามารถต่ำ และลำเอียงต่อกลุ่มนักเรียนชายในระดับความสามารถปานกลาง จำนวน 1 ข้อ ส่วนแบบทดสอบวัดความถนัดทางการเรียนด้านภาษาจำนวน 40 ข้อ มีข้อสอบที่ลำเอียงต่อกลุ่มใดกลุ่มหนึ่งโดยเฉพาะ 11 ข้อ เป็นข้อสอบที่ลำเอียงต่อนักเรียนชาย 3 ข้อ และลำเอียงต่อนักเรียนหญิง 8 ข้อ เมื่อพิจารณาแยกตามระดับความสามารถมีข้อสอบที่มีความลำเอียงในระดับความสามารถปานกลางขึ้นไปมี จำนวน

9 ข้อ ซึ่งวัดความเข้าใจเกี่ยวกับการอ่านคำประพันธ์ บทหรือกรอน อย่างละ 1 ข้อ และวัดความเข้าใจเกี่ยวกับการอ่านข้อความ จำนวน 7 ข้อ จากข้อสอบจำนวน 9 ข้อนี้เป็นข้อที่มีความลำเอียงต่อกลุ่มนักเรียนชาย จำนวน 1 ข้อ และเป็นข้อสอบที่มีความลำเอียงต่อกลุ่มนักเรียนหญิง จำนวน 6 ข้อ และมีข้อสอบที่มีความลำเอียงมากกว่าหนึ่งช่วงความสามารถ จำนวน 2 ข้อ จากผลการวิจัยครั้งนี้ เขาสรุปว่าแบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์และด้านภาษาฉบับนี้ มีความลำเอียงต่อเพศ ดังนั้น เขาเสนอว่าในการที่จะนำแบบทดสอบฉบับนี้ไปทดสอบกับกลุ่มเพศชายและเพศหญิงควรจะพิจารณาข้อที่มีค่าอำนาจจำแนกของทั้ง 2 กลุ่ม ที่แตกต่างกันไม่มาก ไม่ว่าข้อสอบข้อนั้นๆ จะมีค่าอำนาจจำแนกสูงหรือต่ำก็ตาม

ทัศนีย์ พิรมนตรี (2530 : 76-78) ได้ศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของแบบทดสอบวิชาคณิตศาสตร์ โครงการตรวจสอบคุณภาพการศึกษาของสำนักงานทดสอบทางการศึกษาระดับชาตินโยบาย การตรวจศึกษาธิการ โดยเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างกลุ่มนักเรียน ในกรุงเทพมหานครกับกลุ่มนักเรียนในภาคภูมิศาสตร์ ทั้ง 5 ภาค คือ ภาคกลาง ภาคใต้ ภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคตะวันออก ด้วยวิธีวิเคราะห์ 3 วิธี คือ วิธีกำหนดจุดค่าเฉลี่ย วิธีทดสอบความแตกต่างระหว่างกลุ่มด้วยสถิติไคสแควร์ในโมเดลล็อกลิเนียร์ 2 โมเดล คือ โมเดลที่ไม่มีพารามิเตอร์ผลรวมระหว่างระดับคะแนนกับกลุ่ม โมเดลที่ไม่มีพารามิเตอร์ของผลหลักที่เกิดจากกลุ่ม และวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ กลุ่มตัวอย่างที่ใช้ศึกษาเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 จำนวน 7,036 คน

ผลการศึกษา พบว่า เมื่อแยกข้อสอบออกตามค่าความยากที่วิเคราะห์ด้วยวิธีดั้งเดิมมีข้อสอบที่มีความลำเอียง จำนวน 43 ข้อ และไม่ลำเอียง 17 ข้อ เหมือนกันทุกภาค เมื่อเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างกลุ่มนักเรียนในกรุงเทพมหานคร และกลุ่มนักเรียนภาคอื่น ๆ พบว่า การวิเคราะห์ความลำเอียงด้วยวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ พบข้อสอบที่มีความลำเอียงจำนวนมากที่สุด และมีข้อสอบที่มีความลำเอียงซ้ำกันระหว่างกลุ่มนักเรียนทุกๆ ภาคกับนักเรียนในกรุงเทพมหานคร ในแต่ละวิธีมีจำนวนไม่เท่ากัน โดยที่วิธีกำหนดจุดค่าเฉลี่ย และวิธีทดสอบความแตกต่างระหว่างกลุ่มด้วยสถิติไคสแควร์ในโมเดลล็อกลิเนียร์ 2 โมเดล มีจำนวนข้อสอบที่มีความลำเอียงมากที่สุดระหว่างกรุงเทพมหานคร และภาคตะวันออกเฉียงเหนือ ส่วนวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ มีจำนวนข้อสอบที่มีความลำเอียงมากที่สุดระหว่างกรุงเทพมหานคร และภาคตะวันออก

การเปรียบเทียบผลการวิเคราะห์ความลำเอียง ทั้ง 3 วิธี กับความลำเอียงที่เกิดขึ้นในภาคเดียวกันพบข้อสอบที่มีความลำเอียงซ้ำกัน ข้อสอบที่มีความลำเอียงส่วนมากเป็นข้อที่ง่ายสำหรับนักเรียนในกรุงเทพมหานครมากกว่ากลุ่มนักเรียนภาคอื่น ๆ จากการวิเคราะห์ด้วยวิธีกำหนดจุดเดอลตา แต่เป็นข้อที่มีความลำเอียงในเกณฑ์ต่ำ สำหรับวิธีโด่งลักษณะข้อสอบ 3 พารามิเตอร์ และเป็นข้อที่ลำเอียงรูปแบบเดียวกันในการวิเคราะห์ด้วย วิธีทดสอบความแตกต่างระหว่างกลุ่มด้วยสถิติไคสแควร์ในโมเดลล็อกลิเนียร์ ทั้ง 2 โมเดล

พัชรวิ ปิธภัท (2531 : 65-67) ได้ศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบ จากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ด้วยวิธีแปลงค่าความยาก วิธีไคสแควร์ และวิธีโด่งลักษณะข้อสอบ 1 พารามิเตอร์ ตัวแปรที่ศึกษา คือ เพศ และภาคภูมิศาสตร์ กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักเรียนชายและนักเรียนหญิงชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2530 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติในกรุงเทพมหานครและจังหวัดสมุทรสาคร จำนวน 1,948 คน เพื่อวิเคราะห์หาดัชนีความลำเอียงของข้อสอบ แล้วหาสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียงทั้ง 3 วิธี และศึกษาเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบก่อน และหลังการคัดเลือกข้อสอบที่ลำเอียงออกแล้ว

ผลการศึกษา พบว่า การวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีโด่งลักษณะข้อสอบพบจำนวนข้อสอบที่ลำเอียงมากที่สุด รองลงมา คือ วิธีไคสแควร์ และวิธีแปลงค่าความยากพบจำนวนข้อสอบที่ลำเอียงน้อยที่สุด และมีค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงคือเพศ ระหว่างวิธีวิเคราะห์ความลำเอียง 3 วิธี มีค่าอยู่ระหว่าง .1713 ถึง .5618 ค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงต่อกลุ่มนักเรียนกรุงเทพมหานคร กับกลุ่มนักเรียนสมุทรสาครมีค่าอยู่ระหว่าง .1868 ถึง .6009 ส่วนค่าความเชื่อมั่นของแบบทดสอบก่อนและหลังคัดเลือกข้อสอบที่ลำเอียงแล้วแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

สุรศักดิ์ อมรรัตนศักดิ์ (2531 : 200-207) ได้ศึกษาเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่แตกต่างกัน 4 วิธี คือ วิธีวิเคราะห์ความแปรปรวน วิธีแปลงค่าความยากของข้อสอบ วิธีโด่งลักษณะข้อสอบที่มี 1 พารามิเตอร์ และวิธีโด่งลักษณะข้อสอบที่มี 3 พารามิเตอร์ จากแบบทดสอบที่ใช้สอบแข่งขันเพื่อบรรจุเข้ารับราชการเป็นข้าราชการครูในปี พ.ศ. 2529 ซึ่งมี 4 ฉบับ คือ ฉบับที่ 1 แบบทดสอบวิชาเอก (ความรู้เกี่ยวกับการประถมศึกษา) ฉบับที่ 2 แบบทดสอบวิชาการศึกษาและกฎหมาย ฉบับที่ 3 แบบทดสอบวิชาภาษาไทย ฉบับที่ 4 แบบทดสอบวิชาความสามารถทั่วไป เพื่อหาสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงระหว่างวิธี

วิเคราะห์ความลำเอียง 4 วิธี และเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อนและหลัง การศึกษาความลำเอียงของข้อสอบตามวิธีการคิดคะแนนรวมที่แตกต่างกัน 6 วิธี ในด้านจำนวนผู้ ได้รับการคัดเลือก สัดส่วนของชายต่อหญิงที่ได้รับการคัดเลือกและความเชื่อมั่นของแบบทดสอบ กลุ่มตัวอย่างที่ศึกษา คือ ผู้เข้าสอบแข่งขันเพื่อบรรจุเข้ารับราชการเป็นข้าราชการครู ในปี พ.ศ. 2529 จำนวน 2,340 คน เป็นชาย 1,170 คน หญิง 1,170 คน

ผลการศึกษาพบว่า วิธีวิเคราะห์ความลำเอียงแต่ละวิธีพบจำนวนข้อสอบที่มีความลำเอียง ต่างกัน โดยวิธีโดคิงลักษณะข้อสอบที่มี 3 พารามิเตอร์ พบจำนวนข้อสอบที่มีความลำเอียงมากที่สุด รองลงมา ได้แก่ วิธีวิเคราะห์ความแปรปรวน และวิธีแปลงค่าความยากของข้อสอบ พบจำนวนข้อ สอบที่มีความลำเอียงน้อยที่สุด ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงระหว่างวิธี วิเคราะห์ความลำเอียง 4 วิธี พบว่า มีค่าอยู่ระหว่าง .7535 ถึง .9921 สำหรับจำนวนผู้ได้รับ การคัดเลือกเมื่อเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อนและหลังการศึกษาความลำเอียง ของข้อสอบตามวิธีคิดคะแนนดิบกับวิธีรวมคะแนนแบบอื่น ๆ อีก 5 วิธี มีจำนวนผู้ได้รับการคัดเลือก แตกต่างกันประมาณ 4 เปอร์เซนต์ถึง 24 เปอร์เซนต์ ส่วนวิธีแปลงคะแนนดิบเป็นคะแนนมาตรฐาน ที่ปกติกับวิธีรวมคะแนนแบบอื่นๆ อีก 4 วิธี มีจำนวนผู้ได้รับการคัดเลือกต่างกันประมาณ 6 เปอร์เซนต์ ถึง 23 เปอร์เซนต์ ส่วนความเชื่อมั่นของแบบทดสอบหลังจากคัดข้อสอบที่มีความลำเอียงออกพบว่าค่า ความเชื่อมั่นของแบบทดสอบลดลงกว่าแบบทดสอบเดิมเล็กน้อย

สุพันธ์ สุกมลสันต์ (2534) ได้วิเคราะห์ความลำเอียงของข้อสอบภาษาอังกฤษเพื่อคัด เลือกเข้าศึกษาต่อในมหาวิทยาลัย ด้วยวิธีวิเคราะห์ความลำเอียง 3 วิธี ได้แก่ วิธีกำหนดจุด ค่าเดลตา (Delta-Plot Method) วิธีโคสแควร์ชนิดที่แบ่งความสามารถของผู้สอบเป็น 3 ระดับ ได้แก่ กลุ่มที่มีระดับความสามารถต่ำ (ผู้ได้คะแนนรวม 0-40 คะแนน) กลุ่มที่มีระดับความสามารถ ปานกลาง (ผู้ได้คะแนนรวม 41-70 คะแนน) และกลุ่มที่มีระดับความสามารถสูง (ผู้ได้คะแนนรวม 71-100 คะแนน) วิธีการวัดพื้นที่ความแตกต่างระหว่างโดคิงลักษณะข้อสอบที่วิเคราะห์ตามทฤษฎีการ ตอบข้อคำถามแบบ 3 พารามิเตอร์ จากแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยชุด กข และชุด กขค ปี 2531-2533 ซึ่งมีข้อสอบชุดละ 100 ข้อ ตัวแปรที่ศึกษาความลำเอียง คือ เพศและภาค ภูมิศาสตร์ของผู้สอบ ซึ่งแยกออกเป็น 5 ภาค ตามภูมิภาเนาของผู้สอบ ได้แก่ ภาคกลาง ภาค ตะวันออก ภาคตะวันออกเฉียงเหนือ ภาคเหนือ และภาคใต้

เกณฑ์การตัดสินความล่าเอียงของข้อสอบ สำหรับวิธีต่างๆ ที่ สุวัฒน์ สุขมลินต์ กำหนดขึ้น คือ

1. วิธีกำหนดจุดค่าเฉลี่ย ข้อสอบที่ล่าเอียง คือ ข้อที่มีระยะตั้งฉากจากคู่อันดับค่าเฉลี่ยไปยังเส้นแกนหลัก (d) มากกว่า 1.964 Sd. หรือน้อยกว่า -1.964 Sd.
2. วิธีไคสแควร์ ข้อสอบที่ล่าเอียง คือ ข้อสอบที่ข้อสอบต่างกลุ่มกันที่อยู่ในระดับคะแนนเดียวกันมีส่วนการตอบถูกหรือผิดแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ $\alpha = 0.017$
3. วิธีการวัดพื้นที่ความแตกต่างระหว่างโค้งลักษณะข้อสอบ ข้อสอบที่ล่าเอียง คือ ข้อที่มีพื้นที่ความแตกต่างมากกว่า 0.40 โดยแบ่งเป็น 3 ระดับ
 - 3.1 พื้นที่ ที่มากกว่า 0.70 แสดงว่ามีความล่าเอียงมาก
 - 3.2 พื้นที่ที่มีค่าอยู่ระหว่าง 0.40 - 0.70 แสดงว่ามีความล่าเอียงปานกลาง
 - 3.3 พื้นที่ที่มีค่ามากกว่า 0.00 แต่น้อยกว่า 0.40 แสดงว่ามีความล่าเอียงน้อย

ผลการศึกษา พบว่า แบบทดสอบภาษาอังกฤษ ฉบับ กข และ กขค ปี 2531-2533 มีความล่าเอียงต่อเพศในช่วง 7-28 ข้อ และช่วง 4-41 ข้อ โดยมีแนวโน้มที่ล่าเอียงต่อเพศชายมากกว่าเพศหญิง และล่าเอียงต่อภาคภูมิศาสตร์ในช่วง 6-45 ข้อ และ 5-43 ข้อ โดยมีความล่าเอียงต่อผู้สอบจากภาคอื่นมากกว่าภาคกลางประมาณ 2-3 เท่า

การเปรียบเทียบผลของการวิเคราะห์โดยใช้วิธีวิเคราะห์ต่างกัน พบว่า มีจำนวนข้อสอบที่ระบุว่าล่าเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติ และแต่ละวิธีให้ผลที่มีความสัมพันธ์กันอย่างไม่มีนัยสำคัญทางสถิติ การวิเคราะห์ด้วยวิธีของทฤษฎีการตอบข้อคำถามเมื่อใช้เกณฑ์พื้นที่แสดงความล่าเอียงของข้อสอบ มากกว่า 0.40 พบว่า มีจำนวนข้อสอบที่มีความล่าเอียงมากที่สุด รองลงมา ได้แก่ วิธีไคสแควร์ และวิธีกำหนดจุดเฉลี่ย ซึ่งวิธีสุดท้ายนี้พบข้อสอบล่าเอียงน้อยกว่า 2 วิธีแรกประมาณ 3-4 เท่า

นิรมล ชัยสวัสดิ์ (2537 : 57-58) ได้ศึกษาเปรียบเทียบผลการวิเคราะห์ความล่าเอียงของแบบทดสอบวัดความเข้าใจในการอ่านภาษาไทยตามทฤษฎีคลาสสิกอลที่ใช้ วิธีวิเคราะห์ต่างกัน 3 วิธีคือ วิธีแปลงค่าความยาก วิธีไคสแควร์ และวิธีวิเคราะห์ความแปรปรวน จากแบบทดสอบวัดความเข้าใจในการอ่านภาษาไทย จำนวน 50 ข้อ กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักเรียนชายหญิงชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2536 ของโรงเรียนในสังกัดกรมสามัญศึกษา เขตพระโขนง กรุงเทพมหานคร จำนวน 1,006 คน โดยศึกษาเปรียบเทียบจำนวนข้อที่มีความล่าเอียง ค่าความ

เชื่อมั่นที่ใช้สูตรคำนวณแบบ KR-20 และแบบแบ่งครึ่งฉบับ ของแบบทดสอบหลังคิดเลือกข้อสอบที่มีความลำเอียงออกแล้ว

ผลการศึกษา พบว่า การวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากกับวิธีไคสแควร์ และ วิธีแปลงค่าความยากกับวิธีวิเคราะห์ความแปรปรวน มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีไคสแควร์กับวิธีวิเคราะห์ความแปรปรวนพบจำนวนข้อสอบที่มีความลำเอียง แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

สำหรับค่าความเชื่อมั่นของแบบทดสอบที่ใช้สูตรคำนวณทั้งแบบ KR-20 และ แบบแบ่งครึ่งฉบับหลังคิดเลือกข้อสอบที่มีความลำเอียงออกแล้ว จากการวิเคราะห์ด้วยวิธีแปลงค่าความยาก วิธีไคสแควร์และวิธีวิเคราะห์ความแปรปรวนแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

กาญจนา วัฒนสุนทร (2538) ได้ศึกษาเกี่ยวกับการพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ โดยใช้ดัชนีการวิเคราะห์ 4 ตัว คือ พื้นที่ระหว่างโค้งการตอบข้อสอบชนิดมีเครื่องหมาย(SA) และ ไม่มีเครื่องหมาย(UA) จากวิธีใช้ทฤษฎีการตอบข้อคำถามตามแบบจำลอง 2 พารามิเตอร์ α_{MH} จากวิธีของแมนเทิล-แฮนส์เซลล์ และ $B_{S,UB}$ จากวิธี SIBTEST จากข้อมูลเชิงประจักษ์ เพื่อพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ

ข้อมูลที่ใช้ศึกษาเป็นข้อมูลการตอบข้อสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของทบวงมหาวิทยาลัย ปีการศึกษา 2535 โดยใช้ความยาวของแบบทดสอบ 20 , 30 และ 40 ข้อ สำหรับวิชาคณิตศาสตร์ และ 50,60,70 และ 80 ข้อ สำหรับวิชาภาษาอังกฤษ โดยใช้กลุ่มผู้สอบ 6 ขนาด คือ 100, 200, 400, 600, 800 และ 1,000 คน

การพัฒนาเกณฑ์กระทำโดยคำนวณค่าดัชนี ทั้ง 4 ตัว จากข้อมูลการตอบข้อสอบของผู้สอบเพศเดียวกัน เพศละ 50 ค่า สำหรับแต่ละความยาวของแบบทดสอบ และขนาดของกลุ่มผู้สอบ จากนั้นนำค่าดัชนีที่ได้ทั้งหมดมาวิเคราะห์หาค่าเฉลี่ย และกำหนดเกณฑ์จากค่าเฉลี่ย 2 ลักษณะคือ เกณฑ์ที่กำหนด โดยพิจารณาจากค่าเฉลี่ยรวม จากค่าดัชนีทุกข้อโดยไม่พิจารณาความแตกต่างในด้านความยาวของแบบทดสอบและขนาดของกลุ่มผู้สอบ และเกณฑ์ที่พัฒนาจากค่าเฉลี่ยซึ่งพิจารณาถึงความยาวของแบบทดสอบและขนาดของกลุ่มผู้สอบ

ผลจากการศึกษา พบว่า เมื่อนำเกณฑ์ที่กำหนดไว้ไปตัดสินค่าดัชนี ที่ได้จากการวิเคราะห์ความลำเอียงระหว่างผู้สอบเพศหญิงและชาย ไม่ว่าจะใช้เกณฑ์ลักษณะใด ผลด้านจำนวนข้อสอบในรูปของร้อยละของข้อสอบที่มีความลำเอียง และความสอดคล้องของการตัดสินมีความไม่คงที่ทุก ๆ

ความยาวของแบบทดสอบและทุกๆขนาดกลุ่มผู้สอบ โดยที่เกณฑ์ในลักษณะหลังจะลดความไม่คงที่ลงได้ระดับหนึ่ง ทั้งนี้เนื่องจากค่าดัชนีแต่ละตัวที่ได้จากกลุ่มผู้สอบต่างเพศกันมีความไม่คงที่ จะแปรเปลี่ยนไปเมื่อความยาวของแบบทดสอบที่ใช้ และขนาดของกลุ่มผู้สอบเปลี่ยนไป

นอกจากนั้นพบว่าที่ความยาวของแบบทดสอบ 40 ข้อ สำหรับวิชาคณิตศาสตร์ และ 80 ข้อ สำหรับวิชาภาษาอังกฤษ มีจำนวนข้อสอบที่ได้รับการตัดสินว่ามีความลำเอียงมากกว่าในความยาวอื่นๆ มาก อีกทั้งยังพบว่าเกณฑ์ที่กำหนดได้จากการศึกษาครั้งนี้ มีค่าสูงกว่าเกณฑ์ขั้นต่ำที่ควรใช้ในการพิจารณาของค่า UA ในขณะที่ ค่า α_{HH} มีค่าใกล้เคียงกับเกณฑ์ขั้นต่ำที่ควรใช้ในการตัดสินความลำเอียงของข้อสอบจากเกณฑ์ที่พัฒนาได้ครั้งนี้ คือ $|SA| > .40$, $UA > .50$ เมื่อใช้ความยาวของแบบทดสอบ 50 ข้อขึ้นไป และ $|SA| > .80$ $UA > 1.20$ กรณีความยาวของแบบทดสอบสั้นกว่า 50 ข้อ ค่า $.60 > \alpha_{HH} > 1.40$ และ $|SIB| > .06$

งานวิจัยต่างประเทศ

สำหรับงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบในต่างประเทศนั้น มีผู้วิจัยทำวิจัยไว้หลายท่าน ทั้งที่เป็นการวิเคราะห์ความลำเอียงของข้อสอบจากข้อมูลจำลอง และการวิเคราะห์ความลำเอียงของข้อสอบจากข้อมูลจริง ซึ่งผู้วิจัยรวบรวมและสรุปได้ ดังนี้

การวิเคราะห์ความลำเอียงของข้อสอบจากข้อมูลจำลอง

เมอร์ส และกรอสเซน (Berk. 1982 : 164 ; Citing Merz and Grossen. 1979) ได้ศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบ โดยวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยาก วิธีพอยท์ไบซีเรียล วิธีโคสแควร์ การวิเคราะห์องค์ประกอบ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และ วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ โดยใช้ข้อมูลสมมติ

ผลการศึกษา พบว่า การวิเคราะห์ความลำเอียงด้วยวิธีวิเคราะห์ด้วยพอยท์ไบซีเรียล เป็นวิธีที่ให้ผลการวิเคราะห์สอดคล้องกับวิธีอื่นน้อย ส่วนวิธีแปลงค่าความยากให้ผลการวิเคราะห์สอดคล้องกับวิธีอื่นมาก รวมทั้งวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และ 1 พารามิเตอร์ ให้ผลการวิเคราะห์สอดคล้องกับวิธีอื่นมากเช่นกัน วิธีโคสแควร์และวิธีวิเคราะห์องค์ประกอบ ให้ผลการวิเคราะห์สอดคล้องกับวิธีอื่นปานกลาง แต่วิธีแปลงค่าความยากเป็นวิธีที่ตีความเหมาะสมเพราะไม่มีข้อจำกัดในเรื่องจำนวนกลุ่มตัวอย่าง และการคำนวณ

ชูเนอร์ (Shoener. 1984) ได้ศึกษาเปรียบเทียบวิธีการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีการทางสถิติ ที่ใช้ทฤษฎีการตอบข้อคำถามที่มี 3 พารามิเตอร์ และการตัดสินใจของผู้ตัดสินจำนวน 24 คน เป็นผู้ตัดสินผิวดำ 8 คน ผิวขาว 8 คน และเชื้อสายสเปน 8 คน เป็นชายและหญิง จำนวนเท่าๆกัน โดยใช้แบบทดสอบอิงเกณฑ์ กลุ่มตัวอย่างเป็นนักเรียนชายและหญิงระดับชั้นมัธยมศึกษา ที่มีภูมิหลังทางวัฒนธรรมต่างๆ กัน จำนวน 1,064 คน การตัดสินใจของผู้ตัดสินที่มีความรู้ทางด้านจิตศาสตร์และหลักสูตรจิตศาสตร์เป็นอย่างดี และใช้แบบฟอร์มการให้คะแนนเป็นช่วง (rating form) สำหรับการวิเคราะห์ ได้พิจารณาถึงความสอดคล้องของการตัดสินใจของกลุ่มผู้ตัดสินทั้ง 3 กลุ่ม และความสอดคล้องระหว่างวิธีการทางสถิติและการตัดสินใจของผู้ตัดสิน โดยใช้ค่าสถิติของแคปป่า (Kappa)

ผลการศึกษา พบว่า ความสอดคล้องระหว่างการตัดสินใจโดยผู้เชี่ยวชาญต่างกลุ่มไม่มีนัยสำคัญทางสถิติสำหรับทั้งความลำเอียงทางด้านวัฒนธรรมและทางเพศ แต่พบความสอดคล้องกันอย่างมีนัยสำคัญทางสถิติในตัวบ่งชี้บางตัวที่คำนวณขึ้นเพื่อรวมการให้คะแนนของผู้ตัดสิน แต่ไม่พบความสอดคล้องระหว่างวิธีการวิเคราะห์ความลำเอียงทั้ง 2 วิธี การตัดสินข้อสอบที่พบทางสถิติว่ามีความลำเอียงออกไป ไม่ทำให้ลำดับของคนในกลุ่มย่อยและรวมทั้งหมดเปลี่ยนไป นอกจากนี้ทั้งแบบทดสอบฉบับเดิมและแบบทดสอบที่ตัดสินข้อสอบที่มีความลำเอียงออกแล้ว มีความสัมพันธ์อย่างมีนัยสำคัญทางสถิติกับแบบทดสอบวิชาจิตศาสตร์อิงกลุ่มที่เป็นแบบทดสอบมาตรฐาน

รัดเนอร์ และคนอื่นๆ (Rudner and Others. 1980) ได้เปรียบเทียบวิธีวิเคราะห์หาความลำเอียงของข้อสอบ โดยใช้วิธีวิเคราะห์ ที่ต่างกัน 7 วิธี ได้แก่ วิธีแปลงค่าความยากของข้อสอบ 2 วิธี คือ วิธีใช้เส้นแกนหลัก (TID-MA) และวิธีใช้เส้นแกน 45 องศา (TID-45) วิธีไคสแควร์ 2 วิธี คือ วิธีไคสแควร์ที่แบ่งแบบ 5 ช่วง (CHI-5) และวิธีไคสแควร์ที่แบ่งแบบหลายช่วง (CHI-N) และวิธีโค้งลักษณะข้อสอบ 3 วิธี คือ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ (ICC-3) วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ ชนิดค่าสถิติเหมาะสม (ICC-1F) และชนิดความแตกต่างในความง่ายของข้อสอบ (ICC-1E) โดยใช้ข้อมูลสมมติ (Simulate Data) ที่ผู้วิจัยสร้างขึ้นเอง

ผลการศึกษา พบว่า วิธีโค้งลักษณะข้อสอบที่มี 3 พารามิเตอร์ เป็นวิธีที่ค้นพบข้อสอบที่มีความลำเอียงจากข้อมูลสมมติที่ดีที่สุด รองลงมาได้แก่ วิธีวิเคราะห์ด้วยไคสแควร์แบบ 5 ช่วง

ส่วนวิธีแปลงค่าความยากของข้อสอบทั้ง 2 วิธี ผู้วิจัยแนะนำให้ใช้ได้ทั้งทางปฏิบัติ เมื่อพิจารณาค่าสหสัมพันธ์ของจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียงต่างกัน วิธีโด่งลักษณะข้อสอบ 3 พารามิเตอร์ วิธีไคสแควร์ที่แบ่งแบบ 5 ช่วง และวิธีแปลงค่าความยากที่ใช้เส้นแกน 45 องศา เป็นวิธีที่ถูกต้องมากที่สุด คือ มีค่าสหสัมพันธ์ เท่ากับ .80, .73 และ .68 ส่วนวิธีโด่งลักษณะข้อสอบ 1 พารามิเตอร์ ชนิดค่าสถิติเหมาะสม มีความถูกต้องน้อยที่สุด มีค่าสหสัมพันธ์ เท่ากับ .55 วิธีวิเคราะห์ความลำเอียงทุกวิธี ยกเว้นวิธีโด่งลักษณะข้อสอบ 1 พารามิเตอร์ ชนิดค่าสถิติเหมาะสมสามารถค้นพบข้อสอบที่มีความลำเอียงในกรณีที่มีความยากต่างกันเท่านั้น แต่วิธีโด่งลักษณะข้อสอบ 1 พารามิเตอร์ ชนิดค่าสถิติเหมาะสม มีความไวต่อความลำเอียงของข้อสอบที่มีทั้งความยากและค่าอำนาจจำแนกต่างกัน นอกจากนี้ยังพบว่าวิธีไคสแควร์ที่แบ่งแบบ 5 ช่วง มีประสิทธิภาพเท่า ๆ กับวิธีโด่งลักษณะข้อสอบที่มี 3 พารามิเตอร์ และถ้าไม่คำนึงถึงความยาวของแบบทดสอบหรือขนาดของความลำเอียงของข้อสอบแล้ว ทั้งสองวิธีก็สามารถนำมาใช้แทนกันได้

ทิสเซน สไตน์เบอร์ก และวายนอร์ (Thissen, Steinberg and Wainer, 1988) ได้ศึกษาความแตกต่างของโด่งการตอบข้อสอบระหว่างกลุ่มด้วยทฤษฎีการตอบข้อคำถาม เพื่อศึกษาถึงสภาพการณ์ในการทดสอบสมมติฐานความเท่ากันของค่าพารามิเตอร์ ด้วยค่าไคสแควร์แบบอัตราส่วน Likelihood (IRT-LR) โดยใช้อัตราสมมติ (Simulate Data) แล้วจึงทำการเปรียบเทียบการทดสอบสมมติฐานความเท่ากันของค่าพารามิเตอร์ด้วยวิธี IRT-LR กับวิธีของแมนเทิล-ฮันส์เชล

ผลการศึกษา พบว่า วิธีไคสแควร์แบบอัตราส่วน Likelihood (IRT-LR) สามารถใช้วิเคราะห์ความลำเอียงของข้อสอบได้เมื่อมี จำนวนกลุ่มตัวอย่างน้อยๆ คือ ประมาณ 500 คน ความแตกต่างของค่าความยากของข้อสอบระหว่างกลุ่ม มีค่าเท่ากับ 0.3 และความแตกต่างในด้านค่าสัมประสิทธิ์การเดามีค่าเท่ากับ 0.1 โดยจำนวนข้อสอบรวมจะมากหรือน้อยมากก็ไม่มีผลกระทบ ถ้าความแตกต่างมีค่าน้อยกว่านี้ไม่สามารถที่จะใช้วิธีไคสแควร์แบบอัตราส่วน Likelihood (IRT-LR) วิเคราะห์ได้ว่าข้อสอบลำเอียงหรือไม่ แต่ถ้าใช้กลุ่มตัวอย่างจำนวนมาก ก็อาจจะทำการวิเคราะห์ได้

การระบุความลำเอียงของข้อสอบ ใช้วิธีเปรียบเทียบความแตกต่างระหว่างกลุ่มในข้อสอบที่ต้องการวิเคราะห์ความลำเอียงกับข้อสอบรวม ซึ่งมีข้อตกลงว่าข้อสอบรวมนั้นไม่มีความลำเอียง อย่างไรก็ตามควรได้ศึกษาเพื่อหาวิธีเลือกข้อสอบรวม และการทดสอบความลำเอียงของแบบทดสอบทั้งฉบับ สำหรับผลการวิเคราะห์ความลำเอียงของข้อสอบ ด้วยวิธีไคสแควร์แบบอัตราส่วน

Likelihood (IRT-LR) กับวิธีของแมนเทิล-แฮนส์เชล ทั้งสองวิธีให้ผลการวิเคราะห์คล้ายคลึงกัน ข้อเสนอแนะจากการวิจัยครั้งนี้ คือ วิธีของแมนเทิล-แฮนส์เชล อาจจะใช้ในการวิเคราะห์ความความล่าเอียงของแบบทดสอบยาว ๆ หรืออาจใช้เป็นเครื่องมือในการกลั่นกรองข้อสอบก่อนที่จะใช้วิธีไคสแควร์แบบอัตราส่วน Likelihood (IRT-LR) ต่อไป

สวามินาธาน และโรเจอร์ส (Swaminathan and Rogers. 1990) ได้ตรวจสอบความล่าเอียงของข้อสอบ ด้วยวิธีการถดถอยแบบโลจิสต์กับวิธีของแมนเทิล-แฮนส์เชล (MH) จากการจำลองข้อมูลการตอบข้อสอบด้วยโปรแกรม DATAGEN แบบ 3 พารามิเตอร์ และข้อมูลการจำลองความล่าเอียงให้แบบ 2 พารามิเตอร์ จำลองข้อมูลที่มีความล่าเอียง 2 แบบ คือ ความล่าเอียงแบบสม่ำเสมอ (Uniform) และการล่าเอียงแบบไม่สม่ำเสมอ (Nonuniform) โดยให้ค่าอำนาจจำแนกของ 2 กลุ่มเท่ากัน ส่วนค่าความยากของข้อสอบแปรเปลี่ยนไปตามกลุ่มเพื่อให้เกิดระดับความล่าเอียงต่าง ๆ กัน ตามต้องการ แต่ในลักษณะการล่าเอียงแบบไม่สม่ำเสมอ (Nonuniform) นั้นให้ค่าความยากของข้อสอบของ 2 กลุ่มเท่ากัน แต่ค่าอำนาจจำแนกแปรเปลี่ยนไปตามกลุ่ม โดยศึกษาเกี่ยวกับขนาดกลุ่มตัวอย่าง 2 ขนาด คือ 250 และ 500 คน และความยาวของแบบทดสอบ 3 ขนาด คือ 40, 60 และ 80 ข้อ

ผลการศึกษา พบว่า การตรวจสอบความล่าเอียงของข้อสอบ ทั้ง 2 วิธี ให้ผลใกล้เคียงกัน ในกรณีความล่าเอียงแบบสม่ำเสมอ (Uniform) วิธีของแมนเทิล-แฮนส์เชล ตรวจสอบความล่าเอียงของข้อสอบได้ดีกว่าวิธีการถดถอยแบบโลจิสต์เล็กน้อย โดยมีการตรวจสอบความล่าเอียงได้ถูกต้องร้อยละ 75 กรณีใช้กลุ่มตัวอย่าง 250 คน และตรวจสอบได้ถูกต้องร้อยละ 100 ในกรณีใช้กลุ่มตัวอย่าง 500 คน ในทุกความยาวของแบบทดสอบ แต่ในกรณีของความล่าเอียงแบบไม่สม่ำเสมอ (Nonuniform) นั้น วิธีของแมนเทิล-แฮนส์เชล ไม่สามารถตรวจสอบความล่าเอียงได้ ส่วนวิธีการถดถอยแบบโลจิสต์ ตรวจสอบความล่าเอียงได้ถูกต้องประมาณร้อยละ 50 ในกรณีกลุ่มตัวอย่างน้อยและข้อสอบสั้น และตรวจสอบความล่าเอียงได้ถูกต้องประมาณ ร้อยละ 75 ในกรณีกลุ่มตัวอย่างขนาดใหญ่และแบบทดสอบยาว แต่กรณีที่มีความล่าเอียงแบบไม่สม่ำเสมอ โดยที่โค้งลักษณะข้อสอบตัดกันที่ปลายสุดของโค้งด้านใดด้านหนึ่งไม่ว่าจะเป็นด้านความสามารถสูงหรือต่ำก็ตามวิธีของแมนเทิล-แฮนส์เชล จะตรวจสอบความล่าเอียงได้ดีกว่า

ในการตรวจสอบผิดพลาด คือ ระบุว่าข้อสอบมีความล่าเอียง ทั้ง ๆ ที่ ไม่ล่าเอียงนั้น วิธีของแมนเทิล-แฮนส์เชล ดีกว่า โดยตรวจสอบผิดพลาดประมาณร้อยละ 1 ส่วนวิธีการถดถอยแบบ

โลจิสติกส์ตรวจสอบผิดประมาณร้อยละ 1-6 และเสียค่าใช้จ่ายมากกว่าประมาณ 3-4 เท่า แม้ว่าจะมีความยืดหยุ่นและเป็นแบบทั่วไปมากกว่าวิธีของแมนเทล-แฮนส์เชล

ประโยชน์ของวิธีการถดถอยแบบโลจิสต์ อยู่ตรงที่เป็นวิธีการที่มีพื้นฐานเป็นแบบจำลองที่สามารถเพิ่มตัวแปรความสามารถเข้าไปในแบบจำลองได้ ซึ่งทำให้เข้าใจในธรรมชาติของความล่าช้าเชิงไดค็อกว่า

คลอเซอร์ และคนอื่นๆ (Clauser and Others. 1991) ได้ศึกษาถึงอิทธิพลต่าง ๆ ที่มีผลต่อวิธีการตรวจสอบความล่าช้าของข้อสอบ ด้วยวิธีของแมนเทล-แฮนส์เชล ข้อมูลที่ใช้เป็นข้อมูลที่จำลองขึ้นเพื่อใช้ในการศึกษาการตรวจสอบความล่าช้าแบบสม่าเสมอ ข้อมูลที่จำลองขึ้นเป็นแบบทดสอบจำนวน 75 ข้อ ให้มีความล่าช้าอยู่ 16 ข้อ และมีระดับความแตกต่างของความยากหลายระดับ ระหว่างกลุ่มอ้างอิงและกลุ่มเป้าหมาย กลุ่มตัวอย่างที่ใช้ศึกษา มีจำนวน 3,000 คน เป็นกลุ่มอ้างอิง 1,500 คน กลุ่มเป้าหมาย 1,500 คน มีการวิเคราะห์ระดับความยากของแบบทดสอบที่ต่างกันหลายระดับ กับระดับความยากของแบบทดสอบ 5 ระดับ อำนาจจำแนก 4 ระดับ และการกระจายของความสามารถ 2 ลักษณะ

ผลการศึกษา พบว่า สิ่งที่มีอิทธิพลต่อการตรวจสอบความล่าช้าของข้อสอบด้วยวิธีของแมนเทล-แฮนส์เชล คือ ความแตกต่างในระดับความยากระหว่างผู้สอบทั้งสองกลุ่ม รวมทั้งค่าอำนาจจำแนกและค่าความยากของข้อสอบด้วย จากการวิจัยนี้ทำให้ได้ข้อเสนอแนะว่า วิธีของแมนเทล-แฮนส์เชล นี้ใช้ได้ดี สำหรับข้อสอบที่มีค่าอำนาจจำแนกสูง และมีแนวโน้มจะไม่สามารถตรวจสอบความล่าช้าของข้อสอบ ในข้อสอบที่มีค่าความยากสูงได้

เมเซอร์ และคนอื่นๆ (Mazor and Others. 1991) ได้ศึกษาวิธีการตรวจค้นความล่าช้าของข้อสอบด้วยวิธีของแมนเทล-แฮนส์เชล การตรวจค้นความล่าช้าของข้อสอบจะใช้การทดสอบค่า $MH-\chi^2$ มี $df = 1$ จากข้อมูลที่จำลองขึ้นเป็นข้อสอบ จำนวน 5 ชุด ชุดละ 75 ข้อ โดยศึกษากับกลุ่มตัวอย่างที่มีขนาดต่างกัน 5 ขนาด คือ 2,000, 1,000, 500, 200 และ 100 คน เพื่อเปรียบเทียบการตรวจค้นความล่าช้าของข้อสอบ

ผลการศึกษา พบว่า เมื่อใช้จำนวนกลุ่มตัวอย่างขนาด 2,000 คน วิธีนี้วิเคราะห์ความล่าช้าของข้อสอบผิดพลาดร้อยละ 30 และเมื่อใช้จำนวนกลุ่มตัวอย่างตั้งแต่ 500 คนลงไป วิธีนี้จะวิเคราะห์ความล่าช้าของข้อสอบได้ผิดพลาดร้อยละ 50 ส่วนข้อสอบที่ไม่สามารถตรวจค้นความล่าช้าได้ ได้แก่ ข้อสอบที่ยากมากๆ ข้อสอบที่มีค่าความยากต่างกันเล็กน้อยสำหรับกลุ่มผู้สอบ

2 กลุ่ม และเป็นข้อสอบที่มีค่าอำนาจจำแนกต่ำ

สำหรับการวิเคราะห์ความลำเอียงของข้อสอบโดยใช้ข้อมูลจริง มีผู้ศึกษา ดังนี้

พวงจิต อินทสุวรม (Intasuwarn. 1979 : 92) ได้ศึกษาวิธีวิเคราะห์ความลำเอียงของข้อสอบ ระหว่างกลุ่มผู้สอบที่มีเชื้อชาติต่างกัน ด้วยวิธีวิเคราะห์ 3 วิธี คือ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ วิธีโค้งลักษณะข้อสอบตามแบบจำลองของราส์ และวิธีไคสแควร์ การศึกษาครั้งนี้ใช้ข้อมูลจริง จากการใช้แบบทดสอบ ไอ อี เอ (IEA : The International Association for The Evaluation of Educational Achievement) โดยศึกษากับกลุ่มนักเรียนที่มีอายุ 14 ปี ชาวอังกฤษ อเมริกัน และนิวซีแลนด์ จำนวน 8,601 คน

ผลการศึกษา พบว่า วิธีวิเคราะห์ความลำเอียงทั้ง 3 วิธี มีค่าสหสัมพันธ์อยู่ระหว่าง 0.51-0.98 โดยที่วิธีวิเคราะห์แบบไคสแควร์ และวิธีโค้งลักษณะข้อสอบตามแบบจำลองของราส์ มีค่าสหสัมพันธ์สูงถึง .98

เชปาร์ด์ และคนอื่นๆ (Berk. 1982 : 169 ; Citing Sheppard, Camilli and Averill. 1981) ได้ศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบ ระหว่างกลุ่มผู้สอบที่มีเชื้อชาติต่างกัน โดยวิเคราะห์ความลำเอียงด้วยวิธีไคสแควร์ วิธีกำหนดจุดเดคลา หรือวิธีแปลงค่าความยาก วิธีพอยท์ไบซีเรียล วิธีโค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ และวิธีโค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์ (Rasch Model) จากแบบทดสอบ Lorge-Thorndike ซึ่งเป็นแบบทดสอบที่วัดความสามารถทางสมอง แบ่งเป็น 2 ฉบับ คือ ฉบับภาษาและไม่ใช่ภาษา การวิเคราะห์ความลำเอียงของข้อสอบผู้วิจัยใช้ทั้งเกณฑ์ภายนอก และ เกณฑ์ภายใน โดยที่เกณฑ์ภายนอก คือ คะแนนจาก Raven's Coloured Progressive Matrices ส่วนเกณฑ์ภายใน คือ คะแนนรวมจากแบบทดสอบ กลุ่มตัวอย่างที่ใช้มี 3 เชื้อชาติ จำนวน 1,500 คน เป็นนักเรียนผิวขาว ผิวดำ และชิคาโน (Chicano) เชื้อชาติละ 500 คน

ผลการศึกษา พบว่า วิธีโค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ เป็นวิธีที่เหมาะสมในทางทฤษฎีถูกต้อง และตรงที่สุด แต่มีข้อจำกัด คือ ต้องใช้กลุ่มตัวอย่างขนาดใหญ่ ส่วนวิธีอื่นๆ ก็สามารถใช้ได้ ถ้าผลการวิเคราะห์มีความสอดคล้องกับวิธีโค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ ซึ่งเชปาร์ด์และคณะได้เสนอว่า ในทางปฏิบัติสามารถใช้วิธีไคสแควร์ (Chi-square) แทนวิธีโค้งลักษณะข้อสอบแบบ 3 พารามิเตอร์ และวิธีแปลงค่าความยาก (Angoff) สามารถนำมาใช้แทนวิธีโค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์ (Rasch Model) ได้

แฮมเบิลตัน และคนอื่นๆ (Hambleton and Others. 1986) ได้ศึกษาเปรียบเทียบวิธีการวิเคราะห์ความลำเอียงของข้อสอบ 4 วิธี คือ วิธีของแมนเทิล-แฮนส์เชล วิธีการลงจุดค่าความยาก วิธีค่าความแตกต่างของค่าเฉลี่ยกำลังสอง (the root mean squared difference) และวิธีของพื้นที่รวม ซึ่ง 2 วิธีหลังนี้เป็นวิธีการที่ใช้ทฤษฎีการตอบข้อคำถาม แบบทดสอบที่ใช้ในการศึกษา เป็นแบบทดสอบวัดความสามารถในการอ่านของคลีฟแลนด์ มีข้อสอบ จำนวน 75 ข้อ โดยใช้กลุ่มตัวอย่างจำนวน 937 คน เป็นชาย 451 คน และเป็นหญิง 486 คน ใช้คะแนนจุดตัดในการแปลความหมายค่าสถิติที่แสดงความลำเอียงได้จากการจำลองข้อมูลขึ้น

ผลการศึกษา พบว่า วิธีวิเคราะห์ ทั้ง 4 วิธี ให้ผลการวิเคราะห์ข้อสอบที่มีความลำเอียงได้ใกล้เคียงกัน แต่ปัญหาอยู่ที่ความไม่ชัดเจนในการกำหนดจุดตัด ความคลาดเคลื่อนชนิดที่หนึ่ง และการประมาณค่าพารามิเตอร์ที่ไม่ดี ข้อค้นพบที่เห็นได้อย่างเด่นชัด อยู่ที่ความสำคัญของการเลือกใช้ช่วงคะแนนความสามารถในการวัดความลำเอียง นอกจากนี้พบว่า สามารถใช้วิธีของแมนเทิล-แฮนส์เชลแทนวิธีที่ใช้ทฤษฎีการตอบข้อคำถามได้ เนื่องจากประหยัดกว่าทั้งเงินและเวลา

ดูลิทเทิล (Doolittle. 1987) ได้ศึกษาความแตกต่างในการตอบข้อสอบ ระหว่างผู้สอบที่มีเพศต่างกัน จากข้อสอบ 3 ประเภท ได้แก่ การคำนวณทางคณิตศาสตร์ เรขาคณิตและการใช้เหตุผลในการสอบ ACT (Assessment Mathematics Usage Test) ซึ่งมี 8 ฟอรม์ให้นักเรียนที่เป็นกลุ่มตัวอย่าง 8 กลุ่มทำ โดยจับคู่กลุ่มตัวอย่างชายและหญิงซึ่งเป็นนักเรียนชั้นมัธยมศึกษาตอนปลายด้วย Profile ของวิชาวิทยาศาสตร์ที่ได้เรียนมาแล้ว และได้ออกแบบการวิเคราะห์ข้อมูลโดยใช้วิธี Split Plot Factorial Design ชนิด $2 \times 3 \times 8 \times 8$ (เพศ $\times$ ประเภทของข้อสอบ $\times$ ภูมิหลังด้านคณิตศาสตร์ $\times$ ฟอรม์ของแบบทดสอบ) จุดที่ต้องการเน้นในการศึกษาค้นคว้าครั้งนี้ คือ ปฏิสัมพันธ์ระหว่างเพศกับประเภทของข้อสอบ

ผลการศึกษา แสดงให้เห็นว่า นักเรียนชายและนักเรียนหญิงทำข้อสอบคำนวณได้ดีพอๆ กัน แต่นักเรียนหญิงจะทำข้อสอบเรขาคณิตและข้อสอบเชิงเหตุผลได้น้อยกว่านักเรียนชาย ซึ่งต้องใช้กระบวนการวางแผนก่อนการแก้ปัญหา ผลกระทบที่ใหญ่ที่สุดจากการศึกษาค้นคว้าครั้งนี้ คือ ผลกระทบที่เกิดจากภูมิหลังด้านคณิตศาสตร์และเพศของกลุ่มตัวอย่าง

ดูลิทเทิล และเคลียร์ (Doolittle and Cleary. 1987) ได้ศึกษาวิธีวิเคราะห์ความลำเอียงของข้อสอบ โหมดโลจิสติกแบบ 3 พารามิเตอร์ ที่วิเคราะห์ด้วยดัชนี Z ของลินน์ และฮาร์นิสช์ (Linn and Harnisch. 1981) ระหว่างผู้สอบที่มีเพศต่างกัน โดยใช้ข้อสอบในแบบ

ทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ กลุ่มตัวอย่างที่ใช้มีประมาณ 1,300-1,400 คน เป็นหญิงประมาณร้อยละ 55 และแต่ละกลุ่มมีความเท่าเทียมกัน กลุ่มตัวอย่างจะแบ่งเป็นจำนวน 8 กลุ่ม ให้ทำแบบทดสอบ ACT (Assessment Mathematic Test) รวม 8 ฟอรม์ ถ้าดัชนี Z ที่ได้เป็นบวกแสดงว่าข้อสอบข้อนั้นง่ายสำหรับกลุ่มเปรียบเทียบ ถ้าเป็นลบแสดงว่าข้อสอบข้อนั้นยากสำหรับกลุ่มเปรียบเทียบ

ผลการศึกษา พบว่า ดัชนี Z ที่คำนวณได้มีค่าเป็นลบ ในข้อสอบที่วัดด้านเรขาคณิต และการให้เหตุผลเชิงพีชคณิต สำหรับข้อสอบที่วัดด้านเรขาคณิต การให้เหตุผลเชิงพีชคณิต และ เลขคณิต ในทุกฟอรม์ แสดงว่าง่ายสำหรับเพศชายมากกว่าเพศหญิง สำหรับดัชนี Z ที่เหลือส่วนมากมีค่าเป็นบวก

คิม และโคเฮน (Kim and Cohen. 1991) ได้ศึกษาความลำเอียงของข้อสอบที่สร้างขึ้นอย่างจงใจให้ลำเอียง โดยการเปรียบเทียบผลการคำนวณพื้นที่ความแตกต่างระหว่างโค้งที่ได้จากทฤษฎีการตอบข้อคำถาม ของกลุ่มตัวอย่างผิวขาวและผิวดำ 2 วิธี ได้แก่ การคำนวณพื้นที่ภายใต้ความสามารถที่มีขอบเขตจำกัด และที่ไม่มีขอบเขตจำกัด

จำนวนกลุ่มตัวอย่างที่ใช้กลุ่มละ 1,000 คน ซึ่งสุ่มมาจากเดิมที่มีผิวขาว 1,021 คน และผิวดำ 1,008 คน เพราะข้อจำกัด ในการใช้โปรแกรม BILOG ด้วยไมโคร-คอมพิวเตอร์ เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล คือ แบบทดสอบคำศัพท์เป็นแบบเลือกตอบ ชนิด 4 ตัวเลือก จำนวน 50 ข้อ เป็นคำศัพท์ภาษาอังกฤษมาตรฐาน 40 ข้อ อีก 10 ข้อ เป็นคำศัพท์ที่ความง่ายสำหรับกลุ่มผิวดำ มากกว่ากลุ่มผิวขาว

การวิเคราะห์ความเป็นมิติเดียว (Unidimension) ด้วยการวิเคราะห์องค์ประกอบสำคัญ (Principal Components Analysis) พบว่า องค์ประกอบแรกที่ได้ โดยไม่มีการหมุนแกน และใช้ค่าสหสัมพันธ์แบบเตตตราคลอริก แสดงถึงความเป็นมิติเดียวได้พอเพียงในการใช้แบบจำลองทฤษฎีการตอบข้อคำถาม โดยอธิบายความแปรปรวนได้ร้อยละ 25 ในกลุ่มผู้สอบผิวดำ ในขณะที่อธิบายได้ค่อนข้างน้อยในกลุ่มผู้สอบผิวดำ คือ อธิบายได้ร้อยละ 16

การคำนวณพื้นที่ความแตกต่างระหว่างโค้งของกลุ่มตัวอย่าง 2 กลุ่ม ใช้โมเดลการตอบข้อคำถามแบบ 2 พารามิเตอร์ และแบบ 3 พารามิเตอร์ ชนิดบังคับให้ค่าสัมประสิทธิ์การเดาของทั้งสองกลุ่มมีค่าเท่ากัน ($C = C_p = C_r = .23$) และใช้วิธีการวัดพื้นที่ใต้โค้ง 4 วิธี จากค่าพารามิเตอร์ของข้อสอบที่ปรับแล้ว คือ พื้นที่แบบเปิดคิดเครื่องหมาย (Exact Signed Area : ESA)

ได้จากการคำนวณพื้นที่ตลอดช่วงความสามารถโดยไม่จำกัดขอบเขตเพื่อให้ได้พื้นที่ที่แน่นอน พื้นที่แบบเปิดไม่คิดเครื่องหมาย (Exact Unsigned Area : EUA) พื้นที่แบบปิดคิดเครื่องหมาย (Closed-interval Signed Area : CSA) และพื้นที่แบบปิดไม่คิดเครื่องหมาย (Closed-interval Unsigned Area : CUA) กรณีที่เป็นพื้นที่ ที่คิดเครื่องหมายนั้นกลุ่มอ้างอิง (R) คือ กลุ่มผิวค่า ดังนั้น ถ้าค่า CSA หรือ ESA ที่คำนวณได้มีค่าเป็นบวก แสดงว่าข้อสอบข้อนั้นล่าเอียงเข้าข้างกลุ่มผิวค่า

ข้อสอบที่สร้างเพื่อให้เกิดความล่าเอียงเข้าข้างผิวค่า 10 ข้อ มีค่าเป็น 1 อีก 40 ข้อ ที่เหลือมีค่าเป็น 0 แล้วหาค่าสหสัมพันธ์แบบพอยท์ไบซีเรียล ระหว่างค่าดังกล่าวกับการวัดทั้ง 4 วิธี เพื่อผลการตรวจสอบของแต่ละวิธี และหาค่าสหสัมพันธ์ระหว่างวิธีต่าง ๆ เพื่อความคล้ายคลึงกัน

การประมาณค่าอัตราความคลาดเคลื่อนของ การวิเคราะห์ความล่าเอียง กระทำโดยมีข้อตกลงว่าการวัดพื้นที่แต่ละวิธีมีการกระจายเป็นแบบปกติ ค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานของการวัดแต่ละวิธี จะประมาณจากข้อสอบ 40 ข้อ กำหนดค่าวิกฤตแบบทางเดียวเป็น 2 ค่า คือ ที่ระดับ .05 และ .01 ข้อที่มีพื้นที่ใหญ่กว่าค่าวิกฤตจะเป็นข้อสอบที่มีความล่าเอียง การจัดประเภทข้อสอบเช่นนี้ทำให้เกิดความคลาดเคลื่อน 2 ชนิด ได้แก่ ความคลาดเคลื่อนชนิดระบุว่าไม่ล่าเอียง ทั้ง ๆ ที่ข้อสอบนั้นล่าเอียง (FN = False Negatives) และระบุว่าข้อสอบล่าเอียงทั้งๆ ที่ไม่ล่าเอียง (FP = False Positives) และเมื่อใช้ค่าวิกฤตที่ $p = .05$ พบว่ามี FP มากกว่าที่ค่าวิกฤต $p = .01$

ผลการศึกษา พบว่า ในการวิเคราะห์ความล่าเอียงของข้อสอบด้วยการคำนวณพื้นที่แบบช่วงเปิด และแบบช่วงปิดทั้งสองวิธีก็เกือบจะไม่มี ความแตกต่าง เมื่อพิจารณาการใช้พื้นที่แบบไม่คิดเครื่องหมายไม่ว่าจะเป็นแบบช่วงเปิดหรือแบบช่วงปิด มีความไวต่อการตรวจค้นความล่าเอียงของข้อสอบมากกว่าแบบที่คิดเครื่องหมาย นอกจากนั้นยังพบว่าการคำนวณพื้นที่แบบช่วงเปิด พื้นที่ชนิดคิดเครื่องหมายและไม่คิดเครื่องหมาย มีความสัมพันธ์กันมากกว่า พื้นที่ชนิดคิดเครื่องหมายและไม่คิดเครื่องหมายของการคำนวณพื้นที่แบบช่วงปิด

ในด้านของความคลาดเคลื่อนในการวิเคราะห์ความล่าเอียง พบว่า วิธีคำนวณพื้นที่แบบ 3 พารามิเตอร์ โดยใช้พื้นที่แบบคิดเครื่องหมาย มีอัตราความคลาดเคลื่อนต่ำกว่าวิธีอื่นๆ และไม่มี ความคลาดเคลื่อนเลยที่ระดับนัยสำคัญ .01 แต่มีความคลาดเคลื่อนในการระบุว่าข้อสอบล่าเอียง โดยที่ไม่ล่าเอียงจริง จำนวน 2 ข้อ ที่ระดับนัยสำคัญ .05 ส่วนวิธีคำนวณพื้นที่แบบ 3 พารามิเตอร์

แบบไม่มีเครื่องหมายจะมีอัตราความคลาดเคลื่อนสูงกว่าวิธีอื่น ๆ เล็กน้อย

ราชู และคนอื่น ๆ (Raju and Others. 1993) ได้ศึกษาเปรียบเทียบผลการทดสอบ โคลสแควร์ของลอร์ด การทดสอบ Z ของราชู (1990) ในการวัดพื้นที่ความแตกต่างระหว่างโค้ง ลักษณะข้อสอบทั้งพื้นที่ชนิดที่คิดเครื่องหมายและไม่คิดเครื่องหมาย โดยใช้โมเดล 2 พารามิเตอร์ และวิธีโคลสแควร์ของแมนเทล-แฮนส์เชล ในการวิเคราะห์ความลำเอียงของข้อสอบใช้เกณฑ์การตัดสินความลำเอียง ดังนี้คือ χ^2 ของ Lord มีค่าเท่ากับ 13.80 Z มีค่าอยู่ระหว่าง ± 8.30 และโคลสแควร์ของแมนเทล-แฮนส์เชล มีค่าเท่ากับ 10.83 โดยใช้อัตราข้อมูลจากผลการตอบแบบทดสอบ Gater-Mac Ginitie Reading Tests (GMRT) เฉพาะแบบทดสอบย่อยที่วัดคำศัพท์ 45 ข้อ เป็นแบบเลือกตอบชนิด 5 ตัวเลือก กลุ่มตัวอย่างที่ใช้เป็นนักเรียนระดับ 10 และ 12 ในปี 1987 จำนวน 839 คน เป็นนักเรียนผิวดำ 245 คน นักเรียนผิวขาว 436 คน นักเรียนหญิง 440 คน และนักเรียนชาย 399 คน

ผลการศึกษา พบว่า ในการวิเคราะห์ความลำเอียงระหว่างนักเรียนผิวดำ และนักเรียนผิวขาว จากการทดสอบด้วย χ^2 และ Z ทั้งพื้นที่ชนิดคิดเครื่องหมาย และไม่คิดเครื่องหมาย ทั้ง 3 วิธี พบข้อสอบที่มีความลำเอียง 1 ข้อ ได้แก่ ข้อ 41 ส่วนวิธีโคลสแควร์ของแมนเทล-แฮนส์เชล พบข้อสอบที่มีความลำเอียง 2 ข้อ ได้แก่ ข้อ 14 และ 27 ซึ่งข้อสอบทั้ง 3 ข้อ เป็นข้อที่มีค่าความยากต่างกันมากระหว่างกลุ่มผู้สอบผิวดำและผิวขาว

ส่วนผลการวิเคราะห์ความลำเอียงระหว่างนักเรียนชายและนักเรียนหญิง 3 วิธีแรก พบข้อสอบที่มีความลำเอียงซ้ำกัน 4 ข้อ ส่วนวิธีโคลสแควร์ของแมนเทล-แฮนส์เชล พบข้อสอบที่มีความลำเอียง 5 ข้อ ซ้ำกับ 3 วิธีแรก 4 ข้อ คิดเป็นร้อยละ 80 ได้แก่ ข้อ 2, 18, 23, 33 และข้อสอบที่มีความลำเอียงไม่ซ้ำกับ 3 วิธีแรก คือ ข้อ 41 และข้อสอบทุกข้อมีค่าความยากสูงกว่าข้อสอบที่ไม่มีความลำเอียง

จากการค้นพบว่า วิธีโคลสแควร์ของแมนเทล-แฮนส์เชล ค้นพบข้อสอบที่มีความลำเอียงได้มากกว่าวิธีที่ใช้ทฤษฎีการตอบข้อคำถามเช่นนี้ ด้านกับการค้นพบ ของ สวามินาธานและโรเจอร์ส (1990) ที่พบว่าวิธีที่ใช้ทฤษฎีการตอบข้อคำถามสามารถค้นพบข้อสอบที่มีความลำเอียงได้มากกว่าวิธีของแมนเทล-แฮนส์เชล

ข้อควรสังเกตจากการวิจัยนี้ คือ การทดสอบ Z ของราชูอาจให้ผลที่ถูกต้องกว่าการกำหนดเกณฑ์ตัดสินความแตกต่างระหว่างพื้นที่โค้งลักษณะข้อสอบตามอำเภอใจ เพราะอาจจะเกิด

ค่าความคลาดเคลื่อนมาตรฐานในการวัดซึ่งเป็นตัวหนึ่งที่ใช้เข้ามาเกี่ยวข้อง ทำให้ข้อสอบที่มีความแตกต่างของพื้นที่น้อยกว่าเป็นข้อสอบที่มีความลำเอียง และข้อสอบที่มีค่าความแตกต่างของพื้นที่มากกว่าเป็นข้อสอบที่ไม่มีความลำเอียงได้

โคเฮน และคิม (Cohen and Kim, 1993) ได้ศึกษาเปรียบเทียบการใช้ทฤษฎีการตอบข้อคำถามแบบ 2 พารามิเตอร์ โดยใช้การทดสอบ χ^2 ของลอว์รี และการทดสอบความแตกต่างของพื้นที่ระหว่างโค้งลักษณะข้อสอบชนิดคิดเครื่องหมายและไม่คิดเครื่องหมาย โดยใช้สถิติ Z ของราฐ จากข้อมูลจำลองขึ้นตามทฤษฎีการตอบข้อสอบ 2 พารามิเตอร์ ซึ่งจำลองข้อสอบ 2 ขนาด คือ 20 และ 60 ข้อ จำนวนผู้สอบ 2 ขนาด คือ 100 และ 500 คน และมีจำนวนข้อสอบลำเอียงในแบบทดสอบ ที่จำลอง 3 แบบ คือ มีข้อสอบที่ลำเอียงร้อยละ 0, 10 และ 20

ผลการศึกษา พบว่า

1. แบบทดสอบยิ่งยาวการระบุข้อสอบลำเอียงผิดพลาด (FP) จะมากขึ้นและถ้ามีจำนวนข้อสอบลำเอียงอยู่ในแบบทดสอบน้อย จำนวนข้อสอบที่ถูกระบุว่าลำเอียงผิดพลาดจะยิ่งมากขึ้นด้วย

องค์ประกอบที่มีอิทธิพลต่อการระบุข้อสอบลำเอียงผิดพลาด (FP) อีกประการ คือ ระดับนัยสำคัญ (α) ถ้าค่าระดับนัยสำคัญยิ่งมากจะเกิด FP มาก นั่นคือที่ $\alpha = .05$ จะเกิด FP มากกว่าที่ $\alpha = .01$ สำหรับการทดสอบ Z ของราฐเกิดในทุกสถานการณ์ที่ศึกษา แต่การทดสอบ χ^2 ของลอว์รีจะเกิดในบางสถานการณ์เท่านั้น

2. แบบทดสอบที่ยาวกว่า มีร้อยละของข้อสอบลำเอียงมากกว่าจะมีข้อสอบที่ถูกระบุว่าไม่ลำเอียงผิดพลาด (FN) มากกว่า แต่จะเกิดขึ้นกับการทดสอบด้วย Z ของราฐ มากกว่าการทดสอบด้วย χ^2 ของลอว์รี

กรณีที่จำนวนผู้สอบเท่ากับ 100 คน พบว่า ความแตกต่างในการระบุข้อสอบไม่ลำเอียงผิดพลาด สำหรับ Z ของราฐ และ χ^2 ของลอว์รี ไม่แตกต่างกันมากเท่ากับกรณีที่ใช้กลุ่มตัวอย่าง 500 คน ซึ่งมีการระบุข้อสอบไม่ลำเอียงผิดพลาด สำหรับ χ^2 ของลอว์รี น้อยกว่า Z ของราฐ โดยเฉพาะอย่างยิ่งในกรณีที่ข้อสอบลำเอียงในแบบทดสอบมากที่สุด คือ ร้อยละ 20

โคเฮน และคิม ได้ให้ความเห็นว่า การทดสอบด้วย χ^2 ของลอว์รีให้ผลการตรวจสอบข้อสอบลำเอียงได้ดีกว่าการทดสอบ Z ของพื้นที่ ที่คิดเครื่องหมายอาจเป็นเพราะ Z (ESA) ใช้การเปรียบเทียบเฉพาะค่าความยาก (b) ในขณะที่ χ^2 ใช้ทั้งค่าความยากและค่าอำนาจจำแนก (b และ a) แต่ไม่สามารถอธิบายความแตกต่างได้ในกรณีใช้ Z(H) ซึ่งทดสอบพื้นที่ที่ไม่คิดเครื่องหมาย

หมายเหตุเพราะ $Z(H)$ ก็ใช้การทดสอบค่าพารามิเตอร์ทั้ง 2 ตัวเช่นกัน

อย่างไรก็ตามจากผลการวิจัยเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบที่กล่าวมาปรากฏว่าวิธีวิเคราะห์ความลำเอียงของข้อสอบไม่ว่าจะศึกษาจากข้อมูลสมมติ (Simulate Data) หรือศึกษาจากข้อมูลจริง (Empirical Data) ทุกวิธีมีบางส่วนที่สัมพันธ์กัน และมีบางส่วนที่แตกต่างกัน และให้ผลลัพธ์ที่คล้ายคลึงกัน วิธีที่ใช้ทฤษฎีการตอบข้อคำถามเป็นวิธีที่เหมาะสมในทางทฤษฎีให้ผลการตรวจสอบและการประมาณค่าที่ถูกต้อง และตรงที่สุดในการวิเคราะห์ความลำเอียงของข้อสอบ แต่วิธีนี้มีข้อจำกัดในทางปฏิบัติสูงต้องใช้กลุ่มตัวอย่างขนาดใหญ่ และข้อสอบจำนวนมากโดยเฉพาะการคำนวณและการแปลผลค่อนข้างยุ่งยากซับซ้อน ในขณะที่วิธีแปลงค่าความยากของข้อสอบ และวิธีของแมนเทล-แฮนส์เชลนั้น เป็นวิธีที่ง่ายต่อการคำนวณและการแปลผล แม้จะไม่เหมาะสมในทางทฤษฎีมากนักแต่นักวิจัยต่างก็เห็นว่าเป็นวิธีที่ใช้ได้ในทางปฏิบัติ ซึ่งวิธีที่ได้รับการสนับสนุนให้ใช้วิเคราะห์ความลำเอียงของข้อสอบบ่อยที่สุด ได้แก่ วิธีแปลงค่าความยาก วิธีไคสแควร์ วิธีวิเคราะห์ความแปรปรวน และวิธีที่นำมาใช้กันแพร่หลายอีกวิธีหนึ่ง คือ การวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีของแมนเทล-แฮนส์เชล ซึ่งจากผลการวิจัยดังที่กล่าวมาแล้ว เป็นการวิเคราะห์ความลำเอียงของแบบทดสอบ ที่มีความลำเอียงต่อเพศ สภาพภูมิศาสตร์ วัฒนธรรม และเชื้อชาติต่างกัน แต่ผลจากการวิจัยที่ผ่านมายังไม่สามารถระบุได้ชัดเจนว่า วิธีใดจะให้ผลการวิเคราะห์ที่ดีที่สุดและได้ถูกต้องที่สุด

ด้วยเหตุผลดังกล่าว จึงทำให้ผู้วิจัยสนใจที่จะศึกษาวิธีวิเคราะห์ความลำเอียงของข้อสอบ 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีไคสแควร์ ข้อสอบแบบ 3 พารามิเตอร์ และ วิธีของแมนเทล-แฮนส์เชล โดยศึกษากับตัวแปรที่นับถือศาสนาต่างกัน และเปรียบเทียบผลการวิเคราะห์ความลำเอียงทั้ง 3 วิธี ในด้านจำนวนข้อสอบที่มีความลำเอียง และค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว

สมมติฐานของการศึกษาค้นคว้า

1. ข้อสอบที่ใช้วิธีวิเคราะห์ความลำเอียงต่างกัน ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา มีจำนวนข้อสอบที่ลำเอียงแตกต่างกัน
2. ความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วจากการใช้วิธีวิเคราะห์ต่างกันของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา มีค่าแตกต่างกัน

บทที่ 3

วิธีดำเนินการศึกษาค้นคว้า

ประชากร

ประชากรที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษา ในจังหวัดกระบี่ จำนวน 16 โรงเรียน 94 ห้องเรียน มีจำนวนนักเรียนทั้งหมด 4,028 คน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนชายและนักเรียนหญิง ที่นับถือศาสนาพุทธและศาสนาอิสลาม ชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษา ในจังหวัดกระบี่ จำนวน 2,689 คน ซึ่งเลือกมาโดยการสุ่มอย่างง่าย (Simple Random Sampling) โดยมีห้องเรียนเป็นหน่วยการสุ่ม (Sampling Unit) ซึ่งมีลำดับขั้นตอนในการสุ่ม ดังนี้

ขั้นที่ 1 ประมาณขนาดกลุ่มตัวอย่างทั้งหมดด้วยความเชื่อมั่น 95 เปอร์เซ็นต์ ($\alpha = .05$) เมื่อเทียบจากตารางขนาดกลุ่มตัวอย่าง ปรากฏว่า กลุ่มตัวอย่างต่ำสุดที่เป็นตัวแทนของประชากรได้ 364 คน (ล้วน ศาสนศ และ อังคณา ศาสนศ. 2536 : 260 ; อ้างอิงมาจาก Yamane, 1967)

ขั้นที่ 2 สุ่มโรงเรียนจากประชากรทั้งหมด ได้ 12 โรงเรียน จำนวน 58 ห้องเรียน แล้วทำการสุ่มอย่างง่ายอีกครั้งหนึ่ง โดยมีห้องเรียนเป็นหน่วยการสุ่ม โดยสุ่มแบ่งกลุ่มตัวอย่างออกเป็น 2 กลุ่ม คือ

1. กลุ่มตัวอย่างที่ใช้ในการศึกษาเปรียบเทียบ ผลการวิเคราะห์ด้านจำนวนข้อสอบที่มีความลำเอียง จำนวน 42 ห้องเรียน มีจำนวนนักเรียน 1,936 คน เป็นนักเรียนที่นับถือศาสนาพุทธ จำนวน 1,026 คน และเป็นนักเรียนที่นับถือศาสนาอิสลาม จำนวน 910 คน

2. กลุ่มตัวอย่างที่ใช้ในการศึกษาเปรียบเทียบความเชื่อมั่นของแบบทดสอบหลังคิดข้อสอบที่มีความลำเอียงออกแล้ว จำนวน 16 ห้องเรียน มีจำนวนนักเรียน 753 คน ดังรายละเอียดในตาราง 3

ตาราง 3 จำนวนนักเรียนในกลุ่มตัวอย่างที่ใช้ในการวิจัย แยกตามโรงเรียน

โรงเรียน	จำนวนนักเรียนที่ใช้ในการวิจัย	
	เปรียบเทียบจำนวน ข้อสอบที่มีความ ลำเอียง (ห้อง)	เปรียบเทียบค่าความ เชื่อมั่นหลังคัดข้อสอบที่มี ความลำเอียง (ห้อง)
1. หนองทะเลวิทยา	203(5)	-
2. คลองขางประชานุสรณ์	149(4)	-
3. คลองพนสะถุณีศึกษา	254(6)	-
4. คลองหินพิทยาคม	212(4)	-
5. อำมาตย์พานิชกุล	219(4)	169(3)
6. เหนือคลองประชานุรักษ์	270(6)	-
7. เมืองกระบี่	226(5)	45(1)
8. คลองท่อมราษฎร์รังสรรค์	193(4)	92(2)
9. อำฉลิกประชานุสรณ์	210(4)	157(3)
10. ปลายพระยาพิทยาคม	-	86(2)
11. พนมเบญจา	-	120(2)
12. ลำทับประชานุเคราะห์	-	84(3)
รวม	1,936(42)	753(16)

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้ มี 2 ฉบับ ที่ผู้วิจัยสร้างขึ้นเอง

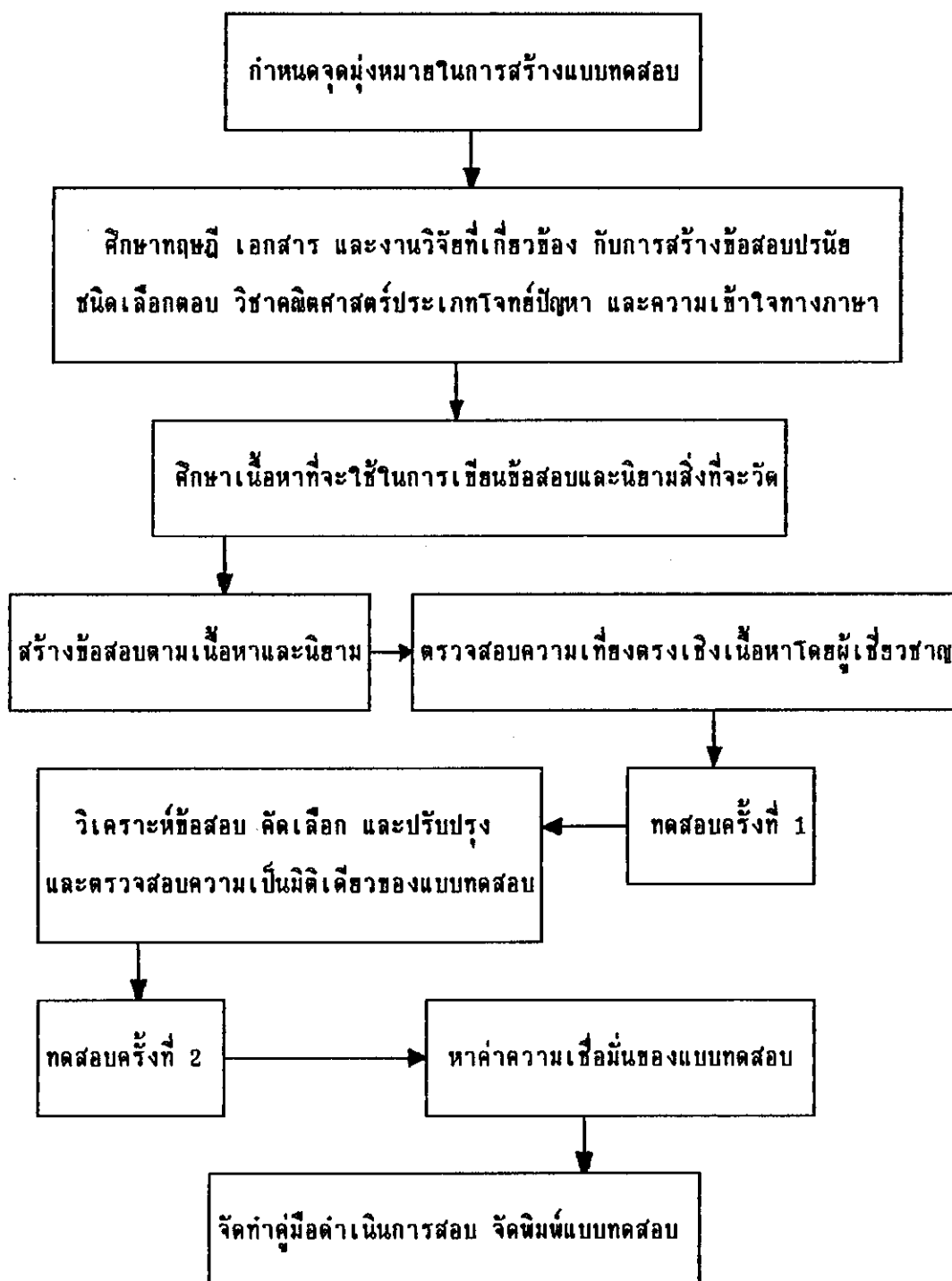
ฉบับที่ 1 แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 (ค 204) ในเรื่องสมการและอสมการ อัตราส่วนและร้อยละ ปริมาตรและพื้นที่ผิว ตามหลักสูตรมัธยมศึกษาตอนต้น พุทธศักราช 2521 (ฉบับปรับปรุง พ.ศ. 2533) เป็นแบบเลือกตอบ ชนิด 5 ตัวเลือก จำนวน 40 ข้อ

ฉบับที่ 2 แบบทดสอบความเข้าใจทางภาษา วิชาภาษาไทย ชั้นมัธยมศึกษาปีที่ 2 เป็นแบบเลือกตอบ ชนิด 5 ตัวเลือก จำนวน 40 ข้อ

วิธีดำเนินการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า

ในการสร้างเครื่องมือสำหรับการศึกษาค้นคว้า ผู้วิจัยได้ดำเนินการสร้างตามขั้นตอน

ดังภาพประกอบ 11



ภาพประกอบ 11 แสดงลำดับขั้นตอนในการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า

รายละเอียดขั้นตอนการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า

จากแผนภูมิค่าเนในการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า มีรายละเอียด ดังนี้

1. กำหนดจุดมุ่งหมายในการสร้างแบบทดสอบ เพื่อสร้างแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 (ค 204) ในเรื่องสมการและอสมการ อัตราส่วนและร้อยละ ปริมาตรและพื้นที่ผิว และแบบทดสอบความเข้าใจทางภาษา เพื่อใช้ในการวิจัยเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบในด้านจำนวนข้อสอบที่มีความลำเอียง ที่ใช้วิธีวิเคราะห์ต่างกัน

2. ศึกษา ทฤษฎี เอกสาร และ งานวิจัยที่เกี่ยวข้องกับวิธีการสร้างแบบทดสอบปรนัย ชนิดเลือกตอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา

3. เขียนนิยามสิ่งที่ต้องการวัด และ ศึกษาเนื้อหาที่จะใช้ในการเขียนข้อสอบ ดังนี้

3.1 เขียนนิยาม เพื่อสร้างแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ ศึกษาเนื้อหา จากหลักสูตร คู่มือครู แบบเรียน และ เอกสารที่เกี่ยวข้องในวิชาคณิตศาสตร์ (ค 204) ชั้นมัธยมศึกษาปีที่ 2 เรื่องสมการและอสมการ อัตราส่วนและร้อยละ ปริมาตรและพื้นที่ผิว ตามหลักสูตรมัธยมศึกษาตอนต้น พุทธศักราช 2521 (ฉบับปรับปรุง พ.ศ. 2533) ของกระทรวงศึกษาธิการ และของสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี (สสวท.) เพื่อกำหนดขอบเขตเนื้อหาวิชา ในการสร้างแบบทดสอบ

3.2 เขียนนิยาม เพื่อสร้างแบบทดสอบความเข้าใจทางภาษา และ ศึกษาเนื้อหา จากคู่มือครู และ เอกสารต่างๆ ในวิชาภาษาไทย ชั้นมัธยมศึกษาปีที่ 2 เพื่อกำหนดสถานการณ์ และ ขอบเขตเนื้อหาวิชา ในการสร้างแบบทดสอบ

4. สร้างแบบทดสอบ

4.1 สร้างแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ เป็นแบบเลือกตอบชนิด 5 ตัวเลือก ตามนิยาม และ เนื้อหาที่กำหนด จำนวน 60 ข้อ

4.2 สร้างแบบทดสอบความเข้าใจทางภาษา เป็นแบบเลือกตอบชนิด 5 ตัวเลือก ตามนิยาม และสถานการณ์ที่กำหนด จำนวน 60 ข้อ

5. นำแบบทดสอบที่สร้างขึ้น ทั้ง 2 ฉบับ จำนวนฉบับละ 60 ข้อ ไปให้ผู้เชี่ยวชาญ ซึ่งมีความรู้ในเนื้อหาและการเขียนข้อสอบในวิชาคณิตศาสตร์ และ ผู้ที่มีความรู้ในเนื้อหาและการเขียนข้อสอบในวิชาภาษาไทย มาแล้วไม่ต่ำกว่า 5 ปี จำนวนวิชาละ 5 คน ดังรายละเอียดในภาคผนวก ก.เป็นผู้ตรวจสอบความเที่ยงตรงเชิงเนื้อหา(Content Validity) ของแบบทดสอบ

เป็นรายชื่อ โดยพิจารณาความสอดคล้องระหว่างข้อสอบแต่ละข้อกับเนื้อหา และ นิยามที่เขียนขึ้น และ ข้อสอบที่ถือว่าวัดได้ตรงตามเนื้อหา และ นิยาม จะต้องได้รับความเห็นชอบจากผู้เชี่ยวชาญ จำนวน 3 ใน 5 คน จึงถือว่าข้อสอบข้อนั้นใช้ได้ ในการวิจัยครั้งนี้ พบว่า ข้อสอบบางข้อต้องปรับปรุงแก้ไขข้อบกพร่องเกี่ยวกับภาษาที่ใช้ หลังจากปรับปรุงแก้ไขข้อบกพร่องแล้ว นำไปให้ผู้เชี่ยวชาญพิจารณาอีกครั้งหนึ่ง ผลปรากฏว่า ข้อสอบทั้ง 2 ฉบับ จำนวนฉบับละ 60 ข้อ ได้รับความเห็นชอบจากผู้เชี่ยวชาญว่ามีความเที่ยงตรงเชิงเนื้อหาทุกข้อ

6. นำแบบทดสอบทั้ง 2 ฉบับ ที่ผ่านการพิจารณาจากผู้เชี่ยวชาญแล้วไปทดลองสอบครั้งที่ 1 ในระหว่างวันที่ 29 เดือนมกราคม พ.ศ. 2539 ถึงวันที่ 7 เดือนกุมภาพันธ์ พ.ศ. 2539 กับกลุ่มตัวอย่างซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ของโรงเรียนสังกัดกรมสามัญศึกษาในจังหวัด กระบี่ โดยสุ่มจากประชากรที่เหลือจากกลุ่มตัวอย่างที่ใช้ในการศึกษาวิจัย จำนวน 5 โรงเรียน มีจำนวนห้องเรียน 9 ห้องเรียน และมีจำนวนนักเรียน 410 คน ซึ่งได้มาโดยวิธีการสุ่มอย่างง่าย (Simple Random Sampling) ใช้ห้องเรียนเป็นหน่วยการสุ่ม (Sampling Unit) ดังรายละเอียด ต่อไปนี้

โรงเรียนอำมาตย์พานิชนุกูล	จำนวน	1	ห้องเรียน	นักเรียน	56 คน
โรงเรียนพนมเบญจา	จำนวน	2	ห้องเรียน	นักเรียน	85 คน
โรงเรียนปลายพระยาวิทยาคม	จำนวน	2	ห้องเรียน	นักเรียน	86 คน
โรงเรียนอ่าวลึกประชาสรรค์	จำนวน	2	ห้องเรียน	นักเรียน	95 คน
โรงเรียนคำทับประธานเคราะห์	จำนวน	2	ห้องเรียน	นักเรียน	88 คน

7. นำผลการสอบของแบบทดสอบทั้ง 2 ฉบับ มาตรวจให้คะแนนโดยใช้หลักว่า ตอบถูก ให้ 1 คะแนน และ ข้อที่ตอบผิดหรือไม่ตอบหรือตอบมากกว่า 1 แห่ง ให้ 0 คะแนน แล้วนำผลการสอบไปวิเคราะห์หาค่าความชากและค่าอำนาจจำแนกของข้อสอบ เป็นรายชื่อและรายชื่อตัวเลือกทั้งตัวถูกและตัวลวงด้วยโปรแกรม Item Analysis เพื่อคัดเลือกข้อสอบที่มีคุณภาพอยู่ในเกณฑ์ คือ ข้อสอบที่มีค่าความชากอยู่ระหว่าง .20 ถึง .80 และ ค่าอำนาจจำแนกตั้งแต่ .20 ขึ้นไป ผลจากการวิเคราะห์ พบว่า แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ จำนวน 60 ข้อ มีค่าความชาก ตั้งแต่ .047 ถึง .909 และ มีค่าอำนาจจำแนก ตั้งแต่ -.073 ถึง .569 ส่วนแบบทดสอบความเข้าใจทางภาษา มีค่าความชาก ตั้งแต่ .032 ถึง .893 และ มีค่าอำนาจจำแนกตั้งแต่ -.099 ถึง .482 ดังรายละเอียดในตาราง 16 และ 17 ภาคผนวก ข. สำหรับการวิจัยครั้งนี้พิจารณา

คัดเลือกข้อสอบไว้ จำนวนฉบับละ 40 ข้อ จากข้อสอบทั้งหมด 60 ข้อ ซึ่งปรากฏค่าความยาก และค่าอำนาจจำแนก ของแบบทดสอบแต่ละฉบับ ดังนี้

แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ ค่าตอบถูกมีค่าความยาก ตั้งแต่ .214 ถึง .794 และ มีค่าอำนาจจำแนก ตั้งแต่ .227 ถึง .569

แบบทดสอบความเข้าใจทางภาษา ค่าตอบถูกมีค่าความยาก ตั้งแต่ .256 ถึง .779 และ มีค่าอำนาจจำแนก ตั้งแต่ .204 ถึง .482

8. นำผลการทดสอบจาก ข้อ 7 ของแบบทดสอบ ทั้ง 2 ฉบับ มาวิเคราะห์องค์ประกอบ (Factor Analysis) เพื่อตรวจสอบความเป็นมิติเดียวของแบบทดสอบ ก่อนที่จะนำไปวิเคราะห์ความลำเอียงของข้อสอบตามทฤษฎีการตอบข้อคำถาม โดยใช้โปรแกรม SPSS/PC⁺ ซึ่งเลือกสกัดองค์ประกอบด้วยวิธีวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis) แล้วหมุนแกนด้วยวิธีแวนแมกซ์ (Varimax) ดังรายละเอียดใน ตาราง 18 และ 19 ภาควนวก ค. ซึ่งปรากฏค่าไอเกน (Eigen Value) ของแบบทดสอบแต่ละฉบับ ดังนี้

แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ มีค่าไอเกนอยู่ระหว่าง 1.00339 ถึง 7.60393 โดยที่ค่าไอเกนขององค์ประกอบที่ 1 มีค่าเท่ากับ 7.60393 ซึ่งสูงกว่าค่าไอเกนขององค์ประกอบที่ 2 เป็นอย่างมาก และ ค่าไอเกนขององค์ประกอบที่ 2 กับองค์ประกอบที่เหลือ มีค่าอยู่ระหว่าง 1.00339 ถึง 1.62804 ซึ่งมีค่าไม่แตกต่างกันมากนัก

แบบทดสอบความเข้าใจทางภาษา มีค่าไอเกนอยู่ระหว่าง 1.01073 ถึง 4.10012 โดยที่ค่าไอเกนขององค์ประกอบที่ 1 มีค่าเท่ากับ 4.10012 ซึ่งสูงกว่าค่าไอเกนขององค์ประกอบที่ 2 เป็นอย่างมาก และ ค่าไอเกนขององค์ประกอบที่ 2 กับองค์ประกอบที่เหลือ มีค่าอยู่ระหว่าง 1.01073 ถึง 1.71896 ซึ่งมีค่าไม่แตกต่างกันมากนัก

ดังนั้น จึงกล่าวได้ว่า แบบทดสอบทั้ง 2 ฉบับ มีความเป็นมิติเดียวพอประมาณ (Warm. 1978 : 104 ; Lord. 1980 : 19) นั่นคือ สามารถนำแบบทดสอบทั้ง 2 ฉบับ ไปวิเคราะห์ความลำเอียงของข้อสอบตามทฤษฎีการตอบข้อคำถามได้

9. นำแบบทดสอบที่คัดเลือกและปรับปรุงแล้ว ทั้ง 2 ฉบับ จำนวนฉบับละ 40 ข้อ ไปทดสอบ ครั้งที่ 2 ในระหว่างวันที่ 12-15 เดือนกุมภาพันธ์ พ.ศ. 2539 กับกลุ่มตัวอย่างซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 จำนวน 306 คน ที่เลือกมาโดยการสุ่มอย่างง่าย (Simple Random Sampling) ใช้ห้องเรียนเป็นหน่วยการสุ่ม (Sampling Unit) ดังรายละเอียดต่อไปนี้

โรงเรียนอำมาตย์พานิชนุกูล	จำนวน	2	ห้องเรียน	นักเรียน	99 คน
โรงเรียนเมืองกระบี่	จำนวน	1	ห้องเรียน	นักเรียน	43 คน
โรงเรียนเหนือคลองประชาราษฎร์	จำนวน	2	ห้องเรียน	นักเรียน	84 คน
โรงเรียนคลองท่อมราษฎร์รังสรรค์	จำนวน	2	ห้องเรียน	นักเรียน	80 คน

เพื่อหาค่าความเชื่อมั่นของแบบทดสอบ ทั้ง 2 ฉบับ โดยใช้สูตร KR-20 ของคูเดอร์-ริชาร์ดสัน ผลการวิเคราะห์ ปรากฏว่า แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ มีค่าความเชื่อมั่นเท่ากับ .837 ส่วนแบบทดสอบความเข้าใจทางภาษา มีค่าความเชื่อมั่น เท่ากับ .731

10. นำแบบทดสอบ ทั้ง 2 ฉบับ ตามข้อ 9 จำนวนฉบับละ 40 ข้อ จัดพิมพ์เป็นรูปเล่ม และ จัดทำคู่มือดำเนินการสอบ เพื่อนำไปใช้ในการวิจัยครั้งนี้

ลักษณะของเครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 (ค 204) ในเรื่องสมการและอสมการ อัตราส่วนและร้อยละ ปริมาตรและพื้นที่ผิว และ แบบทดสอบความเข้าใจทางภาษา เป็นแบบเลือกตอบ ชนิด 5 ตัวเลือก จำนวน 2 ฉบับ ๆ ละ 40 ข้อ ที่ผู้วิจัยสร้างขึ้น ซึ่งมีลักษณะดังนี้

ตัวอย่างข้อสอบโจทย์ปัญหาทางคณิตศาสตร์

(0) นิยมมีเหรียญบาทอยู่จำนวนหนึ่ง พี่สาวให้อีก 4 เหรียญ รวมเป็นเหรียญบาทมากกว่า 10 เหรียญ เดิมนิยมมีเหรียญบาท อย่างน้อยที่สุดกี่เหรียญ?

ก. 5 เหรียญ

ข. 6 เหรียญ

ค. 7 เหรียญ

ง. 8 เหรียญ

จ. 9 เหรียญ

เฉลย ค

(00) นมสดผสมน้ำในอัตราส่วน 4:1 ถ้าหากต้องการน้ำนม จำนวน 60 ลิตร

ต้องใช้นมสดเท่าไร?

- ก. 40 ลิตร
- ข. 42 ลิตร
- ค. 44 ลิตร
- ง. 46 ลิตร
- จ. 48 ลิตร

เฉลย จ

(000) รูปทรงลูกบาศก์ที่มีปริมาตร 27 ลูกบาศก์เซนติเมตร จะมีพื้นที่ผิวทั้งหมดเท่าไร?

- ก. 27 ตารางเซนติเมตร
- ข. 36 ตารางเซนติเมตร
- ค. 48 ตารางเซนติเมตร
- ง. 54 ตารางเซนติเมตร
- จ. 81 ตารางเซนติเมตร

เฉลย ง

ตัวอย่างข้อสอบวัดความเข้าใจทางภาษา

(0) "อันคนดีมิใช่ดีด้วยมีทรัพย์ มิใช่นับโคตรเหง้าเผ่าพงศ์"

คนดีนี้ดีด้วยการงานนานา ที่อุตสาหกรรมตั้งใจใฝ่ต้องงาน"

คำกล่าวข้างต้นนี้ มีจุดมุ่งหมายใด?

- ก. ให้คำแนะนำ
- ข. ให้คำสั่งสอน
- ค. ให้คำตักเตือน
- ง. ให้คำประกาศ
- จ. ให้คำโฆษณาชวนเชื่อ

เฉลย ข

(00) “สอนหนังสือสังฆราช” มีความเหมาะสมคล้ายกับข้อใด?

- ก. ผงเข้าตาตัวเอง
- ข. ปลอบลื่อเข้าป่า
- ค. อัฐยาสซื้อขนมยาส
- ง. คำน้าพริกละลายแม่น้ำ
- จ. เอมะพร้าวหัวไปขายสวน

เฉลย จ

การตรวจให้คะแนน

การตรวจให้คะแนน ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา ชนิด 5 ตัวเลือก ใช้เกณฑ์การตรวจให้คะแนน ดังนี้คือ ถ้าตอบถูกให้ 1 คะแนน ตอบผิดหรือไม่ตอบหรือตอบมากกว่า 1 แห่ง ในข้อนั้น ๆ ให้ 0 คะแนน

วิธีดำเนินการเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลเพื่อทำการวิจัยในครั้งนี้ ผู้วิจัยได้ดำเนินการตามลำดับขั้นตอน ดังนี้

1. นำหนังสือจากบัณฑิตวิทยาลัยไปติดต่อขอความร่วมมือกับโรงเรียนที่ถูกเลือกเป็นกลุ่มตัวอย่าง เพื่อกำหนด วัน เวลา ในการดำเนินการสอบ
2. จัดเตรียมแบบทดสอบให้เพียงพอกับจำนวนนักเรียนที่จะทำการสอบในแต่ละครั้ง
3. วางแผนดำเนินการสอบ โดยผู้วิจัยเป็นผู้ดำเนินการสอบด้วยตนเอง
4. ชี้แจงให้นักเรียนที่เป็นกลุ่มตัวอย่างทราบวัตถุประสงค์ของการสอบ และ ขอความร่วมมือในการสอบ เพื่อให้ได้ผลการสอบตามความเป็นจริงมากที่สุด
5. นำแบบทดสอบทั้ง 2 ฉบับ ๆ ละ 60 ข้อ ไปทดสอบกับกลุ่มตัวอย่างครั้งที่ 1 จำนวน 410 คน
6. นำผลการสอบในข้อ 5 มาตรวจให้คะแนน แล้ววิเคราะห์หาค่าความยาก และ ค่าอำนาจจำแนก ของข้อสอบเป็นรายข้อ แล้วคัดเลือกข้อสอบไว้ ฉบับละ 40 ข้อ

7. นำผลการสอบใน ข้อ 5 ของแบบทดสอบ ทั้ง 2 ฉบับ ฉบับละ 60 ข้อ มาตรวจให้คะแนนแล้ว ตรวจสอบความเป็นมิติเดียวของแบบทดสอบ

8. นำแบบทดสอบที่มีคุณภาพจาก ข้อ 6 ทั้ง 2 ฉบับ ไปทดสอบกับกลุ่มตัวอย่างครั้งที่ 2 จำนวน 306 คน เพื่อหาค่าความเชื่อมั่นของแบบทดสอบทั้งฉบับ

9. นำแบบทดสอบที่มีคุณภาพตาม ข้อ 6 ทั้ง 2 ฉบับ จำนวนฉบับละ 40 ข้อ ไปทดสอบกับกลุ่มตัวอย่างที่ใช้ในการวิจัย กลุ่มที่ 1 จำนวน 1,936 คน เพื่อศึกษาผลการวิเคราะห์ความลำเอียงของข้อสอบในด้านจำนวนข้อสอบที่มีความลำเอียง

10. นำผลการสอบใน ข้อ 9 มาวิเคราะห์ทางสถิติ และทดสอบสมมติฐานของการวิจัย

11. นำแบบทดสอบที่คัดข้อสอบที่มีความลำเอียงออกแล้ว ด้วยวิธีวิเคราะห์ความลำเอียงแต่ละวิธีของแบบทดสอบทั้ง 2 ฉบับ ไปทดสอบกับกลุ่มตัวอย่างที่ใช้ในการวิจัย กลุ่มที่ 2 จำนวน 753 คน ซึ่งใช้ในการเปรียบเทียบผลการวิเคราะห์ค่าความเชื่อมั่นของแบบทดสอบ

สถิติที่ใช้ในการวิเคราะห์ข้อมูล

1. สถิติพื้นฐาน คือ ค่าเฉลี่ย และ ค่าความเบี่ยงเบนมาตรฐาน

2. สถิติที่ใช้ในการหาคุณภาพของแบบทดสอบ เพื่อใช้สร้างแบบทดสอบในการวิจัย ได้แก่

2.1 ค่าความยาก และ ค่าอำนาจจำแนกของแบบทดสอบ ทั้งตัวถูกและตัวหลง ด้วยโปรแกรม Item Analysis

2.2 หาค่าความเชื่อมั่นก่อนคัดเลือกข้อสอบที่มีความลำเอียงออก จากสูตร KR-20

3. สถิติที่ใช้ในการวิเคราะห์ความลำเอียงของข้อสอบ

3.1 วิธีแปลงค่าความยากของข้อสอบ (Transformed Item Difficulty) มีวิธีดำเนินการ ดังนี้ (Angoff. 1982 : 96)

3.1.1 คำนวณค่าความยาก (p) ของข้อสอบทุกข้อจากผู้สอบแต่ละกลุ่ม

3.1.2 แปลงค่า p เป็นคะแนนมาตรฐาน (Z) แล้วเปลี่ยนให้อยู่ในรูปของเดลตา (Δ -Values) โดยใช้สมการ $\Delta = 4Z + 13$

3.1.3 นำค่าความยากในรูปของเดลตาที่คำนวณได้ในข้อ 3.1.2 มาเขียนกราฟ แล้วคำนวณค่าระยะห่างตั้งฉากจากจุดศูนย์กลางเดลตาของข้อสอบแต่ละข้อ ไปยังเส้นแกนหลักจากสูตร

$$d_i = \frac{ax_i - y_i + b}{\sqrt{a^2 + 1}}$$

- เมื่อ d_i แทน ระยะห่างตั้งฉากจากจุดอันดับค่าเฉลี่ยไปยังเส้นแกนหลัก
 x_i แทน ค่าเฉลี่ย ของกลุ่มที่นับถือศาสนาพุทธ สำหรับข้อที่ i
 y_i แทน ค่าเฉลี่ย ของกลุ่มที่นับถือศาสนาอิสลาม สำหรับข้อที่ i
 a แทน ความชันของเส้นแกนหลัก
 b แทน ค่าคงที่ของการตัดแกน y

ซึ่ง a, b คำนวณจากสูตร ดังนี้

$$a = \frac{(S_y^2 - S_x^2) \pm \sqrt{(S_y^2 - S_x^2)^2 + 4r_{xy}^2 S_x^2 S_y^2}}{2r_{xy} S_x S_y}$$

$$b = M_y - aM_x$$

- เมื่อ S_x^2 แทน ความแปรปรวนของค่าเฉลี่ย กลุ่มที่นับถือศาสนาพุทธ
 S_y^2 แทน ความแปรปรวนของค่าเฉลี่ย กลุ่มที่นับถือศาสนาอิสลาม
 S_x แทน ความเบี่ยงเบนมาตรฐานของค่าเฉลี่ย กลุ่มที่นับถือศาสนาพุทธ
 S_y แทน ความเบี่ยงเบนมาตรฐานของค่าเฉลี่ย กลุ่มที่นับถือศาสนาอิสลาม
 r_{xy} แทน สหสัมพันธ์ระหว่างค่าเฉลี่ยของผู้สอบกลุ่มที่นับถือศาสนาพุทธกับ
 กลุ่มที่นับถือศาสนาอิสลาม
 M_x, M_y แทน ค่าเฉลี่ยของค่าเฉลี่ยของกลุ่มที่นับถือศาสนาพุทธและกลุ่มที่นับถือ
 ศาสนาอิสลาม ตามลำดับ

3.1.4 คำนวณค่าความเบี่ยงเบนมาตรฐาน จากสูตร

$$Sd = \sqrt{1 - r_{xy}}$$

3.1.5 เกณฑ์ที่ใช้ในการตัดสินความลำเอียง คือ ระยะห่างจากเส้นแกนหลัก ถ้ามากกว่า 3Sd หรือ น้อยกว่า -3Sd ถือว่าข้อสอบข้อนั้นเป็นข้อสอบที่มีความลำเอียง (Angoff ; Citing in Berk. 1982 : 107)

3.1.6 ระยะห่างที่คำนวณได้ในข้อ 3.1.3 คือ ดัชนีความลำเอียงของข้อสอบ

3.2 วิธีวัดลักษณะข้อสอบที่มี 3 พารามิเตอร์ (Item Characteristic Curve Approach) มีวิธีดำเนินการ ดังนี้

3.2.1 ตรวจสอบความเป็นมิติเดียว (Unidimension) ของแบบทดสอบตาม ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อคำถาม (Item Response Theory) โดยนำข้อมูลที่ได้จากการตอบมาวิเคราะห์ องค์ประกอบ (Factor Analysis) โดยใช้โปรแกรม SPSS/PC+ ซึ่งเลือกสกัดองค์ประกอบด้วยวิธีวิเคราะห์องค์ประกอบหลัก (Principal Component Analysis : PC) แล้วหมุนแกนด้วยวิธีแวนแมกซ์ (Varimax) ซึ่งการตรวจสอบคุณสมบัติการ วัดเพียงมิติเดียวของแบบทดสอบด้วยวิธีการวิเคราะห์องค์ประกอบ จะใช้หลักการพิจารณาจากค่า โยเกน (Eigen Value) ถ้าค่าโยเกนขององค์ประกอบที่ 1 มีค่าสูงกว่าค่าโยเกนขององค์ประกอบที่ 2 อย่างมาก คือ มีค่ามากกว่า 3 เท่า ขององค์ประกอบที่ 2 และ ค่าโยเกนของ องค์ประกอบที่ 2 มีค่าใกล้เคียงกับองค์ประกอบที่เหลือ แสดงว่า แบบทดสอบนั้นมีการวัดเพียงมิติเดียว (Warm. 1978 : 104 ; Lord. 1980 : 21)

3.2.2 ประมวลค่าความสามารถของผู้สอบ และ ค่าพารามิเตอร์ ของข้อสอบ โดยใช้โปรแกรม BILOG3

3.2.3 นำค่าพารามิเตอร์ที่คำนวณได้ในข้อ 3.2.2 มาคำนวณหาโอกาสในการตอบข้อสอบถูก ของผู้สอบแต่ละกลุ่ม โดยใช้ค่าพารามิเตอร์ที่ปรับจาก ข้อ 3.2.2 จากสูตร

$$P_i(\theta_j) = \frac{c_i + (1 - c_i) \frac{e^{Da_i(\theta - b_i)}}{1 + e^{Da_i(\theta - b_i)}}}{1 + e^{Da_i(\theta - b_i)}} ; i = 1, 2, 3, \dots, n$$

เมื่อ θ_j	แทน	ค่าความสามารถของผู้สอบคนที่ j
$P_i(\theta_j)$	แทน	ค่าความน่าจะเป็นที่ผู้สอบ ซึ่งมีระดับความสามารถ θ_j จะตอบข้อสอบข้อที่ i ได้ถูกต้อง
c_i	แทน	พารามิเตอร์ค่าการเดาของข้อสอบข้อที่ i
b_i	แทน	พารามิเตอร์ค่าความยากของข้อสอบข้อที่ i
a_i	แทน	พารามิเตอร์ค่าอำนาจจำแนกของข้อสอบข้อที่ i
D	แทน	ค่าคงที่ ซึ่งมีค่าเท่ากับ 1.702...
e	แทน	ค่าคงที่ ซึ่งมีค่าเท่ากับ 2.7182818...

3.2.4 นำค่าโอกาสที่จะตอบข้อสอบถูกต้องที่คำนวณได้ในข้อ 3.2.3 มาเขียนกราฟแสดงโค้งลักษณะข้อสอบ ของแต่ละกลุ่มไว้บนกราฟเดียวกัน

3.2.5 เปรียบเทียบโค้งลักษณะข้อสอบแต่ละข้อ สำหรับกลุ่มผู้สอบ 2 กลุ่ม ว่าข้อสอบข้อใดบ้างเป็นข้อสอบที่มีความลำเอียง โดยใช้สูตร

$$CUA = \int_{\theta_1}^{\theta_2} [P_H(\theta) - P_F(\theta)] d\theta$$

3.2.6 เกณฑ์ที่ใช้ตัดสินความลำเอียง ใช้ค่า $CUA \geq .40$ เป็นค่าในการกำหนดว่าข้อสอบมีความลำเอียง (ทัศนีย์ พิรมนตรี. 2530 : 9 ; อ้างอิงมาจาก Rudner. 1977)

3.2.7 ค่า CUA ที่คำนวณได้ในข้อ 3.2.5 คือ คำนวณความลำเอียงของข้อสอบ

3.3 วิธีของแมนเทล-แฮนส์เซิล โดยใช้สูตรการคำนวณ ดังนี้

$$\alpha_{MH} = \frac{\frac{\sum_{j=1}^S A_j D_j}{T_j}}{\frac{\sum_{j=1}^S B_j C_j}{T_j}}$$

เมื่อ A_j แทน จำนวนผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบได้ถูกต้อง
 B_j แทน จำนวนผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบไม่ถูกต้อง
 C_j แทน จำนวนผู้สอบกลุ่มเปรียบเทียบกับที่ตอบข้อสอบได้ถูกต้อง
 D_j แทน จำนวนผู้สอบกลุ่มเปรียบเทียบกับที่ตอบข้อสอบไม่ถูกต้อง
 T_j แทน จำนวนผู้สอบทั้ง 2 กลุ่มที่ได้คะแนนรวมจากการสอบอยู่ในช่วง
 ระดับคะแนน $j = A_j + B_j = C_j + D_j$

3.3.1 สถิติที่ใช้ทดสอบค่าความแตกต่างจาก 1.00 อย่างมีนัยสำคัญทางสถิติที่ระดับขึ้นแห่งความเป็นอิสระ เท่ากับ 1 ใช้สูตรการคำนวณ ดังนี้คือ

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

เมื่อ $E(A_j)$ และ $VAR(A_j)$ เป็นค่าความคาดหวังและค่าความแปรปรวน
ของ A ที่ระดับคะแนน j

$$E(A_j) = \frac{N_{Rj} M_{1j}}{T_j}$$

$$VAR(A_j) = \frac{N_{Rj} N_{Fj} M_{1j} M_{0j}}{T_j^2 (T_j - 1)}$$

4. สถิติที่ใช้ในการทดสอบความแตกต่างของจำนวนข้อสอบที่มีความลำเอียง จากการ
วิเคราะห์ความลำเอียงด้วยวิธีต่างกัน ใช้สูตร (ลิวน ศาสยศ และอังคณา ศาสยศ. 2536 : 93)

$$Z = \frac{P_1 - P_2}{S_{P_1 - P_2}}$$

เมื่อ Z แทน ค่าการแจกแจงของ Z
 P_1 แทน สัดส่วนของข้อสอบที่ลำเอียงที่วิเคราะห์ด้วยวิธีที่ 1 คำนวณจาก
จำนวนข้อสอบที่ลำเอียง ที่วิเคราะห์ด้วยวิธีที่ 1 หารด้วย
จำนวนข้อสอบทั้งหมด

$$P_1 = \frac{f_1}{n_1}$$

P_e แทน สัดส่วนของข้อสอบที่ล่าเอียงที่วิเคราะห์ด้วยวิธีที่ 2 คำนวณจาก
จำนวนข้อสอบที่ล่าเอียง ที่วิเคราะห์ด้วยวิธีที่ 2 หารด้วย
จำนวนข้อสอบทั้งหมด

$$P_e = \frac{f_e}{n_e}$$

$S_{P_1 - P_e}$ แทน ค่าความคลาดเคลื่อนมาตรฐาน ของความแตกต่างระหว่าง
สัดส่วนที่วิเคราะห์ความล่าเอียง ด้วยวิธีวิเคราะห์ 2 วิธี
ซึ่งมีสูตร ดังนี้

$$S_{P_1 - P_e} = \sqrt{pq \frac{1}{n_1} + \frac{1}{n_e}}$$

เมื่อ P แทน สัดส่วนของข้อสอบที่มีความล่าเอียงที่วิเคราะห์ด้วยวิธีวิเคราะห์
ทั้ง 2 วิธี ซึ่งคำนวณมาจากจำนวนข้อสอบที่ล่าเอียง ที่วิเคราะห์
วิธีที่ 1 และ วิธีที่ 2 หารด้วยจำนวนข้อสอบของทั้ง 2 วิธี

$$P = \frac{f_1 + f_e}{n_1 + n_e}$$

และ $q = 1 - p$

5. ค่าความเชื่อมั่น ของแบบทดสอบ ใช้สูตรของ คูเดอร์ ริชาร์ดสัน 20 (KR-20)

(ลั่วน ศาสยศ และ อังคณา ศาสยศ. 2536 : 168)

$$r_{kk} = \frac{n}{n-1} \left[1 - \frac{\sum pq}{S^2_{\bar{u}}} \right]$$

เมื่อ r_{kk}	แทน	ความเชื่อมั่นของแบบทดสอบทั้งฉบับ
n	แทน	จำนวนข้อสอบของแบบทดสอบ
p	แทน	สัดส่วนของผู้ตอบถูกในข้อหนึ่ง ๆ
q	แทน	สัดส่วนของผู้ตอบผิดในข้อหนึ่ง ๆ หรือ คือ $1-p$
$S^2_{\bar{u}}$	แทน	ความแปรปรวนของแบบทดสอบทั้งฉบับ

6. ค่าความเชื่อมั่นของแบบทดสอบ ด้วยวิธีแบ่งครึ่งฉบับ (Split-half)

(ลั่วน ศาสยศ และ อังคณา ศาสยศ. 2536 : 166)

$$r_{kk} = \frac{2r_{\frac{1}{2}\frac{1}{2}}}{1+r_{\frac{1}{2}\frac{1}{2}}}$$

เมื่อ r_{kk}	แทน	ความเชื่อมั่นของแบบทดสอบทั้งฉบับ
$r_{\frac{1}{2}\frac{1}{2}}$	แทน	ความเชื่อมั่นของแบบทดสอบครึ่งฉบับโดยใช้สูตร Pearson Product-Moment Coefficient

7. สถิติใช้ทดสอบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดข้อสอบที่มีความลำเอียงออกแล้วจากการวิเคราะห์ความลำเอียงด้วยวิธีต่างกัน ทดสอบด้วยไคสแควร์

(Chi-Square : χ^2) ใช้สูตรการคำนวณ ดังนี้ (Snedecor and Cochran. 1967 : 187)

$$\chi^2 = \sum (n_i - 3) Z_i^2 - \frac{\left[\sum (n_i - 3) Z_i \right]^2}{\sum (n_i - 3)}$$

เมื่อ	χ^2	แทน	ค่าไคสแควร์ (Chi-Square)
	Z_i	แทน	คะแนนมาตรฐานซึ่งแปลงจากสัมประสิทธิ์ความเชื่อมั่นด้วยตาราง Fisher-Z
	n_i	แทน	จำนวนนักเรียนในแต่ละกลุ่ม
	k	แทน	จำนวนแบบทดสอบ
	df	แทน	degrees of freedom ; df = K-1

8. ถ้าพบว่าความเชื่อมั่นที่ทดสอบ มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติแล้ว
ตรวจสอบความแตกต่างเป็นรายคู่ โดยใช้สูตรทดสอบความแตกต่างของคะแนนมาตรฐาน (Z)
(Ferguson. 1981 : 196)

$$Z = \frac{Z_{r1} - Z_{r2}}{\sqrt{\frac{1}{(N_1-3)} + \frac{1}{(N_2-3)}}}$$

เมื่อ	Z	แทน	คะแนนมาตรฐานของโค้งปกติ
Z_{r1} และ Z_{r2}		แทน	คะแนนมาตรฐานที่แปลงจากค่าสัมประสิทธิ์ความเชื่อมั่นด้วย ตาราง Fisher-Z
N_1	แทน		จำนวนนักเรียนในกลุ่มที่นับถือศาสนาพุทธ
N_2	แทน		จำนวนนักเรียนในกลุ่มที่นับถือศาสนาอิสลาม

บทที่ 4

ผลการวิเคราะห์ข้อมูล

ในการเสนอผลการวิเคราะห์ข้อมูลจากการศึกษาครั้งนี้เพื่อความสะดวกและความเข้าใจตรงกัน ในการนำเสนอ และ การแปลความหมายของผลการวิเคราะห์ข้อมูล ผู้วิจัยจึงได้กำหนดสัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล ดังต่อไปนี้

สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล

$\bar{X}$	แทน	ค่าเฉลี่ย
S	แทน	ความเบี่ยงเบนมาตรฐาน
N	แทน	จำนวนนักเรียนในกลุ่มตัวอย่าง
K	แทน	จำนวนข้อของแบบทดสอบ
P	แทน	ค่าความยากง่ายของข้อสอบที่ได้จากการวิเคราะห์โดยใช้ทฤษฎีดั้งเดิม
r	แทน	ค่าอำนาจจำแนกของข้อสอบที่ได้จากการวิเคราะห์โดยใช้ทฤษฎีดั้งเดิม
a	แทน	ค่าอำนาจจำแนกของข้อสอบที่ได้จากการวิเคราะห์โดยใช้ทฤษฎีการตอบข้อคำถาม
b	แทน	ค่าความยากของข้อสอบที่ได้จากการวิเคราะห์โดยใช้ทฤษฎีการตอบข้อคำถาม
c	แทน	ค่าสัมประสิทธิ์การเดาของข้อสอบที่ได้จากการวิเคราะห์โดยใช้ทฤษฎีการตอบข้อคำถาม
d	แทน	ค่าที่แสดงถึงความลำเอียงของข้อสอบจากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากของข้อสอบ
α_{MH}	แทน	ค่าที่แสดงถึงความลำเอียงของข้อสอบจากการวิเคราะห์ความลำเอียงด้วยวิธีของแมนเทิล-แฮนส์เชล ที่ใช้ค่าแอลฟา
χ^2_{MH}	แทน	ค่าที่แสดงถึงความลำเอียงของข้อสอบจากการวิเคราะห์ความลำเอียงด้วยวิธีของแมนเทิล-แฮนส์เชล ที่ใช้ค่าไคสแควร์
TID	แทน	การวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากของข้อสอบ

MH	แทน	การวิเคราะห์ความล่าเอียงด้วยวิธีของแมนเทิล-แฮนส์เซล
ICC	แทน	การวิเคราะห์ความล่าเอียงด้วยวิธีโค้งลักษณะข้อสอบ
r_{L}	แทน	ค่าความเชื่อมั่นของแบบทดสอบที่คำนวณโดยใช้สูตรแบบแบ่งครึ่งฉบับ
KR-20	แทน	ค่าความเชื่อมั่นของแบบทดสอบที่คำนวณโดยใช้สูตรแบบ KR-20
χ^2	แทน	ค่าไค-สแควร์
Z	แทน	ค่านัยสำคัญของการทดสอบด้วยสถิติ Z-test
Z_r	แทน	คะแนนมาตรฐานจากค่าสัมประสิทธิ์ของความเชื่อมั่นด้วยตาราง Fisher-Z

การเสนอผลการวิเคราะห์ข้อมูล

ผลการวิเคราะห์ข้อมูลในการศึกษานี้ ผู้วิจัยได้เสนอเป็นลำดับ ดังนี้

1. ผลการหาค่าสถิติพื้นฐานของค่าพารามิเตอร์ของแบบทดสอบโจทซ์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา
2. ผลการหาค่าดัชนีความล่าเอียงของข้อสอบจากแบบทดสอบโจทซ์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษาจากการวิเคราะห์ความล่าเอียงที่แตกต่างกัน 3 วิธี
3. ผลการเปรียบเทียบจำนวนข้อสอบที่มีความล่าเอียง จากวิธีวิเคราะห์ความล่าเอียงที่แตกต่างกัน 3 วิธี ของแบบทดสอบโจทซ์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา
4. ผลการหาค่าความเชื่อมั่นของแบบทดสอบโจทซ์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษาหลังจกตัดข้อสอบที่มีความล่าเอียงออกแล้ว จากการวิเคราะห์ความล่าเอียงด้วยวิธีที่ต่างกัน โดยใช้สูตรการคำนวณแบบ KR-20 และแบบแบ่งครึ่งฉบับ
5. ผลการเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบโจทซ์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา หลังจกตัดข้อสอบที่มีความล่าเอียงออกแล้ว จากการวิเคราะห์ความล่าเอียงด้วยวิธีที่ต่างกัน

ผลการวิเคราะห์ข้อมูล

1. ผลการหาค่าสถิติพื้นฐานของค่าพารามิเตอร์ ของแบบทดสอบโจทก์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา

หลังจากนำแบบทดสอบโจทก์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา จำนวน 2 ฉบับๆละ 40 ข้อ ไปทดสอบกับนักเรียนกลุ่มตัวอย่าง จำนวน 1,936 คน เป็นนักเรียนที่นับถือศาสนาพุทธ จำนวน 1,026 คน และเป็นนักเรียนที่นับถือศาสนาอิสลาม จำนวน 910 คน ต่อจากนั้น ผู้วิจัยนำผลการสอบของนักเรียนแต่ละคนจากแบบทดสอบแต่ละฉบับ ไปวิเคราะห์ตามทฤษฎีการตอบข้อคำถามโดยใช้โปรแกรม BILOG3 ปรากฏว่า ได้ค่าพารามิเตอร์ของข้อสอบ คือ ค่าอำนาจจำแนก(a) ค่าความยาก(b) และ ค่าสัมประสิทธิ์การเดา(c) ดังแสดงรายละเอียดในตาราง 20 และ 21 ภาคผนวก ง แล้วนำค่าพารามิเตอร์ที่ได้ มาคำนวณหาค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย และ ค่าความเบี่ยงเบนมาตรฐาน ดังปรากฏผลการวิเคราะห์ในตาราง 4

ตาราง 4 ค่าสถิติพื้นฐานของค่าพารามิเตอร์ของแบบทดสอบโจทก์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา ของกลุ่มนักเรียนที่นับถือศาสนาพุทธ และ ศาสนาอิสลาม

แบบทดสอบ	ค่าพารามิเตอร์	a		b		c	
	กลุ่มผู้สอบ	$\bar{X}$	S	$\bar{X}$	S	$\bar{X}$	S
โจทก์ปัญหาทางคณิต-ศาสตร์	ศาสนาพุทธ	1.4840	.8169	1.2656	.6095	.1980	.0655
	ศาสนาอิสลาม	1.1331	.4116	1.3512	.6848	.1744	.0617
	รวมทั้งหมด	1.3304	.6617	1.3042	.6322	.1843	.0670
ความเข้าใจทางภาษา	ศาสนาพุทธ	.6131	.3456	1.1836	1.6615	.1640	.0697
	ศาสนาอิสลาม	.4992	.3547	1.1684	1.6404	.1280	.0461
	รวมทั้งหมด	.5401	.3772	1.2316	1.7353	.1344	.0678

ผลการวิเคราะห์ ค่าสถิติพื้นฐานของค่าพารามิเตอร์ ของแบบทดสอบแต่ละฉบับจาก ตาราง 4 ปรากฏผล ดังนี้

แบบทดสอบจิตวิทยาปัญหาทางคณิตศาสตร์ ในกลุ่มผู้สอบที่นับถือศาสนาพุทธ มีค่าอำนาจจำแนก(a) ของข้อสอบสูงกว่าหรือจำแนกได้ดีกว่า กลุ่มผู้สอบที่นับถือศาสนาอิสลาม ส่วนค่าความยาก(b) ของข้อสอบ ในกลุ่มผู้สอบที่นับถือศาสนาอิสลามมีค่าสูงกว่าหรือยากกว่า กลุ่มผู้สอบที่นับถือศาสนาพุทธ และ ค่าสัมประสิทธิ์การเดา(c) นั้น ในกลุ่มผู้สอบที่นับถือศาสนาพุทธมีค่าสูงกว่ากลุ่มผู้สอบที่นับถือศาสนาอิสลาม สำหรับค่าความเบี่ยงเบนมาตรฐานของค่าอำนาจจำแนกของข้อสอบ ในกลุ่มผู้สอบที่นับถือศาสนาพุทธมีค่าสูงกว่า หรือ มีการกระจายมากกว่า กลุ่มผู้สอบที่นับถือศาสนาอิสลาม แต่ค่าความเบี่ยงเบนมาตรฐานของค่าความยากของข้อสอบ ในกลุ่มผู้สอบที่นับถือศาสนาอิสลามมีค่าสูงกว่าหรือกระจายมากกว่ากลุ่มผู้สอบที่นับถือศาสนาพุทธ และ ค่าความเบี่ยงเบนมาตรฐานของค่าสัมประสิทธิ์การเดาในกลุ่มผู้สอบที่นับถือศาสนาพุทธ มีค่าสูงกว่ากลุ่มผู้สอบที่นับถือศาสนาอิสลาม นั่นคือ แบบทดสอบฉบับนี้มีค่าอำนาจจำแนกสูงสามารถจำแนกผู้ที่มีความรู้ กับผู้ที่ไม่มีความรู้ของผู้สอบแต่ละกลุ่มได้ และ ค่าอำนาจจำแนกของข้อสอบในกลุ่มผู้สอบแต่ละกลุ่มมีค่าใกล้เคียงกัน ส่วนค่าความยากของข้อสอบนั้นมีค่าใกล้เคียงกันด้วยและเป็นแบบทดสอบที่ค่อนข้างยาก ซึ่งยากมากในกลุ่มผู้สอบที่นับถือศาสนาอิสลาม ส่วนค่าสัมประสิทธิ์การเดานั้นมีค่าใกล้เคียงกัน แสดงว่า โอกาสที่ผู้ไม่มีความรู้ของกลุ่มผู้สอบแต่ละกลุ่มจะทำข้อสอบถูกมีค่าใกล้เคียงกัน สำหรับการกระจายของค่าอำนาจจำแนกของข้อสอบ และ ค่าสัมประสิทธิ์การเดาในกลุ่มผู้สอบที่นับถือศาสนาพุทธมีการกระจายมากกว่าในกลุ่มผู้สอบที่นับถือศาสนาอิสลาม ส่วนการกระจายของค่าความยากของข้อสอบ ในกลุ่มผู้สอบที่นับถือศาสนาอิสลามมีการกระจายมากกว่ากลุ่มผู้สอบที่นับถือศาสนาพุทธ

ส่วนแบบทดสอบความเข้าใจทางภาษา ในกลุ่มผู้สอบที่นับถือศาสนาพุทธมีค่าอำนาจจำแนกของข้อสอบ(a) ค่าความยากของข้อสอบ(b) และค่าสัมประสิทธิ์การเดา(c) สูงกว่าในกลุ่มผู้สอบที่นับถือศาสนาอิสลาม ส่วนค่าความเบี่ยงเบนมาตรฐานของค่าอำนาจจำแนกของข้อสอบ ในกลุ่มผู้สอบที่นับถือศาสนาอิสลามมีค่าสูงกว่ากลุ่มผู้สอบที่นับถือศาสนาพุทธ และค่าความเบี่ยงเบนมาตรฐานของค่าความยากของข้อสอบ และค่าสัมประสิทธิ์การเดา ในกลุ่มผู้สอบที่นับถือศาสนาพุทธมีค่าสูงกว่ากลุ่มผู้สอบที่นับถือศาสนาอิสลาม แสดงว่า แบบทดสอบฉบับนี้สามารถจำแนกผู้ที่มีความรู้กับผู้ที่ไม่มีความรู้ของกลุ่มผู้สอบที่นับถือศาสนาพุทธได้ดีกว่ากลุ่มผู้สอบที่นับถือศาสนาอิสลาม ส่วนค่าความยากของข้อสอบมีค่าความยากพอเหมาะ และแต่ละกลุ่มมีค่าความยากของข้อสอบใกล้เคียงกัน สำหรับ

ค่าสัมประสิทธิ์การเดานั้นแต่ละกลุ่มมีค่าใกล้เคียงกัน ซึ่งแสดงว่า โอกาสที่ผู้ไม่มีความรู้ในกลุ่มผู้สอบที่นับถือศาสนาพุทธจะทำข้อสอบถูก มีมากกว่ากลุ่มผู้สอบที่นับถือศาสนาอิสลาม สำหรับการกระจายของค่าอำนาจจำแนกของข้อสอบของผู้สอบแต่ละกลุ่มมีค่าต่ำ และใกล้เคียงกัน ส่วนการกระจายของค่าความยากของข้อสอบนั้น มีการกระจายในช่วงกว้าง และการกระจายของค่าสัมประสิทธิ์การเดาของผู้สอบแต่ละกลุ่มมีค่าต่ำและใกล้เคียงกันด้วย

เมื่อพิจารณาเปรียบเทียบระหว่างแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ กับแบบทดสอบความเข้าใจทางภาษา พบว่า ในกลุ่มผู้สอบที่นับถือศาสนาพุทธมีค่าอำนาจจำแนกของข้อสอบสูงกว่าในกลุ่มผู้สอบที่นับถือศาสนาอิสลาม ทั้ง 2 ฉบับ โดยที่ค่าอำนาจจำแนกของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์จะมีค่าสูงกว่าค่าอำนาจจำแนกของแบบทดสอบความเข้าใจทางภาษา ส่วนค่าความยากของข้อสอบ ทั้ง 2 ฉบับ มีค่าใกล้เคียงกัน โดยที่แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ในกลุ่มผู้สอบที่นับถือศาสนาอิสลามจะมีค่าความยากของข้อสอบสูงกว่าในกลุ่มผู้สอบที่นับถือศาสนาพุทธ แต่แบบทดสอบความเข้าใจทางภาษานั้น ในกลุ่มผู้สอบที่นับถือศาสนาพุทธจะมีค่าความยากของข้อสอบสูงกว่ากลุ่มผู้สอบที่นับถือศาสนาอิสลาม แสดงว่า แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ยากสำหรับกลุ่มผู้สอบที่นับถือศาสนาอิสลาม ส่วนแบบทดสอบความเข้าใจทางภาษายากสำหรับกลุ่มผู้สอบที่นับถือศาสนาพุทธ สำหรับค่าสัมประสิทธิ์การเดานั้น ในกลุ่มผู้สอบที่นับถือศาสนาพุทธจะมีค่าสัมประสิทธิ์การเดาของข้อสอบสูงกว่ากลุ่มผู้สอบที่นับถือศาสนาอิสลาม ทั้ง 2 ฉบับ ซึ่งแสดงว่า โอกาสที่ผู้ไม่มีความรู้ของกลุ่มผู้สอบที่นับถือศาสนาพุทธ จะตอบข้อสอบได้ถูกต้อง มีค่ามากกว่าผู้ที่ไม่มีความรู้ของกลุ่มผู้สอบที่นับถือศาสนาอิสลาม ส่วนค่าความเบี่ยงเบนมาตรฐานของค่าอำนาจจำแนกของข้อสอบในแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ ในกลุ่มผู้สอบที่นับถือศาสนาพุทธจะสูงกว่ากลุ่มผู้สอบที่นับถือศาสนาอิสลาม แต่แบบทดสอบความเข้าใจทางภาษา ในกลุ่มผู้สอบที่นับถือศาสนาอิสลามจะสูงกว่ากลุ่มผู้สอบที่นับถือศาสนาพุทธ และ ค่าความเบี่ยงเบนมาตรฐานของค่าความยากของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ ในกลุ่มผู้สอบที่นับถือศาสนาอิสลามจะสูงกว่ากลุ่มผู้สอบที่นับถือศาสนาพุทธ ส่วนแบบทดสอบความเข้าใจทางภาษา ในกลุ่มผู้สอบที่นับถือศาสนาพุทธจะสูงกว่ากลุ่มผู้สอบที่นับถือศาสนาอิสลาม ค่าความเบี่ยงเบนมาตรฐานของค่าสัมประสิทธิ์การเดาในกลุ่มผู้สอบที่นับถือศาสนาพุทธมีค่าสูงกว่า กลุ่มผู้สอบที่นับถือศาสนาอิสลาม ทั้ง 2 ฉบับ

2. ผลการหาค่าดัชนีความลำเอียงของข้อสอบ จากแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา จากวิธีวิเคราะห์ความลำเอียงที่แตกต่างกัน 3 วิธี

ผู้วิจัยได้นำผลการตอบแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา มาวิเคราะห์ความลำเอียงด้วยวิธีวิเคราะห์ต่างกัน 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีโด่งลักษณะข้อสอบ และ วิธีของแมนเทิล-แฮนส์เชล ได้ผลการวิเคราะห์ดังแสดงในตาราง 5 ถึง ตาราง 10 ดังนี้

2.1 วิธีแปลงค่าความยากของข้อสอบ

วิธีการแปลงค่าความยากของข้อสอบนี้ หมายถึง การแปลงจากค่าความยาก (P) ที่อยู่ในรูปคะแนนมาตรฐาน Z จากคะแนนมาตรฐาน Z แปลงให้เป็นค่าความยากมาตรฐาน(Delta) ซึ่งมีค่าเฉลี่ยเท่ากับ 13 มีค่าความเบี่ยงเบนมาตรฐานเท่ากับ 4 ข้อสอบที่มีค่าเดลตาสูง หมายถึง ข้อสอบข้อนั้นยาก ส่วนข้อสอบที่มีค่าเดลตาต่ำ หมายถึง ข้อสอบข้อนั้นง่าย การตัดสินใจว่าข้อสอบนั้นลำเอียงหรือไม่ จะพิจารณาจากระยะตั้งฉากจากจุดศูนย์กลางค่าเดลตาถึงเส้นแกนหลัก เป็นค่าดัชนีความลำเอียง ถ้าระยะห่างมากกว่า $3sd$ หรือน้อยกว่า $-3sd$ จะสรุปว่าข้อสอบข้อนั้นมีความลำเอียง สำหรับการวิจัยครั้งนี้ ข้อสอบจะมีความลำเอียงก็ต่อเมื่อ มีค่าระยะตั้งฉากจากจุดศูนย์กลางค่าเดลตาถึงเส้นแกนหลักมากกว่า $+ .4665$ หรือน้อยกว่า $- .4665$ ในแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ มากกว่า $+ .5433$ หรือน้อยกว่า $- .5433$ ในแบบทดสอบความเข้าใจทางภาษา

การเสนอผลการวิเคราะห์หาดัชนีความลำเอียงของข้อสอบจากแบบทดสอบทั้ง 2 ฉบับ นั้น จะเสนอค่าดัชนีความลำเอียงที่คำนวณได้ ดังแสดงผลการวิเคราะห์ในตาราง 5

ตาราง 5 ค่าดัชนีความลำเอียงของข้อสอบด้วยวิธีการแปลงค่าความยากของข้อสอบ จากแบบทดสอบจิตวิทยาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา ระหว่างกลุ่มนักเรียนที่นับถือศาสนาพุทธ และ ศาสนาอิสลาม

ข้อ	แบบทดสอบจิตวิทยาทางคณิตศาสตร์	แบบทดสอบความเข้าใจทางภาษา
	ดัชนีความลำเอียงระหว่างศาสนา	ดัชนีความลำเอียงระหว่างศาสนา
1.	-.0493	-.4550
2.	-.1047	-.8058*
3.	-.3438	-.0271
4.	-.3810	-.2323
5.	.1606	-.2157
6.	-.2056	-.1957
7.	-.1169	.2403
8.	-.0617	.1474
9.	.2584	.7263*
10.	.1895	.3312
11.	-.0230	.4285
12.	.2166	.5173
13.	.0983	-.5625*
14.	.2094	.1947
15.	.2898	-.0115
16.	-.1510	.0044
17.	.1317	-.3110

ตาราง 5 (ต่อ)

ข้อ	แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์	แบบทดสอบความเข้าใจทางภาษา
	ดัชนีความลำเอียงระหว่างศาสนา	ดัชนีความลำเอียงระหว่างศาสนา
18.	.1909	-.3461
19.	-.1028	-.1360
20.	.0180	.0277
21.	-.1548	-.3669
22.	.1343	-.3505
23.	.2326	.3486
24.	.1131	.2939
25.	-.2217	.1560
26.	.2030	-.1114
27.	-.3695	-.0645
28.	.2512	-.3806
29.	.2422	.1697
30.	-.0783	-.3894
31.	-.0250	.1926
32.	-.1523	.1884
33.	-.0783	.0309
34.	.0385	.2180
35.	.1451	.0797
36.	.1221	-.1938
37.	-.1086	.3089

ตาราง 5 (ต่อ)

ข้อ	แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์	แบบทดสอบความเข้าใจทางภาษา
	ดัชนีความลำเอียงระหว่างศาสนา	ดัชนีความลำเอียงระหว่างศาสนา
38.	-.2468	.2981
39.	-.1754	-.2573
40.	-.0950	.4563

* หมายถึง ข้อสอบที่มีความลำเอียง คือ มีดัชนีความลำเอียงมากกว่า 3Sd หรือ น้อยกว่า -3Sd

ผลการวิเคราะห์ความลำเอียงด้วยวิธีที่แปลงค่าความชากของข้อสอบ จากแบบทดสอบแต่ละฉบับ จากตาราง 5 ปรากฏผล ดังนี้

ดัชนีความลำเอียงของข้อสอบ จากแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ ซึ่งหาได้จากระยะห่างตั้งฉากจากคู่อันดับค่าเฉลี่ยไปยังเส้นแกนหลัก (d) มีค่าอยู่ระหว่าง.0180 ถึง.3810 ซึ่งการวิเคราะห์ความลำเอียง ด้วยวิธีที่แปลงค่าความชากของข้อสอบ สำหรับแบบทดสอบฉบับนี้ ข้อสอบที่มีความลำเอียงจะต้องมีค่าดัชนีความลำเอียงมากกว่า +.4665 หรือ น้อยกว่า -.4665 และผลการวิเคราะห์ พบว่า ไม่มีข้อสอบที่มีความลำเอียงระหว่างนักเรียนที่นับถือศาสนาพุทธ กับนักเรียนที่นับถือศาสนาอิสลาม

สำหรับดัชนีความลำเอียงของข้อสอบจากแบบทดสอบความเข้าใจทางภาษา ซึ่งหาได้จากระยะห่างตั้งฉากจากคู่อันดับค่าเฉลี่ยไปยังเส้นแกนหลัก (d) มีค่าอยู่ระหว่าง.0044 ถึง.8058 ซึ่งการวิเคราะห์ความลำเอียงด้วยวิธีที่แปลงค่าความชากของข้อสอบ สำหรับแบบทดสอบฉบับนี้ ข้อสอบที่มีความลำเอียงจะต้องมีค่าดัชนีความลำเอียง มากกว่า +.5433 หรือ น้อยกว่า -.5433 ผลการวิเคราะห์ พบว่า มีข้อสอบที่มีความลำเอียง 3 ข้อ คือ ข้อ 2,9,และ 13 คิดเป็นร้อยละ 7.50 ของข้อสอบทั้งฉบับ ข้อสอบที่มีความลำเอียงมีค่าดัชนีความลำเอียง อยู่ระหว่าง -.5625 ถึง -.8058 เป็นข้อสอบที่ค่าเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาพุทธ 2 ข้อ คือ ข้อ 2 และ

13 และเป็นข้อสอบที่ล่าเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาอิสลาม 1 ข้อ คือ ข้อ 9 ส่วนข้อสอบที่ไม่มีความล่าเอียง มี 37 ข้อ มีค่าดัชนีความล่าเอียงอยู่ระหว่าง .0044 ถึง .5173

2.2 วิธีวัดลักษณะข้อสอบ

วิธีนี้จะวิเคราะห์หาดัชนีความล่าเอียงของข้อสอบ โดยพิจารณาจากค่าความแตกต่างของพื้นที่ใต้โค้งของข้อสอบแต่ละข้อระหว่างกลุ่มผู้สอบ 2 กลุ่ม โดยใช้โปรแกรมBILOG3 ประมาณค่าพารามิเตอร์ของข้อสอบเป็นรายข้อจากผู้สอบแต่ละกลุ่ม แล้วนำมาเข้าสู่หาคำนวณหาพื้นที่ใต้โค้ง ถ้าพื้นที่ใต้โค้งที่คำนวณได้มีค่ามากกว่า .40 จะสรุปว่าข้อสอบข้อนั้นมีความล่าเอียง

การเสนอผลการวิเคราะห์หาดัชนีความล่าเอียงของข้อสอบจากแบบทดสอบทั้ง 2 ฉบับนั้น จะเสนอค่าดัชนีความล่าเอียงที่คำนวณได้ ดังแสดงผลการวิเคราะห์ในตาราง 6

ตาราง 6 ค่าดัชนีความล่าเอียงของข้อสอบ ด้วยวิธีวัดลักษณะข้อสอบ จากแบบทดสอบโจทยปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา ระหว่างกลุ่มนักเรียนที่นับถือศาสนาพุทธ และ ศาสนาอิสลาม

ข้อ	แบบทดสอบโจทยปัญหาทางคณิตศาสตร์	แบบทดสอบความเข้าใจทางภาษา
	ดัชนีความล่าเอียงระหว่างศาสนา	ดัชนีความล่าเอียงระหว่างศาสนา
1.	.3537	.2292
2.	.3268	.9995*
3.	.0543	.2269
4.	.1635	.0564
5.	.2909	.6887*
6.	.3228	.1445
7.	.3738	.6008*

ตาราง 6 (ต่อ)

ข้อ	แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์	แบบทดสอบความเข้าใจทางภาษา
	ดัชนีความลำเอียงระหว่างศาสนา	ดัชนีความลำเอียงระหว่างศาสนา
8.	.3018	.9673*
9.	.1881	.6121*
10.	.3098	1.2717*
11.	.5481*	.5107*
12.	.2587	.7059*
13.	.1889	.6064*
14.	.2907	.6447*
15.	.4521*	.4955*
16.	.8951*	.3124
17.	.0720	.1360
18.	.0681	.8221*
19.	.3309	.1650
20.	.6734*	2.1535*
21.	.4755*	.2427
22.	.2572	.2349
23.	.6073*	.4457*
24.	1.0955*	.3921
25.	1.3305*	.3626
26.	.1663	.4717*
27.	.7744*	.3189

ตาราง 6 (ต่อ)

ข้อ	แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์	แบบทดสอบความเข้าใจทางภาษา
	ดัชนีความลำเอียงระหว่างศาสนา	ดัชนีความลำเอียงระหว่างศาสนา
28.	3.7451 [*]	.2995
29.	1.1920 [*]	.3260
30.	3.9147 [*]	.4325 [*]
31.	.4373 [*]	.2434
32.	.6517 [*]	.5287 [*]
33.	.5341 [*]	.2968
34.	1.0689 [*]	.4856 [*]
35.	.2639	.4910 [*]
36.	.2640	.4751 [*]
37.	.5195 [*]	.2303
38.	.2124	.3347
39.	.4278 [*]	.0992
40.	.8770 [*]	1.0547 [*]

* หมายถึง ข้อที่มีความลำเอียง คือ มีดัชนีความลำเอียงมากกว่า .40

ผลการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีที่ได้อธิบายไว้ข้างต้นของข้อสอบ 3 พารามิเตอร์ของแบบทดสอบแต่ละฉบับ จากตาราง 6 ปรากฏผล ดังนี้

ดัชนีความลำเอียงของข้อสอบจากแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ ซึ่งคำนวณจากความแตกต่างของพื้นที่ใต้โค้งของข้อสอบแต่ละข้อระหว่างกลุ่มผู้สอบ 2 กลุ่ม มีค่าอยู่ระหว่าง .0543 ถึง 3.9147 และ มีข้อสอบที่มีความลำเอียง 19 ข้อ คิดเป็นร้อยละ 47.50 ของข้อสอบทั้งฉบับ

ข้อสอบที่มีความลำเอียง มีค่าดัชนีความลำเอียงอยู่ระหว่าง .4278 ถึง 3.9147 โดยแยกเป็นข้อสอบที่ลำเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาพุทธ 11 ข้อ เป็นข้อสอบที่ลำเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาอิสลาม 8 ข้อ ข้อสอบที่มีดัชนีความลำเอียงสูงสุด คือ ข้อ 30 ซึ่งมีค่าดัชนีความลำเอียง เท่ากับ 3.9147 เป็นข้อสอบที่มีความลำเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาอิสลาม ส่วนข้อสอบที่มีดัชนีความลำเอียงต่ำสุด คือ ข้อ 39 ซึ่งมีค่าดัชนีความลำเอียงเท่ากับค่าเอียงมี 21 ข้อ มีค่าดัชนีความลำเอียง อยู่ระหว่าง .0543 ถึง .3738

ดัชนีความลำเอียงของข้อสอบจากแบบทดสอบความเข้าใจทางภาษา ซึ่งคำนวณจากความแตกต่างของพื้นที่ใต้โค้งของข้อสอบแต่ละข้อระหว่างกลุ่มผู้สอบ 2 กลุ่ม มีค่าอยู่ระหว่าง .0992 ถึง 2.1535 และมีข้อสอบที่มีความลำเอียง 21 ข้อ คิดเป็นร้อยละ 52.50 ของข้อสอบทั้งฉบับ ข้อสอบที่มีความลำเอียงนั้นมีดัชนีความลำเอียงอยู่ระหว่าง .4457 ถึง 2.135 โดยแยกเป็นข้อสอบที่ลำเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาพุทธ 10 ข้อ เป็นข้อสอบที่ลำเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาอิสลาม 11 ข้อ ข้อสอบที่มีดัชนีความลำเอียงสูงสุด คือ ข้อ 20 ซึ่งมีค่าดัชนีความลำเอียงเท่ากับ 2.1535 เป็นข้อสอบที่ลำเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาพุทธ ส่วนข้อสอบที่มีดัชนีความลำเอียงต่ำสุด คือ ข้อ 23 ซึ่งมีค่าดัชนีความลำเอียงเท่ากับ .4457 เป็นข้อสอบที่ลำเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาอิสลาม สำหรับข้อสอบ ที่ไม่มีความลำเอียง มี 19 ข้อ มีค่าดัชนีความลำเอียง อยู่ระหว่าง .0564 ถึง .3921

2.3 วิธีของแมนเทิล-แฮนส์เฟล

วิธีนี้จะวิเคราะห์หาดัชนีความลำเอียงของข้อสอบ โดยพิจารณาเปรียบเทียบกับผลการตอบข้อสอบแต่ละข้อระหว่างกลุ่มผู้สอบ 2 กลุ่ม คือ กลุ่มอ้างอิง และ กลุ่มเปรียบเทียบกับ การตัดสินใจข้อสอบข้อนี้ว่าลำเอียงหรือไม่จะพิจารณาจากค่า α_{MH} หรือ χ^2_{MH} ถ้า α_{MH} มีค่ามากกว่า 1.00 หรือ χ^2_{MH} มีนัยสำคัญทางสถิติที่ระดับ .05 จะสรุปว่า ข้อสอบข้อนี้มีความลำเอียง

การเสนอผลการวิเคราะห์หาดัชนีความลำเอียงของข้อสอบ จากแบบทดสอบทั้ง 2 ฉบับ จะเสนอค่าดัชนีความลำเอียง (α_{MH} ; χ^2_{MH}) ที่คำนวณได้ ดังแสดงผลการวิเคราะห์ในตาราง 7

ตาราง 7 ค่าดัชนีความจำเริญของข้อสอบ ด้วยวิธีของแมนเทล-แฮนส์เซด (α_{MH} ; χ^2_{MH})

จากแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา ระหว่างกลุ่ม
นักเรียนที่นับถือศาสนาพุทธ และ ศาสนาอิสลาม

ข้อ	แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์		แบบทดสอบความเข้าใจทางภาษา	
	α_{MH}	χ^2_{MH}	α_{MH}	χ^2_{MH}
1.	1.13 [*]	1.22	1.38 [*]	8.09 [*]
2.	1.13 [*]	1.29	1.71 [*]	28.11 [*]
3.	1.19 [*]	1.84	1.00	.00
4.	1.31 [*]	6.07 [*]	1.18 [*]	2.40
5.	.84	2.20	1.14 [*]	1.67
6.	1.18 [*]	2.28	1.14 [*]	1.36
7.	1.08 [*]	.43	.85	2.69
8.	.91	.33	.93	.54
9.	.99	.00	.64	18.72 [*]
10.	.86	1.70	.81	2.74
11.	1.03 [*]	.04	.76	7.51 [*]
12.	.89	1.20	.74	9.88 [*]
13.	1.03 [*]	.05	1.42 [*]	11.73 [*]
14.	.83	2.65	.89	.87
15.	.78	4.72 [*]	.99	.00
16.	1.04 [*]	.08	.99	.00
17.	.88	1.16	1.20 [*]	2.94
18.	.98	.01	1.21 [*]	2.19

ตาราง 7 (ต่อ)

ข้อ	แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์		แบบทดสอบความเข้าใจทางภาษา	
	α_{MH}	χ^2_{MH}	α_{MH}	χ^2_{MH}
19.	1.05*	.15	1.09*	.64
20.	1.02*	.01	.97	.06
21.	1.12*	1.06	1.24*	3.95*
22.	.93	.33	1.27*	4.98*
23.	.84	2.35	.78	6.15*
24.	.92	.47	.83	3.15
25.	1.14*	1.08	.88	1.46
26.	.81	3.06	1.07*	.39
27.	1.29*	5.65*	1.09*	.51
28.	.85	2.34	1.31*	6.85*
29.	.86	1.85	.90	.88
30.	1.08*	.48	1.26*	5.32*
31.	1.03	.07	.88	1.06
32.	.94	.12	.88	1.50
33.	1.09*	.58	.99	.00
34.	.95	.17	.86	1.81
35.	.83	1.61	.94	.26
36.	.92	.63	1.14*	1.63
37.	.98	.02	.84	2.32
38.	1.16*	1.59	.80	4.24*

ตาราง 7 (ต่อ)

ข้อ	แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์		แบบทดสอบความเข้าใจทางภาษา	
	α_{MH}	χ^2_{MH}	α_{MH}	χ^2_{MH}
39.	1.14*	1.48	1.20*	2.45
40.	1.09*	.52	.76	7.80*

*หมายถึง ข้อที่มีความลำเอียง คือ มีดัชนีความลำเอียง $\alpha_{MH} > 1.00$

และ χ^2_{MH} มีนัยสำคัญทางสถิติที่ระดับ.05

ผลการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีของแมนเทิล-ฮันส์เชลของแบบทดสอบแต่ละฉบับ จากตาราง 7 ปรากฏผล ดังนี้

ดัชนีความลำเอียงของข้อสอบจากแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ ซึ่งพิจารณาจากค่า α_{MH} และค่า χ^2_{MH} ซึ่งมีค่า α_{MH} อยู่ระหว่าง .81 ถึง 1.31 และมีค่า χ^2_{MH} อยู่ระหว่าง .00 ถึง 6.07 จากการพิจารณาค่า $\alpha_{MH} > 1.00$ หรือค่า χ^2_{MH} มีนัยสำคัญทางสถิติที่ระดับ .05 พบว่า มีข้อสอบที่มีความลำเอียง 21 ข้อ คิดเป็นร้อยละ 52.50 ของข้อสอบทั้งฉบับ ข้อสอบที่มีความลำเอียงนั้นมีค่าดัชนีความลำเอียง (α_{MH}) อยู่ระหว่าง 1.02 ถึง 1.31 และมีค่า χ^2_{MH} อยู่ระหว่าง 4.72 ถึง 6.07 โดยแยกเป็นข้อสอบที่ลำเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาพุทธ 1 ข้อ เป็นข้อสอบที่ลำเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาอิสลาม 20 ข้อ ข้อสอบที่มีค่า α_{MH} สูงสุด คือข้อ 4 ซึ่งมีค่า α_{MH} เท่ากับ 1.31 และข้อสอบที่มี χ^2_{MH} สูงสุด คือข้อ 4 เช่นเดียวกัน ซึ่งมีค่า χ^2_{MH} เท่ากับ 6.07 เป็นข้อสอบที่ลำเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาอิสลาม ส่วนข้อสอบที่มี α_{MH} ต่ำสุด คือ ข้อ 20 ซึ่งมีค่า α_{MH} เท่ากับ 1.02 เป็นข้อสอบที่ลำเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาอิสลาม และข้อที่มี χ^2_{MH} ต่ำสุด คือข้อ 15 ซึ่งมีค่า χ^2_{MH} เท่ากับ 4.72 เป็นข้อสอบที่ลำเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาพุทธ สำหรับข้อสอบที่ไม่มีความลำเอียง มีจำนวน 19 ข้อ มีค่า α_{MH} อยู่ระหว่าง .78 ถึง .99 และมีค่า χ^2_{MH} อยู่ระหว่าง .00 ถึง 3.06

ดัชนีความล่าเอียงของข้อสอบจากแบบทดสอบความเข้าใจทางภาษา ซึ่งพิจารณาจากค่า α_{MH} และค่า χ^2_{MH} โดยมีค่า α_{MH} อยู่ระหว่าง .64 ถึง 1.71 และมีค่า χ^2_{MH} อยู่ระหว่าง .00 ถึง 28.11 จากการพิจารณาค่า $\alpha_{MH} > 1.00$ หรือค่า χ^2_{MH} มีนัยสำคัญทางสถิติที่ระดับ .05 พบว่า มีข้อสอบที่มีความล่าเอียง 23 ข้อ คิดเป็นร้อยละ 57.50 ของข้อสอบทั้งฉบับ ข้อสอบที่มีความล่าเอียงนั้นมีค่าดัชนีความล่าเอียง (α_{MH}) อยู่ระหว่าง 1.07 ถึง 1.71 และมีค่า χ^2_{MH} อยู่ระหว่าง 3.95 ถึง 28.11 โดยแยกเป็นข้อสอบที่ล่าเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาพุทธ 6 ข้อ เป็นข้อสอบที่ล่าเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาอิสลาม 17 ข้อ ข้อสอบที่มีค่า α_{MH} สูงสุด คือ ข้อ 2 ซึ่งมีค่า α_{MH} เท่ากับ 1.71 และข้อสอบที่มีค่า χ^2_{MH} สูงสุด คือ ข้อ 2 เช่นเดียวกัน ซึ่งมีค่า χ^2_{MH} เท่ากับ 28.11 เป็นข้อสอบที่ล่าเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาอิสลาม ส่วนข้อสอบที่มีค่า α_{MH} ต่ำสุด คือ ข้อ 26 ซึ่งมีค่า α_{MH} เท่ากับ 1.07 และข้อสอบที่มีค่า χ^2_{MH} ต่ำสุด คือ ข้อ 21 ซึ่งมีค่า χ^2_{MH} เท่ากับ 3.95 เป็นข้อสอบที่ล่าเอียงเข้าสู่กลุ่มนักเรียนที่นับถือศาสนาอิสลาม ทั้ง 2 ข้อ สำหรับข้อสอบที่ไม่มีความล่าเอียง มีจำนวน 17 ข้อ ซึ่งมีค่า α_{MH} อยู่ระหว่าง .64 ถึง 1.00 และ χ^2_{MH} อยู่ระหว่าง .00 ถึง 3.15

จากผลการวิเคราะห์ความล่าเอียงของข้อสอบ ระหว่างกลุ่มนักเรียนที่นับถือศาสนาพุทธ กับกลุ่มนักเรียนที่นับถือศาสนาอิสลาม ทั้ง 3 วิถี พบว่า วิถีวิเคราะห์ความล่าเอียงของข้อสอบบางวิถีค้นพบข้อสอบที่มีความล่าเอียงได้ซ้ำกัน และ บางวิถีค้นพบข้อสอบที่มีความล่าเอียงได้ต่างกันเพื่อให้เห็นจำนวนข้อสอบที่มีความล่าเอียงชัดเจนขึ้น ผู้วิจัยจึงได้สรุปและเปรียบเทียบจำนวนข้อสอบที่มีความล่าเอียงในวิถีวิเคราะห์ความล่าเอียงทุกวิถี ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา ดังแสดงผลในตาราง 8

ตาราง 8 สรุปจำนวนข้อสอบที่มีความลำเอียงในแต่ละวิธีการวิเคราะห์ความลำเอียง ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา

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

ผลการวิเคราะห์จากตาราง 8 แสดงให้เห็นว่า การวิเคราะห์ความลำเอียงของข้อสอบทั้ง 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ และ วิธีของแมนเทิล-แฮนส์เชลของแบบทดสอบทั้ง 2 ฉบับ พบจำนวนข้อสอบที่มีความลำเอียงสอดคล้องกัน โดยที่การวิเคราะห์ความลำเอียงด้วยวิธีของแมนเทิล-แฮนส์เชล มีข้อสอบที่มีความลำเอียงมากที่สุด รองวิธีลงมาเป็นวิธีโค้งลักษณะข้อสอบ และวิธีแปลงค่าความยากของข้อสอบ พบจำนวนข้อสอบที่มีความ

ค่าเฉลี่ยน้อยที่สุด ทั้งในแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา

จากการวิเคราะห์ความล่าช้าของข้อสอบ ระหว่างกลุ่มนักเรียนที่นับถือศาสนาพุทธและกลุ่มนักเรียนที่นับถือศาสนาอิสลามของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ พบว่า วิธียแปลงค่าความยากของข้อสอบ (TID) ไม่มีข้อสอบที่มีความล่าช้าเอียง จากเกณฑ์ที่กำหนดในการวิจัยครั้งนี้ ส่วนวิธีโด่งลักษณะข้อสอบ (ICC) มีข้อสอบที่มีความล่าช้าเอียง จำนวน 19 ข้อ และวิธีของแมนเทิล-แฮนส์เชล (MH) มีข้อสอบที่มีความล่าช้าเอียง จำนวน 21 ข้อ ซึ่งจากการวิเคราะห์ความล่าช้าเอียงของข้อสอบ ทั้ง 3 วิธี ไม่พบข้อสอบที่มีความล่าช้าเอียงซ้ำกันในทุกวิธีวิเคราะห์ แต่พบข้อสอบที่มีความล่าช้าเอียงซ้ำกันระหว่างวิธีโด่งลักษณะข้อสอบกับวิธีของแมนเทิล-แฮนส์เชล จำนวน 12 ข้อ ได้แก่ ข้อ 11, 15, 16, 20, 21, 25, 27, 30, 31, 33, 39 และ ข้อ 40 และมีข้อสอบที่มีความล่าช้าเอียง 1 วิธี จำนวน 16 ข้อ นอกนั้นเป็นข้อสอบที่ไม่มีความล่าช้าเอียงในทุกวิธีวิเคราะห์จากเกณฑ์ที่กำหนดในการวิจัยครั้งนี้ จำนวน 12 ข้อ

ส่วนการวิเคราะห์ความล่าช้าเอียงของข้อสอบ ระหว่างกลุ่มนักเรียนที่นับถือศาสนาพุทธและกลุ่มนักเรียนที่นับถือศาสนาอิสลาม ของแบบทดสอบความเข้าใจทางภาษา พบว่า วิธียแปลงค่าความยากของข้อสอบ (TID) มีข้อสอบที่มีความล่าช้าเอียง จำนวน 3 ข้อ วิธียโด่งลักษณะข้อสอบ (ICC) มีข้อสอบที่มีความล่าช้าเอียงจำนวน 21 ข้อ และ วิธีของแมนเทิล-แฮนส์เชล มีข้อสอบที่มีความล่าช้าเอียงจำนวน 23 ข้อ ซึ่งจากการวิเคราะห์ความล่าช้าเอียงของข้อสอบทั้ง 3 วิธี พบว่า มีข้อสอบที่มีความล่าช้าเอียงซ้ำกัน ทั้ง 3 วิธีวิเคราะห์ จำนวน 3 ข้อ ได้แก่ ข้อ 2, 9, 13 ข้อสอบที่มีความล่าช้าเอียงซ้ำกันระหว่างวิธีโด่งลักษณะข้อสอบกับวิธีของแมนเทิล-แฮนส์เชล จำนวน 12 ข้อ ได้แก่ ข้อ 2, 5, 9, 11, 12, 13, 18, 23, 26, 30, 36 และข้อ 40 และข้อสอบที่มีความล่าช้าเอียง 1 วิธีวิเคราะห์ จำนวน 20 ข้อ นอกนั้นเป็นข้อสอบที่ไม่มีความล่าช้าเอียง ในทุกวิธีวิเคราะห์จากเกณฑ์ที่กำหนดในการวิจัยครั้งนี้ จำนวน 8 ข้อ

3. ผลการเปรียบเทียบจำนวนข้อสอบที่มีความล่าช้าเอียง จากวิธีวิเคราะห์ความล่าช้าเอียง 3 วิธี ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา

จากผลการวิเคราะห์ความล่าช้าเอียงของข้อสอบ ทั้ง 3 วิธี คือ วิธียแปลงค่าความยากของข้อสอบ (TID) วิธียโด่งลักษณะข้อสอบ (ICC) และ วิธีของแมนเทิล-แฮนส์เชล (MH) พบว่า วิธีวิเคราะห์บางวิธีพบข้อสอบที่มีความล่าช้าเอียงซ้ำกัน บางวิธีพบข้อสอบที่มีความล่าช้าเอียงแตกต่างกัน ดังนั้น เพื่อให้ทราบว่าวิธีวิเคราะห์ความล่าช้าเอียง ทั้ง 3 วิธี มีจำนวนข้อสอบที่มีความล่าช้าเอียง

แตกต่างกับหรือไม่ ผู้วิจัยจึงนำผลการวิเคราะห์ความลำเอียงของ ทั้ง 3 วิธี ของแบบทดสอบ แต่ละฉบับมาเปรียบเทียบความแตกต่างของจำนวนข้อที่มีความลำเอียง จากการวิเคราะห์ความลำเอียงด้วยวิธีที่ต่างกัน โดยใช้สถิติการทดสอบความแตกต่างด้วย Z-test ดังแสดงผลการทดสอบนัยสำคัญทางสถิติ ในตาราง 9 และ ตาราง 10

ตาราง 9 ผลการทดสอบความแตกต่างของจำนวนข้อสอบที่มีความลำเอียง ระหว่างวิธีวิเคราะห์ความลำเอียงต่างกันของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์

วิธีวิเคราะห์ความลำเอียง	TID	ICC	MH
TID	-	4.9895***	5.3354***
ICC		-	.4472
MH			-

*** $P < .001$

$Z_{.001} = 3.291$

ผลการวิเคราะห์จากตาราง 9 แสดงให้เห็นว่า จำนวนข้อสอบที่มีความลำเอียงจากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากของข้อสอบกับวิธีโด่งลักษณะข้อสอบ และวิธีแปลงค่าความยากของข้อสอบกับวิธีของแมนเทิล-แฮนส์เชล แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับสถิติ นั่นคือ วิธีแปลงค่าความยากของข้อสอบกับวิธีโด่งลักษณะข้อสอบ และ วิธีแปลงค่าความยากของข้อสอบกับวิธีของแมนเทิล-แฮนส์เชล มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกัน ส่วนวิธีโด่งลักษณะข้อสอบกับวิธีของแมนเทิล-แฮนส์เชล มีจำนวนข้อสอบที่มีความลำเอียงไม่แตกต่างกัน

ตาราง 10 ผลการทดสอบความแตกต่างของข้อสอบที่มีความลำเอียง ระหว่างวิธีวิเคราะห์ความลำเอียงต่างกันของแบบทดสอบความเข้าใจทางภาษา

วิธีวิเคราะห์ความลำเอียง	TID	ICC	MH
TID	-	4.3902***	4.7755***
ICC		-	.4496
MH			-

*** $P < .001$

$Z_{.001} = 3.291$

ผลการวิเคราะห์ จากตาราง 10 แสดงให้เห็นว่า จำนวนข้อสอบที่มีความลำเอียงจากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากของข้อสอบกับวิธีโด่งลักษณะข้อสอบ และวิธีแปลงค่าความยากของข้อสอบกับวิธีของแมนเทิล-แฮนส์เชล แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.001 ส่วนวิธีโด่งลักษณะข้อสอบกับวิธีของแมนเทิล-แฮนส์เชล แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ นั่นคือ วิธีแปลงค่าความยากของข้อสอบกับวิธีโด่งลักษณะข้อสอบ และ วิธีแปลงค่าความยากของข้อสอบกับวิธีของแมนเทิล-แฮนส์เชล มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกัน ส่วนวิธีโด่งลักษณะข้อสอบกับวิธีของแมนเทิล-แฮนส์เชล มีจำนวนข้อสอบที่มีความลำเอียงไม่แตกต่างกัน

4. ผลการหาค่าความเชื่อมั่นของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา หลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วจากการวิเคราะห์ความลำเอียง ด้วยวิธีต่างกัน 3 วิธี โดยใช้สูตรการคำนวณแบบ KR-20 และ แบบแบ่งครึ่งฉบับ

ผู้วิจัยได้นำ แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา หลังจากคัดเลือกข้อสอบที่มีความลำเอียงออก จากการวิเคราะห์ความลำเอียงด้วยวิธีต่างกัน คือ วิธีแปลงค่าความยากของข้อสอบ (TID) วิธีโด่งลักษณะข้อสอบ (ICC) และวิธีของแมนเทิล-แฮนส์เชล (MH) ไปทดสอบกับนักเรียนกลุ่มตัวอย่าง จำนวน 753 คน แล้วนำคะแนนผลการสอบ

ของแบบทดสอบแต่ละฉบับ มาคำนวณหาค่าความเชื่อมั่น โดยใช้สูตรการคำนวณแบบ KR-20 และ ใช้สูตรการคำนวณแบบแบ่งครึ่งฉบับ ดังแสดงผลในตาราง 11 และ ตาราง 12

ตาราง 11 ค่าความเชื่อมั่นของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา หลังจากคัดข้อสอบที่มีความลำเอียงออก ด้วยวิธีแปลงค่าความยากของข้อสอบ (TID) วิธีโด่งลักษณะข้อสอบ (ICC) และวิธีของแมนเทิล-แฮนส์เชล (MH) ซึ่งคำนวณโดยใช้สูตรแบบ KR-20

วิธีวิเคราะห์ ความลำเอียง	แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์			แบบทดสอบความเข้าใจทางภาษา		
	K	KR-20	SE _{mean}	K	KR-20	SE _{mean}
TID	40	0.761	<u>+2.770</u>	37	0.719	<u>+2.737</u>
ICC	21	0.841	<u>+1.959</u>	19	0.671	<u>+1.952</u>
MH	19	0.843	<u>+1.801</u>	17	0.627	<u>+1.808</u>

ผลการวิเคราะห์จากตาราง 11 แสดงให้เห็นว่า ค่าความเชื่อมั่นของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์หลังจากคัดข้อสอบที่มีความลำเอียงออก ด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโด่งลักษณะข้อสอบ และ วิธีของแมนเทิล-แฮนส์เชล มีค่าเท่ากับ 0.761, 0.841 และ 0.843 ตามลำดับ ซึ่งแบบทดสอบฉบับที่คัดข้อสอบที่มีความลำเอียงออกด้วยวิธีของแมนเทิล-แฮนส์เชล มีค่าความเชื่อมั่นสูงสุด คือ มีค่าเท่ากับ 0.843 รองลงมา เป็นแบบทดสอบฉบับที่คัดข้อสอบที่มีความลำเอียงออกด้วยวิธีโด่งลักษณะข้อสอบ คือ มีค่าเท่ากับ 0.841 และ แบบทดสอบฉบับที่คัดข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบมีค่าความเชื่อมั่นต่ำสุด คือ มีค่าเท่ากับ 0.761 ส่วนค่าความคลาดเคลื่อนมาตรฐานในการวัดของแบบทดสอบหลังจากคัดข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบวิธีโด่งลักษณะข้อสอบและวิธีของแมนเทิล-แฮนส์เชล

มีค่าเท่ากับ ± 2.770 , ± 1.959 และ ± 1.801 ตามลำดับ โดยที่แบบทดสอบฉบับที่คัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบ มีค่าความคลาดเคลื่อนมาตรฐานในการวัดสูงสุด คือ มีค่าเท่ากับ ± 2.770 รองลงมา เป็นแบบทดสอบฉบับที่คัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีโค้งลักษณะข้อสอบ คือ มีค่าเท่ากับ ± 1.959 และ แบบทดสอบฉบับที่คัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีของแมนเทิล-แฮนส์เชล มีค่าความคลาดเคลื่อนมาตรฐานในการวัดต่ำสุด คือ มีค่าเท่ากับ ± 1.801

ส่วนค่าความเชื่อมั่นของแบบทดสอบความเข้าใจทางภาษา หลังจากคัดข้อสอบที่มีความล่าเอียงออก ด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ และ วิธีของแมนเทิล-แฮนส์เชลมีค่าเท่ากับ 0.719, 0.671 และ 0.627 ตามลำดับ โดยที่แบบทดสอบฉบับที่คัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบมีความเชื่อมั่นสูงสุด คือ มีค่าเท่ากับ 0.719 รองลงมา เป็นแบบทดสอบฉบับที่คัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีโค้งลักษณะข้อสอบ คือ มีค่าเท่ากับ 0.671 และแบบทดสอบฉบับที่คัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีของแมนเทิล-แฮนส์เชล มีค่าความเชื่อมั่นต่ำสุด คือ มีค่าเท่ากับ 0.627 ส่วนค่าความคลาดเคลื่อนมาตรฐานในการวัดของแบบทดสอบหลังจากคัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ และ วิธีของแมนเทิล-แฮนส์เชล มีค่าเท่ากับ ± 2.737 , ± 1.952 และ ± 1.808 ตามลำดับ ซึ่งแบบทดสอบฉบับที่คัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบมีความคลาดเคลื่อนมาตรฐานในการวัดสูงสุด คือ มีค่าเท่ากับ ± 2.737 รองลงมา เป็นแบบทดสอบฉบับที่คัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีโค้งลักษณะข้อสอบ คือ มีค่าเท่ากับ ± 1.952 และ แบบทดสอบฉบับที่คัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีของแมนเทิล - แฮนส์เชล มีค่าความคลาดเคลื่อนมาตรฐานในการวัดต่ำสุด คือ มีค่าเท่ากับ ± 1.808

ตาราง 12 ค่าความเชื่อมั่นของแบบทดสอบโจทซ์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษาหลังจากคัดข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบ (TID) วิธีโด่งลักษณะข้อสอบ (ICC) และวิธีของแมนเทิล-แฮนส์เชล (MH) โดยใช้สูตรการคำนวณแบบแบ่งครึ่งฉบับ

วิธีวิเคราะห์ ความลำเอียง	แบบทดสอบโจทซ์ปัญหาทางคณิตศาสตร์			แบบทดสอบความเข้าใจทางภาษา		
	K	r_{ICC}	SE_{mean}	K	r_{ICC}	SE_{mean}
TID	40	0.781	± 2.653	37	0.730	± 2.681
ICC	21	0.829	± 2.029	19	0.727	± 1.781
MH	19	0.855	± 1.730	17	0.672	± 1.694

ผลการวิเคราะห์จากตาราง 12 แสดงให้เห็นว่า ค่าความเชื่อมั่นของแบบทดสอบโจทซ์ปัญหาทางคณิตศาสตร์หลังจากคัดข้อสอบที่มีความลำเอียงออก ด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโด่งลักษณะข้อสอบ และวิธีของแมนเทิล-แฮนส์เชล มีค่าเท่ากับ 0.781, 0.829 และ 0.855 ตามลำดับ โดยที่แบบทดสอบฉบับที่คัดข้อสอบที่มีความลำเอียงออกด้วยวิธีของแมนเทิล-แฮนส์เชล มีค่าความเชื่อมั่นสูงสุด คือ มีค่าเท่ากับ 0.855 รองลงมาเป็นแบบทดสอบฉบับที่คัดข้อสอบที่มีความลำเอียงออกด้วยวิธีโด่งลักษณะข้อสอบ คือ มีค่าเท่ากับ 0.829 และ แบบทดสอบฉบับที่คัดข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบมีค่าความเชื่อมั่นต่ำสุด คือ มีค่าเท่ากับ 0.781 ส่วนค่าความคลาดเคลื่อนมาตรฐานในการวัดของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโด่งลักษณะข้อสอบ และวิธีของแมนเทิล-แฮนส์เชล มีค่าเท่ากับ ± 2.653 , ± 2.029 และ ± 1.730 ตามลำดับ ซึ่งแบบทดสอบฉบับที่คัดข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบ มีค่าความคลาดเคลื่อนมาตรฐานในการวัดสูงสุด คือ มีค่าเท่ากับ ± 2.653 รองลงมา เป็นแบบทดสอบฉบับที่คัดข้อสอบที่มีความลำเอียงออกด้วยวิธีโด่งลักษณะข้อสอบ คือ มีค่าเท่ากับ ± 2.029 และแบบทดสอบฉบับที่คัดข้อสอบ

ที่มีความล่าเอียงออกด้วยวิธีของแมนเทล-แฮนส์เชล มีค่าความคลาดเคลื่อนมาตรฐานในการวัดต่ำสุด คือ มีค่าเท่ากับ $+1.730$

ส่วนค่าความเชื่อมั่นของแบบทดสอบความเข้าใจทางภาษา หลังจากคัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีแปลงค่าความชากของข้อสอบ วิธีโค้งลักษณะข้อสอบ และ วิธีของแมนเทล-แฮนส์เชล มีค่าเท่ากับ 0.730 , 0.727 และ 0.672 ตามลำดับ ซึ่งแบบทดสอบฉบับที่คัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีแปลงค่าความชากของข้อสอบ มีค่าความเชื่อมั่นสูงสุด คือ มีค่าเท่ากับ 0.730 รองลงมา เป็นแบบทดสอบฉบับที่คัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีโค้งลักษณะข้อสอบ คือ มีค่าเท่ากับ 0.727 และแบบทดสอบฉบับที่คัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีของแมนเทล-แฮนส์เชล มีค่าความเชื่อมั่นต่ำสุด คือ มีค่าเท่ากับ 0.672 ส่วนค่าความคลาดเคลื่อนมาตรฐานในการวัดของแบบทดสอบหลังจากคัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีแปลงค่าความชากของข้อสอบ วิธีโค้งลักษณะข้อสอบ และ วิธีของแมนเทล-แฮนส์เชล มีค่าเท่ากับ $+2.681$, $+1.781$ และ $+1.694$ ตามลำดับ ซึ่งแบบทดสอบฉบับที่คัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีแปลงค่าความชากของข้อสอบ มีค่าความคลาดเคลื่อนมาตรฐานในการวัดสูงสุด คือ มีค่าเท่ากับ $+2.681$ รองลงมา เป็นแบบทดสอบฉบับที่คัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีโค้งลักษณะข้อสอบ คือ มีค่าเท่ากับ $+1.781$ และ แบบทดสอบฉบับที่คัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีของแมนเทล-แฮนส์เชล มีค่าความคลาดเคลื่อนมาตรฐานในการวัดต่ำสุด คือ มีค่าเท่ากับ $+1.694$

5. ผลการเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา หลังจากคัดข้อสอบที่มีความล่าเอียงออก จากการวิเคราะห์ความล่าเอียงด้วยวิธีที่แตกต่างกัน 3 วิธี

เพื่อให้ทราบว่า ค่าความเชื่อมั่นของ แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา เมื่อคัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีวิเคราะห์ความล่าเอียงต่างกัน จะมีค่าแตกต่างกันหรือไม่ ผู้วิจัยจึงนำค่าความเชื่อมั่นของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา หลังจากคัดข้อสอบที่มีความล่าเอียงออกด้วยวิธีแปลงค่าความชากของข้อสอบ วิธีโค้งลักษณะข้อสอบ และวิธีของแมนเทล-แฮนส์เชล ที่คำนวณได้มาแปลงเป็นคะแนนมาตรฐานตามวิธีของฟิชเชอร์ (Fisher's Z Transformation) แล้วนำไปทดสอบความแตกต่าง โดยใช้ ไค-สแควร์ (X^2) ดังแสดงผลในตาราง 13 ถึง ตาราง 15

ตาราง 13 ผลการทดสอบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบโจทก์ปัญหาทาง
คณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา หลังจากคัดข้อสอบที่มีความลำเอียงออก
ด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ และวิธีของแมนเทล-แฮนส์เชล
ที่คำนวณโดยใช้สูตรแบบ KR-20

แบบทดสอบ	วิธีวิเคราะห์ความลำเอียง	K	KR-20	Z_r	χ^2
โจทก์ปัญหาทาง คณิตศาสตร์	TID	40	0.761	.9973	8.4943*
	ICC	21	0.841	1.2225	
	MH	19	0.843	1.2254	
ความเข้าใจ ทางภาษา	TID	37	0.719	.8982	3.3780
	ICC	19	0.671	.8122	
	MH	17	0.627	.7332	

* มีนัยสำคัญทางสถิติที่ระดับ .05

$$\chi^2_{.05,2} = 5.991$$

ผลการวิเคราะห์จากตาราง 13 แสดงให้เห็นว่า ค่าความเชื่อมั่นของแบบทดสอบโจทก์
ปัญหาทางคณิตศาสตร์ หลังจากคัดข้อสอบที่มีความลำเอียงออกด้วยวิธีต่างกัน คือ วิธีแปลงค่าความ
ยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ และ วิธีของแมนเทล-แฮนส์เชล มีค่าแตกต่างกันอย่างมี
นัยสำคัญทางสถิติที่ระดับ .05 แสดงว่า มีค่าความเชื่อมั่นอย่างน้อย 1 คู่ที่มีความแตกต่างกัน ส่วน
ค่าความเชื่อมั่นของแบบทดสอบความเข้าใจทางภาษา หลังจากคัดข้อสอบที่มีความลำเอียงออกด้วย
วิธีต่างกัน คือ วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ และ วิธีของแมนเทล-
แฮนส์เชลมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ นั่นคือ แบบทดสอบความเข้าใจทางภาษา
หลังจากคัดเลือกที่มีความลำเอียงออก ด้วยวิธีต่างกัน มีค่าความเชื่อมั่นของแบบทดสอบที่คำนวณ

โดยวิธีสุทธแบบ KR-20 ไม่แตกต่างกัน

ดังนั้น เพื่อให้ทราบว่าค่าความเชื่อมั่นของแบบทดสอบโจทก์ปัญหาทางคณิตศาสตร์หลังจาก คัดข้อสอบที่มีความลำเอียงออกด้วยวิธีวิเคราะห์ความลำเอียงต่างกัน คือ วิธีแปลงค่าความยาก ของข้อสอบ วิธีโค้งลักษณะข้อสอบ และ วิธีของแมนเทล-แฮนส์เชล คู่ใดบ้างที่แตกต่างกัน ผู้วิจัยจึง นำค่าความเชื่อมั่นของแบบทดสอบ ไปทดสอบความแตกต่าง เป็นรายคู่โดยวิธีสุทธการทดสอบความ แตกต่างของคะแนนมาตรฐาน (Z-test) ดังแสดงผล ในตาราง 14

ตาราง 14 ผลการทดสอบความแตกต่างของค่าความเชื่อมั่นรายคู่ของแบบทดสอบโจทก์ปัญหาทาง คณิตศาสตร์ หลังจากคัดข้อสอบที่มีความลำเอียงออก ด้วยวิธีวิเคราะห์ความลำเอียงที่แตกต่างกัน 3 วิธี

วิธีวิเคราะห์ความลำเอียง	TID ICC MH			
	Z_r	.9973	1.2225	1.2254
TID	.9973	-	2.5078*	2.5401*
ICC	1.2225		-	.0323
MH	1.2254			-

$$Z_{.05} = 1.960$$

* มีนัยสำคัญทางสถิติที่ระดับ .05

ผลการวิเคราะห์จากตาราง 14 แสดงให้เห็นว่า ค่าความเชื่อมั่นของแบบทดสอบโจทก์ ปัญหาทางคณิตศาสตร์ หลังจากคัดข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบ กับ วิธีโค้งลักษณะข้อสอบ และ แบบทดสอบฉบับที่คัดข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความ ยากของข้อสอบ กับ วิธีของแมนเทล-แฮนส์เชล มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ส่วนค่าความเชื่อมั่นของแบบทดสอบฉบับที่คัดข้อสอบที่มีความลำเอียงออกด้วยวิธีโด่งลักษณะข้อสอบ กับ วิธีของแมนเทล-แฮนส์เชล มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

ส่วนค่าความเชื่อมั่นของแบบทดสอบความเข้าใจทางภาษา หลังจากคัดข้อสอบที่มีความลำเอียงออกด้วยวิธีวิเคราะห์ความลำเอียงต่างกัน คือ วิธีแปลงค่าความยากของข้อสอบวิธีโด่ง ลักษณะข้อสอบ และ วิธีของแมนเทล-แฮนส์เชล มีค่าความเชื่อมั่นไม่แตกต่างกัน ดังนั้น ผู้วิจัยจึงไม่ทำการทดสอบความแตกต่างของค่าความเชื่อมั่นเป็นรายคู่ของแบบทดสอบความเข้าใจทางภาษา

ตาราง 15 ผลการทดสอบความแตกต่างของค่าความเชื่อมั่นของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา หลังจากคัดข้อสอบที่มีความลำเอียงออก ด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโด่งลักษณะข้อสอบ และวิธีของแมนเทล-แฮนส์เชล ที่ คำนวณจากสูตร แบบแบ่งครึ่งฉบับ

แบบทดสอบ	วิธีวิเคราะห์ความลำเอียง	K	r_{oe}	Z_r	χ^2
โจทย์ปัญหาทาง คณิตศาสตร์	TID	40	0.781	1.0463	5.8808
	ICC	21	0.829	1.1697	
	MH	19	0.855	1.2634	
ความเข้าใจ ทางภาษา	TID	37	0.730	.9290	2.0009
	ICC	19	0.727	.9168	
	MH	17	0.672	.8134	

$$\chi^2_{.05,2} = 5.991$$

ผลการวิเคราะห์จากตาราง 15 แสดงให้เห็นว่า ค่าความเชื่อมั่นของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ หลังจากคัดข้อสอบที่มีความลำเอียงออก ด้วยวิธีวิเคราะห์ความลำเอียง

ต่างกัน คือ วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ และวิธีของแมนเทล-แฮนส์เชล มีค่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ส่วนค่าความเชื่อมั่นของแบบทดสอบความเข้าใจทางภาษาหลังจากคัดข้อสอบที่มีความลำเอียงออก ด้วยวิธีวิเคราะห์ความลำเอียงต่างกัน คือ วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ และ วิธีของแมนเทล-แฮนส์เชล มีค่าแตกต่างกัน อย่างไม่มีนัยสำคัญทางสถิติ เช่นเดียวกัน นั่นคือ การวิเคราะห์ความลำเอียงของข้อสอบทั้ง 3 วิธีนี้ มีผลต่อค่าความเชื่อมั่นซึ่งคำนวณ โดยใช้สูตรแบบแบ่งครึ่งฉบับ ของแบบทดสอบปัจจัยปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา หลังจากคัดข้อสอบที่มีความลำเอียงออก แล้วไม่แตกต่างกัน

บทที่ 5

สรุปผล อภิปรายผล และข้อเสนอแนะ

ความมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียง ระหว่างวิธีวิเคราะห์ความลำเอียงต่างกันของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา
2. เพื่อเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบ หลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว ระหว่างวิธีวิเคราะห์ความลำเอียงต่างกัน ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนชายและนักเรียนหญิง ที่นับถือศาสนาพุทธและศาสนาอิสลาม ชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษา ในจังหวัดกระบี่ จำนวน 2,689 คน ซึ่งเลือกมาโดยการสุ่มอย่างง่าย (Simple Random Sampling) โดยมีห้องเรียนเป็นหน่วยการสุ่ม (Sampling Unit)

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้ มี 2 ฉบับ ที่ผู้วิจัยสร้างขึ้นเอง

ฉบับที่ 1 แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 (ค 204) ในเรื่องสมการและอสมการ อัตราส่วนและร้อยละ ปริมาตรและพื้นที่ผิว ตามหลักสูตรมัธยมศึกษาตอนต้น พุทธศักราช 2521 (ฉบับปรับปรุง พ.ศ. 2533) เป็นแบบเลือกตอบชนิด 5 ตัวเลือก จำนวน 40 ข้อ

ฉบับที่ 2 แบบทดสอบความเข้าใจทางภาษา ชั้นมัธยมศึกษาปีที่ 2 เป็นแบบเลือกตอบชนิด 5 ตัวเลือก จำนวน 40 ข้อ

วิธีดำเนินการเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลเพื่อทำการวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการตามลำดับขั้นตอน ดังนี้

1. นำหนังสือจากบัณฑิตวิทยาลัยไปติดต่อขอความร่วมมือ กับโรงเรียนที่ถูกเลือกเป็นกลุ่มตัวอย่าง เพื่อกำหนดวัน เวลา ในการดำเนินการสอบ

2. จัดเตรียมแบบทดสอบให้เพียงพอกับจำนวนนักเรียนที่จะทำการสอบในแต่ละครั้ง

3. วางแผนดำเนินการสอบ โดยผู้วิจัยเป็นผู้ดำเนินการสอบด้วยตนเอง

4. ชี้แจงให้นักเรียนที่เป็นกลุ่มตัวอย่าง ทราบวัตถุประสงค์ของการสอบและขอความร่วมมือในการสอบ

5. นำแบบทดสอบทั้งสองฉบับไปทดสอบกับกลุ่มตัวอย่าง ครั้งที่ 1 จำนวน 410 คน

6. นำผลการสอบในข้อ 5 มาตรวจให้คะแนนแล้ววิเคราะห์หาค่าความยาก และ ค่าอำนาจจำแนกเป็นรายข้อ โดยใช้โปรแกรม Item Analysis แล้วคัดเลือกข้อสอบไว้ฉบับละ 40 ข้อ

7. นำแบบทดสอบที่มีคุณภาพจากข้อ 6 ทั้ง 2 ฉบับ ไปทดสอบกับกลุ่มตัวอย่างกลุ่มที่ 2 จำนวน 306 คน เพื่อหาค่าความเชื่อมั่นของแบบทดสอบทั้งฉบับ

8. นำแบบทดสอบที่มีคุณภาพจากข้อ 7 ทั้ง 2 ฉบับ จำนวนฉบับละ 40 ข้อ ไปทดสอบกับกลุ่มตัวอย่างที่ใช้ในการวิจัย กลุ่มที่ 1 จำนวน 1,936 คน เป็นนักเรียนที่นับถือศาสนาพุทธ 1,026 คน และเป็นนักเรียนที่นับถือศาสนาอิสลาม 910 คน เพื่อศึกษาผลการวิเคราะห์ความลำเอียงของข้อสอบที่ใช้วิธีวิเคราะห์ต่างกัน ในด้านจำนวนข้อสอบที่มีความลำเอียง

9. นำแบบทดสอบที่คัดเลือกข้อสอบ ที่มีความลำเอียงออกแล้ว ทั้ง 2 ฉบับ ไปทดสอบกับกลุ่มตัวอย่างที่ใช้ในการวิจัยกลุ่มที่ 2 จำนวน 753 คน ซึ่งใช้ในการเปรียบเทียบผลการวิเคราะห์ความเชื่อมั่นของแบบทดสอบ

การเสนอผลการวิเคราะห์ข้อมูล

ผลการวิเคราะห์ข้อมูลในการศึกษานี้ ผู้วิจัยได้เสนอเป็นลำดับ ดังนี้

1. ผลการหาค่าสถิติพื้นฐานของค่าพารามิเตอร์ของแบบทดสอบโดยปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา

2. ผลการหาค่าดัชนีความลำเอียงของข้อสอบ จากแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา จากวิธีวิเคราะห์ความลำเอียงที่แตกต่างกัน 3 วิธี

3. ผลการเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียง จากวิธีวิเคราะห์ความลำเอียง 3 วิธี ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา

4. ผลการหาค่าความเชื่อมั่นของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา หลังจากคัดข้อสอบที่มีความลำเอียงออก จากการวิเคราะห์ความลำเอียงด้วยวิธีที่ต่างกัน โดยใช้สูตรการคำนวณแบบ KR - 20 และแบบแบ่งครึ่งฉบับ

5. ผลการเปรียบเทียบค่าความเชื่อมั่น ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา หลังจากคัดข้อสอบที่มีความลำเอียงออกจากการวิเคราะห์ความลำเอียงด้วยวิธีที่ต่างกัน

สรุปผลการวิเคราะห์ข้อมูล

ผลการวิเคราะห์ข้อมูลในการศึกษาค้นคว้า สรุปได้ดังนี้

1. ผลการหาค่าสถิติพื้นฐานของค่าพารามิเตอร์ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา ปรากฏว่า

แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ในกลุ่มผู้สอบที่นับถือศาสนาพุทธมีค่าอำนาจจำแนกของข้อสอบสูงกว่ากลุ่มผู้สอบที่นับถือศาสนาอิสลาม ส่วนค่าความยากของข้อสอบในกลุ่มผู้สอบที่นับถือศาสนาอิสลามมีค่าสูงกว่ากลุ่มผู้สอบที่นับถือศาสนาพุทธ และค่าสัมประสิทธิ์การเดานั้นในกลุ่มผู้สอบที่นับถือศาสนาพุทธ มีค่าสัมประสิทธิ์การเดาสูงกว่ากลุ่มผู้สอบที่นับถือศาสนาอิสลาม ส่วนค่าความเบี่ยงเบนมาตรฐานของค่าอำนาจจำแนกของข้อสอบ ในกลุ่มผู้สอบที่นับถือศาสนาพุทธจะมีค่าสูงกว่ากลุ่มผู้สอบที่นับถือศาสนาอิสลาม และค่าความเบี่ยงเบนมาตรฐานของค่าความยากของข้อสอบ ในกลุ่มผู้สอบที่นับถือศาสนาอิสลาม จะมีค่าสูงกว่ากลุ่มผู้สอบที่นับถือศาสนาพุทธ ส่วนค่าความเบี่ยงเบนมาตรฐานของค่าสัมประสิทธิ์การเดานั้นในกลุ่มผู้สอบที่นับถือศาสนาพุทธจะมีค่าสูงกว่ากลุ่มผู้สอบที่นับถือศาสนาอิสลาม

ส่วนแบบทดสอบความเข้าใจทางภาษา ในกลุ่มผู้สอบที่นับถือศาสนาพุทธมีค่าอำนาจจำแนกของข้อสอบ ค่าความยากของข้อสอบ และค่าสัมประสิทธิ์การเดาสูงกว่าในกลุ่มผู้สอบที่นับถือศาสนาอิสลาม สำหรับค่าความเบี่ยงเบนมาตรฐานของค่าอำนาจจำแนกของข้อสอบ ในกลุ่มผู้สอบที่นับถือ

ศาสนาอิสลาม มีค่าสูงกว่าในกลุ่มผู้สอบที่นับถือศาสนาพุทธ ส่วนค่าความเบี่ยงเบนมาตรฐานของค่าความยากของข้อสอบ และค่าสัมประสิทธิ์การเดานั้น ในกลุ่มผู้สอบที่นับถือศาสนาพุทธมีค่าสูงกว่าในกลุ่มผู้สอบที่นับถือศาสนาอิสลาม

2. จำนวนข้อสอบที่มีความลำเอียงในแบบทดสอบโจทยปัญหาทางคณิตศาสตร์ จากการใช้วิธีวิเคราะห์ที่ต่างกัน 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ และวิธีของแมนเทิล-แฮนส์เชล ในแต่ละวิธีวิเคราะห์มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกัน คือ วิธีของแมนเทิล-แฮนส์เชล มีข้อสอบที่มีความลำเอียงมากที่สุด จำนวน 21 ข้อ รองลงมาเป็นวิธีโค้งลักษณะข้อสอบ มีข้อสอบที่มีความลำเอียง จำนวน 19 ข้อ และ วิธีแปลงค่าความยากของข้อสอบ ไม่มีข้อสอบที่มีความลำเอียง โดยที่จำนวนข้อสอบที่มีความลำเอียงที่วิเคราะห์แบบวิธีแปลงค่าความยากของข้อสอบกับวิธีโค้งลักษณะข้อสอบ และ วิธีวิเคราะห์ความลำเอียงแบบวิธีแปลงค่าความยากของข้อสอบกับวิธีของแมนเทิล-แฮนส์เชล มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ที่ระดับ .001 ส่วนจำนวนข้อสอบที่มีความลำเอียงที่วิเคราะห์แบบวิธีโค้งลักษณะข้อสอบกับวิธีของแมนเทิล-แฮนส์เชล มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

ส่วนจำนวนข้อสอบที่มีความลำเอียง ในแบบทดสอบความเข้าใจทางภาษาจากการใช้วิธีวิเคราะห์ความลำเอียงที่แตกต่างกัน 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ และ วิธีของแมนเทิล-แฮนส์เชล ในแต่ละวิธีมีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกัน คือ วิธีของแมนเทิล-แฮนส์เชล มีข้อสอบที่มีความลำเอียงมากที่สุด จำนวน 23 ข้อ รองลงมาเป็นวิธีโค้งลักษณะข้อสอบ มีข้อสอบที่มีความลำเอียง จำนวน 21 ข้อ และ วิธีแปลงค่าความยากของข้อสอบมีข้อสอบที่มีความลำเอียงน้อยที่สุด จำนวน 3 ข้อ โดยที่จำนวนข้อสอบที่มีความลำเอียงที่วิเคราะห์แบบวิธีแปลงค่าความยากของข้อสอบกับวิธีโค้งลักษณะข้อสอบ และ วิธีวิเคราะห์ความลำเอียงแบบวิธีแปลงค่าความยากของข้อสอบกับวิธีของแมนเทิล-แฮนส์เชลมีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนจำนวนข้อสอบที่มีความลำเอียงที่วิเคราะห์แบบวิธีโค้งลักษณะข้อสอบกับวิธีของแมนเทิล-แฮนส์เชล มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

3. ความเชื่อมั่นที่คำนวณโดยวิธีสูตรแบบ KR-20 ของแบบทดสอบปัจจัยปัญหาทางคณิตศาสตร์หลังจากคิดข้อสอบที่มีความลำเอียงออกแล้วจากการวิเคราะห์ความลำเอียงด้วยวิธีที่ต่างกัน ปรากฏว่า แบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วยวิธีของแมนเทล-แฮนส์เชล มีความเชื่อมั่นสูงสุด เท่ากับ .843 รองลงมาเป็นแบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วยวิธีโค้งลักษณะข้อสอบ มีความเชื่อมั่น เท่ากับ .841 และ แบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบมีความเชื่อมั่นต่ำสุดเท่ากับ .761 โดยที่ความเชื่อมั่นของแบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วย วิธีแปลงค่าความยากของข้อสอบ กับ วิธีโค้งลักษณะข้อสอบ และ ความเชื่อมั่นของแบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบ กับ วิธีของแมนเทล-แฮนส์เชล มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนความเชื่อมั่นของแบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วยวิธีโค้งลักษณะข้อสอบกับวิธีของแมนเทล-แฮนส์เชล มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ เมื่อพิจารณาความเชื่อมั่นที่คำนวณโดยวิธีสูตรแบบแบ่งครึ่งฉบับ ของแบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกแล้ว ปรากฏว่า แบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วยวิธีของแมนเทล-แฮนส์เชล มีความเชื่อมั่นสูงสุดเท่ากับ .855 รองลงมาเป็นแบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วยวิธีโค้งลักษณะข้อสอบมีความเชื่อมั่น เท่ากับ .829 และ แบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบมีความเชื่อมั่นต่ำสุด เท่ากับ .781 โดยที่ค่าความเชื่อมั่นของแบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ และ วิธีของแมนเทล-แฮนส์เชล มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

ส่วนความเชื่อมั่นที่คำนวณโดยวิธีสูตรแบบ KR-20 ของแบบทดสอบความเข้าใจทางภาษาหลังจากคิดข้อสอบที่มีความลำเอียงออกแล้ว จากการวิเคราะห์ความลำเอียงด้วยวิธีที่แตกต่างกัน ปรากฏว่า แบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบ มีความเชื่อมั่นสูงสุด เท่ากับ .719 รองลงมาเป็นแบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วยวิธีโค้งลักษณะข้อสอบ มีความเชื่อมั่น เท่ากับ .671 และแบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วยวิธีของแมนเทล-แฮนส์เชล มีความเชื่อมั่นต่ำสุด เท่ากับ .627 เมื่อพิจารณาค่าความเชื่อมั่นที่คำนวณโดยวิธีสูตรแบบแบ่งครึ่งฉบับ ของแบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบ มีความเชื่อมั่นสูงสุด เท่ากับ .730

รองลงมาเป็นแบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วยวิธีโค้งลักษณะข้อสอบ มีค่าความเชื่อมั่นเท่ากับ .727 และ แบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วยวิธีของแมนเทล-แฮนส์เชลมีค่าความเชื่อมั่นต่ำสุด เท่ากับ .672 โดยที่ความเชื่อมั่นที่คำนวณโดยใช้สูตรแบบ KR-20 และ แบบแบ่งครึ่งฉบับของแบบทดสอบหลังจากคิดข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ และ วิธีของแมนเทล-แฮนส์เชล มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

อภิปรายผล

จากการศึกษาค้นคว้าครั้งนี้ สามารถอภิปรายผลได้ ดังนี้

1. จำนวนข้อสอบที่มีความลำเอียงของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา จากการวิเคราะห์ความลำเอียง 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ และวิธีของแมนเทล-แฮนส์เชล ซึ่งวิธีวิเคราะห์ความลำเอียงแต่ละวิธีจะพบจำนวนข้อสอบที่มีความลำเอียงแตกต่างกัน วิธีของแมนเทล - แฮนส์เชล มีจำนวนข้อสอบที่มีความลำเอียงมากที่สุด ทั้งแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา รองลงมาเป็นวิธีโค้งลักษณะข้อสอบ ส่วนวิธีแปลงค่าความยากของข้อสอบ จะพบจำนวนข้อสอบที่มีความลำเอียงน้อยที่สุด ทั้งนี้อาจเนื่องมาจากเกณฑ์ที่ใช้ในการตัดสินความลำเอียงของข้อสอบซึ่งเป็นเกณฑ์ที่ค่อนข้างต่ำ โดยที่จำนวนข้อสอบที่มีความลำเอียง ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา ที่วิเคราะห์ด้วยวิธีแปลงค่าความยากของข้อสอบ กับ วิธีโค้งลักษณะข้อสอบ และ วิธีแปลงค่าความยากของข้อสอบกับวิธีของแมนเทล-แฮนส์เชล มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ซึ่งสอดคล้องกับสมมติฐานของการวิจัย ข้อที่ 1 และ งานวิจัยของ ทศนีย์ ภิรมนตรี (2530:49-64) ที่ศึกษาวิธีวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ พบว่า จำนวนข้อสอบที่มีความลำเอียงจากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากของข้อสอบ กับ วิธีโค้งลักษณะข้อสอบ มีจำนวนข้อที่มีความลำเอียงแตกต่างกัน นอกจากนี้ยังสอดคล้องกับงานวิจัยของ สุรศักดิ์ อมรรัตนศักดิ์ (2531:197) ซึ่งศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบจากการวิเคราะห์ความลำเอียง ด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ ก็พบว่า จำนวนข้อสอบที่มีความลำเอียงจากการวิเคราะห์ความลำเอียงด้วย

วิธีแปลงค่าความยากของข้อสอบกับวิธีใดจึงลักษณะข้อสอบ มีจำนวนข้อที่มีความลำเอียงแตกต่างกัน

ส่วนวิธีวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีใดจึงลักษณะข้อสอบกับวิธีของแมนเทล-แฮนส์เชลนั้น มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ทั้งแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ แบบทดสอบความเข้าใจทางภาษา ซึ่งผลที่ได้ไม่สอดคล้องกับสมมติฐานของการวิจัยในครั้งนี้ ทั้งนี้เนื่องจาก วิธีวิเคราะห์ความลำเอียงทั้งสองวิธีค้นพบจำนวนข้อสอบที่มีความลำเอียงใกล้เคียงกัน กล่าวคือ แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ ที่วิเคราะห์ความลำเอียงด้วยวิธีใดจึงลักษณะข้อสอบ มีจำนวนข้อสอบที่มีความลำเอียง 19 ข้อ วิธีของแมนเทล-แฮนส์เชล มีจำนวนข้อสอบที่มีความลำเอียง 21 ข้อ ส่วนแบบทดสอบความเข้าใจทางภาษานั้น วิธีที่วิเคราะห์ความลำเอียงด้วยวิธีใดจึงลักษณะข้อสอบ มีจำนวนข้อสอบที่มีความลำเอียง 21 ข้อ วิธีของแมนเทล-แฮนส์เชล มีจำนวนข้อสอบที่มีความลำเอียง 23 ข้อ ดังนั้น เมื่อนำจำนวนข้อที่มีความลำเอียงมาทดสอบนัยสำคัญทางสถิติ จึงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และยังสอดคล้องกับงานวิจัยของ สวามินาธานและโรเจอร์ส (Swaminathan, H. and Rogers, H. Jane 1990) ที่ได้ตรวจสอบความลำเอียงของข้อสอบด้วยวิธีการใช้ทฤษฎีการตอบข้อคำถามกับวิธีของแมนเทล-แฮนส์เชล ก็พบว่า ทั้ง 2 วิธี ให้ผลใกล้เคียงกัน และในกรณีที่กลุ่มตัวอย่างขนาดใหญ่ วิธีของแมนเทล-แฮนส์เชล สามารถตรวจสอบความลำเอียงของข้อสอบได้ดีกว่า และยังสอดคล้องกับงานวิจัยของราชู และ คนอื่นๆ (Raju and Others. 1993) ซึ่งได้ศึกษาเปรียบเทียบผลการตรวจสอบความลำเอียง ในการวัดพื้นที่ความแตกต่างระหว่างโดเมนลักษณะข้อสอบกับวิธีของแมนเทล-แฮนส์เชล พบว่า วิธีของแมนเทล-แฮนส์เชล ค้นพบข้อสอบที่มีความลำเอียงได้มากกว่าวิธีโดเมนลักษณะข้อสอบ แต่ทั้งสองวิธีพบจำนวนข้อสอบที่มีความลำเอียงได้ใกล้เคียงกัน และสามารถที่จะเลือกใช้วิธีใดวิธีหนึ่งแทนกันได้

ผลจากการศึกษา แสดงว่า วิธีวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา ระหว่างกลุ่มผู้สอบที่นับถือศาสนาพุทธและศาสนาอิสลาม ที่นำมาใช้ในการศึกษานี้ ค้นพบจำนวนข้อสอบที่มีความลำเอียงได้สอดคล้องกัน

2. ความเชื่อมั่นที่คำนวณโดยใช้สูตรแบบ KR-20 ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ หลังจากคัดข้อสอบที่มีความลำเอียงออกแล้ว จากการวิเคราะห์ความลำเอียง 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีโดเมนลักษณะข้อสอบ และ วิธีของแมนเทล-แฮนส์เชล นั้น วิธีแปลงค่าความยากของข้อสอบกับวิธีโดเมนลักษณะข้อสอบ และ วิธีแปลงค่าความยากของข้อสอบกับ

วิธีของแมนเทล-แฮนส์เชล มีค่าความเชื่อมั่นแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งสอดคล้องกับสมมติฐานของการวิจัยข้อที่ 2 ส่วนวิธีไค้ลักษณะข้อสอบกับวิธีของแมนเทล-แฮนส์เชล มีค่าความเชื่อมั่นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ซึ่งไม่สอดคล้องกับสมมติฐานของการวิจัย ทั้งนี้อาจเนื่องมาจาก จำนวนข้อสอบหลังจากคัดข้อสอบที่มีความลำเอียงออกแล้วมีจำนวนข้อใกล้เคียงกัน จึงมีผลทำให้ค่าความเชื่อมั่นของแบบทดสอบแต่ละฉบับไม่แตกต่างกัน

สำหรับความเชื่อมั่นของแบบทดสอบความเข้าใจทางภาษา หลังจากคัดข้อสอบที่มีความลำเอียงออก จากการวิเคราะห์ความลำเอียงต่างกัน 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีไค้ลักษณะข้อสอบ และ วิธีของแมนเทล - แฮนส์เชล มีค่าความเชื่อมั่นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และผลที่ได้ไม่สอดคล้องกับสมมติฐานของการวิจัย แต่สอดคล้องกับงานวิจัยของ สุรศักดิ์ อมรรัตนศักดิ์ (2531 : 197) ที่ศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีไค้ลักษณะข้อสอบ พบว่า ค่าความเชื่อมั่นของแบบทดสอบหลังคัดข้อสอบที่มีความลำเอียงออกแล้วมีค่าไม่แตกต่างกัน

เมื่อพิจารณาค่าความเชื่อมั่นที่คำนวณโดยใช้สูตรแบบแบ่งครั้งฉบับ ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา หลังจากคัดข้อสอบที่มีความลำเอียงออกจากการวิเคราะห์ความลำเอียง 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีไค้ลักษณะข้อสอบ และ วิธีของแมนเทล-แฮนส์เชล มีค่าความเชื่อมั่นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ซึ่งผลที่ได้ไม่สอดคล้องกับสมมติฐานของการวิจัย แต่สอดคล้องกับงานวิจัยของ สุรศักดิ์ อมรรัตนศักดิ์ (2531:197) ที่ศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีไค้ลักษณะข้อสอบ พบว่า ค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดข้อสอบที่มีความลำเอียงออกแล้วมีค่าไม่แตกต่างกัน

ผลจากการศึกษาครั้งนี้ แสดงว่า แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา เมื่อคัดข้อสอบที่มีความลำเอียงออกจากรีการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีไค้ลักษณะข้อสอบ วิธีของแมนเทล-แฮนส์เชล แล้วนำแบบทดสอบทั้ง 2 ฉบับ มาคำนวณหาค่าความเชื่อมั่นโดยใช้สูตรการคำนวณแบบแบ่งครั้งฉบับ ของแบบทดสอบหลังคัดข้อสอบที่มีความลำเอียงออกด้วยวิธีวิเคราะห์ความลำเอียงที่ต่างกัน 3 วิธี ดังกล่าว แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ แสดงว่า การวิเคราะห์ความลำเอียงของข้อสอบทั้ง 3 วิธี มีผลต่อค่าความเชื่อมั่นของแบบทดสอบไม่แตกต่างกัน

ข้อเสนอแนะในการวิจัย

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 จากการศึกษาวิจัยครั้งนี้พบว่า วิธียึดลักษณะข้อสอบ และวิธีของแมนเทล - แชนส์เซลล์ ให้ผลการวิจัยชัดเจนสมบูรณ์กว่าและมีความไวต่อความลำเอียงได้ดีกว่า วิธีแปลงค่าความยากของข้อสอบ และในการตรวจสอบข้อสอบที่มีความลำเอียง ระหว่างวิธีวิเคราะห์ความลำเอียงแบบวิธียึดลักษณะข้อสอบกับวิธีของแมนเทล-แชนส์เซลล์นั้น จะเลือกใช้วิธีวิเคราะห์ความลำเอียงวิธีใดก็ได้ ทั้งนี้เพราะทั้ง 2 วิธี ค้นพบข้อสอบที่มีความลำเอียงไม่แตกต่างกัน

1.2 ในการสร้างแบบทดสอบทุกฉบับ ควรกำหนดให้การวิเคราะห์ความลำเอียงเป็นขั้นตอนหนึ่งของการสร้างแบบทดสอบ ก่อนที่จะนำแบบทดสอบไปใช้ เพื่อจะช่วยลดความลำเอียงที่อาจเกิดขึ้นจากกลุ่มผู้สอบที่แตกต่างกัน และจะทำให้แบบทดสอบมีคุณภาพดีขึ้นได้ผลการทดสอบมีความยุติธรรมมากขึ้น

1.3 ในการพิจารณาว่าข้อสอบข้อใดมีความลำเอียงหรือไม่ และจะตัดสินใจว่าควรจะให้ข้อสอบที่มีความลำเอียงหรือไม่นั้น ขึ้นอยู่กับจุดประสงค์ของการทดสอบซึ่งข้อสอบที่ลำเอียงนั้น ไม่ได้ลำเอียงตลอดไปในทุกสถานการณ์และก็ได้ให้หมายความว่า เป็นข้อสอบที่ไม่ดีหรือขาดคุณภาพทางการวัด แต่เป็นการเสนอเพื่อพิจารณาข้อสอบเหล่านี้ต่อไป ซึ่งผลจากการวิเคราะห์ความลำเอียงของข้อสอบครั้งนี้ แม้จะพบว่าข้อสอบบางข้อที่มีความลำเอียงทางการวัดระหว่างกลุ่มนักเรียนที่นับถือศาสนาพุทธ กับ นักเรียนที่นับถือศาสนาอิสลาม ก็ยังสามารถนำไปใช้ทดสอบกับกลุ่มใดกลุ่มหนึ่งโดยเฉพาะได้เป็นอย่างดี หรืออาจจะเป็นข้อสอบที่ดี ถ้านำไปใช้ทดสอบกับกลุ่มนักเรียนที่ไม่ได้คำนึงถึงความแตกต่างระหว่างศาสนา

2. ข้อเสนอแนะในการทำวิจัย

2.1 ควรมีการศึกษาเปรียบเทียบ ผลการวิเคราะห์ความลำเอียงของข้อสอบวัดผลสัมฤทธิ์ทางการเรียน เนื้อหาวิชาอื่นๆ และระดับชั้นอื่นๆ

2.2 ควรมีการศึกษาเปรียบเทียบ ผลการวิเคราะห์ความลำเอียงของข้อสอบกับตัวแปรอื่นๆ เช่น ภาษากัน เชื้อชาติ ระดับความสามารถทางการเรียน เป็นต้น

2.3 ความรู้การศึกษาเปรียบเทียบ ผลการวิเคราะห์ความลำเอียงของข้อสอบ ของแบบทดสอบวัดความถนัดในด้านต่าง ๆ แบบทดสอบวัดบุคลิกภาพ แบบทดสอบที่กลุ่มโรงเรียนใช้ร่วมกัน หรือ แบบทดสอบที่ใช้วัดผลสัมฤทธิ์ในโรงเรียนสังกัดสำนักงานการประถมศึกษาแห่งชาติ เป็นต้น

2.4 ความรู้การศึกษาเปรียบเทียบ ผลการวิเคราะห์ความลำเอียงของข้อสอบ ในเนื้อหาวิชาเดียวกันที่มีระดับความยากง่ายของเนื้อหาไม่เท่ากัน

2.5 ความรู้การศึกษาเปรียบเทียบ ผลการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีวิเคราะห์ความลำเอียงวิธีอื่นๆ เช่น วิธีโค้งลักษณะข้อสอบแบบ 1 พารามิเตอร์ วิธีการเปรียบเทียบความแตกต่างของพื้นที่แบบคิดเครื่องหมาย หรือพื้นที่แบบไม่คิดเครื่องหมาย วิธี SIBTEST วิธีไคสแควร์ เป็นต้น

உதாரணம்

บรรณานุกรม

- กาญจนา วัฒนสุนทร. การพัฒนาเกณฑ์ตัดสินข้อสอบล่าเอียงทางเพศ. วิทยานิพนธ์ ค.ค.
กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2538. อัดสำเนา.
- ชวาล แพรัตกุล. เทคนิคการวัดผล. พิมพ์ครั้งที่ 5. กรุงเทพฯ : ไทยวัฒนาพานิช, 2516.
- ชัยชัย เผ่าพงษ์. การวิเคราะห์ความล่าเอียงของข้อสอบจากแบบทดสอบวัดความถนัดทางการ
เรียนด้านคณิตศาสตร์และภาษาในระดับชั้นมัธยมศึกษาตอนต้น. วิทยานิพนธ์ ศษ.ม.
ขอนแก่น : มหาวิทยาลัยขอนแก่น, 2527. อัดสำเนา.
- ทัศนีย์ พิรมนตรี. การวิเคราะห์ความล่าเอียงของแบบสอบวิชาคณิตศาสตร์โครงการตรวจสอบ
คุณภาพการศึกษาชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526. วิทยานิพนธ์ ค.ม. กรุงเทพฯ :
จุฬาลงกรณ์มหาวิทยาลัย, 2530. อัดสำเนา.
- นิรมล ชัยสวัสดิ์. การเปรียบเทียบผลการวิเคราะห์ความล่าเอียงของแบบทดสอบความเข้าใจ
ในการอ่านภาษาไทยตามทฤษฎีคลาสสิกของที่ใช้วิธีวิเคราะห์ต่างกัน. วิทยานิพนธ์ กศ.ม.
กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2537. อัดสำเนา.
- ผจงจิต อินทสุวรรณ. "Latent Trait Theory," วารสารการวัดผลการศึกษา. 3 : 51-69 ;
มกราคม-เมษายน 2525.
- _____. ทฤษฎีการตอบข้อคำถาม. กรุงเทพฯ : สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัย
ศรีนครินทรวิโรฒ ประสานมิตร, 2528.
- _____. "แนวทางการศึกษาความล่าเอียงของข้อคำถาม," วารสารการวัดผลการศึกษา.
9 : 54-62 ; พฤษภาคม-สิงหาคม, 2530.
- พัชรี ปิยภัณท์. การวิเคราะห์ล่าเอียงของข้อสอบจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนชั้น
ประถมศึกษาปีที่ 6. วิทยานิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ
ประสานมิตร, 2531. อัดสำเนา.
- ล้วน สายยศ และ อังคณา สายยศ. หลักการวิจัยทางการศึกษา. พิมพ์ครั้งที่ 3 กรุงเทพฯ :
ศูนย์ส่งเสริมวิชาการ, 2536.
- สงบ ลักษณะ. "ความสามารถกับการตอบข้อสอบ," วารสารการวัดผลการศึกษา. 4 :
47-54, 88-93 ; พฤษภาคม-สิงหาคม, 2525.

สมเกียรติ ปดิษฐพร. การสร้างแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน (Test construction).

มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2525.

สุวัฒน์ สุขมลสันต์. การวิเคราะห์ข้อทดสอบแนวใหม่ด้วยคอมพิวเตอร์. กรุงเทพฯ : วิทยพัฒน์, 2538.

_____. การวิเคราะห์ความลำเอียงของข้อสอบภาษาอังกฤษเข้ามหาวิทยาลัยปี 2531-2533.

สถาบันภาษา กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2534.

สุรศักดิ์ อมรรัตนศักดิ์. การศึกษาเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่แตกต่างกัน 4 วิธี. วิทยานิพนธ์ ค.ด. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2531.
อัครสำเนา.

สำนักงานคณะกรรมการการศึกษาแห่งชาติ, สำนักงานกฤษฎิกมนตรี. รายงานวิจัยองค์ประกอบ
บางประการที่มีอิทธิพลต่อสัมฤทธิ์ผลทางการเรียนของนักเรียนของชั้นมัธยมศึกษา.

กรุงเทพฯ : ม.ป.พ., 2524.

อนันต์ ศรีโสภกา. การวัดและการประเมินผลการศึกษา. กรุงเทพฯ : ไทยวัฒนาพานิช, 2524.

อวยพร วิบูลย์กาญจน์. การเปรียบเทียบผลวิเคราะห์แบบสอบอุปมาอุปไมยด้วยคลาสสิกอลโมเดล
กับราล์ฟโมเดล. วิทยานิพนธ์ ค.ม. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2526.
อัครสำเนา.

Allen, M.J. and Yen, W.M. Introduction to Measurement Theory. Monterey,
Calif Brooks/Cole, 1979.

Anastasi, Anne. Differential Psychology. : New York, Macmillan
Company, 1976.

Angoff, W.H. "Use of Difficulty and Discrimination Indices for
Detecting Item Bias," In Handbook of Methods for Detecting Test
Bias. Baltimore and London : The Johns Hopkins University
Press, 1982.

- Angoff, William H. and Ford, Susan F. "Item-Race Interaction on a Test of Scholastic Aptitude," Journal of Educational Measurement. 10(2) : 95-105 ; Summer, 1973.
- Baghi, H. and Ferrara, Steven F. A Comparison of IRT, Delta Plot, and Mantel-Haenszel Techniques for Detecting Differential Item Functioning Across Subpopulations in the Maryland Test of Citizenship Skills. Paper presented at the Annual Meeting of the American Educational Research Association. San Francisco, California, March 27-31 : 1990.
- _____. Detecting Differential Item Functioning Using IRT and Mantel-Haenszel Techniques : Implementing Procedures and Comparing Results. 14-17 ; February, 1990.
- Baker, Frank B. "A Criticism of Scheuneman's Item Bias Technique," Journal of Educational Measurement. 18(1) : 59-62 ; Spring, 1981.
- Berk, R.A. (Ed.) Handbook of Methods for Detecting Test Bias. Baltimore and London Md: The Johns Hopkins University Press, 1982.
- Breland, Hunter M. and Ironson, Gial H. "DeFunis Reconsidered : A Comparative Analysis of Alternative Admissions Strategies," Journal of Educational Measurement. 13(1) : 89-99 ; Spring, 1976.

- Burrill, L.E. "Comparative Studies of Item Bias Method," In Handbook of Methods for Detecting Test Bias. Baltimore and London : The Johns Hopkins University Press, 1982.
- Camilli, Gregory and Shepard, Lorrie A. Methods for Identifying Biased Test Items. California : Sage Publication, 1994.
- Clauser, Brian E., Mazor, Kathleen and Hambleton, Ronald K. "Influence of the Criterion Variable on the Identification of Differentially Functioning Test Items Using the Mantel-Haenszel Statistics," Applied Psychological Measurement. 15(4) : 353-359 ; December, 1991.
- Cleary, T.A. and Hiltion, T.L. "An Investigation of Item Bias," Educational and Psychological Measurement. 28 : 61-75, 1968.
- Cohen, Allan and Kim, Seock-Ho. "A Comparison of Lord's χ^2 and Raju's Area Measures in Detection of DIF," Applied Psychological Measurement. 17(1) : 39-52 ; Spring, 1993.
- Cohen, Allen S., Kim, Seock-Ho and Subkoviak, Michael J. "Influence of Prior Distributions on Detection of DIF," Journal of Educational Measurement. 28(1) : 49-59 ; Spring, 1991.
- Cronbach, L.J. Essentials of Psychological Testing. New York: Harper & Row Publishers, 1970.

Doolittle, Allen E. Gender. Differences in Performance on Categories of Mathematics Achievement Test Items. Unpublished Doctoral Dissertation, The University of Iowa, 1987.

Doolittle, Allen E. Gender and Cleary, T. Anne. "Gender-Based Differential Item Performance in Mathematics Achievement Items," Journal of Educational Measurement. 24(2) : 157-166 ; Summer, 1987.

Ferguson, G.A. Statistical Analysis in Psychology and Education. New York : McMrav-Hill, 1981.

Haebara, T. A Method for Investigating Item Bias Using Birnbuam's Three-parameter Logistic Model. Iowa Testing Programs Occasional Papers, The University of Iowa, 1979.

Hambleton, R.K., and Cook, L.L. "Latent Trait Models and Their Use in the Analysis of Educational Test Data," Journal of Educational Measurement. 14 : 75-96 ; Summer, 1977.

Hambleton, Ronald K. and Others. Identifying Potentially Biased Test Items : A Comparison of the Mantel-Haenszel Statistic and Several Item Response Theory Methods. Paper Presented at the Annual Meeting of the American Educational Research Association. San Francisco, California, April 16-20 ; 1986.

Hambleton, Ronald K. and Swaminathan, H. Item Response Theory Principles and Applications. Boston : Kluwer-Nijhoff Publishing, 1985.

Hambleton, Ronald K., Swaminathan, H. and Rogers, H. Jane, Fundamentals of Item Response Theory. Newbury Park, California, Sage Publications, 1991.

Holland, Paul W. and Thayer, Dorothy T. Differential Item Derformance and the Mantel-Haenszel Procedure. In Wainer, Howard and Braun, Henty I. (Eds.) Test Validity. Hillsdale, New Jersey : Lawrence Erlbanm Associates, 1988.

Holland, Paul W. and Wainer, H. (Eds.). Differential Item Functioning. Hillsdale NJ : Erlbanm, 1993.

Hulin, Charles L., Drasgow, Fritz and Parson, Charles K. Item Response Theory Application to Psychological Measurement. Illinois : Dow Jones Irwin, 1983.

Intasuwan, P. A Comparison of Three Approaches For Determining Item Bias in Gross-National Testing. Doctor's Thesis The University of Pittsburgh, 1979.

Ironson, G. H. and Subkoviak, M. J. "A Comparison of Several Methods of Assessing Item Bias," Journal of Educational Measurement. 16 : 209-225 ; 1979.

- Jensen, Arthur. R. Bias in Mental Testing. New York : A division of Macmillan Publishing, 1980.
- Kim, Seock-Ho and Cohen, Allen S. "A Comparison of Two Area Measures for Detecting Differential Item Functioning," Applied Psychological Measurement. 15(3) : 269-278 ; September, 1991.
- Laksana, S. Application of Analysis of Variance Approach and ITBS Form 7. The University of Iowa, 1979.
- Laksana, S. and Coffman, W.E. A Comparison of an ANOVA Approach an ICC Approach for Assessing Item Bias in an Achievement Test. Iowa Testing Programs Occasional Papers, The University of Iowa, 1980.
- Lautenschlager, Gary J. and Park, Dong-Gun. "IRT Item Bias Detection Procedure : Issues of Model Misspecification, Robustness, and Parameter Linking," Applied Psychological Measurement. 12(4) : 365-376 ; December, 1988.
- Linacre, John Michael. The Practical Realization of The Standard Error of the Mantel-Haenszel Statistic. Paper Presented at the Annual Meeting of the American Educational Research Association. New Orleans, Los Angeles, April, 5-9 ; 1988.
- Linn, Robert L. "Selection Bias : Multiple Meanings," Journal of Educational Measurement. 21(1) : 33-47 ; Spring, 1984.

- Linn, Robert L. and Harnisch Delwyn L. "Interactions Between Item Context and Group Membership on Achievement Test Items," Journal of Educational Measurement. 18(2) : 109-118 ; summer, 1981.
- Linn, Robert L. and Levine, Michael V., Hastings, Nicholas C. and Wardrop James L. "Item Bias in a Test of Reading Comprehension," Applied Psychological Measurement. 5(2) : 159-173 ; 1981.
- Lord, F.M. "Practical Applications of Item Characteristic Curve Theory," Journal of Educational Measurement. 14 : 17-138 ; Summer, 1977.
- Lord, F.M. Applications of Item Response Theory to Practical Testing Problems. Hillsdale New Jersey : Lawrence Erlbaum Associates, 1980.
- Lord, F.M. and Novick, M.R. Statistical Theory of Mental Test Scores. Reading, Massachusetta : Addinon-Wesley, 1968.
- Marascuilo, Leonard A. and Slaughter, Robert E. "Statistical Procedures for Identifying Possible Sources of Item Biased on χ^2 Statistics," Journal of Educational Measurement. 18(4) : 229-248 ; Winter, 1981.

- Mazor, Kathleen M. and Others. The Effect of Sample Size on the Functioning of the Mantel-Haenszel Statistic. Paper Presented at the Annual Meeting of the National Council on Measurement in Education Chicago, Illinois, April, 1991.
- McCauley, Cynthia D. and Mendoza, Jorge. "A simulation Study of Item Bias Using a Two-Parameter Item Response Model. Applied Psychological Measurement. 9(4) : 389-400 ; 1985.
- Mehrens, W.A. and Lehmann, I.J. Measurement and Evaluation in Education and Psychology. 2nd ed, New York : 1978.
- Mellenbergh, G. "Contingency Table Models for Assessing Item Bias," Journal of Educational Statistics. 7 : 105-118 ; 1982.
- Osterlind, S.J. Test Item Bias. Beverly Hills : California, Sara Miller McCune, 1983.
- Popham, W.J. Modern Educational Measurement. Englewood Cliffs, New Jersey : Prentice-Hall, 1981.
- Raju, Drasgow F. and Slinde, Jeffrey A. "An Empirical Comparison of the Area Methods, Lord's Chi-Square Test and the Mantel-Haenszel Technique for Assessing Differential Item Functioning," Educational and Psychological Measurement. 53, 1993.

- Raju, Nambury S. "Determining the Significance of Signed and Unsigned Areas Between Two Item Response Function," Applied Psychological Measurement. 14(2) : 197-207 ; June, 1990
- Raju, Nambury S. Bode, Rita K. and Larsen, Vernon S. "An Empirical Assessment of the Mantel-Haenszel Statistic for Studying Differential Item Performance," Applied Measurement in Education. 2(1) : 1-13 ; 1989.
- Rogers, Helen Jane. A Logistic Regression Procedure for Detecting Item Bias. Unpublished Doctoral Dissertation, University of Massachusetts, 1989.
- Rudner, Lawrence M., Getson, Pamela R. and Knight, David L. "A Monte Carlo Comparison of Seven Biased Item Detection Techniques," Journal of Educational Measurement. 17 : 1-10 ; Spring, 1980.
- Ryan, Katherine. "The Performance of the Mantel-Haenszel Procedure Across Samples and Matching Criteria," Journal of Educational Measurement. 28(4) : 325-337 ; Winter, 1991.
- Scheunemen, Janice D. "A Response to Baker's Criticism," Journal of Educational Measurement. 18(1) : 63-66 ; Spring, 1981.
- Scheunemen, Janice D. "A Posteriori Analysis of Biased Items," In Handbook of Methods for Detecting Test Bias. Ronald A. Berk, ed., Baltimore and London, The John Hopkins University Press, 1982.

- Schoener, John Edwin. A Comparison of Statistical and Judgmental Methods For Identifying Item Bias (Latent Trait, Test Construction). Unpublished Doctoral Dissertation, City University of New York, 1984.
- Shepard, L.A. "Definitions of Bias," In Handbook of Methods for Detecting Test Bias. Baltimore and London : The Johns Hopkins University Press, 1982.
- Shepard, L. A., Camilli, G. and Averill, M. "Test Item Bias," Journal of Education Statistics. 6 : 371-375 ; 1981.
- Shepard, L. A., Camilli, G. and Williams, D. M. "Accounting for Statistical Artifacts in Item Bias Research," Journal of Educational Measurement. 21(2) : 93-128 ; Summer, 1984.
- _____. "Validity of Approximation Techniques for Detecting Item Bias," Journal of Educational Measurement. 22(2) : 77-105 ; Summer, 1985.
- Snedecor, and Cochran. Statistical Method. Iowa : Iowa State University Press, 1967.
- Subkoviak, M.J., et al. "Empirical Comparison of Selected Item Bias Detection Procedures With Bias Manipulation," Journal of Educational Measurement. 21(1) : 49-58 ; Spring, 1984.

Sudweeks, Richard, R. and Tolmen, Richard R. The Use of Empirical Versus Subjective Procedures for Identifying Science Test Items Which Function Differentially for Females and Males.
Paper Presented at the Annual Meeting of the National Association for Research in Science Teaching. Atlanta, Georgia, 8-11 ; April ; 1990.

Swaminathan, Hariharan and Rogers, H. Jane. "Detecting Differential Item Functioning Using Logistic Regression Procedures," Journal of Educational Measurement. 27(4) : 361-370 ; Winter, 1990.

Thissen, David, Steinberg, Lynne and Wainer Howard. "Use of Item Response Theory in the Study of Group Differences in Trace Lines," In Test Validity. Hillsdale, New Jersey : Lawrence Erlbaum Associates, 1988.

Warm, T.A. A Primer of Item Response Theory. Oklahoma : U.S. Coast Guard Institute, 1978.

ภาคผนวก

ภาคผนวก ก
รายชื่อผู้เข้าร่วมงาน

ผู้เชี่ยวชาญในการพิจารณาความเที่ยงตรงเชิงเนื้อหาของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์

- | | | |
|--------------------|-------------|---------------------------------------|
| 1. อาจารย์ระวีวรรณ | พันธุ์พานิช | มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร |
| 2. อาจารย์ชวลิต | รวชอาจิณ | มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร |
| 3. อาจารย์เรวดี | อินทสระระ | มหาวิทยาลัยศรีนครินทรวิโรฒ ภาควิชา |
| 4. อาจารย์ณัฐ | ณ นคร | มหาวิทยาลัยสุโขทัยธรรมมาธิราช |
| 5. อาจารย์สุนันท์ | ฉิมวิชัย | โรงเรียนอัสสัมชัญ ลำปาง |
| 6. อาจารย์กัญญา | โทดามา | โรงเรียนอัสสัมชัญ ลำปาง |

ผู้เชี่ยวชาญในการพิจารณาความเที่ยงตรงเชิงเนื้อหาของแบบทดสอบความเข้าใจทางภาษา

- | | | |
|--------------------|-------------|---------------------------------------|
| 1. รศ. วิทยา | วิศาลาภรณ์ | มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร |
| 2. อาจารย์ระวีวรรณ | พันธุ์พานิช | มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร |
| 3. ผศ. เชาวนา | ชวลิตดำรง | มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร |
| 4. ผศ. จันทิมา | พรหมโชติกุล | มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร |
| 5. อาจารย์เรวดี | อินทสระระ | มหาวิทยาลัยศรีนครินทรวิโรฒ ภาควิชา |

ผู้เชี่ยวชาญในการพิจารณาคัดเลือกสถานการณ์เพื่อใช้สร้างแบบทดสอบความเข้าใจทางภาษา

- | | | |
|-----------------------|----------------|------------------------|
| 1. อาจารย์เอื้อนทิพย์ | บัณฑูรจินดา | โรงเรียนวัดบวรนิเวศ |
| 2. อาจารย์รัชนี้ | ตันประเสริฐ | โรงเรียนวัดบวรนิเวศ |
| 3. อาจารย์ลาวัลย์ | พหลยุทธ | โรงเรียนวังม่วงวิทยาคม |
| 4. อาจารย์พวงแก้ว | ศรีเลิศชัย | โรงเรียนมักกะสันพิทยา |
| 5. อาจารย์มณฑา | สุวรรณสัมพันธ์ | โรงเรียนมักกะสันพิทยา |

ภาคผนวก ข

**ค่าความชุก (p) และค่าอำนาจจำแนก (r) รวบรวม
ของแบบทดสอบใจทึ่ปัญหาทางคณิตศาสตร์ และ
แบบทดสอบความเข้าใจทางภาษา**

ตาราง 16 ค่าความชาก (p) ค่าอำนาจจำแนก (r) เฉพาะตัวถูกของแบบทดสอบโงทย์ปัญหา
ทางคณิตศาสตร์จากการทดสอบครั้งที่ 1

ข้อ	p	r	ข้อ	p	r	ข้อ	p	r
1*	0.747	0.414	21	0.909	0.290	41*	0.794	0.419
2	0.467	-0.073	22	0.875	0.354	42	0.840	0.215
3*	0.506	0.440	23*	0.418	0.360	43	0.167	0.170
4	0.236	0.160	24*	0.263	0.359	44	0.307	0.079
5*	0.425	0.411	25*	0.491	0.538	45*	0.600	0.480
6*	0.762	0.401	26*	0.364	0.412	46	0.880	0.182
7*	0.494	0.321	27*	0.730	0.263	47*	0.445	0.278
8*	0.612	0.451	28*	0.595	0.454	48*	0.759	0.385
9	0.845	0.329	29*	0.791	0.377	49*	0.398	0.338
10*	0.538	0.437	30*	0.787	0.352	50*	0.214	0.445
11*	0.329	0.569	31*	0.489	0.543	51*	0.538	0.467
12*	0.752	0.227	32*	0.649	0.478	52*	0.521	0.241
13	0.211	0.118	33*	0.693	0.485	53	0.079	-0.013
14	0.840	0.355	34*	0.582	0.507	54	0.064	0.218
15*	0.511	0.514	35	0.199	0.249	55*	0.560	0.354
16*	0.727	0.454	36*	0.376	0.332	56	0.138	0.143
17	0.391	-0.028	37*	0.359	0.271	57	0.047	0.072
18	0.049	0.063	38	0.108	0.361	58*	0.729	0.241
19*	0.747	0.407	39*	0.609	0.454	59	0.098	0.147
20*	0.432	0.341	40*	0.428	0.373	60*	0.759	0.423

* ข้อสอบที่คัดเลือกไว้สำหรับการวิจัย

ตาราง 17 ค่าความซาก (p) ค่าอำนาจจำแนก (r) เฉพาะตัวถูกของแบบทดสอบความเข้าใจ
ทางภาษา จากการทดสอบครั้งที่ 1

ข้อ	p	r	ข้อ	p	r	ข้อ	p	r
1*	0.746	0.315	21*	0.324	0.328	41	0.195	0.080
2*	0.749	0.287	22	0.193	0.020	42	0.639	0.336
3	0.039	0.101	23	0.032	0.106	43*	0.359	0.267
4	0.107	-0.002	24*	0.344	0.309	44	0.173	0.075
5*	0.741	0.217	25	0.395	0.158	45*	0.468	0.287
6*	0.688	0.275	26*	0.471	0.349	46	0.454	0.120
7*	0.622	0.255	27	0.088	0.029	47	0.317	0.219
8	0.110	-0.099	28*	0.256	0.295	48	0.232	0.053
9*	0.471	0.379	29*	0.500	0.271	49*	0.698	0.482
10	0.107	-0.008	30*	0.349	0.265	50*	0.732	0.296
11*	0.532	0.307	31*	0.276	0.289	51*	0.332	0.230
12	0.202	0.055	32*	0.759	0.264	52*	0.653	0.451
13*	0.476	0.233	33*	0.705	0.346	53	0.176	-0.012
14*	0.737	0.252	34*	0.485	0.445	54*	0.288	0.351
15*	0.481	0.380	35*	0.768	0.406	55*	0.485	0.313
16	0.112	-0.084	36	0.893	0.370	56	0.166	0.095
17*	0.532	0.340	37*	0.394	0.304	57*	0.305	0.204
18*	0.437	0.246	38*	0.779	0.265	58*	0.715	0.288
19*	0.583	0.354	39*	0.727	0.227	59*	0.276	0.232
20*	0.480	0.259	40	0.039	0.049	60*	0.356	0.246

* ข้อสอบที่คัดเลือกไว้สำหรับการวิจัย

ภาคผนวก ค

ผลการวิเคราะห์องค์ประกอบเพื่อตรวจสอบความเป็นมิติเดียว

ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์

และแบบทดสอบความเข้าใจทางภาษา จากการทดสอบครั้งที่ 1

ตาราง 18 ค่าไอเกน เพอร์เซ็นต์ของความแปรปรวน และเพอร์เซ็นต์ของความ แปรปรวนสะสม
จากการนำแบบทดสอบปัจจัยปัญหาทางคณิตศาสตร์มาวิเคราะห์องค์ประกอบ

องค์ประกอบที่	ค่าไอเกน	เพอร์เซ็นต์ของความ แปรปรวน	เพอร์เซ็นต์ของความ แปรปรวนสะสม
1	7.60393	19.0	19.0
2	1.62804	4.1	23.1
3	1.33616	3.3	26.4
4	1.20001	3.0	29.4
5	1.11157	2.8	32.2
6	1.08129	2.7	34.9
7	1.04543	2.6	37.5
8	1.00777	2.5	40.0
9	1.00339	2.5	42.5

ตาราง 19 ค่าไอเกน เปอร์เซ็นต์ของความแปรปรวน และเปอร์เซ็นต์ของความ แปรปรวนสะสม
จากการนำแบบทดสอบความเข้าใจทางภาษา มาวิเคราะห์องค์ประกอบ

องค์ประกอบที่	ค่าไอเกน	เปอร์เซ็นต์ของความ แปรปรวน	เปอร์เซ็นต์ของความ แปรปรวนสะสม
1	4.10012	10.3	10.3
2	1.71896	4.2	14.5
3	1.38865	3.5	18.0
4	1.25987	3.1	21.2
5	1.22831	3.1	24.2
6	1.20308	3.0	27.2
7	1.18146	3.0	30.2
8	1.14604	2.9	33.2
9	1.12303	2.8	35.9
10	1.08844	2.7	38.6
11	1.06892	2.7	41.3
12	1.05538	2.6	43.9
13	1.02388	2.6	46.5
14	1.01073	2.5	49.0

ภาคผนวก ง

ค่าพารามิเตอร์ ค่าอำนาจจำแนก (a) ค่าความชาก (b) และ
ค่าสัมประสิทธิ์การเคา (c) ของแบบทดสอบจิตวิทยา
ทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา

ตาราง 20 ค่าพารามิเตอร์ของแบบทดสอบใจที่ปัญหาทางคณิตศาสตร์

ข้อ	ศาสนาพุทธ (N=1,026)			ศาสนาอิสลาม (N=910)			รวม (N=1,936)		
	a	b	c	a	b	c	a	b	c
1	.749	-.027	.251	.673	.120	.199	.722	.104	.244
2	1.014	1.175	.311	1.097	.963	.217	.998	1.084	.255
3	.690	1.473	.075	.854	1.478	.050	.733	1.444	.046
4	.955	.334	.149	.901	.546	.089	.894	.402	.098
5	.865	1.325	.118	.845	1.199	.100	.806	1.252	.095
6	.786	.881	.167	.997	1.228	.197	.869	1.061	.181
7	1.146	1.525	.219	1.066	1.279	.140	1.051	1.409	.175
8	1.151	1.815	.080	.897	1.897	.064	1.012	1.835	.068
9	.326	.042	.218	.387	-.075	.162	.330	-.127	.166
10	1.277	1.349	.160	1.093	1.337	.159	1.235	1.364	.164
11	.557	.490	.069	.753	.810	.137	.591	.547	.057
12	.995	.175	.190	.690	.185	.166	.829	.179	.175
13	.623	1.819	.247	.675	3.358	.324	.576	2.370	.277
14	3.223	1.700	.208	1.207	1.828	.206	2.068	1.780	.215
15	2.765	1.425	.171	.821	1.984	.180	1.435	1.684	.179
16	1.250	1.186	.152	1.060	1.352	.132	1.133	1.274	.139
17	1.865	1.592	.213	1.279	1.758	.219	1.592	1.660	.216
18	.402	.600	.131	.546	.466	.100	.428	.451	.083
19	1.354	.836	.153	1.415	1.152	.188	1.407	1.022	.177
20	1.385	.230	.248	1.034	.242	.181	1.157	.216	.203

ตาราง 20 (ต่อ)

ข้อ	ศาสนาพุทธ (N=1,026)			ศาสนาอิสลาม (N=910)			รวม (N=1,936)		
	a	b	c	a	b	c	a	b	c
21	1.026	.474	.178	.797	.628	.119	.896	.534	.141
22	1.165	1.803	.226	1.327	1.665	.221	1.211	1.749	.222
23	2.179	1.314	.230	1.285	1.189	.198	1.841	1.292	.255
24	2.047	1.345	.221	1.823	1.443	.230	2.063	1.402	.227
25	1.495	1.382	.155	1.955	1.399	.137	1.557	1.406	.141
26	1.908	1.705	.178	1.883	1.548	.181	1.935	1.648	.180
27	3.136	1.958	.311	1.361	2.187	.243	2.618	2.003	.282
28	1.101	1.912	.240	.846	1.412	.165	.942	1.730	.212
29	4.131	1.935	.238	1.968	1.806	.240	3.638	1.821	.242
30	1.246	.801	.242	.830	.784	.144	1.025	.839	.206
31	2.132	1.570	.270	1.196	1.472	.207	1.927	1.560	.253
32	2.153	1.883	.100	1.707	1.680	.072	1.974	1.792	.087
33	1.318	1.276	.335	1.181	1.103	.251	1.205	1.193	.292
34	1.880	1.436	.174	1.452	1.819	.202	1.937	1.599	.192
35	1.263	2.594	.106	.699	2.602	.070	.813	2.668	.085
36	1.792	1.401	.276	1.982	1.664	.284	2.200	1.658	.285
37	2.474	1.745	.207	1.127	1.656	.146	1.984	1.724	.191
38	.746	1.959	.189	.918	2.328	.190	.823	2.148	.193
39	1.454	1.082	.272	1.094	1.308	.226	1.366	1.226	.261
40	1.334	1.104	.243	1.604	1.248	.238	1.423	1.201	.241

ตาราง 21 ค่าพารามิเตอร์ของแบบทดสอบความเข้าใจทางภาษา

ข้อ	ศาสนาพุทธ (N=1,026)			ศาสนาอิสลาม (N=910)			รวม (N=1,936)		
	a	b	c	a	b	c	a	b	c
1	.345	-1.757	.173	.367	-1.296	.126	.350	-1.556	.143
2	.989	.459	.346	.461	.478	.124	.758	.599	.289
3	.275	-2.098	.169	.347	-1.891	.114	.288	-2.173	.124
4	.392	-.983	.159	.387	-.904	.110	.373	-1.041	.115
5	.491	1.188	.267	.266	1.297	.149	.347	1.243	.206
6	.635	1.183	.097	.687	1.305	.107	.643	1.228	.093
7	.489	.269	.125	.295	.044	.125	.377	.114	.106
8	.369	3.011	.273	.295	1.548	.112	.247	2.366	.154
9	.418	-.073	.149	.362	-1.078	.105	.367	-.638	.100
10	.514	6.230	.152	.367	3.886	.103	.336	5.146	.119
11	.387	1.875	.113	.413	1.004	.068	.365	1.353	.060
12	.277	2.662	.209	.176	1.764	.149	.184	2.342	.149
13	.651	1.645	.232	.385	2.363	.134	.533	1.951	.197
14	.712	1.919	.106	.394	2.562	.092	.510	2.226	.095
15	.553	-.138	.123	.375	-.294	.107	.449	-.290	.087
16	.416	1.551	.121	.511	1.066	.080	.416	1.270	.075
17	.380	2.229	.124	.346	2.605	.097	.322	2.437	.088
18	.875	2.380	.154	.367	4.247	.105	.606	2.996	.134
19	.450	1.588	.177	.444	1.357	.117	.422	1.453	.133
20	.543	-	.276	.311	4.306	.195	.387	5.621	.254

ตาราง 21 (ต่อ)

ข้อ	ศาสนาพุทธ (N=1,026)			ศาสนาอิสลาม (N=910)			รวม (N=1,936)		
	a	b	c	a	b	c	a	b	c
21	.600	2.119	.209	.481	2.389	.150	.597	2.186	.191
22	.565	-.321	.122	.637	-.064	.125	.593	-.223	.109
23	1.465	.929	.253	1.053	.590	.206	1.332	.812	.247
24	1.333	.612	.201	1.875	.305	.162	1.573	.472	.188
25	2.016	.595	.322	1.833	.232	.225	2.154	.461	.292
26	.557	2.345	.217	.369	3.580	.219	.477	2.875	.227
27	.775	-.529	.072	.611	-.610	.068	.693	-.602	.046
28	.654	-.219	.083	.541	.013	.088	.570	-.179	.061
29	.424	2.960	.174	.354	2.725	.143	.355	2.956	.150
30	.420	.835	.085	.313	1.682	.105	.347	1.080	.067
31	.908	-.594	.073	.767	-.766	.096	.828	-.710	.059
32	.690	-.312	.083	.470	-.595	.095	.568	-.485	.065
33	.379	2.711	.193	.302	2.628	.144	.289	2.898	.155
34	.806	-.337	.111	.572	-.618	.116	.695	-.458	.108
35	.433	-	.156	.278	-	.274	.303	-	.253
36	.493	1.257	.113	.333	1.992	.114	.384	1.493	.090
37	.345	3.111	.098	.278	3.220	.101	.273	3.227	.073
38	.760	.055	.094	.662	-.312	.048	.682	-.177	.041
39	.432	2.670	.133	.400	2.945	.107	.381	2.888	.110
40	.307	3.949	.222	.282	1.862	.114	.228	2.882	.124

ภาคผนวก จ

**ค่าเฉลี่ยของข้อสอบแต่ละข้อระหว่างนักเรียนที่นับถือศาสนาพุทธกับนักเรียน
ที่นับถือศาสนาอิสลาม ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และ
แบบทดสอบความเข้าใจทางภาษา**

**ภาพประกอบการกระจายของคู่อันดับค่าเฉลี่ยของข้อสอบแต่ละข้อระหว่าง
นักเรียนที่นับถือศาสนาพุทธกับนักเรียนที่นับถือศาสนาอิสลาม ของแบบทดสอบ
โจทย์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา**

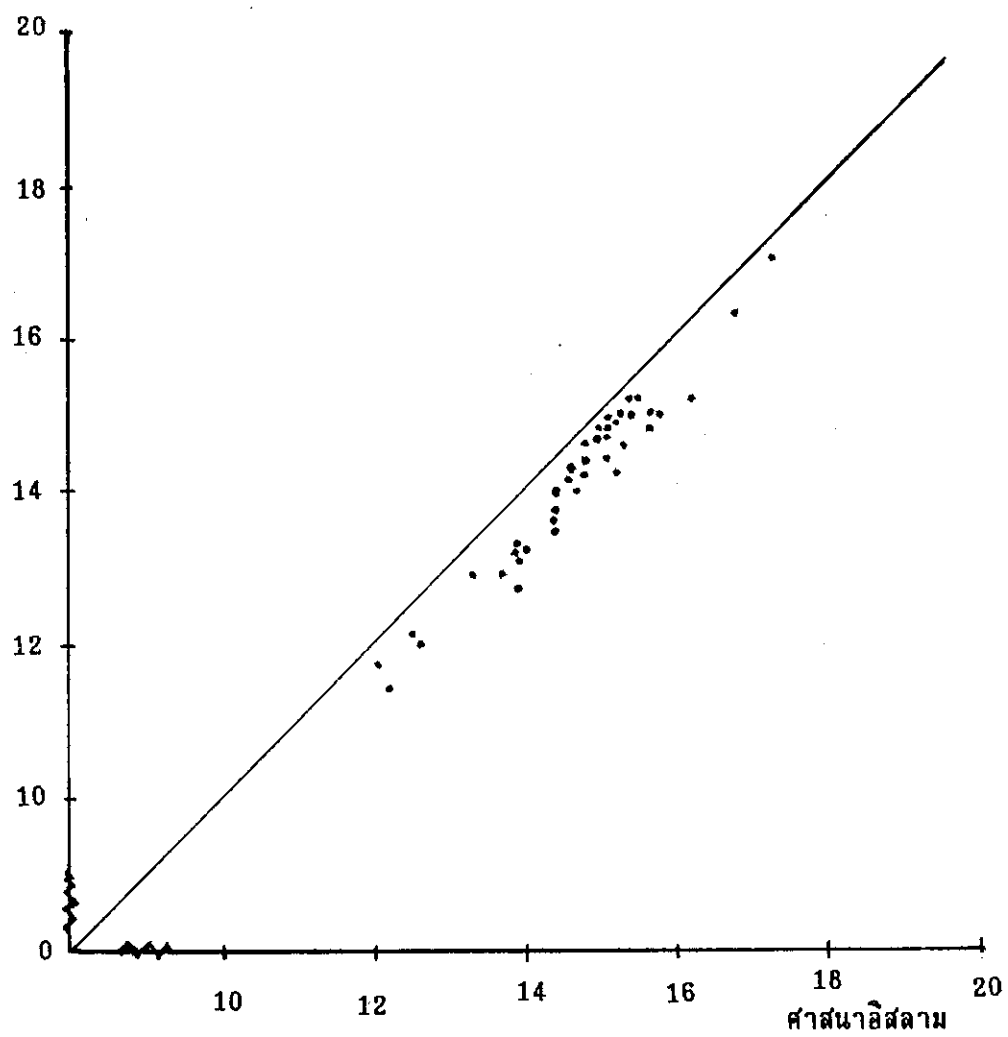
ตาราง 22 แสดงค่าเฉลี่ยของข้อสอบแต่ละข้อระหว่างกลุ่มนักเรียนที่นับถือศาสนาพุทธ กับกลุ่มนักเรียนที่นับถือศาสนาอิสลาม ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษา

ข้อ	แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์		แบบทดสอบความเข้าใจทางภาษา	
	ศาสนาพุทธ	ศาสนาอิสลาม	ศาสนาพุทธ	ศาสนาอิสลาม
1	11.44	12.20	10.00	10.72
2	13.16	13.92	11.92	13.12
3	15.20	16.20	10.08	10.12
4	12.76	13.92	10.84	11.24
5	14.80	15.12	13.08	13.44
6	13.60	14.48	14.68	15.00
7	14.40	15.16	12.72	12.44
8	16.32	16.88	14.20	14.04
9	11.76	12.08	12.08	11.12
10	14.80	15.08	17.04	16.60
11	13.32	13.96	14.72	14.16
12	12.16	12.52	14.04	13.36
13	14.04	14.48	14.24	15.08
14	15.08	15.32	16.00	15.76
15	15.28	15.40	12.04	12.12
16	14.60	15.36	14.40	14.44
17	14.80	15.16	15.00	15.48
18	12.96	13.32	16.28	16.80
19	14.00	14.72	14.12	14.36

ตาราง 22 (ต่อ)

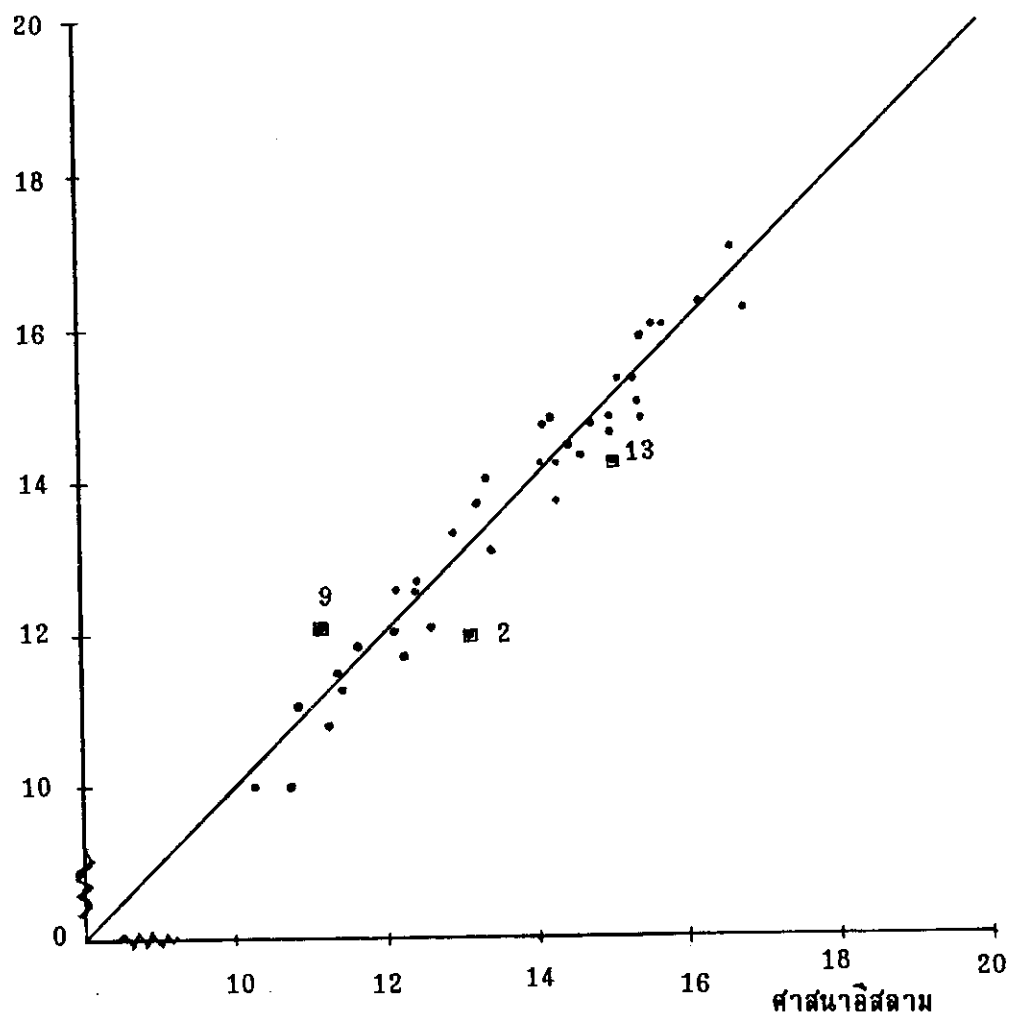
ข้อ	แบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์		แบบทดสอบความเข้าใจทางภาษา	
	ศาสนาพุทธ	ศาสนาอิสลาม	ศาสนาพุทธ	ศาสนาอิสลาม
20	12.04	12.68	15.36	15.36
21	12.92	13.76	14.84	15.40
22	14.72	15.08	11.72	12.28
23	14.36	14.60	13.72	13.28
24	14.48	14.88	13.32	12.96
25	15.00	15.84	12.56	12.40
26	15.28	15.52	14.84	15.04
27	14.20	15.28	11.32	11.48
28	14.68	14.88	12.04	12.64
29	14.96	15.16	15.36	15.16
30	13.24	13.96	13.72	14.32
31	14.28	14.88	11.00	10.80
32	16.44	17.12	11.80	11.60
33	13.24	13.96	14.76	14.76
34	15.00	15.48	11.56	11.32
35	17.08	17.32	16.28	16.20
36	14.20	14.60	14.32	14.64
37	15.08	15.76	15.88	15.48
38	14.88	15.76	12.52	12.16
39	13.52	14.40	15.60	16.00
40	13.76	14.48	14.84	14.24

ศาสนาพุทธ



ภาพประกอบ 12 แสดงการกระจายของคู่อันดับค่าเดลตาระหว่างกลุ่มนักเรียนที่ นับถือศาสนาพุทธ กับกลุ่มนักเรียนที่นับถือศาสนาอิสลามของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์จำนวน 40 ข้อ

ศาสนาพุทธ



☐ ข้อสอบที่มีความลำเอียง

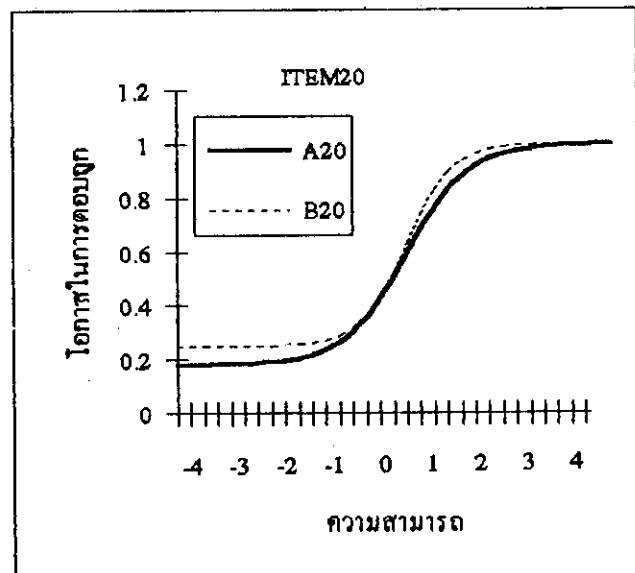
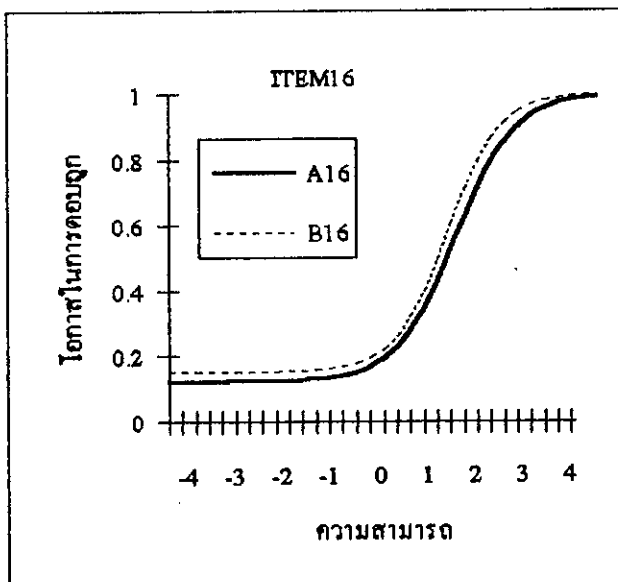
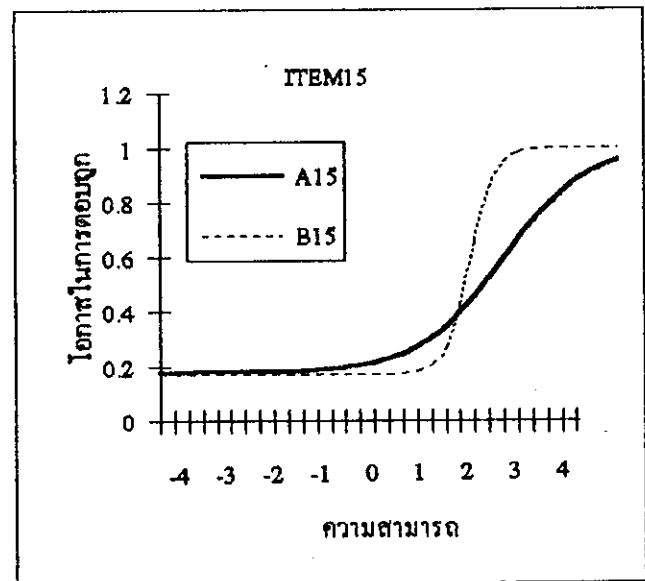
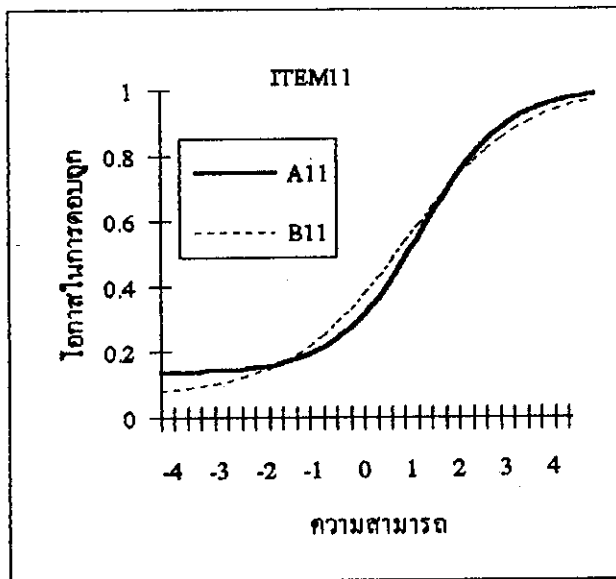
ภาพประกอบ 13 แสดงการกระจายของคู่อันดับค่าเฉลยระหว่างกลุ่มนักเรียนที่นับถือศาสนาพุทธ กับกลุ่มนักเรียนที่นับถือศาสนาอิสลามของแบบทดสอบความเข้าใจทางภาษา จำนวน 40 ข้อ

ภาคผนวก จ

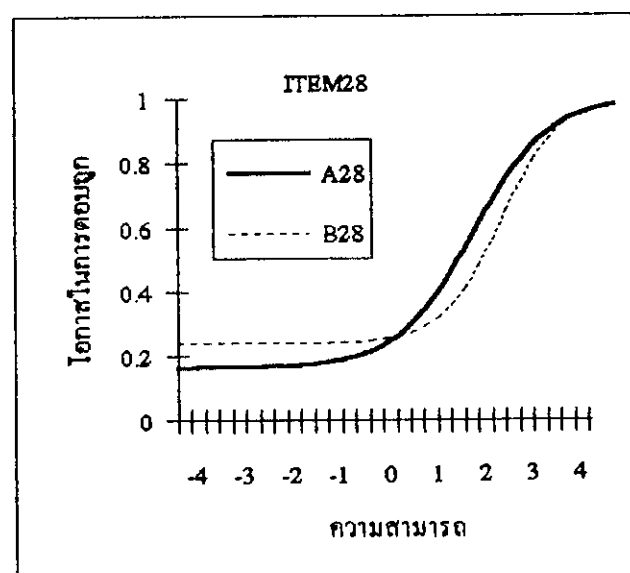
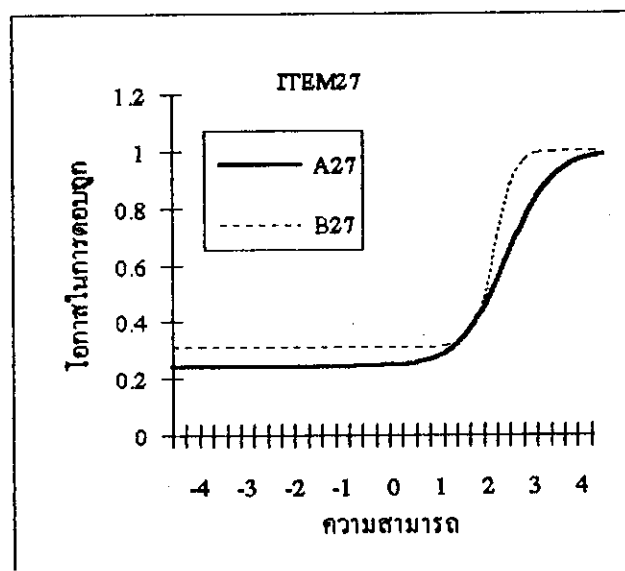
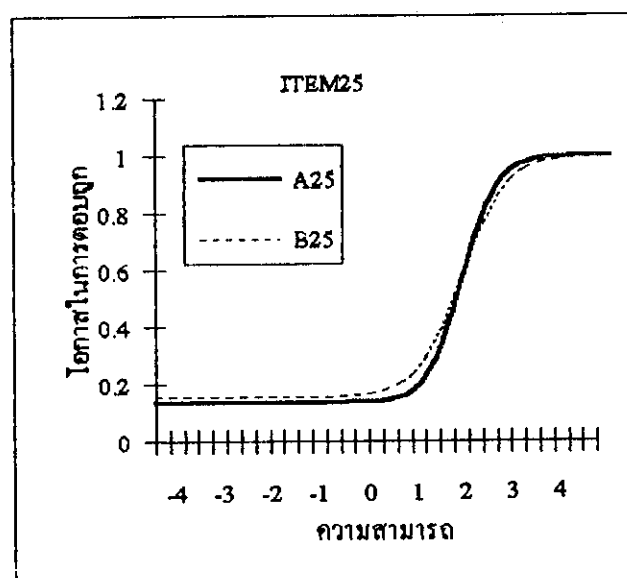
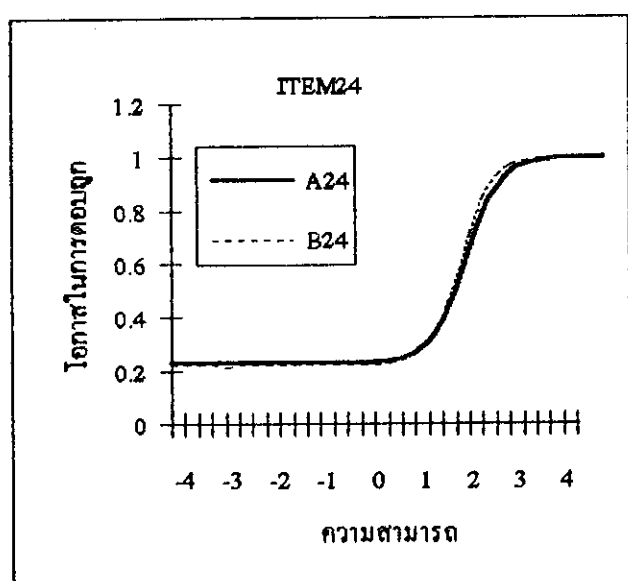
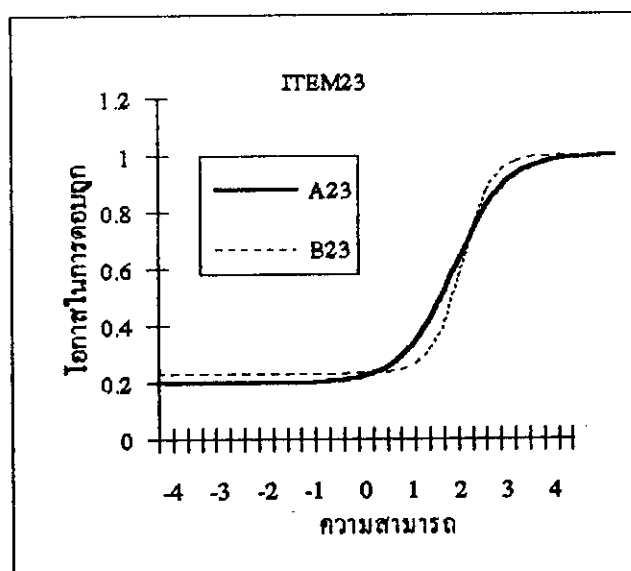
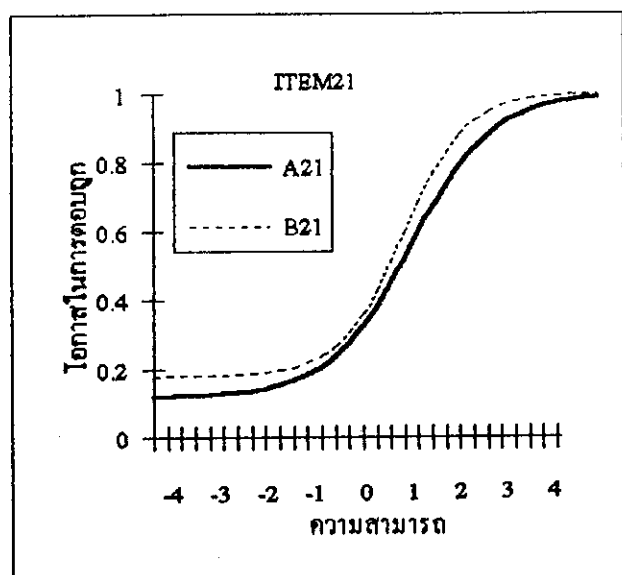
ติดตั้งลักษณะข้อสอบรายข้อระหว่างกลุ่มนักเรียนที่นับถือศาสนาพุทธ

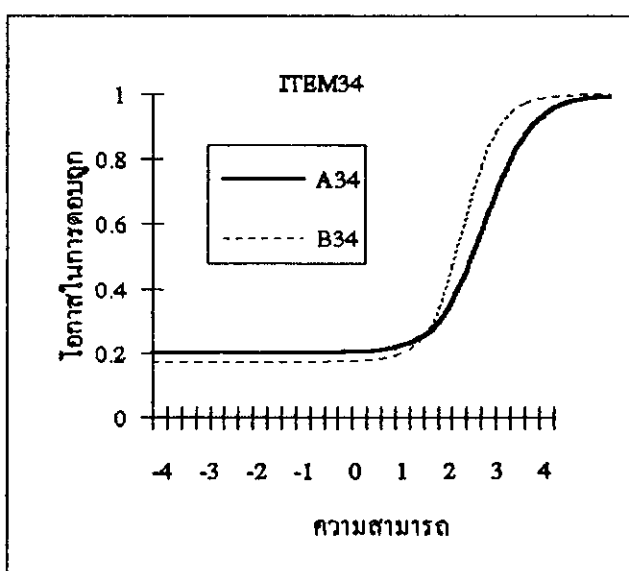
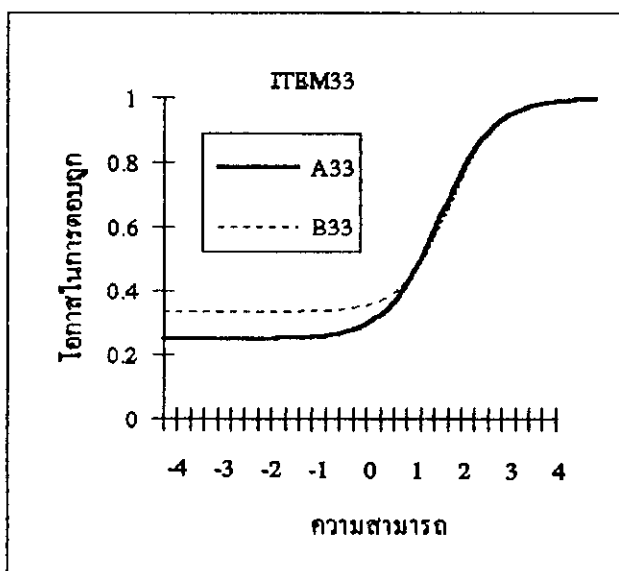
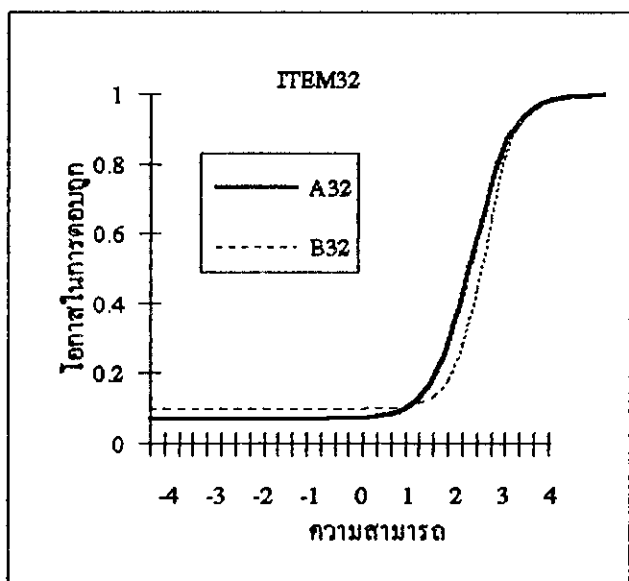
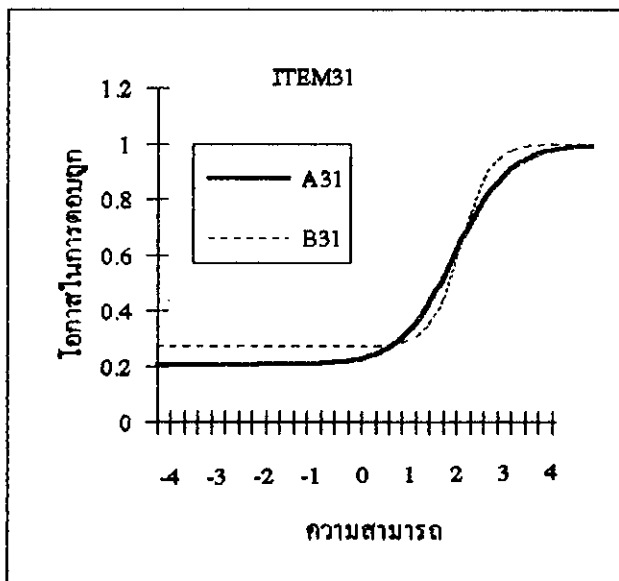
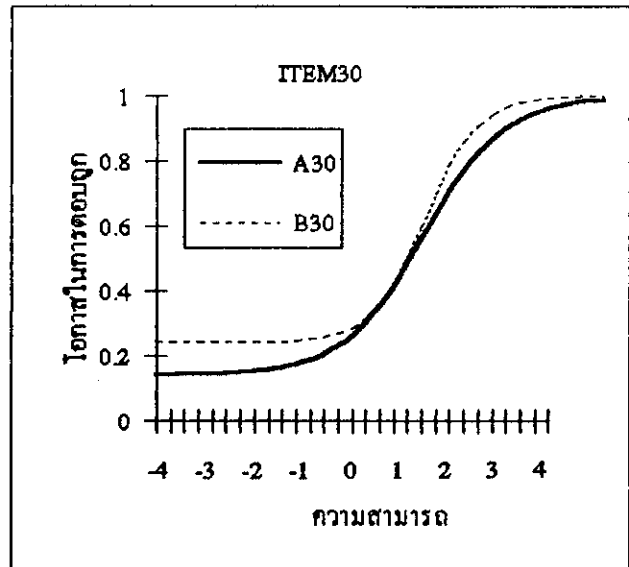
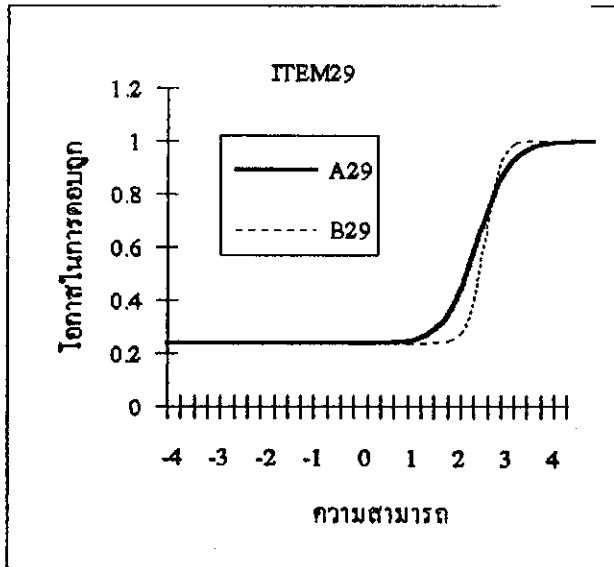
กับกลุ่มนักเรียนที่นับถือศาสนาอิสลาม ของแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ จำนวน 40 ข้อ

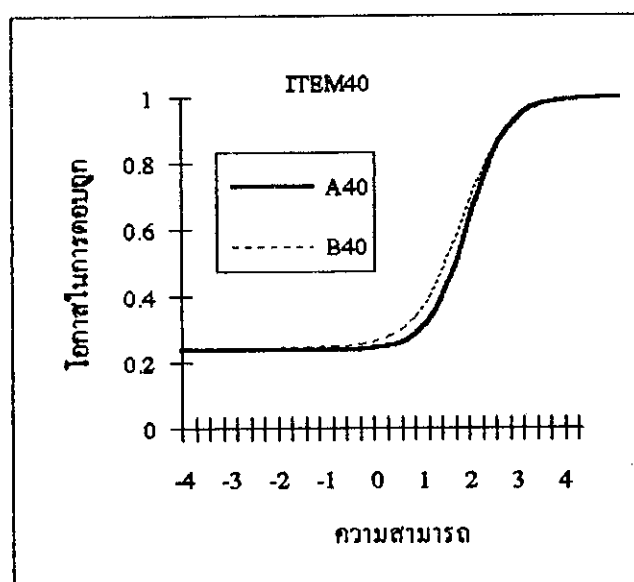
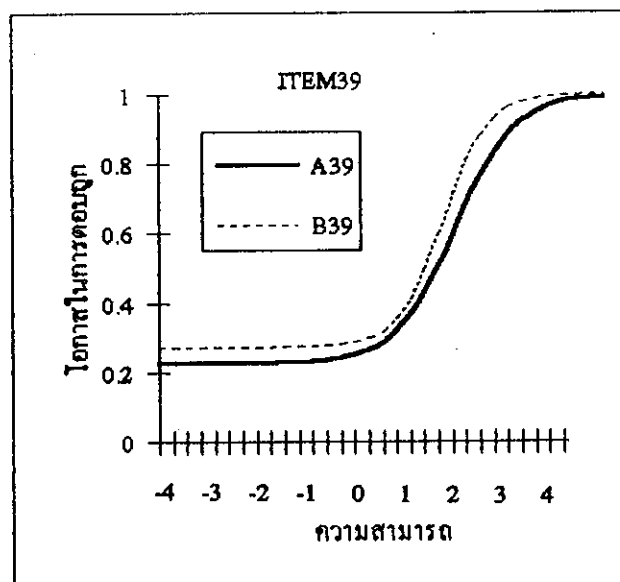
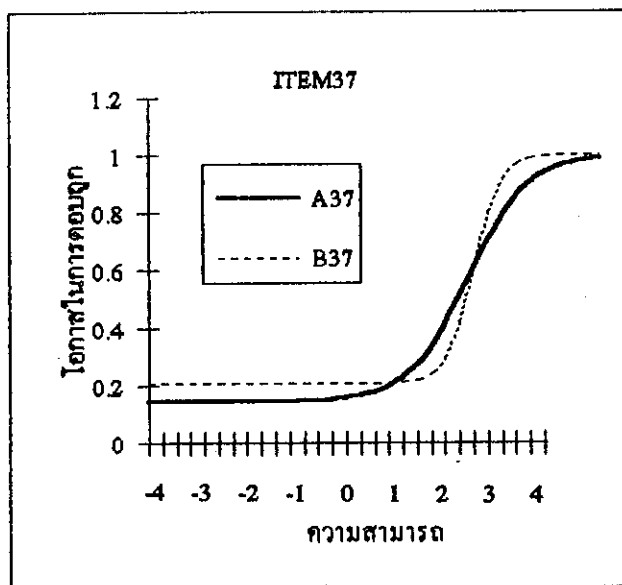
ข้อสอบที่มีความลำเอียงของแบบทดสอบใจหายปัญหาทางคณิตศาสตร์



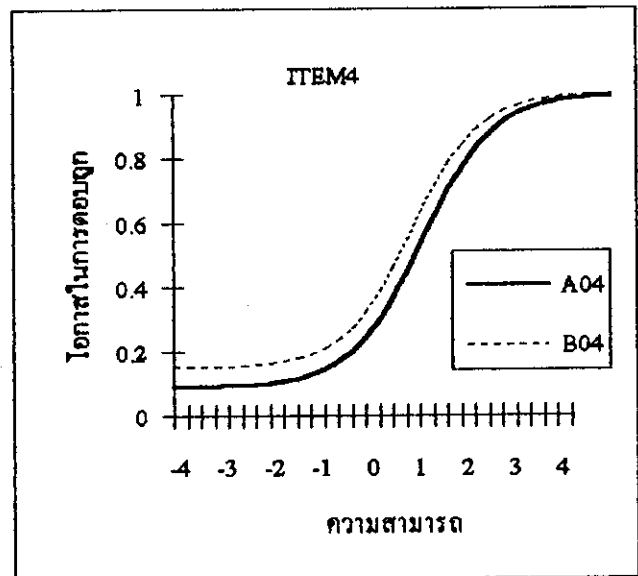
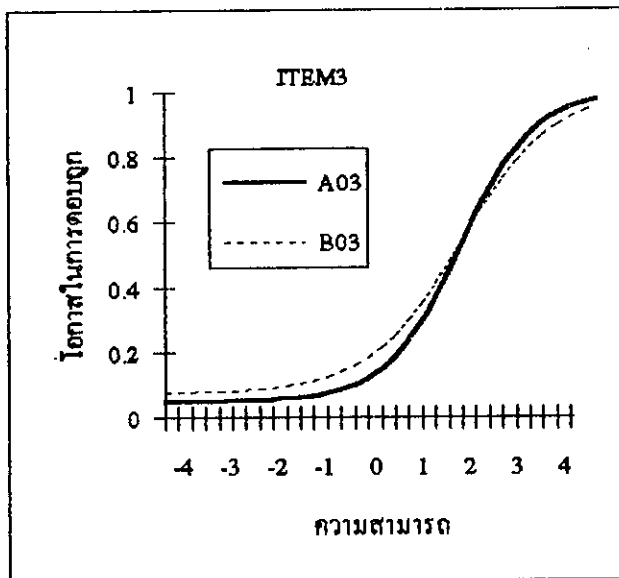
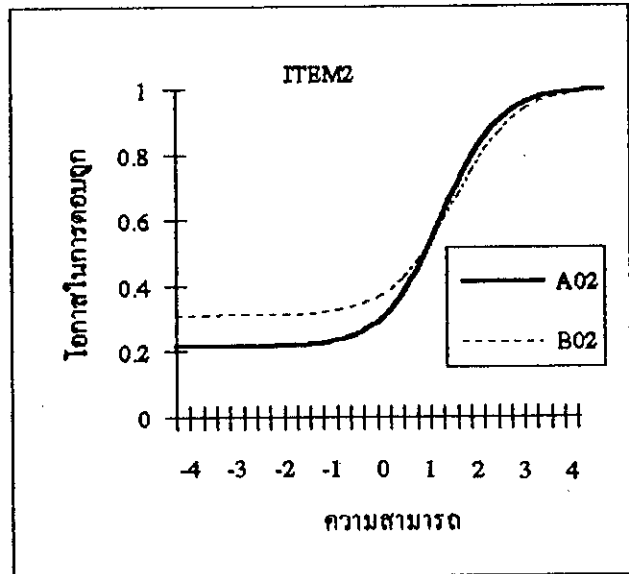
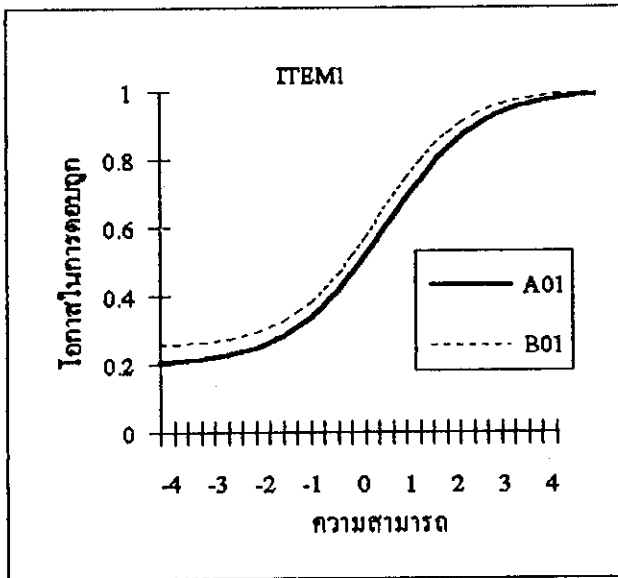
- A กลุ่มนักเรียนที่นับถือศาสนาอิสลาม
- - - B กลุ่มนักเรียนที่นับถือศาสนาพุทธ

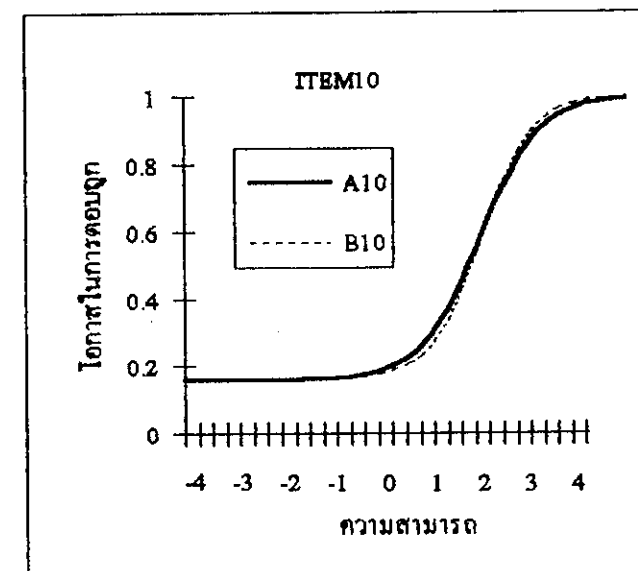
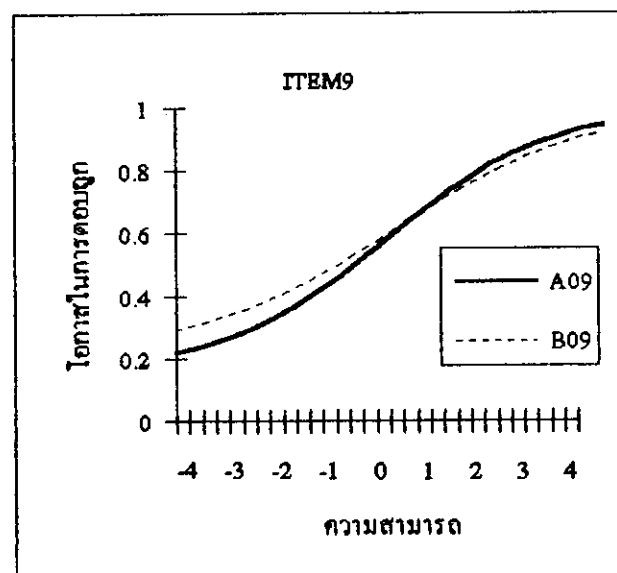
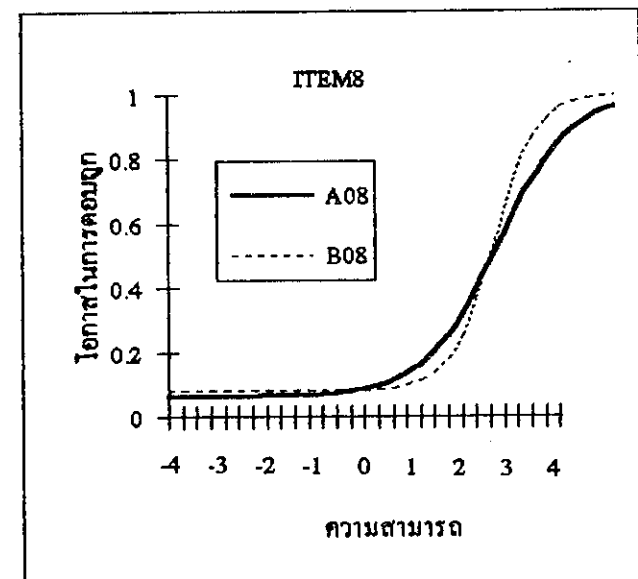
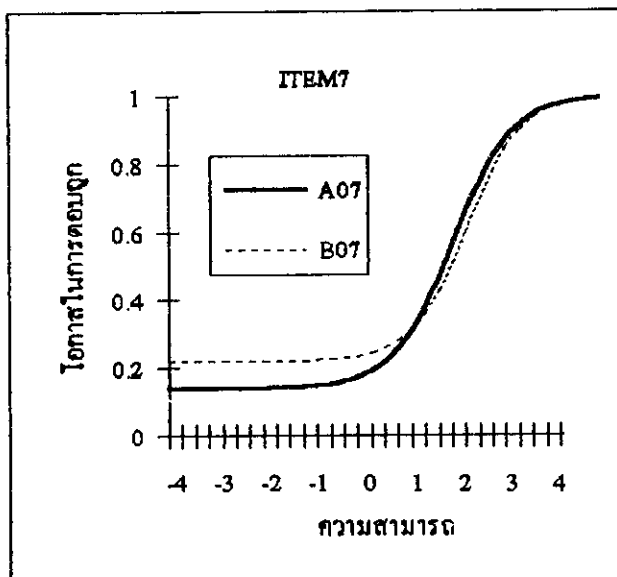
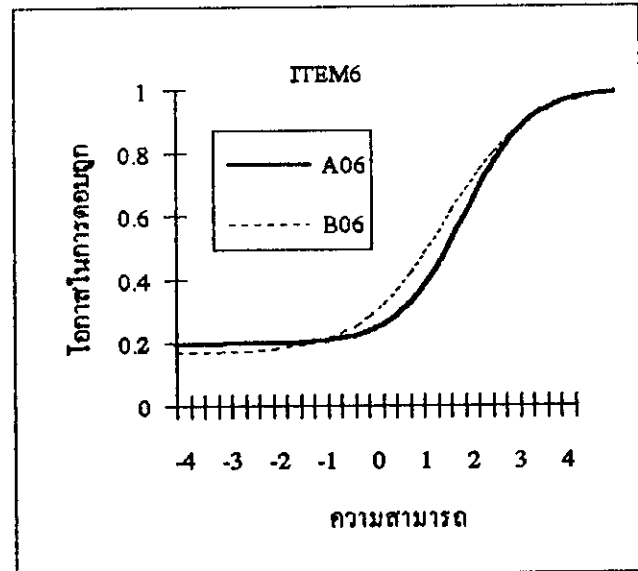
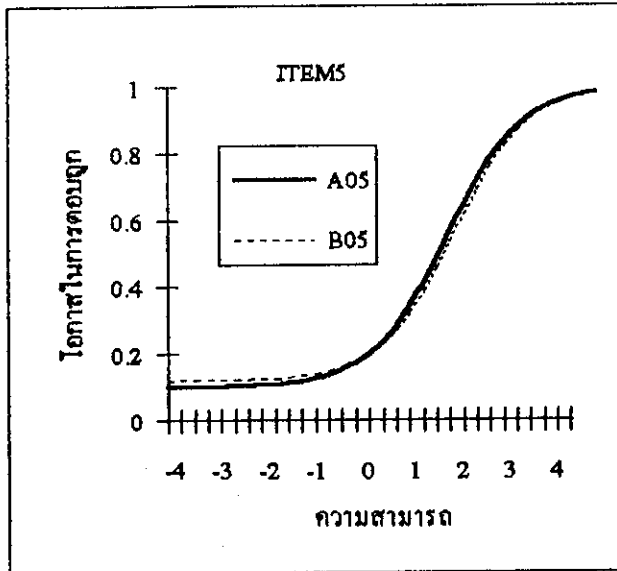


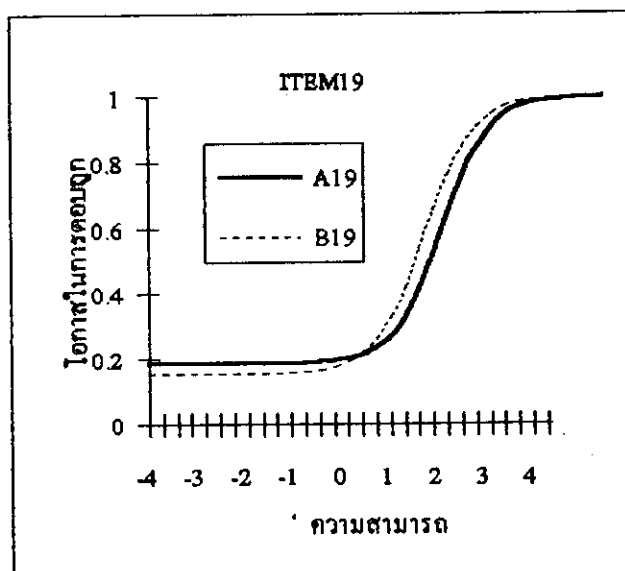
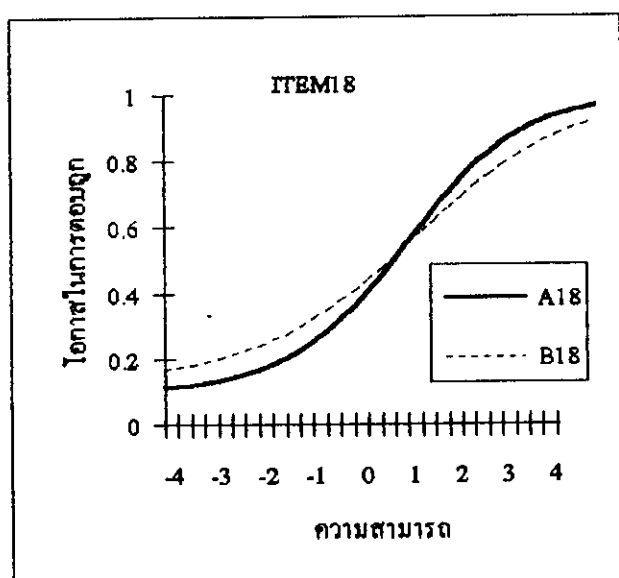
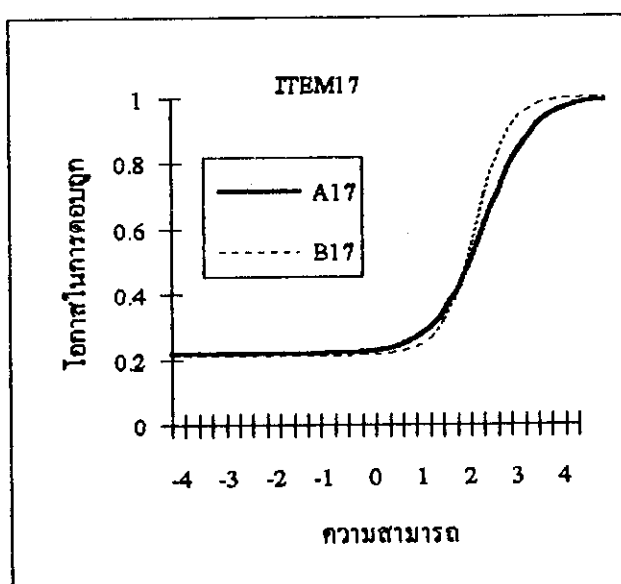
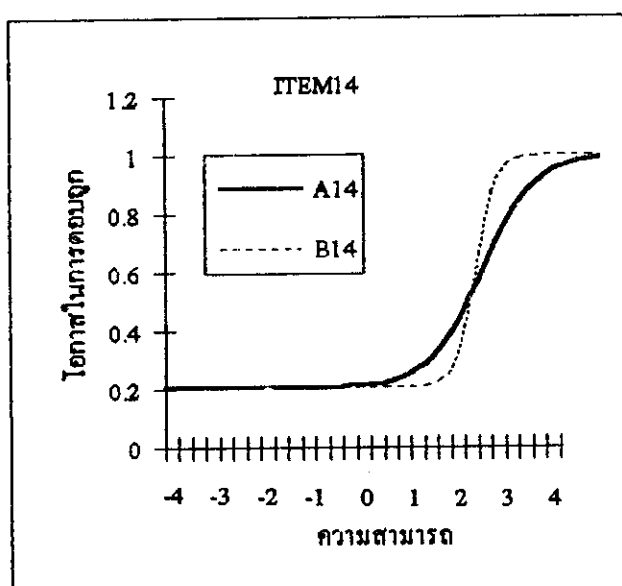
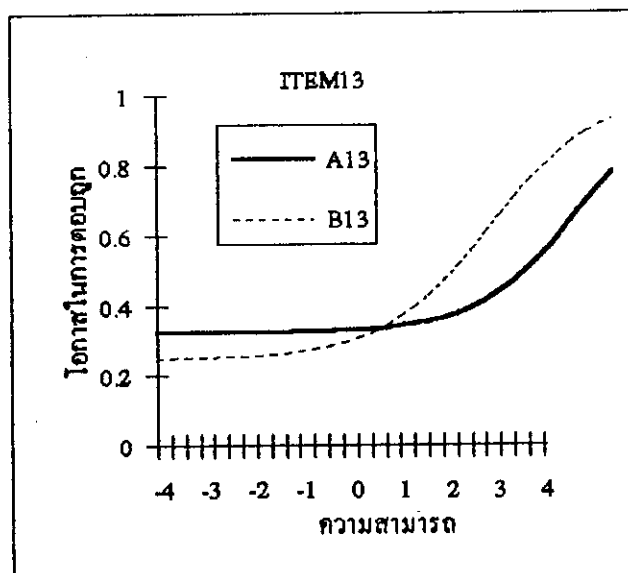
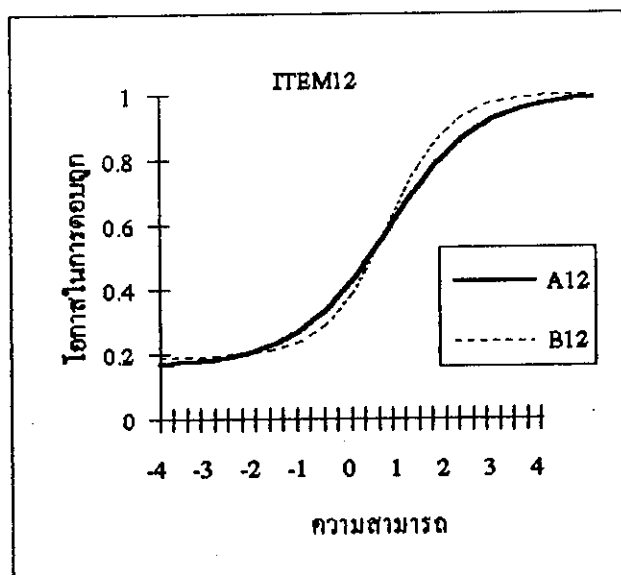


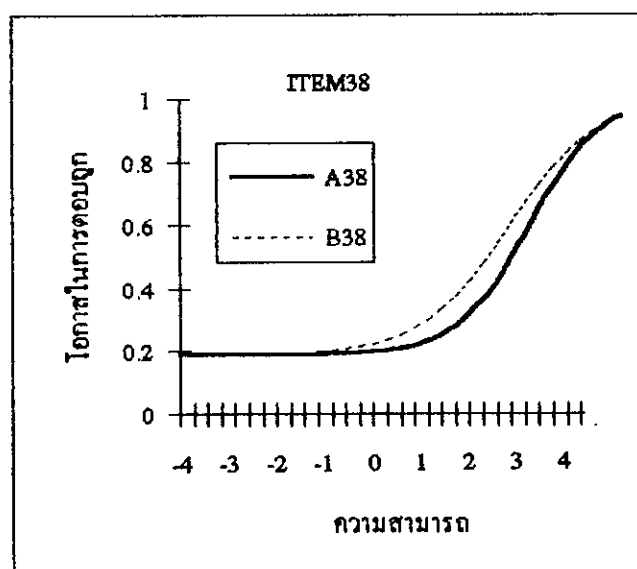
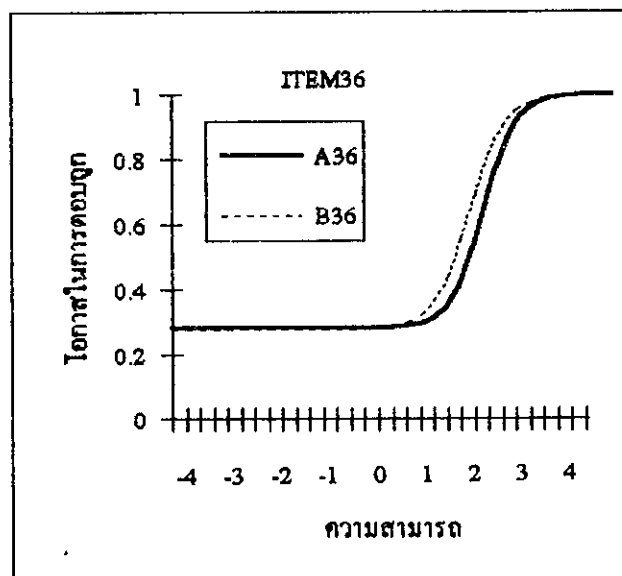
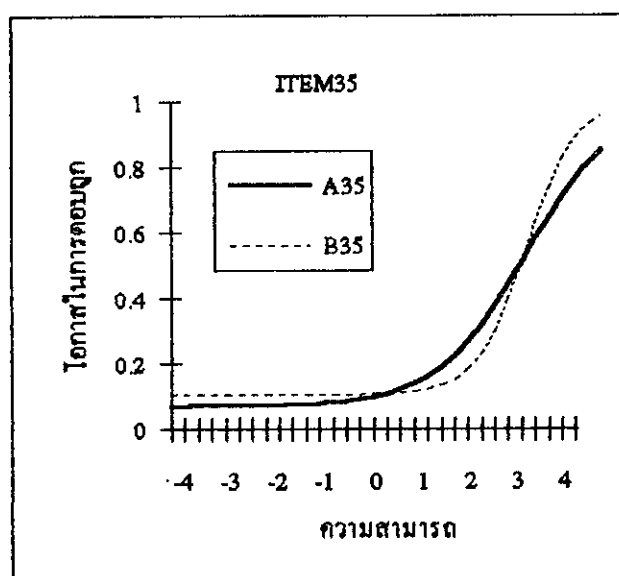
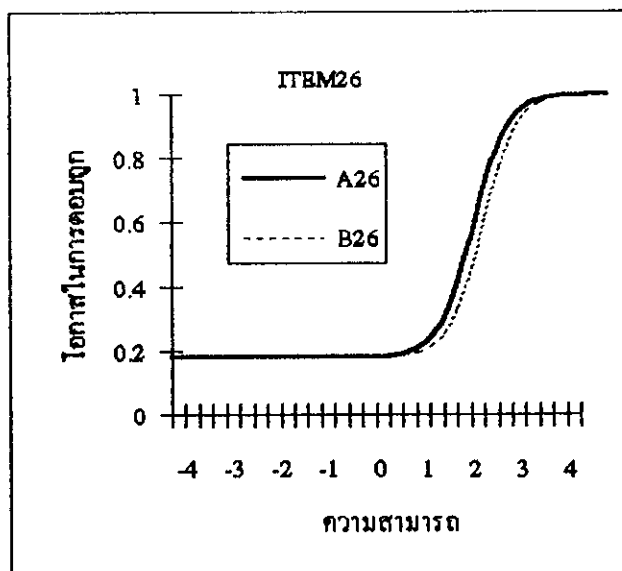
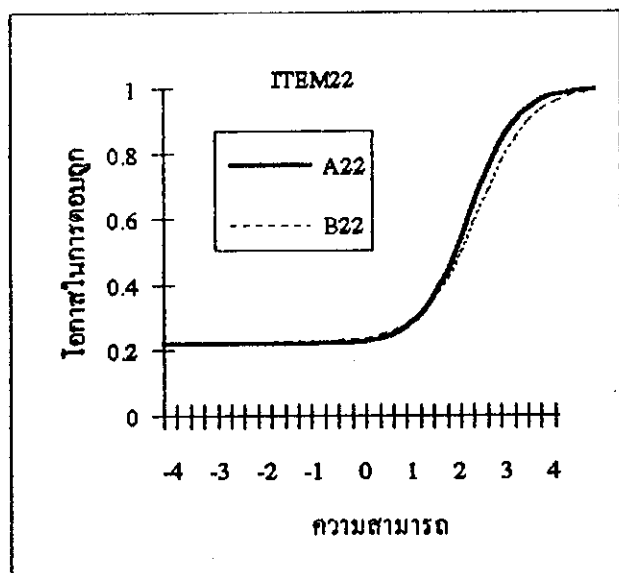


ข้อสอบที่ไม่มีความลำเอียงของแบบทดสอบโดยปัญหาคณิตศาสตร์





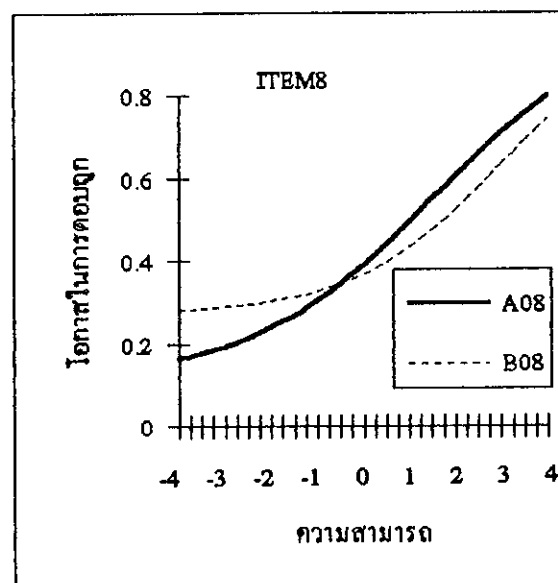
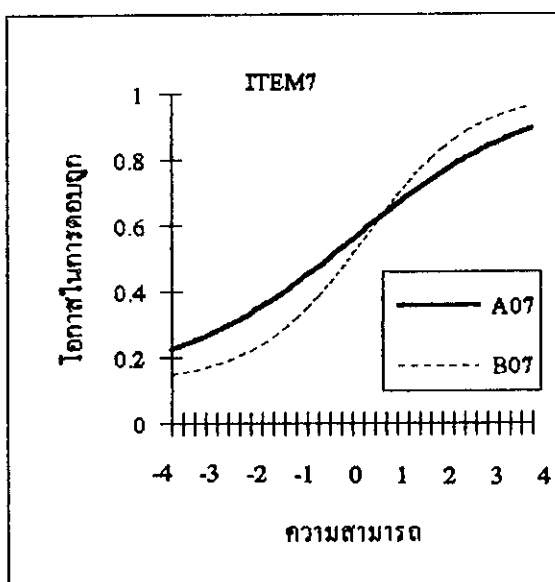
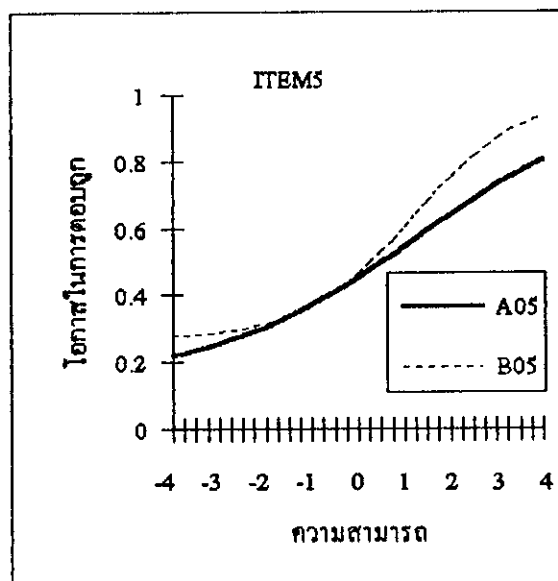
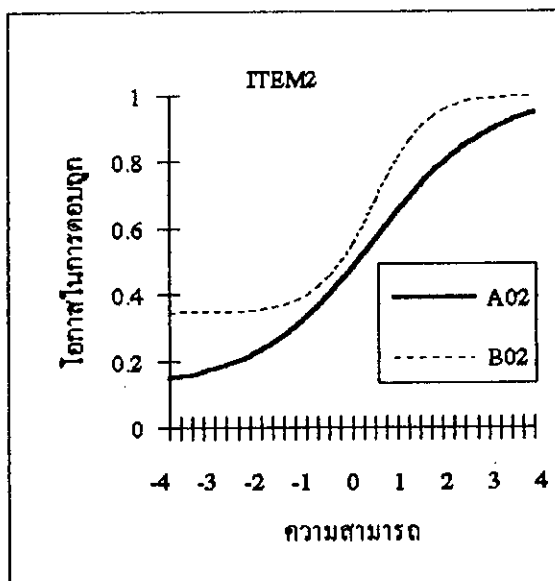




ภาคผนวก ข

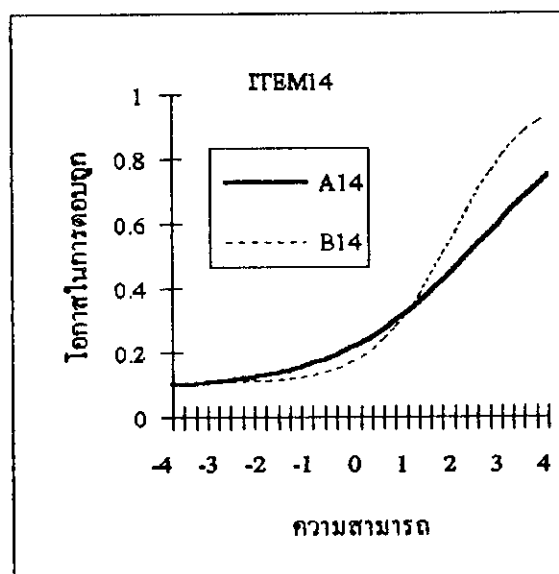
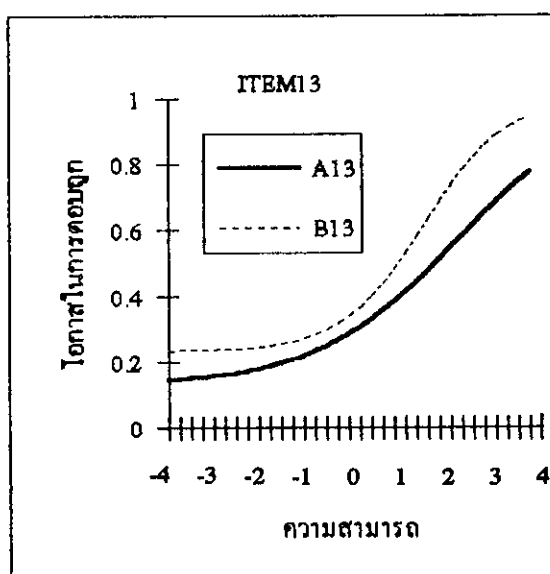
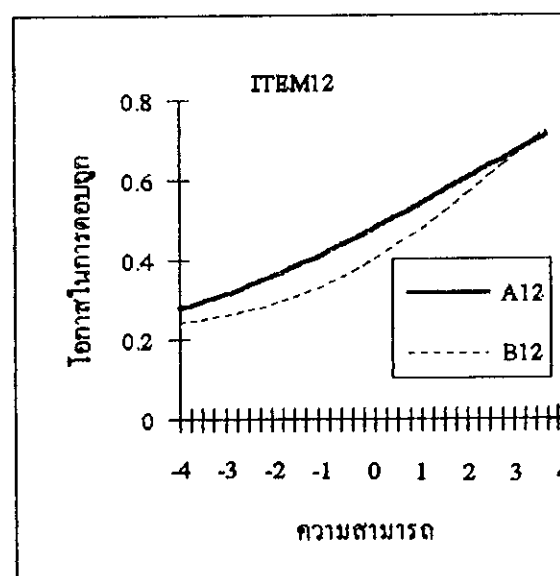
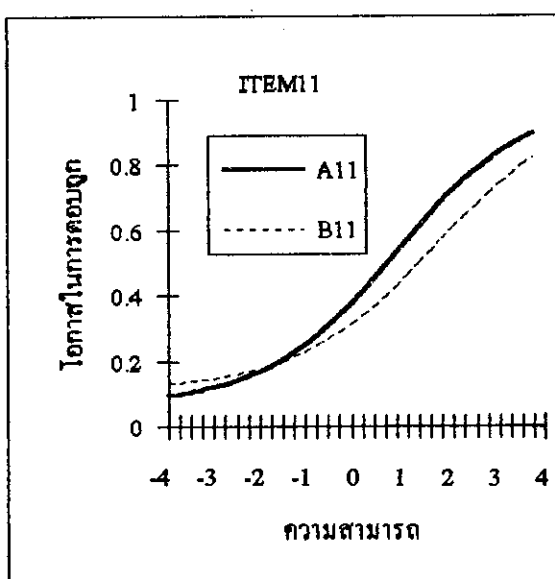
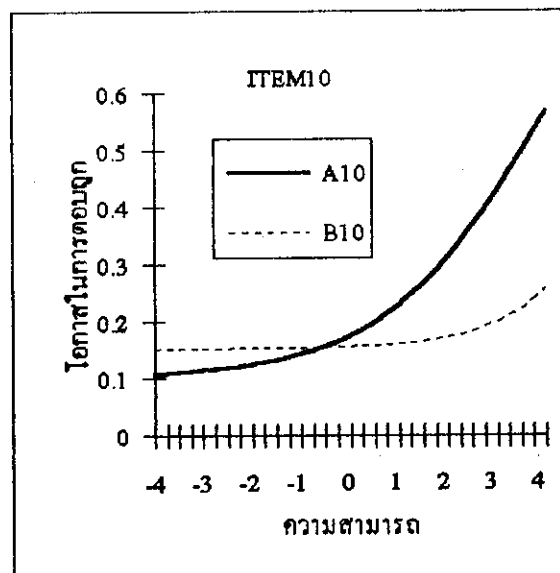
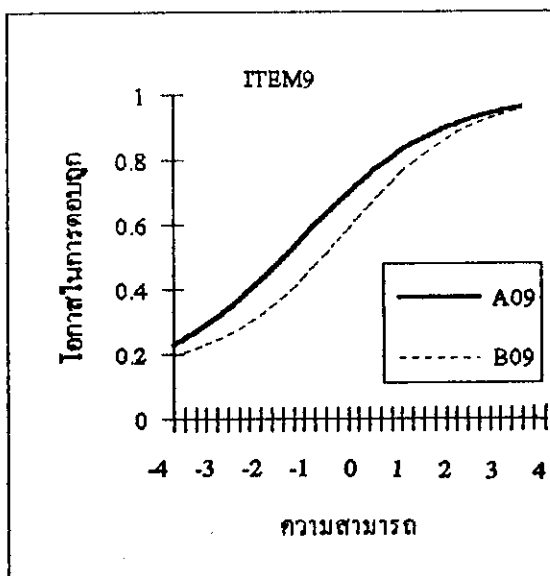
**โด่งลักษณะข้อสอบรายข้อระหว่างกลุ่มนักเรียนที่นับถือศาสนาพุทธ
กับกลุ่มนักเรียนที่นับถือศาสนาอิสลามของแบบทดสอบความเข้าใจทางภาษา จำนวน 40 ข้อ**

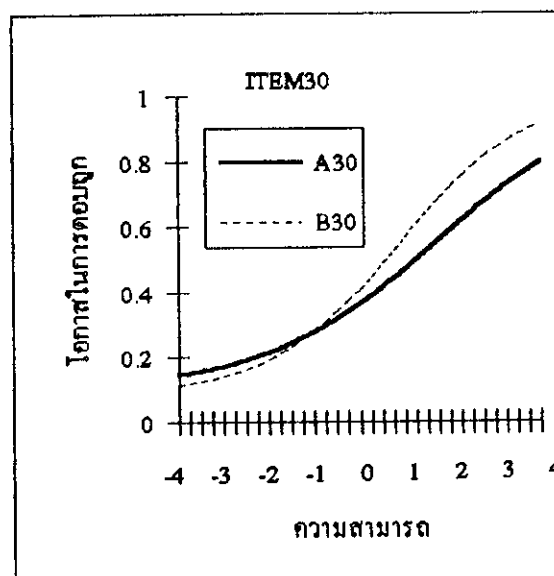
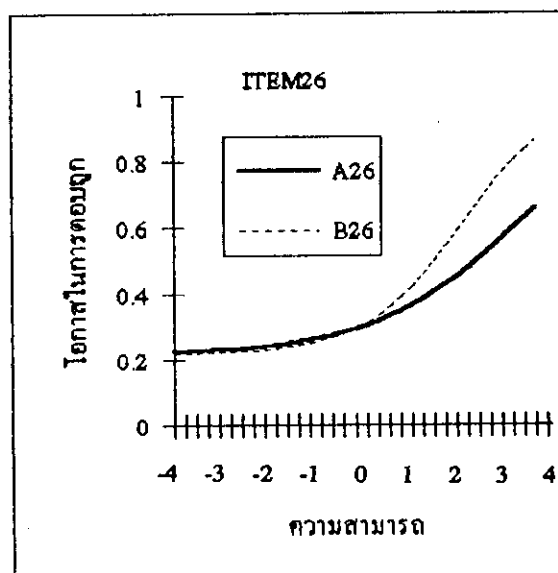
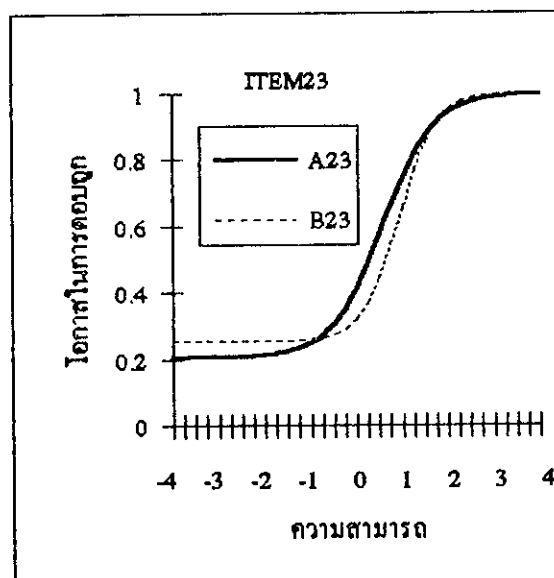
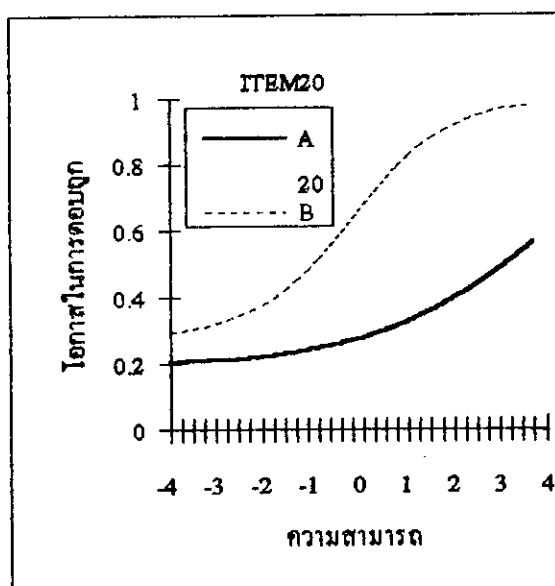
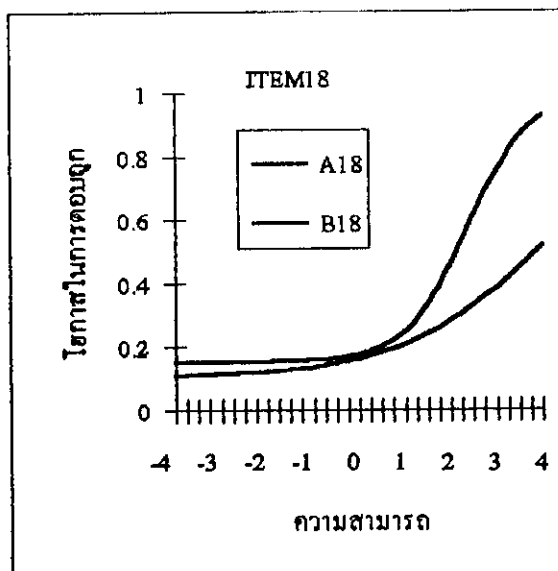
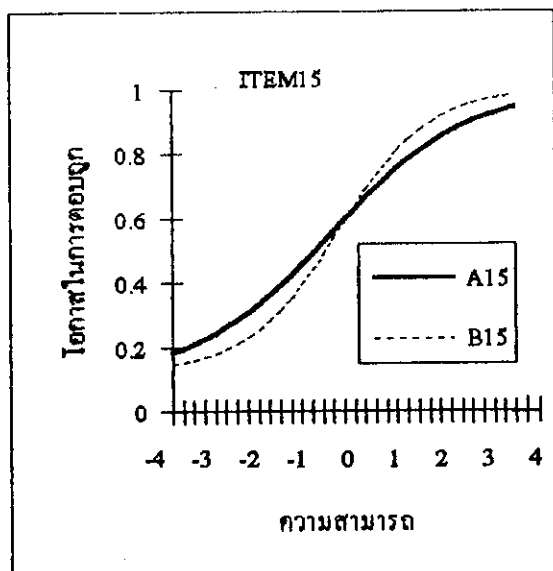
ข้อสอบที่มีความลำเอียงของแบบทดสอบวัดความเข้าใจทางภาษา



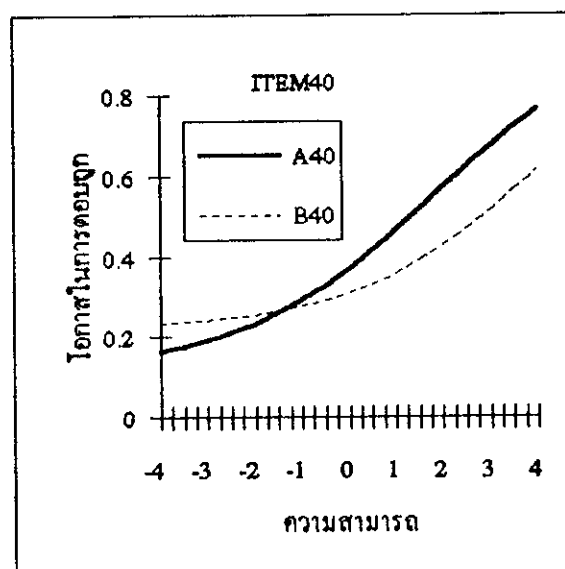
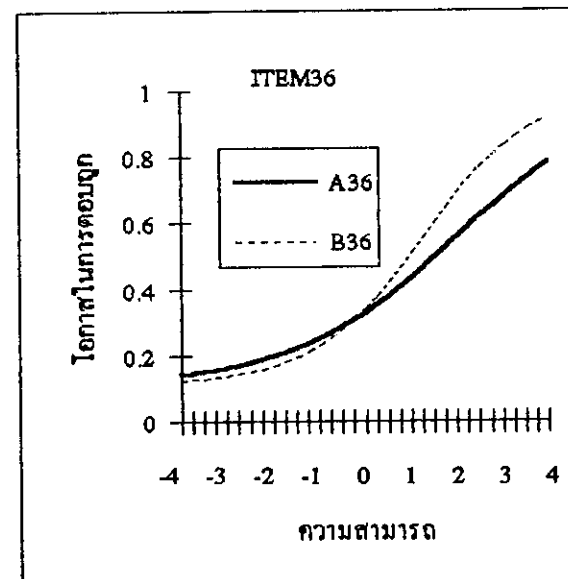
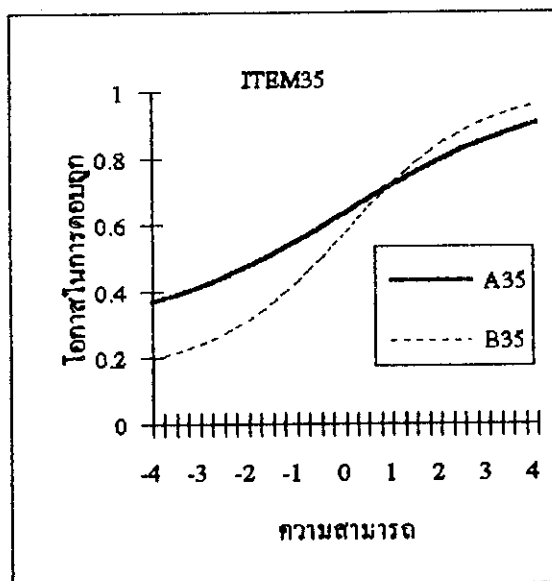
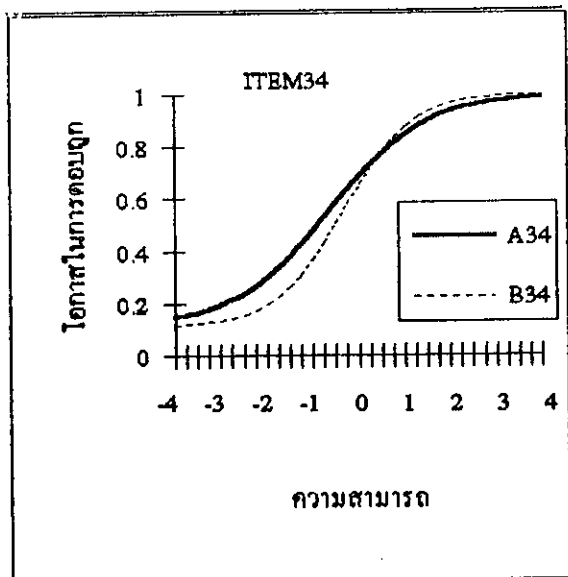
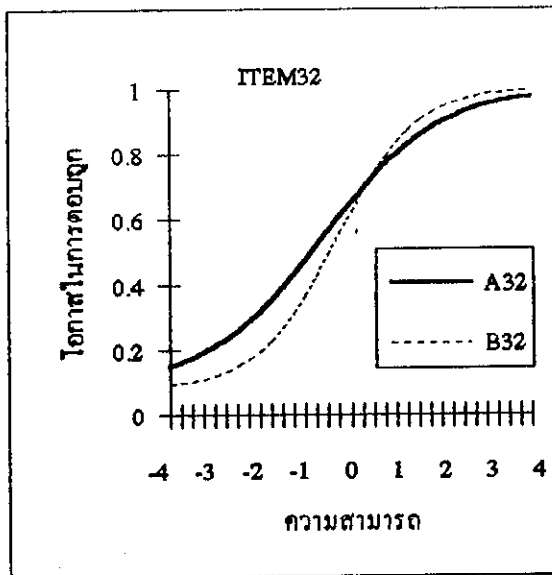
— A กลุ่มนักเรียนที่นับถือศาสนาอิสลาม
- - B กลุ่มนักเรียนที่นับถือศาสนาพุทธ

Sheet2

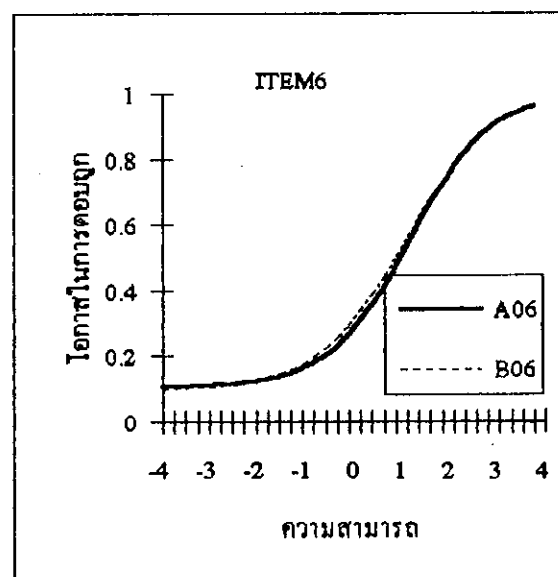
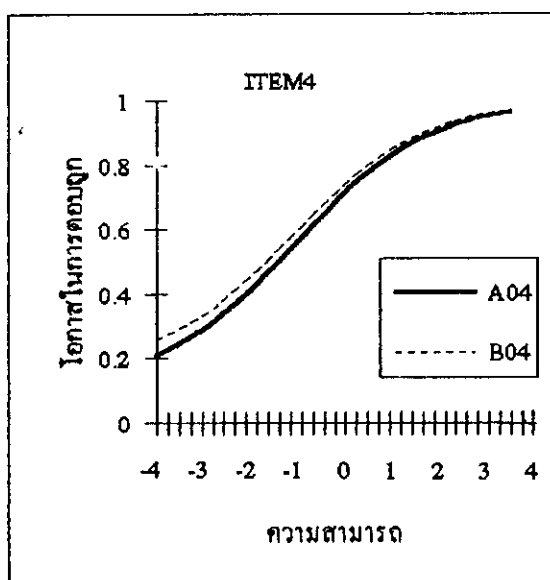
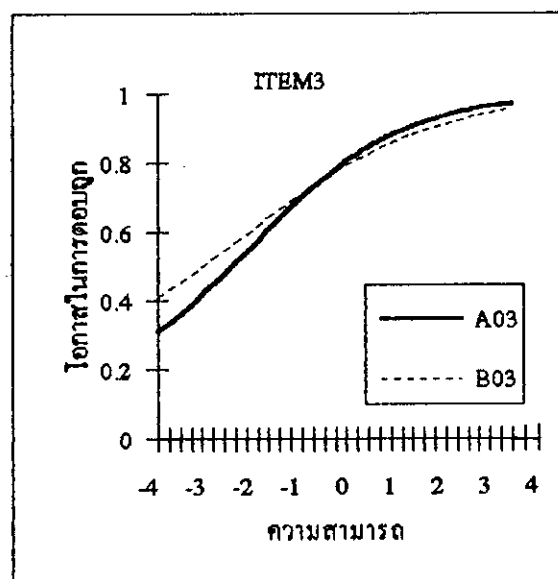
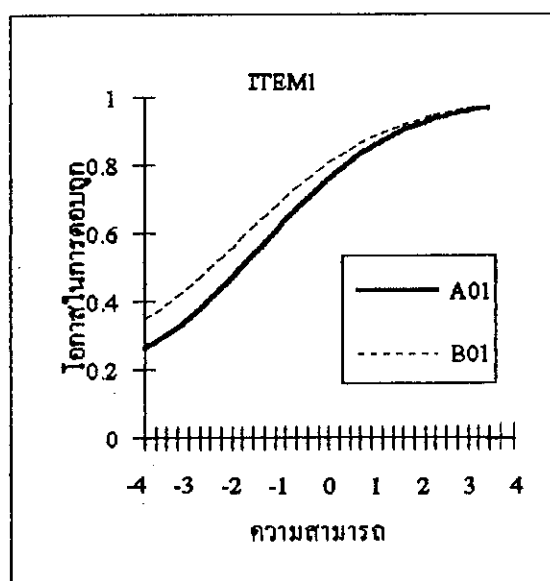


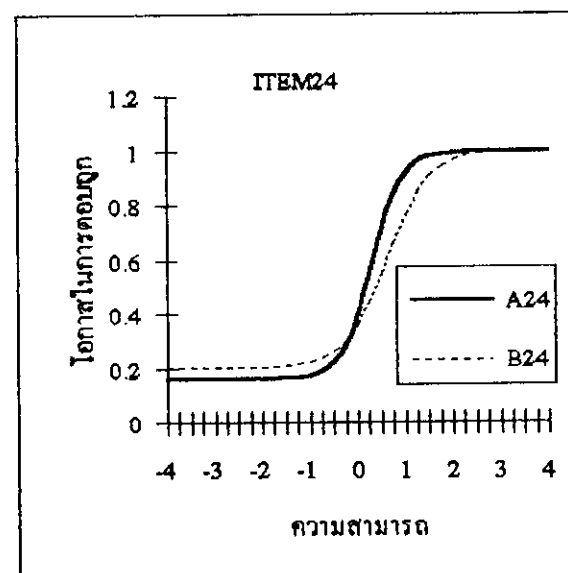
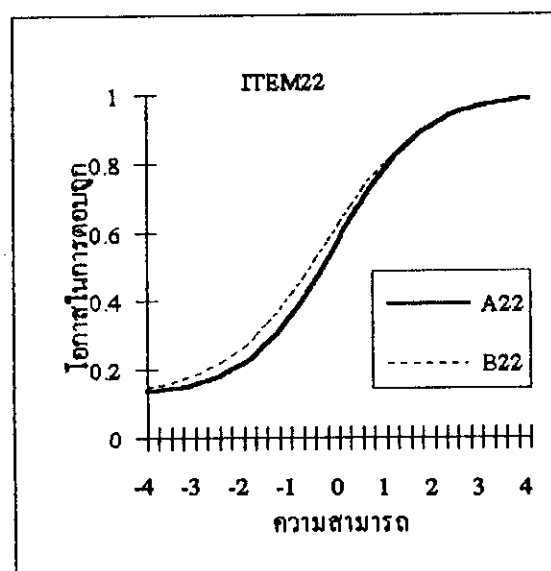
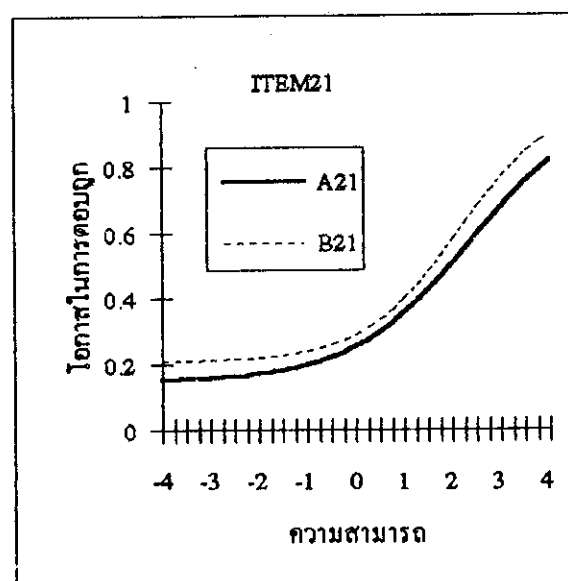
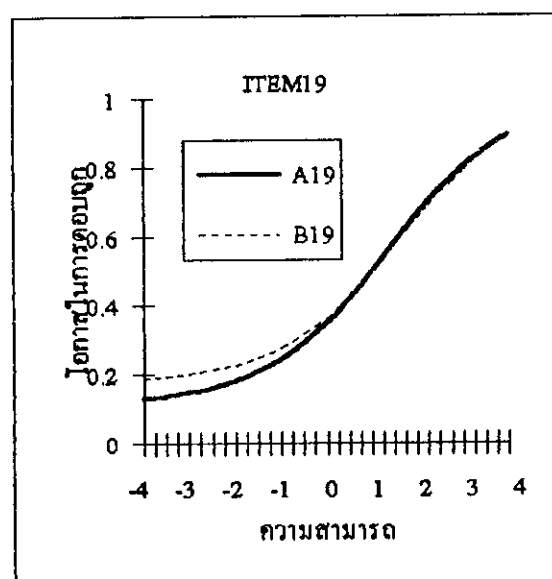
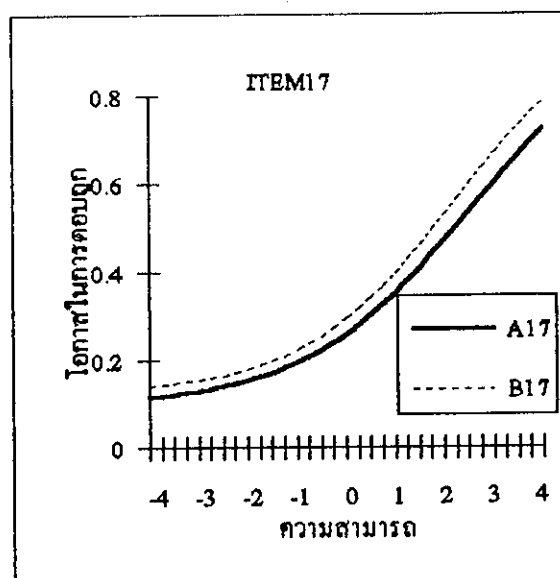
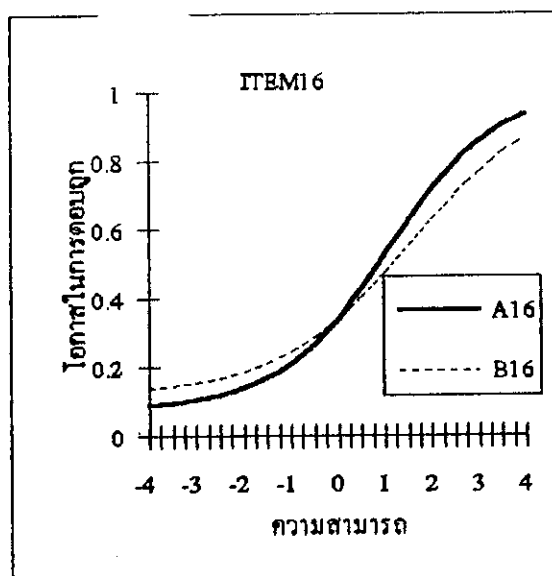


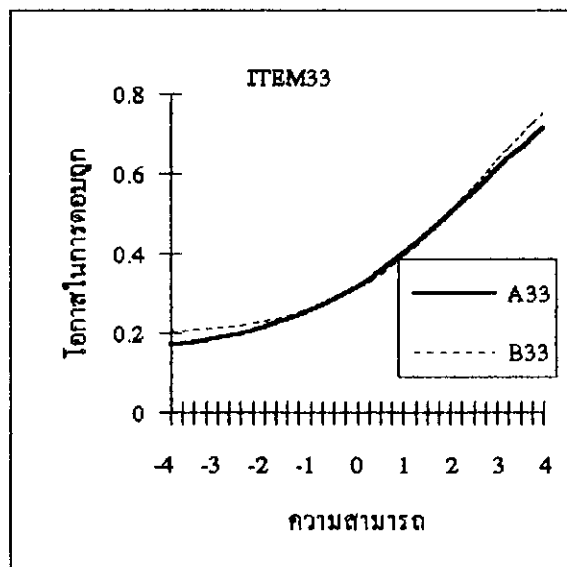
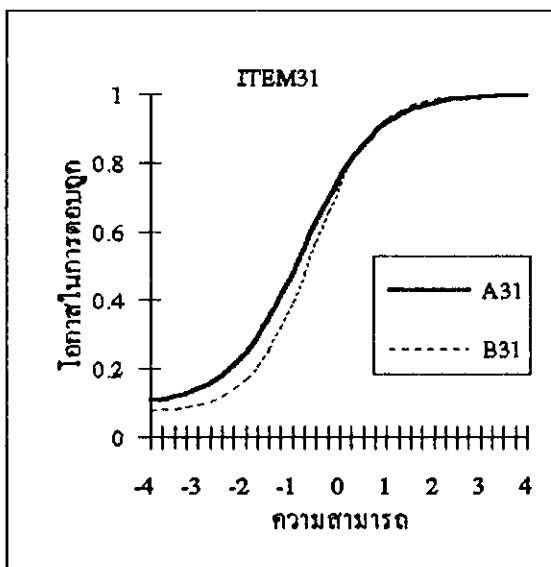
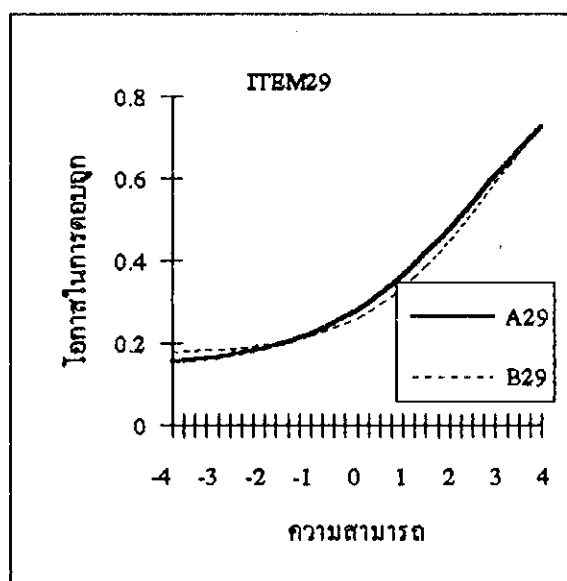
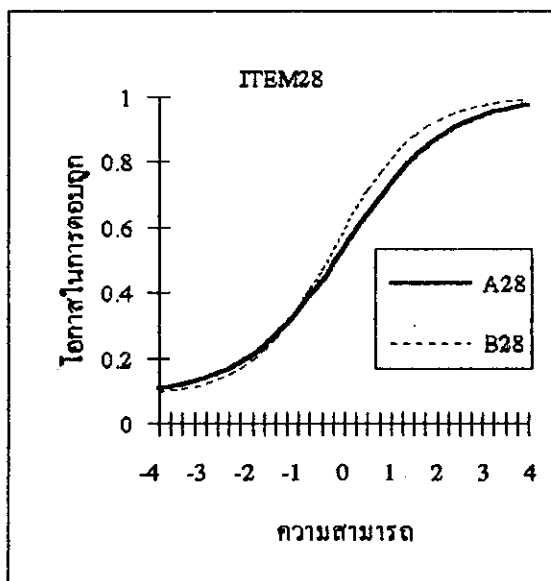
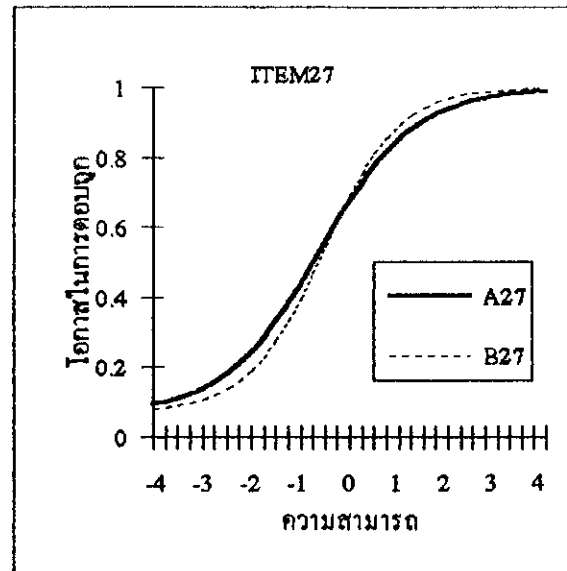
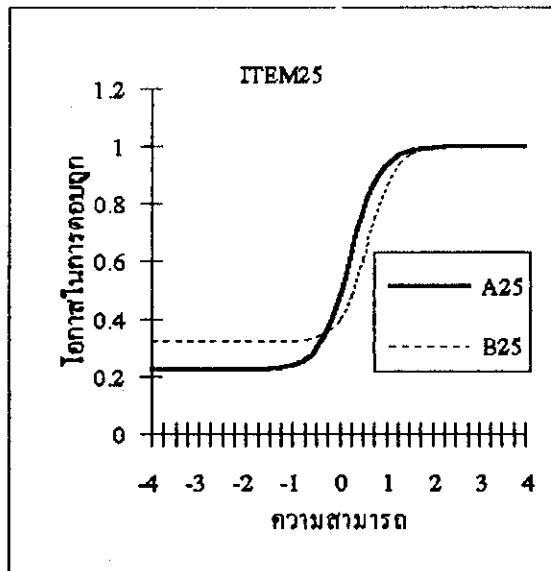
Sheet2

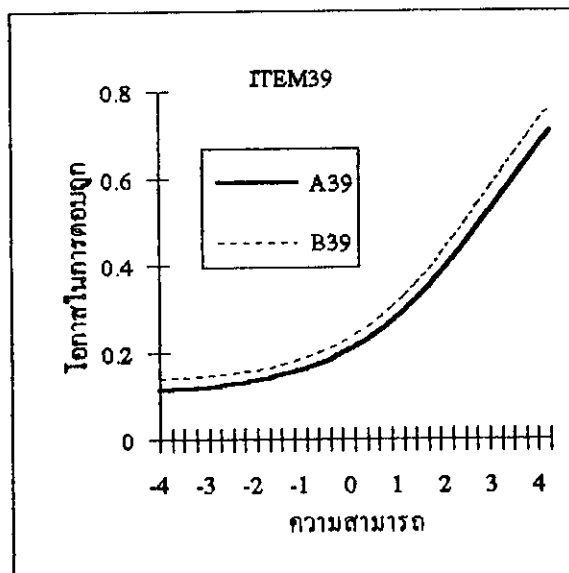
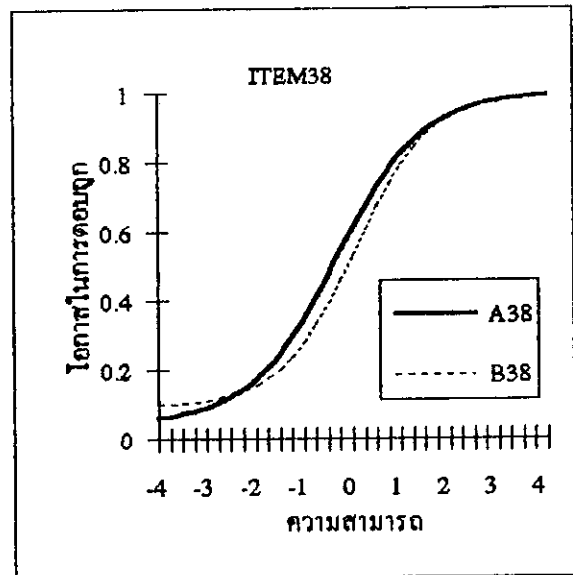
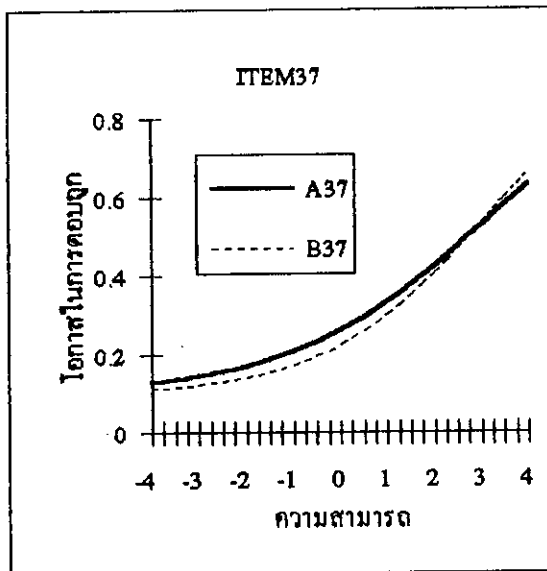


ข้อสอบที่ไม่มีความลำเอียงของแบบทดสอบวัดความเข้าใจทางภาษา









ประวัติย่อของผู้วิจัย

ชื่อ	นางสาวจิตสุดา	ชื่อสกุล	ธราพร
เกิดวันที่	11 เดือน กุมภาพันธ์	พุทธศักราช	2508
สถานที่เกิด		อำเภอฉวาง	จังหวัดนครราชสีมา
สถานที่อยู่ปัจจุบัน		บ้านเลขที่ 36 หมู่ที่ 5 ตำบลละอาย อำเภอฉวาง	จังหวัดนครราชสีมา
ตำแหน่งหน้าที่การงาน		อาจารย์ 1	ระดับ 4
สถานที่ทำงานปัจจุบัน		โรงเรียนบ้านคลองแห้ง ตำบลอ่าวนาง อำเภอเมือง	จังหวัดกระบี่ 81000

ประวัติการศึกษา

พ.ศ. 2525	มัธยมศึกษาตอนปลาย (วิทย์-คณิต)
	จากโรงเรียนฉวางรัชดาภิเษก จังหวัดนครราชสีมา
พ.ศ. 2529	ค.บ. (วิชาเอกคณิตศาสตร์)
	จากวิทยาลัยครูนครราชสีมา
พ.ศ. 2539	กศ.ม. (การวัดผลการศึกษา)
	จากมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

การเปรียบเทียบการวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบ
โจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา

บทคัดย่อ

ของ

จิตสุภา ขวาท

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา

สิงหาคม 2539

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบโจทก์ปัญหาทางคณิตศาสตร์ และแบบทดสอบความเข้าใจทางภาษาของนักเรียนที่นับถือศาสนาพุทธและศาสนาอิสลามโดยใช้วิธีวิเคราะห์ความลำเอียงต่างกัน 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีโด่งลักษณะข้อสอบและวิธีของแมนเทล-แฮนส์เชล เพื่อเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียง และ เปรียบเทียบค่าความเชื่อมั่นที่คำนวณโดยวิธีสูตรแบบ KR-20 และวิธีสูตรแบบแบ่งครึ่งฉบับ ของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออก กลุ่มตัวอย่างที่ใช้ศึกษาเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษาในจังหวัดกระบี่ จำนวน 2,689 คน ซึ่งเลือกมาโดยการสุ่มอย่างง่าย

ผลการวิจัย พบว่า วิธีวิเคราะห์ความลำเอียงของข้อสอบ จากแบบทดสอบโจทก์ปัญหาทางคณิตศาสตร์และแบบทดสอบความเข้าใจทางภาษา แต่ละวิธีพบจำนวนข้อสอบที่มีความลำเอียงแตกต่างกัน ซึ่งวิธีของแมนเทล-แฮนส์เชล พบจำนวนข้อสอบที่มีความลำเอียงมากที่สุด และวิธีแปลงค่าความยากของข้อสอบพบจำนวนข้อสอบที่มีความลำเอียงน้อยที่สุด โดยที่การวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากของข้อสอบกับวิธีโด่งลักษณะข้อสอบ และวิธีแปลงค่าความยากของข้อสอบกับวิธีของแมนเทล-แฮนส์เชล มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .001 ส่วนวิธีโด่งลักษณะข้อสอบกับวิธีของแมนเทล-แฮนส์เชล มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

สำหรับค่าความเชื่อมั่นของแบบทดสอบโจทก์ปัญหาทางคณิตศาสตร์ที่คำนวณโดยวิธีสูตร แบบ KR-20 หลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกจากผลการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากของข้อสอบกับวิธีโด่งลักษณะข้อสอบ และวิธีแปลงค่าความยากของข้อสอบ กับวิธีของแมนเทล-แฮนส์เชล มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และวิธีโด่งลักษณะข้อสอบกับวิธีของแมนเทล-แฮนส์เชล มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ส่วนค่าความเชื่อมั่นของแบบทดสอบที่คำนวณโดยวิธีสูตรแบบแบ่งครึ่งฉบับ หลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกจากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโด่งลักษณะข้อสอบ และวิธีของแมนเทล-แฮนส์เชล มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

ค่าความเชื่อมั่นของแบบทดสอบความเข้าใจทางภาษาที่คำนวณโดยวิธีสูตรแบบ KR-20 และแบบแบ่งครึ่งฉบับ หลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกจากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากของข้อสอบ วิธีโด่งลักษณะข้อสอบ และ วิธีของแมนเทล-แฮนส์เชล มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

THE COMPARISONS OF THE ITEM BIAS ANALYSES OF MATHEMATICAL
PROBLEM SOLVING AND LANGUAGE COMPREHENSION TESTS

AN ABSTRACT

BY

JITSUDA TARAPORN

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

August 1996

The purpose of this research was to compare the Mathematical Problem Solving and Language Comprehension item bias analysis with different three approaches; Transformed Item Difficulty (TID), Item Characteristic Curve Method (IRT) and Mantel-Haenzel Method (MH). The comparison of the item biasness with respect to the numbers of the biased items and KR-20 and Split-Half reliabilities of the test after the deletion of biased items. The sample of the study were 2,689 Mathayom Suksa 2 Buddhist and Itsalam Students selected by simple random sampling from Secondary School ; Krabi.

The results of the study show the different amount of biased items detected by each techniques. Mantel-Haenzel Method gave the largest number of biased items while Transformed item Difficulty (TID) gave the less. The different between Transformed Item Difficulty (TID) and Mantel-Haenzel Methods (MH), Transformed Item Difficulty (TID) and Item Characteristic Curve Method (IRT) were significant at .001, Item Characteristic Curve Method (IRT) and Mantel-Haenzel Method (MH) were not different at .05 .

The reliabilities coefficient of the test with out biased items by different method were significant different between Transformed Item Difficulty (TID) and Item Characteristic Curve Method (IRT), Transformed Item Difficulty (TID) and Mantel-Haenzel Method (MH). Item Characteristic Curve Method (IRT) and Mantel-Haenzel Method (MH) were insignificant, no different was found among three approaches in Split-Half reliability. The Language Comprehension test reliabilities were not significant.