

371. 26013

8767 W

8.3

ผลการตรวจสอบความลำเอียงข้อสอบต่อการศึกษาความเที่ยงตรงเชิงพยากรณ์
ของแบบทดสอบคัดเลือกที่คิดคะแนนต่างกัน

14 ส.ค. 2540

ปริญญาณพนธ์

ของ

เรวดี อินทสระระ

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาคุณวุฒิ สาขาการทดสอบและวัดผลการศึกษา

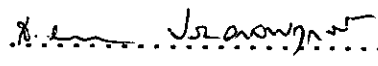
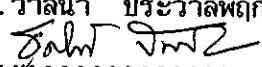
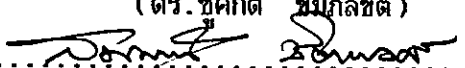
มิถุนายน 2539

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

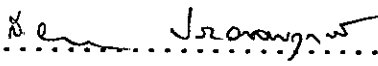
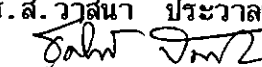
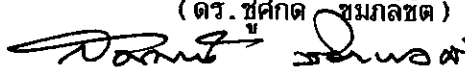
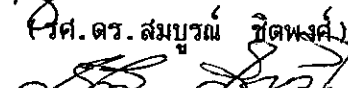
B. 52961

คณะกรรมการควบคุมและคณะกรรมการสอบได้พิจารณาปริญญานิพนธ์ฉบับนี้แล้ว เห็นสมควร
รับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาการศึกษาคุณวุฒิปบัณฑิต สาขาการทดสอบและวัดผล
การศึกษา ของมหาวิทยาลัยศรีนครินทรวิโรฒได้

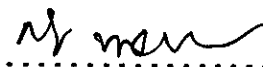
คณะกรรมการควบคุม

.......... ประธาน
(รศ.ดร.ส. วาสนา ประवालพฤษ์)
.......... กรรมการ
(ดร. ชัคกี้ ชุมลิต)
.......... กรรมการ
(รศ.ดร. สมบูรณ์ ชิตพงศ์)

คณะกรรมการสอบ

.......... ประธาน
(รศ.ดร.ส. วาสนา ประवालพฤษ์)
.......... กรรมการ
(ดร. ชัคกี้ ชุมลิต)
.......... กรรมการ
(รศ.ดร. สมบูรณ์ ชิตพงศ์)
.......... กรรมการที่แต่งตั้งเพิ่มเติม
(ผศ.ดร. สนั่น สิทธิวงษ์)

บัณฑิตวิทยาลัยอนุมัติให้รับปริญญานิพนธ์ฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษาคุณวุฒิปบัณฑิต สาขาการทดสอบและวัดผลการศึกษา ของมหาวิทยาลัยศรีนครินทรวิโรฒ

.......... คณบดีบัณฑิตวิทยาลัย
(ดร.ศิริยา พูลสุวรรณ)
วันที่ ๒๔ เดือน มิถุนายน พ.ศ. ๒๕๓๙

ผู้วิจัยสำนึกในพระมหากรุณาธิคุณเป็นล้นเกล้า
ล้นกระหม่อมหาที่สุดมิได้ ที่ได้พระราชทาน
"ทุนสิรินธร"
เพื่อสนับสนุนการศึกษาในระดับปริญญาเอก

ขอขอบพระคุณ
มหาวิทยาลัยศรีนครินทรวิโรฒ ภาคใต้
ที่พิจารณาให้
"ทุนพัฒนาอาจารย์"
เพื่อสนับสนุนในการทำปริญญานิพนธ์

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดีด้วยความกรุณาจาก
รองศาสตราจารย์ ดร.ส.วาสนา ประवालพุกษ์ ดร.ชูศักดิ์ ชัมภลิต
รองศาสตราจารย์ ดร.สมบูรณ์ ชิตพงศ์ และผู้ช่วยศาสตราจารย์ ดร.สนั่น สิทธิวัง
ที่ได้ให้คำปรึกษาและคำแนะนำ ตลอดจนแก้ไขข้อบกพร่องต่าง ๆ ผู้วิจัยกราบ
ขอบพระคุณเป็นอย่างสูง

ขอขอบพระคุณ ผู้ช่วยศาสตราจารย์ ดร.สุนทร ไส่ถิกพันธุ์
ผู้ช่วยศาสตราจารย์ ดร.สุจิตรา จรจิตร คุณพัชรี สุขสง คุณสัมพันธ์ ลัมเจริญชาติ
คุณประภารัตน์ ฉายศิริพันธ์ คุณสุพัชรีย์ อังคสุวรรณ คุณเยาวนาถ อุทัยมกุล และ
คุณกัลยา อนุพัฒน์ ที่ให้ความอนุเคราะห์ในการใช้ข้อมูล

กราบขอบพระคุณคณาจารย์ภาควิชาวัดผลและวิจัยการศึกษาและสำนักทดสอบ
ทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร ที่ได้ให้การ
อบรมสั่งสอนและอนุเคราะห์ข้อมูลในการวิจัย

ขอขอบคุณ ดร.กาญจนา วิธสุนทร ดร.วรรณดี แสงประทีปทอง และ
อาจารย์ณินี ณ นคร ที่ได้ให้ความอนุเคราะห์ในด้านโปรแกรมการวิเคราะห์ข้อมูล

ขอขอบคุณ อาจารย์อักษรประเสริฐ เศรษฐประเสริฐ อาจารย์นิคม ลัม
วชิรานันท์ ผู้ช่วยศาสตราจารย์ เสริม ทศศรี อาจารย์นันทินี มลิวัลย์ อาจารย์
ศิริพร ชนะสิทธิ์ อาจารย์อัจฉรินทร์ พันฤทธธำ อาจารย์วรวิรัตน์ สุขำพันธ์
อาจารย์สะโอด คีวนาง อาจารย์พรทิพย์ เพิ่มคำ อาจารย์ชะบิตะ เป็ญหมัด
อาจารย์อุษา ช่วยมั่ง อาจารย์วรัญญา แสงอุบล อาจารย์วรัญญา เหลือง
ธรรมชาติ อาจารย์ปัทมา ทองไสว อาจารย์อาภรณ์ กาญจนกิจโสภณ และ
อาจารย์จิตสุดา ธราพร ที่ได้ให้การช่วยเหลือในการเก็บรวบรวมข้อมูล

ขอขอบคุณ คุณสมกิจ อาจารย์จินดา กิจพูนวงศ์ อาจารย์บุญเรือง
อาจารย์รัตนา ศรีเหรียญ ที่ได้ให้ความช่วยเหลือในการวิเคราะห์ข้อมูล คุณศิริรัตน์
แสงเรืองรอง ที่ช่วยเหลือในการจัดพิมพ์ อาจารย์บังอร บัวเมือง อาจารย์
คมศักดิ์ ชื่นชม อาจารย์รุ่งทิพย์ ชื่นชม อาจารย์จ่านงค์ จันทฤกษ์ คุณฉัตรศิริ
ปิยะพิมลสิทธิ์ อาจารย์ระวี แสงพลบ รุ่งพี เพื่อน ๆ รุ่นน้องชาววัดผลการศึกษา
และ พ.ด.ด.สมนึก กระโหมวงศ์ ที่ได้ให้การช่วยเหลือทั้งกำลังใจและกำลังใจ
ด้วยดีเสมอมา

ขอขอบพระคุณผู้บริหารคณาจารย์และเจ้าหน้าที่มหาวิทยาลัยศรีนครินทรวิโรฒ
ภาคใต้ ผู้ช่วยศาสตราจารย์ ดร.บันลือ ถิ่นพังงา รองศาสตราจารย์ ประดิษฐ์
มีสุข ดร.วีระชัย ดร.พรทิพย์ เสมากดี ผู้ช่วยศาสตราจารย์ ดร.สนั่น ลีทธิวัง
และอาจารย์นรา บุรณรัช ที่ได้ให้การช่วยเหลือและสนับสนุนให้ผู้วิจัยได้มีโอกาส
ลาศึกษาต่อ ได้รับทุนอุดหนุนการศึกษาและทำวิจัย

กราบขอบพระคุณพ่อ แม่ ญาติพี่น้อง และครู อาจารย์ ที่ได้ให้การอบรม
สั่งสอน สนับสนุนการศึกษา และให้กำลังใจแก่ผู้วิจัยตลอดมา

คุณค่าและประโยชน์ใด ๆ ของปริญญานิพนธ์ฉบับนี้ขอมอบแด่ผู้มีพระคุณ
ดังกล่าวข้างต้น

เรวดี อินทสระระ

สารบัญ

บทที่	หน้า
1 บทนำ.....	1
ความเป็นมาและความสำคัญของปัญหา.....	1
จุดมุ่งหมายของการวิจัย.....	4
สมมติฐานในการวิจัย.....	4
ความสำคัญของการวิจัย.....	4
ขอบเขตของการวิจัย.....	5
นิยามศัพท์เฉพาะ.....	6
2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	7
ความเป็นมาของการศึกษาความลำเอียงของข้อสอบ.....	7
วิธีการตรวจสอบความลำเอียงของข้อสอบ.....	9
การตรวจสอบความลำเอียงของข้อสอบโดยใช้ทฤษฎีการตอบ ข้อคำถาม.....	9
การตรวจสอบความลำเอียงของข้อสอบตามวิธีของแมนเทล- แฮนส์เชล (MH).....	16
การตรวจสอบความลำเอียงของข้อสอบวิธีการทดสอบ SIB (SIBTEST).....	21
แหล่ง/สาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบ.....	26
วิธีการคิดคะแนน.....	27
งานวิจัยที่เกี่ยวข้องกับการตรวจสอบความลำเอียง.....	28
งานวิจัยที่เกี่ยวข้องกับวิธีการคิดคะแนน.....	76
3 วิธีดำเนินการวิจัย.....	79
แหล่งข้อมูลในการวิจัย.....	79
วิธีดำเนินการวิจัย.....	80
เครื่องมือที่ใช้ในการวิจัย.....	81
สถิติที่ใช้ในการวิจัย.....	82

4	ผลการวิเคราะห์ข้อมูล.....	83
	สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล.....	83
	การนำเสนอผลการวิเคราะห์ข้อมูล.....	84
	ผลการตรวจสอบความเป็นมิติเดียวของแบบทดสอบแต่ละวิชา.....	85
	ค่าสถิติพื้นฐานของผลการสอบและคุณภาพของแบบทดสอบแต่ละวิชา.....	89
	วิเคราะห์ความลำเอียงของข้อสอบแต่ละวิชา.....	145
	ผลการทดสอบความแตกต่างของจำนวนข้อสอบที่ลำเอียงจากการ วิเคราะห์ด้วย 3 วิธีของแบบทดสอบแต่ละวิชา.....	186
	ความสัมพันธ์ของผลการสอบที่ให้คะแนนที่-ปกติ และให้คะแนนน้ำหนัก ความสามารถที่ใช้ข้อสอบทั้งหมดกับใช้เฉพาะข้อสอบที่ปราศจาก ความลำเอียง.....	237
	ประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการเรียนของการให้คะแนน ที่-ปกติ และคะแนนน้ำหนักความสามารถกับการใช้จำนวนข้อสอบ ทั้งหมด กับเฉพาะข้อสอบที่ปราศจากความลำเอียง.....	240
5	สรุปผล อภิปรายผล และข้อเสนอแนะ.....	248
	จุดมุ่งหมายของการวิจัย.....	248
	สมมติฐานในการวิจัย.....	248
	วิธีดำเนินการวิจัย.....	248
	การวิเคราะห์ข้อมูล.....	249
	สรุปผลการวิจัย.....	250
	อภิปรายผล.....	251
	ข้อเสนอแนะในการวิจัย.....	253
	บรรณานุกรม.....	254
	ภาคผนวก.....	267
	ประวัติย่อของผู้วิจัย.....	275

บัญชีตาราง

ตาราง		หน้า
1	ข้อมูลผลการสอบของกลุ่มตัวอย่างกลุ่ม R และ F ที่มีคะแนนรวมอยู่ในช่วงคะแนน j	17
2	สัดส่วนการตอบข้อสอบของกลุ่มประชากร 2 กลุ่ม.....	17
3	จำนวนข้อมูลที่ใช้ในการศึกษาดัชนีความลำเอียงและการตัดสินผลการสอบ.....	79
4	ค่าไอเกนและร้อยละของความแปรปรวนขององค์ประกอบที่ 1-4 ของแบบทดสอบวิชาภาษาไทย ก สังคมศึกษา ก ภาษาอังกฤษ กข ภาษาไทย กข สังคมศึกษา กข และภาษาอังกฤษ กขค ทั้ง 12 ฉบับ	86
5	ค่าสถิติพื้นฐานของผลการสอบของแบบทดสอบแต่ละวิชา.....	90
6	ค่าพารามิเตอร์ของแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 1.....	93
7	ค่าพารามิเตอร์ของแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 2.....	97
8	ค่าพารามิเตอร์ของแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 1.....	101
9	ค่าพารามิเตอร์ของแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 2.....	106
10	ค่าพารามิเตอร์ของแบบทดสอบวิชาภาษาอังกฤษ กข ฉบับที่ 1.....	111
11	ค่าพารามิเตอร์ของแบบทดสอบวิชาภาษาอังกฤษ กข ฉบับที่ 2.....	115
12	ค่าพารามิเตอร์ของแบบทดสอบวิชาภาษาไทย กข ฉบับที่ 1.....	119
13	ค่าพารามิเตอร์ของแบบทดสอบวิชาภาษาไทย กข ฉบับที่ 2.....	123
14	ค่าพารามิเตอร์ของแบบทดสอบวิชาสังคมศึกษา กข ฉบับที่ 1.....	127
15	ค่าพารามิเตอร์ของแบบทดสอบวิชาสังคมศึกษา กข ฉบับที่ 2.....	132
16	ค่าพารามิเตอร์ของแบบทดสอบวิชาภาษาอังกฤษ กขค ฉบับที่ 1.....	137
17	ค่าพารามิเตอร์ของแบบทดสอบวิชาภาษาอังกฤษ กขค ฉบับที่ 2.....	141
18	ค่าดัชนีความลำเอียงของทั้ง 3 วิธีของแบบทดสอบวิชา ภาษาไทย ก	145
19	ค่าดัชนีความลำเอียงของทั้ง 3 วิธีของแบบทดสอบวิชา สังคมศึกษา ก	148
20	ค่าดัชนีความลำเอียงของทั้ง 3 วิธีของแบบทดสอบวิชา ภาษาอังกฤษ กข.....	152
21	ค่าดัชนีความลำเอียงของทั้ง 3 วิธีของแบบทดสอบวิชา ภาษาไทย กข.....	154

[illegible]

36	แสดงค่า Q จากการทดสอบความแตกต่างของจำนวนข้อสอบที่ ไม่ลำเอียงทั้ง 3 วิธีของแบบทดสอบแต่ละวิชา.....	186
37	แสดงค่าไคสแควร์ในการทดสอบความแตกต่างของจำนวนข้อสอบ ที่ลำเอียงของการวิเคราะห์แต่ละวิธี.....	187
38	แสดงสาเหตุความลำเอียงของข้อสอบวิชาภาษาไทย ก.....	189
39	แสดงสาเหตุความลำเอียงของข้อสอบวิชาสังคมศึกษา ก.....	198
40	แสดงสาเหตุความลำเอียงของข้อสอบวิชาภาษาอังกฤษ กข.....	211
41	แสดงสาเหตุความลำเอียงของข้อสอบวิชาภาษาไทย กข.....	217
42	แสดงสาเหตุความลำเอียงของข้อสอบวิชาสังคมศึกษา กข.....	220
43	แสดงสาเหตุความลำเอียงของข้อสอบวิชาภาษาอังกฤษ กขค.....	229
44	สาเหตุอันดับ 1 ของความลำเอียงของข้อสอบระหว่างเพศชาย และเพศหญิง จำแนกตามรายวิชาและผู้ระบุสาเหตุ.....	235
45	แสดงค่าความสัมพันธ์ลำดับที่ของผลการสอบที่ให้คะแนน 2 วิธี และใช้ จำนวนข้อสอบทั้งหมด กับใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง ในผู้สอบสายวิทยาศาสตร์.....	237
46	แสดงค่าความสัมพันธ์ลำดับที่ของผลการสอบที่คิดคะแนน 2 วิธี และใช้ จำนวนข้อสอบทั้งหมด กับใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง ในผู้สอบสายศิลปศาสตร์.....	239
47	ค่าสัมประสิทธิ์ตัวทำนาย ค่าความคลาดเคลื่อนมาตรฐานของ สัมประสิทธิ์ตัวทำนาย ค่าความคลาดเคลื่อนมาตรฐานของการทำนาย ค่าสัมประสิทธิ์สหสัมพันธ์หาคู่ ค่าสัดส่วนความแปรปรวนที่ทำนายได้ และค่าทดสอบสัมประสิทธิ์ตัวทำนายในการทำนายผลสัมฤทธิ์ทางการ เรียนด้วยคะแนนสอบคัดเลือก 3 วิชา ที่คิดคะแนนแบบทวิ-ปกติ และ ใช้จำนวนข้อสอบทั้งหมด.....	240
48	ค่าสัมประสิทธิ์ตัวทำนาย ค่าความคลาดเคลื่อนมาตรฐานของ สัมประสิทธิ์ตัวทำนาย ค่าความคลาดเคลื่อนมาตรฐานของการทำนาย ค่าสัมประสิทธิ์สหสัมพันธ์หาคู่ ค่าสัดส่วนความแปรปรวนที่ทำนายได้ และค่าทดสอบสัมประสิทธิ์ตัวทำนายในการทำนายผลสัมฤทธิ์ทางการ เรียนด้วยคะแนนสอบคัดเลือก 3 วิชา ที่คิดคะแนนแบบทวิ-ปกติ และ ใช้ข้อสอบเฉพาะที่ปราศจากความลำเอียง.....	242

- 49 คำสัมประสิทธิ์ตัวทำนายน ค่าความคลาดเคลื่อนมาตรฐานของ
สัมประสิทธิ์ตัวทำนายน ค่าความคลาดเคลื่อนมาตรฐานของการทำนายน
คำสัมประสิทธิ์สหสัมพันธ์พหุคูณ ค่าสัดส่วนความแปรปรวนที่ทำนายนได้
และค่าทดสอบสัมประสิทธิ์ตัวทำนายนในการทำนายนผลสัมฤทธิ์ทางการ
เรียนด้วยคะแนนสอบคัดเลือก 3 วิชา ที่คิดคะแนนน้ำหนักความ
สามารถ และใช้จำนวนข้อสอบทั้งหมด..... 244
- 50 คำสัมประสิทธิ์ตัวทำนายน ค่าความคลาดเคลื่อนมาตรฐานของ
สัมประสิทธิ์ตัวทำนายน ค่าความคลาดเคลื่อนมาตรฐานของการทำนายน
คำสัมประสิทธิ์สหสัมพันธ์พหุคูณ ค่าสัดส่วนความแปรปรวนที่ทำนายนได้
และค่าทดสอบสัมประสิทธิ์ตัวทำนายนในการทำนายนผลสัมฤทธิ์ทางการ
เรียนด้วยคะแนนสอบคัดเลือก 3 วิชา ที่คิดคะแนนน้ำหนักความ
สามารถ และใช้ข้อสอบเฉพาะที่ปราศจากความลำเอียง..... 245
- 51 คำสัมประสิทธิ์ตัวทำนายน ค่าความคลาดเคลื่อนมาตรฐานของ
สัมประสิทธิ์ตัวทำนายน ค่าความคลาดเคลื่อนมาตรฐานของการทำนายน
คำสัมประสิทธิ์สหสัมพันธ์พหุคูณ ค่าสัดส่วนความแปรปรวนที่ทำนายนได้
และค่าทดสอบสัมประสิทธิ์ตัวทำนายนที่ทำนายนผลสัมฤทธิ์ทางการเรียน
ในแต่ละสายที่คิดคะแนนและจำนวนข้อสอบที่ต่างกัน..... 246

บัญชีภาพประกอบ

ภาพประกอบ	หน้า
1 แสดงไค้ังการตอบข้อสอบที่มีความลำเอียง 3 ลักษณะสำหรับกลุ่ม A และ B.....	11
2 แสดงความแตกต่างระหว่างกลุ่มประชากรย่อยในความสามารถ เป้าหมาย (θ).....	24
3 แสดงค่าสถิติที่แบ่งตามระดับความสามารถตามแบบทดสอบย่อย.....	25
4 แสดงค่าไอเกนของ 4 องค์ประกอบของแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 1 และฉบับที่ 2.....	87
5 แสดงค่าไอเกนของ 4 องค์ประกอบของแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 1 และฉบับที่ 2.....	87
6 แสดงค่าไอเกนของ 4 องค์ประกอบของแบบทดสอบวิชาภาษาอังกฤษ กข ฉบับที่ 1 และฉบับที่ 2.....	88
7 แสดงค่าไอเกนของ 4 องค์ประกอบของแบบทดสอบวิชาภาษาไทย กข ฉบับที่ 1 และฉบับที่ 2.....	88
8 แสดงค่าไอเกนของ 4 องค์ประกอบของแบบทดสอบวิชาสังคมศึกษา กข ฉบับที่ 1 และฉบับที่ 2.....	89
9 แสดงค่าไอเกนของ 4 องค์ประกอบของแบบทดสอบวิชาภาษาอังกฤษ กขค ฉบับที่ 1 และฉบับที่ 2.....	89
10 ไค้ังลักษณะของแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 1.....	96
11 ไค้ังสารสนเทศของแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 1.....	96
12 ไค้ังลักษณะของแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 2.....	100
13 ไค้ังสารสนเทศของแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 2.....	100
14 ไค้ังลักษณะของแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 1.....	105
15 ไค้ังสารสนเทศของแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 1.....	105
16 ไค้ังลักษณะของแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 2.....	110
17 ไค้ังสารสนเทศของแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 2.....	110
18 ไค้ังลักษณะของแบบทดสอบวิชาภาษาอังกฤษ กข ฉบับที่ 1.....	114
19 ไค้ังสารสนเทศของแบบทดสอบวิชาภาษาอังกฤษ กข ฉบับที่ 1.....	114
20 ไค้ังลักษณะของแบบทดสอบวิชาภาษาอังกฤษ กข ฉบับที่ 2.....	118

21	ไค้งสารสนเทศของแบบทดสอบวิชาภาษาอังกฤษ กข ฉบับที่ 2.....	118
22	ไค้งลักษณะของแบบทดสอบวิชาภาษาไทย กข ฉบับที่ 1.....	122
23	ไค้งสารสนเทศของแบบทดสอบวิชาภาษาไทย กข ฉบับที่ 1.....	122
24	ไค้งลักษณะของแบบทดสอบวิชาภาษาไทย กข ฉบับที่ 2.....	126
25	ไค้งสารสนเทศของแบบทดสอบวิชาภาษาไทย กข ฉบับที่ 2.....	126
26	ไค้งลักษณะของแบบทดสอบวิชาสังคมศึกษา กข ฉบับที่ 1.....	131
27	ไค้งสารสนเทศของแบบทดสอบวิชาสังคมศึกษา กข ฉบับที่ 1.....	131
28	ไค้งลักษณะของแบบทดสอบวิชาสังคมศึกษา กข ฉบับที่ 2.....	136
29	ไค้งสารสนเทศของแบบทดสอบวิชาสังคมศึกษา กข ฉบับที่ 2.....	136
30	ไค้งลักษณะของแบบทดสอบวิชาภาษาอังกฤษ กขค ฉบับที่ 1.....	140
31	ไค้งสารสนเทศของแบบทดสอบวิชาภาษาอังกฤษ กขค ฉบับที่ 1.....	140
32	ไค้งลักษณะของแบบทดสอบวิชาภาษาอังกฤษ กขค ฉบับที่ 2.....	144
33	ไค้งสารสนเทศของแบบทดสอบวิชาภาษาอังกฤษ กขค ฉบับที่ 2.....	144

บทที่ 1

บทนำ

ความเป็นมาและความสำคัญของปัญหา

นโยบายการกระจายโอกาสทางการศึกษาในแผนพัฒนาการศึกษาแห่งชาติ ฉบับที่ 7 (พุทธศักราช 2535-2539) ให้เร่งกระจายโอกาสทางการศึกษาทุกระดับ และประเภทของการศึกษาอย่างกว้างขวางและเป็นธรรมแก่ผู้ด้อยโอกาสทางเศรษฐกิจ สังคมและกลุ่มชนต่างวัฒนธรรม ตลอดจนจัดการศึกษาเพื่อส่งเสริมผู้มีความสามารถ พิเศษ โดยเฉพาะในมาตรการข้อ 4 ที่ว่าให้ปรับปรุงและพัฒนารูปแบบการรับหรือ คัดเลือกเข้าศึกษาในระดับและประเภทการศึกษาต่าง ๆ โดยเฉพาะในระดับอุดม ศึกษาเพื่อกระจายโอกาสการเข้ารับการศึกษากับผู้ด้อยโอกาสทางเศรษฐกิจและสังคม อย่างเป็นธรรม (สำนักงานคณะกรรมการการศึกษาแห่งชาติ. 2534 : 5) ดังนั้น มหาวิทยาลัยสงขลานครินทร์ ซึ่งเป็นมหาวิทยาลัยที่ตั้งอยู่ในเขตพื้นที่ภาคใต้ มีภาระ หน้าที่จัดการศึกษาให้กับนักศึกษาและบุคคลทั่วไป และเพื่อให้การปฏิบัติการกิจสอดคล้อง กับนโยบายการกระจายโอกาสทางการศึกษาดตามมาตรการในข้อ 4 นั้น สถาบันการ ศึกษาได้รับอนุมัติให้สามารถรับสมัครสอบคัดเลือกบุคคลเข้าเป็นนักศึกษาในชั้นปีที่ 1 ประเภทรับตรง โดยเปิดรับเฉพาะนักเรียนที่กำลังศึกษาหรือสำเร็จการศึกษาจาก สถานศึกษาในเขตพื้นที่ 14 จังหวัดในภาคใต้ ซึ่งในแต่ละปีการศึกษา มหาวิทยาลัย ได้รับจัดสรรคัดเลือกเข้าศึกษาประมาณ 50 % ของจำนวนนักศึกษาที่จะเข้าศึกษาใน ชั้นปีที่ 1 การดำเนินการคัดเลือกนั้น ซึ่งมหาวิทยาลัยสามารถดำเนินการได้เองทั้งหมด ตั้งแต่การรับสมัครสอบ การจัดทำข้อสอบ การดำเนินการสอบ และการตัดสินผลการ สอบ เนื่องจากจำนวนนักเรียนที่สมัครสอบคัดเลือกประเภทรับตรงในแต่ละปีมีจำนวน มาก โดยอัตราส่วนที่รับเข้าศึกษาต่อจำนวนผู้สมัครมีถึง 1 ต่อ 8 สำหรับการสอบ คัดเลือกของมหาวิทยาลัยสงขลานครินทร์ ในปีการศึกษา 2538 (สารมหาวิทยาลัย สงขลานครินทร์. 2538 : 24) จากข้อจำกัดเกี่ยวกับจำนวนนักเรียนที่จะรับเข้า ศึกษาต่อนี้เอง มหาวิทยาลัยจึงจำเป็นต้องคัดเลือกบุคคลที่มีความรู้พื้นฐานที่ สามารถที่จะศึกษาต่อในสาขาวิชาต่าง ๆ ไปได้จนสำเร็จ และในการวัดความรู้ พื้นฐาน ควรที่จะพิจารณาถึงความถูกต้องแม่นยำและยุติธรรมจากผลการวัด (อนันต์ จันท์กวี. 2529 : 12, 14) ซึ่งจะต้องอาศัยเครื่องมือวัดที่มีคุณภาพเชื่อถือได้ และ

จะต้องสามารถวัดคุณลักษณะ (Trait) ที่ต้องการจะวัดได้ตรงตามเกณฑ์ นั่นคือ เครื่องมือนั้นจะต้องมีความเที่ยงตรง (Validity) (วิรัช วรรณรัตน์. 2532 : 62) ซึ่งเป็นคุณสมบัติที่สำคัญมากประกอบด้วยความเที่ยงตรงเชิงเนื้อหา (Content Validity) ความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) ความเที่ยงตรงตามสภาพ (Concurrent Validity) และความเที่ยงตรงเชิงพยากรณ์ (Predictive Validity) และในการตรวจสอบความเที่ยงตรงทั้งหมดนี้ สามารถจัดกระทำด้วยวิธีการหลาย ๆ รูปแบบ อาทิเช่น ความเที่ยงตรงเชิงเนื้อหาสามารถใช้วิธีการให้ผู้เชี่ยวชาญพิจารณาความสอดคล้องของข้อสอบกับเนื้อหาที่จะวัด ซึ่งในแบบทดสอบคัดเลือกจะจัดกระทำก่อนดำเนินการจัดพิมพ์ข้อสอบ ความเที่ยงตรงเชิงพยากรณ์เป็นคุณสมบัติที่สำคัญของแบบทดสอบคัดเลือก เพราะว่าถ้าแบบทดสอบคัดเลือกมีความเที่ยงตรงเชิงพยากรณ์สูงแล้วก็มั่นใจได้ว่าแบบทดสอบนั้นสามารถคัดเลือกผู้ที่เข้าศึกษาแล้ว สามารถศึกษาได้สำเร็จตามต้องการ ในการตรวจสอบความเที่ยงตรงเชิงพยากรณ์นั้น วิธีการคิดคะแนนจากผลการสอบในแบบทดสอบแต่ละฉบับที่จะนำมาประมวลเข้าด้วยกันเพื่อการตัดสินใจนั้น เป็นตัวแปรที่สำคัญอีกอย่างหนึ่งที่จะบอกได้ว่าควรคิดคะแนนแบบใดที่จะให้ผลการตัดสินใจดีเยี่ยมมากที่สุด และคุณภาพของแบบทดสอบในความเที่ยงตรงเชิงพยากรณ์สูงด้วย ซึ่งวิธีการคิดคะแนนมีหลายวิธี เช่นการใช้คะแนนดิบไปประมวลผลแล้วตัดสินใจ แต่วิธีนี้จะไม่นิยมใช้เนื่องจากคะแนนดิบแต่ละรายวิชาไม่สามารถเปรียบเทียบกันได้เนื่องจากฐานคะแนนอยู่คนละมาตรกัน การแปลงคะแนนดิบให้เป็นคะแนนมาตรฐาน หรือการให้น้ำหนักคะแนนแต่ละข้อตามความสามารถ (อ) ตามทฤษฎีการตอบข้อคำถาม ซึ่งลอร์ด (Lord. 1980 : 74) ได้เปรียบเทียบการให้คะแนนผลการสอบความถนัดทางการเรียนโดยใช้น้ำหนักที่เหมาะสมของคะแนนความสามารถกับแบบ 0-1 พบว่าการให้คะแนนตามน้ำหนักที่เหมาะสมของความเหมาะสมที่สุด ซึ่งสอดคล้องกับการศึกษาของ ชู และชูศักดิ์ (ชูศักดิ์ ชัมภลลิขิต. 2525 : 33 ; อ้างอิงมาจาก Hsu and Choosak Khampalikit. 1982 : 19-23) ที่พบว่าค่าความเชื่อมั่นและค่าอำนาจจำแนกของแบบทดสอบมีค่าสูงขึ้น เมื่อใช้ค่าน้ำหนักคะแนนความสามารถแทนคะแนนรวมของ 0-1 ส่วนจำแนกสุขหลาย (2530 : 172) ได้ศึกษาผลของวิธีการคิดคะแนนต่อความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกเข้าศึกษาต่อโดยใช้วิธีการคิดคะแนนแบบน้ำหนักความสามารถ คะแนนถ่วงน้ำหนักจากสมการการถดถอย คะแนนมาตรฐานที่-ปกติ และคะแนนดิบ พบว่าไม่สามารถเฉพาะเจาะจงได้ว่าวิธีใดเหมาะสมที่สุดหรือมีประสิทธิภาพในการทำนายมากที่สุด

แบบทดสอบคัดเลือกลักษณะนั้น นอกจากจะตรวจสอบความเที่ยงตรงเชิงพยากรณ์ที่จะพิจารณาความสามารถในการทำนายผลการเรียนแล้ว การพิจารณาถึงความยุติธรรมที่เกิดกับผู้สอบต่างกลุ่ม โดยใช้คะแนนการสอบของแต่ละกลุ่มมาเปรียบเทียบกันเพื่อพิจารณาถึงโอกาสของความได้เปรียบเสียเปรียบกันของโอกาสในการตอบถูกทั้ง ๆ ที่ทั้งสองกลุ่มมีความสามารถเท่ากัน นั่นคือความลำเอียงของข้อสอบมีผู้เสนอมากมายพอจะสรุปได้ 3 ประเภทคือ

1. การวัดความเบี่ยงเบนสัมพัทธ์ (Relative Deviation) ของข้อสอบแต่ละข้อจากแนวโน้มเข้าสู่ส่วนกลาง เช่น การวิเคราะห์ความแปรปรวน วิธีแปลงค่าความยากของแองกอฟฟ์

2. การประเมินความเที่ยงตรงเชิงโครงสร้าง เช่น การวิเคราะห์องค์ประกอบ

3. การประมาณค่าโอกาสในการตอบข้อสอบถูกของผู้ที่มีความสามารถระดับเดียวกัน เช่น วิธีโคสแควร์ การใช้ทฤษฎีการตอบข้อคำถาม วิธีแมนเทิล-แฮนส์เซล และวิธี SIBTEST

จากการศึกษาของเพิร์ลแมน และคนอื่น ๆ (Perlman and Others. 1988) แฮมเบิลตันและคนอื่น ๆ (Hambleton and Others. 1986) เชปเพิร์ด คามิลลี และวิลเลียม (Shepard, Camilli and Williams. 1984) และ รัดเนอร์ เกทสัน และไนท์ (Rudner, Getson and Knight. 1980) ที่ได้วิเคราะห์ความลำเอียงของข้อสอบโดยใช้ข้อมูลจำลองในลักษณะต่าง ๆ กัน ส่วนใหญ่พบว่าวิธีวิเคราะห์ความลำเอียงโดยใช้ทฤษฎีการตอบข้อคำถามชนิด 2-3 พารามิเตอร์สามารถวิเคราะห์ความลำเอียงได้สอดคล้องมากที่สุด และ กาญจนา วัฒนสุนทร (2538) ได้พัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ โดยใช้วิธีวิเคราะห์ความลำเอียง วิธีใช้ทฤษฎีการตอบข้อคำถาม 2 พารามิเตอร์ วิธีแมนเทิล-แฮนส์เซล และวิธี SIBTEST พบว่าไม่ว่าจะใช้เกณฑ์ในลักษณะใด จำนวนร้อยละของข้อสอบที่ลำเอียงกับความสอดคล้องของการตัดสินความลำเอียงจะไม่คงที่กับความยาวของแบบทดสอบและจำนวนผู้สอบ ทั้งนี้เนื่องมาจากดัชนีความลำเอียงแต่ละตัวที่ได้จากผู้สอบต่างประเทศไม่มีความคงที่นั่นเอง

เพื่อเป็นแนวทางในการพัฒนาแบบทดสอบคัดเลือกรูปแบบที่เป็นแบบทดสอบที่ผู้สร้างมีโอกาสในการตรวจสอบก่อนการทดสอบน้อยมาก เนื่องจากข้อจำกัดด้านเวลา และเป็นการป้องกันการรั่วไหล อีกทั้งเพื่อให้ผลการคัดเลือกมีความถูกต้อง แม่นยำ และยุติธรรมแก่ผู้สอบทุกคนทุกกลุ่ม ผู้วิจัยจึงสนใจที่จะศึกษาคูณภาพของแบบทดสอบคัดเลือกรูปแบบ

เพื่อเข้าศึกษาในระดับชั้นปีที่ 1 ประเภทรับตรงของมหาวิทยาลัยสงขลานครินทร์ โดยศึกษาความเที่ยงตรงเชิงพยากรณ์จากการคิดคะแนนมาตรฐานที่-ปกติ และคะแนนน้ำหนักความสามารถ ตรวจสอบความลำเอียงของข้อสอบในแต่ละวิชา ด้วยวิธีการใช้ทฤษฎีการตอบข้อคำถาม วิธีแมนเทิล-แฮนส์เซล และวิธี SIBTEST ระหว่างเพศชายและหญิง และศึกษาสาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบแต่ละข้อ ในแต่ละกลุ่มที่ตรวจสอบพบว่ามีความลำเอียง ด้วยวิธีการทางสถิติทั้ง 3 วิธี

จุดมุ่งหมายของการวิจัย

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่วิเคราะห์ความลำเอียงด้วยวิธีต่างกันโดยศึกษา

1. จำนวนข้อสอบที่ลำเอียงต่อเพศชายและหญิงในวิชาต่าง ๆ เมื่อใช้วิธีการวิเคราะห์ความลำเอียงด้วยวิธีการต่างกัน
2. ความสอดคล้องของการตัดสินผลการสอบที่ใช้วิธีการคิดคะแนนต่างกัน
3. สาเหตุของความลำเอียงของข้อสอบที่ลำเอียงต่อเพศชายและหญิง

สมมติฐานในการวิจัย

1. เมื่อใช้วิธีการวิเคราะห์ความลำเอียงด้วยวิธีการใช้ทฤษฎีการตอบข้อคำถาม วิธีแมนเทิล-แฮนส์เซล และวิธี SIBTEST ระหว่างเพศชายและหญิงต่างตัดสินความลำเอียงของข้อสอบแตกต่างกัน
2. การตัดสินผลการสอบที่คิดคะแนนมาตรฐานที่-ปกติ และคะแนนน้ำหนักความสามารถจากข้อสอบทั้งหมดกับคะแนนมาตรฐานที่-ปกติ และคะแนนน้ำหนักความสามารถจากข้อสอบที่ปราศจากความลำเอียงระหว่างเพศชายและหญิงไม่สอดคล้องกัน

ความสำคัญของการวิจัย

ได้แนวทางในการสร้างข้อสอบคัดเลือกที่มีประสิทธิภาพยิ่งขึ้นโดย

1. ใช้วิธีการตรวจสอบความลำเอียงที่สามารถค้นพบความลำเอียงระหว่างเพศได้ดีที่สุด

2. ใช้วิธีการคิดคะแนนที่จะใช้ประมวลผลเพื่อการตัดสินผลการสอบคัดเลือก ที่ถูกต้องและยุติธรรมมากที่สุด

3. ใช้ข้อสรุปถึงสาเหตุของความลำเอียงของข้อสอบระหว่างเพศชายและหญิง

ขอบเขตของการวิจัย

1. แหล่งข้อมูลที่ใช้ในการศึกษาค้างนี้คือ

1.1 ดัชนีความลำเอียง ความสอดคล้องของการตัดสินผลการสอบ และสาเหตุของความลำเอียง ศึกษาจากนักเรียนที่เข้าสอบคัดเลือกเข้าเป็นนักศึกษาชั้นปีที่ 1 ประเภทรับตรง ปีการศึกษา 2538 ของมหาวิทยาลัยสงขลานครินทร์ โดยแบ่งกลุ่มที่ศึกษาความลำเอียงเป็นเพศชายและเพศหญิง

1.2 ความเที่ยงตรงเชิงพยากรณ์ ศึกษาจากนักศึกษาที่ได้รับการคัดเลือกเข้าศึกษาในชั้นปีที่ 1 ปีการศึกษา 2538 ของมหาวิทยาลัยสงขลานครินทร์

2. ตัวแปรที่ใช้ในการศึกษา

2.1 ตัวแปรอิสระ ได้แก่

2.1.1 วิธีการตรวจสอบความลำเอียงได้แก่

2.1.1.1 วิธีการใช้ทฤษฎีการตอบข้อคำถาม

2.1.1.2 วิธีแมนเทิล-แฮนส์เซล

2.1.1.3 วิธี SIBTEST

2.1.2 กลุ่มที่ศึกษาความลำเอียง ได้แก่ เพศชายและหญิง

2.1.3 วิธีการให้คะแนน ได้แก่

2.1.3.1 คะแนนมาตรฐานที่-ปกติ

2.1.3.2 คะแนนน้ำหนักความสามารถ

2.2 ตัวแปรตาม ได้แก่

2.2.1 ดัชนีความลำเอียงของข้อสอบ

2.2.2 ค่าสัมประสิทธิ์สหสัมพันธ์แบบลำดับ

2.2.3 ความเที่ยงตรงเชิงพยากรณ์

3. เครื่องมือที่ใช้ในการศึกษาค้างนี้ เป็นแบบทดสอบวิชา ภาษาไทย ก สังคมศึกษา ก ภาษาอังกฤษ กข สังคมศึกษา กข ภาษาไทย กข และภาษา

อังกฤษ กษค ที่ใช้สอบคัดเลือกบุคคลเข้าเป็นนักศึกษาชั้นปีที่ 1 ประเภทรับตรง ปีการศึกษา 2538 ของมหาวิทยาลัยสงขลานครินทร์

นิยามศัพท์เฉพาะ

1. ดัชนีความลำเอียงของข้อสอบ หมายถึง ขนาดของความแตกต่างของโอกาสในการตอบข้อสอบข้อนั้น ๆ ถูกของบุคคลต่างกลุ่ม ดัชนีความลำเอียงคำนวณจาก 3 วิธีคือ

1.1 วิธีใช้ทฤษฎีการตอบข้อคำถาม หมายถึง วิธีที่พิจารณาขนาดพื้นที่ที่แตกต่างกันระหว่างโค้งการตอบของข้อสอบข้อนั้นของแต่ละกลุ่มย่อย

1.2 วิธีแมนเทล - แฮนส์เซล หมายถึง วิธีที่พิจารณาค่าของ χ^2_{MH} ที่ได้จากการคำนวณตามแนวคิดของแมนเทล - แฮนส์เซล

1.3 วิธี SIBTEST หมายถึง วิธีที่พิจารณาค่าของเบต้า (β) ที่ได้จากการทดสอบความแตกต่างของโอกาสในการตอบถูกของแต่ละกลุ่มย่อย

2. กลุ่มย่อย หมายถึง กลุ่มของผู้สอบที่แบ่งตามเพศ (ชาย-หญิง)

3. การคิดคะแนน หมายถึง การให้น้ำหนักหรือคะแนนในผลการตอบข้อสอบ เพื่อนำไปประเมินผลการสอบคัดเลือก ซึ่งใช้ 2 วิธี คือ

3.1 คะแนนมาตรฐานที่-ปกติ หมายถึง ระบบการคิดคะแนนที่แปลงคะแนนดิบให้เป็นคะแนนมาตรฐานที่-ปกติก่อน แล้วจึงจะนำไปประมวลผล

3.2 คะแนนน้ำหนักความสามารถ หมายถึง ระบบการคิดคะแนนที่คิดตามค่าความสามารถ (θ) ของข้อสอบแต่ละข้อตามทฤษฎีการตอบข้อคำถาม

4. ผู้ระบุน้ำหนักของความลำเอียง หมายถึง บุคคลที่มีความรู้ในเนื้อหาวิชาของข้อสอบในแต่ละฉบับนั้น ๆ ซึ่งได้แก่ อาจารย์ผู้สอน นักวัดผลการศึกษา หรือนักศึกษาที่เรียนในสาขาวิชาที่เกี่ยวข้อง

5. ประสิทธิภาพในการทำนาย หมายถึง ค่าร้อยละของความแปรปรวนของคะแนนสอบในแต่ละวิชา เมื่อพิจารณาเกรดเฉลี่ยภาคเรียนที่ 1. เป็นเกณฑ์

เอกสารและงานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องกับการศึกษาครั้งนี้ ผู้วิจัยได้ศึกษาใน
เนื้อหาต่าง ๆ ดังต่อไปนี้

1. ความเป็นมาของการศึกษาความลำเอียงของข้อสอบ
2. วิธีการตรวจสอบความลำเอียงของข้อสอบ
 - 2.1 โดยใช้ทฤษฎีการตอบข้อคำถาม
 - 2.2 วิธีของแมนเทิล - แฮนส์เซล (MH)
 - 2.3 วิธี SIBTEST
3. แหล่ง/สาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบ
4. วิธีการคิดคะแนน
5. งานวิจัยที่เกี่ยวข้อง
 - 5.1 งานวิจัยที่เกี่ยวข้องกับการตรวจสอบความลำเอียง
 - 5.2 งานวิจัยที่เกี่ยวข้องกับการคิดคะแนน

ความเป็นมาของการศึกษาความลำเอียงของข้อสอบ

การศึกษาความลำเอียงของข้อสอบนั้น เริ่มมาจากการสังเกตผลการสอบ
คัดเลือก ซึ่งพบว่าไม่เป็นไปตามสัดส่วนสถิติปัญหาหรือโครงสร้างของประชากร กรณี
ที่ทำให้มีการตื่นตัวมากในเรื่องความลำเอียงของข้อสอบคือในปี ค.ศ. 1971 มาร์โค
เดอฟูนีส (Marco Defunis) และคนอื่น ๆ ซึ่งถูกปฏิเสธจากโรงเรียนกฎหมายของ
มหาวิทยาลัยวอชิงตัน ได้ฟ้องร้องว่าเขาได้คะแนนการสอบสูงกว่าผู้ที่ได้รับการคัดเลือก
บางคน และได้ยื่นฎีกาฟ้องร้อง ชาร์ลส์ โอเดการ์ด (Charles Odegaurd)
เพื่อให้พิจารณาทบทวนการคัดเลือกนักศึกษาใหม่และจากนั้นเป็นต้นมา การพิจารณา
ตรวจสอบความลำเอียงของข้อสอบระหว่างผู้สอบกลุ่มย่อยก็ได้รับการปฏิบัติกันจนถึง
ปัจจุบัน (Breland and Ironson. 1976 : 89) ซึ่งการตรวจสอบความลำเอียง
จะทำการตรวจสอบก่อนและหลังจากการทดลองใช้ การตรวจสอบความลำเอียงของ
ข้อสอบก่อนนำไปทดลองใช้จะใช้การตรวจสอบโดยผู้เชี่ยวชาญ โดยพิจารณาถึงรูปแบบ

ของข้อสอบ เนื้อหา คำที่ใช้และอื่น ๆ เพื่อไม่ให้เกิดความลำเอียง ส่วนการตรวจสอบความลำเอียงของข้อสอบหลังจากการทดลองใช้ผู้เสนอหลายวิธีดังนี้

1. การพิจารณาความแตกต่างกันในค่าเฉลี่ยระหว่างกลุ่ม
2. การวิเคราะห์องค์ประกอบ เป็นการดูความสัมพันธ์ระหว่างคะแนนย่อย ๆ ในแบบทดสอบสำหรับกลุ่มย่อย
3. การเปรียบเทียบค่าความยากเป็นการพิจารณาลำดับที่สัมพันธ์ของค่าความยากที่ได้จากการวิเคราะห์แยกกลุ่ม พิจารณาปฏิสัมพันธ์ระหว่างข้อสอบและกลุ่มผู้สอบ ถ้าข้อสอบข้อใดมีลักษณะของความสัมพันธ์สำหรับบางกลุ่มแตกต่างไปจากข้ออื่น ๆ ในแบบทดสอบเดียวกันจะถือว่าข้อสอบข้อนั้นมีความลำเอียง ซึ่งมีวิธีการทำได้ทั้งการวิเคราะห์ความแปรปรวนและการใช้แผนภูมิเส้น
4. การใช้ทฤษฎีการตอบข้อคำถาม เป็นการเปรียบเทียบความน่าจะเป็นในการตอบข้อสอบของผู้สอบกลุ่มย่อย โดยพิจารณาถึงค่าพารามิเตอร์ของข้อสอบและผู้ตอบตามแบบจำลองที่เลือกใช้ทั้ง 1, 2 และ 3 พารามิเตอร์
5. การวิเคราะห์ด้วยค่าโคสแควร์ เป็นวิธีการที่แปลงค่าความสามารถของสมาชิกในกลุ่มเป็นช่วง ๆ แล้วใส่ข้อมูลในตาราง 2 ทาง (จำนวนกลุ่ม x การตอบถูก - ผิด) จากนั้นคำนวณค่าโคสแควร์แต่ละตารางและเมื่อรวมค่าทุกตาราง จะได้ค่าโคสแควร์ที่มีขึ้นแห่งความเป็นอิสระเท่ากับจำนวนกลุ่ม ถ้ากลุ่มผู้สอบ 2 กลุ่ม มีคะแนนเฉลี่ยต่างกันแล้วจะเกิดผลที่ความลำเอียงได้
6. เทคนิคลอกลีเนียร์ (Loglinear) หรือแบบจำลองโลจิสต์ที่เบี่ยงเบนจากโคสแควร์ และได้มีการนำมาขยายโดยการทำซ้ำ หลังจากการจัดข้อสอบที่ลำเอียงออกไปแล้วในแต่ละรอบของการวิเคราะห์
7. วิธีของแมนเทล - แฮนส์เซล (Mantel - Haenszel) เป็นการศึกษาตัวแปรประเภทแบ่งสองในกลุ่มที่แตกต่างกันด้วยตัวแปรบลิอค และนำมาใช้เป็นตัวบ่งชี้ความลำเอียงของข้อสอบ
8. วิธีการตอบข้อสอบแบบเทียม ซึ่งเสนอโดย ลินน์ และ ฮาร์นิสซ์ (Linn and Harnisch. 1981) โดยวิเคราะห์ตามทฤษฎีการตอบข้อสอบ 3 พารามิเตอร์ และรวมกลุ่มย่อยเข้าด้วยกันแล้วแบ่งความสามารถเป็นช่วง พิจารณาความแตกต่างระหว่างความน่าจะเป็นของการตอบตามที่คาดหวัง ซึ่งประมาณได้ด้วยทฤษฎีการตอบข้อคำถามกับความน่าจะเป็นของการตอบถูกจากข้อมูลเชิงประจักษ์พิจารณาเป็นช่วง ๆ ความสามารถ แล้วนำมารวมกันเป็นดัชนีรวมของความลำเอียงของแต่ละข้อ ให้ได้ในกรณีกลุ่มตัวอย่างน้อย ทำให้ใช้ทฤษฎีการตอบข้อคำถามแบบเต็มรูปไม่ได้

9. วิธี SIBTEST เป็นวิธีดัดแปลงจากวิธีทฤษฎีการตอบข้อคำถาม สามารถใช้ได้ทั้งเพื่อการตรวจสอบความลำเอียงของข้อสอบรายข้อ การตรวจสอบความลำเอียงของกลุ่มข้อสอบหรือแบบทดสอบทั้งฉบับ

การวิจัยในครั้งนี้ ผู้วิจัยใช้วิธีการตรวจสอบความลำเอียง 3 วิธี คือ วิธี การใช้ทฤษฎีการตอบคำถาม วิธีของแมนเทิล-แฮนส์เซล และวิธี SIBTEST ซึ่งมีรายละเอียดดังนี้

วิธีการตรวจสอบความลำเอียงของข้อสอบ

การตรวจสอบความลำเอียงของข้อสอบโดยใช้ทฤษฎีการตอบข้อคำถาม

นิยามของความลำเอียงของข้อสอบในแง่ของทฤษฎีการตอบข้อคำถามก็คือ ข้อสอบจะลำเอียงถ้าฟังก์ชันการตอบข้อสอบนั้น ๆ ของประชากรย่อย 2 กลุ่มไม่เป็นเส้นเดียวกัน

จากนิยามดังกล่าว การตรวจสอบความลำเอียงของข้อสอบตามแนวทฤษฎีการตอบข้อคำถามสามารถกระทำได้โดยการเปรียบเทียบฟังก์ชันการตอบข้อสอบของผู้สอบกลุ่มย่อยที่ต้องการจะศึกษา และในการเปรียบเทียบอาจกระทำดังนี้

1. แบบจำลองโลจิสติก 1 พารามิเตอร์ ซึ่งกำหนดให้ค่าการเดา = 0 และค่าอำนาจจำแนกของทุกข้อ = 1.00 มีค่าความยากแปรผันไปตามกลุ่ม

เริ่มด้วยการทดสอบความเหมาะสมของแบบจำลองกับข้อมูลทั้ง 2 กลุ่ม เป็นรายข้อด้วย H_i ซึ่งมีการกระจายแบบไคสแควร์ (χ^2) มี $df = N - 1$ ตัดข้อที่ไม่เหมาะสมออก แล้ววิเคราะห์ความลำเอียงรายข้อด้วย Z_i

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

U_{ij} = คะแนนจากการตอบข้อสอบแบบ 0 - 1 ของคนที่ j ข้อ ที่ i

$$P_i(\theta) = \frac{1}{1 + \exp[-D(\theta - b_i)]}$$

$$Z_i = \frac{\hat{b}_{iA} - \hat{b}_{iB}}{\sqrt{SE^2_{\hat{b}_{iA}} + SE^2_{\hat{b}_{iB}}}}$$

$$SE^2_{\hat{b}_i} = \sqrt{I_{ai} / (I_{ai} I_{bi} I_{aibi}^2)}$$

ค่า Z_i สูง แสดงว่า ข้อสอบนั้นยากสำหรับกลุ่มหนึ่งมากกว่าอีกกลุ่มหนึ่ง หรืออาจจะทดสอบด้วยค่า t ซึ่งคำนวณได้จากสูตร

$$t = \frac{\hat{b}_{iA} - \hat{b}_{iB}}{\sqrt{SE^2_{\hat{b}_{iA}} + SE^2_{\hat{b}_{iB}}}}$$

ในกรณีนี้ ดราบา (Berk. 1982 ; citing Draba.n.d.) เสนอให้ใช้ค่า $t > 2.4$ เป็นเกณฑ์ในการตัดสินความลำเอียงของข้อสอบ

2. แบบจำลองโลจิสติก 2 พารามิเตอร์ กำหนดค่าการเดา = 0 ส่วนค่าความยากและค่าอำนาจจำแนกผันแปรไปตามกลุ่มย่อย วิธีการเริ่มด้วยการประมาณค่าพารามิเตอร์แยกตามกลุ่มย่อย A และ B ด้วยวิธี Maximum Likelihood แล้วเปรียบเทียบได้ถึงการตอบข้อสอบด้วยค่า F

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

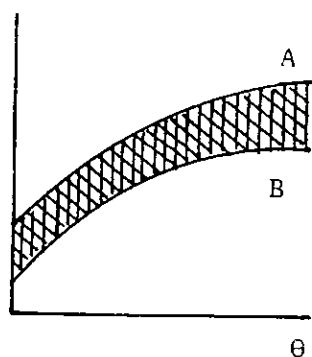
ค่าชั้นแห่งความเป็นอิสระ คือ $2, (J_A + J_B - 4)$ เมื่อ J เป็นจำนวนช่วงความสามารถ การทดสอบค่า F จะบอกว่าเส้นถดถอยรวมสามารถอธิบายความแปรปรวนของสัดส่วนการตอบของเส้นถดถอยจากแต่ละกลุ่มได้เท่ากันหรือไม่ ถ้าเท่ากัน (ไม่มีนัยสำคัญ) ก็ไม่จำเป็นต้องใช้เส้นถดถอยแยกกันเพราะความสัมพันธ์ระหว่างความสามารถและการตอบข้อสอบของทั้งสองกลุ่มเป็นเช่นเดียวกัน และข้อสอบทำหน้าที่ในการวัดความสามารถของคน 2 กลุ่มนี้ได้เหมือนกัน

3. แบบจำลองโลจิสติก ชนิด 3 พารามิเตอร์ เป็นแบบที่ให้ค่าการเดา ค่าความยาก และค่าอำนาจจำแนกของข้อสอบแปรเปลี่ยนไปตามกลุ่มย่อย

วิธีนี้เริ่มจากการประมาณค่าพารามิเตอร์รวมของทุกคนทุกกลุ่ม แล้วแปลงค่ามาตรฐานบนค่าประมาณความยาก จากนั้นแยกประมาณค่าความยากและอำนาจจำแนกตามกลุ่มย่อย โดยใช้ค่าการเดาที่ประมาณได้จากการรวมทุกกลุ่ม การทำค่าความยากให้เป็นมาตรฐานต้องตัดข้อที่มีค่าความยากสูงมากหรือต่ำมากออก

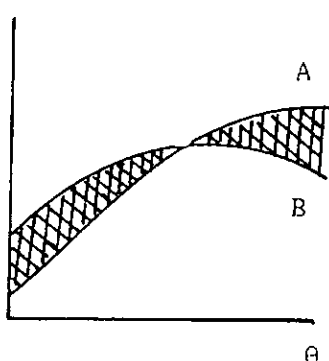
การวิเคราะห์ความลำเอียงของข้อสอบตามโมเดล 3 พารามิเตอร์ นั้น ลินน์ ลีไวน์ ฮาสตังส์ และวอร์ดรอป (Hulin, Drasgow and Parson. 1983 : 176 ; citing Linn and Others. 1981) ได้พัฒนาการทดสอบความลำเอียงจากความแตกต่างระหว่างโค้งลักษณะของข้อสอบโดยตรง โดยพิจารณาจากพื้นที่ระหว่างโค้ง 2 โค้งเป็นตัวแสดงความลำเอียง โดยมีลักษณะของพื้นที่อยู่ 3 ลักษณะ ดังภาพประกอบ 1

P (A)



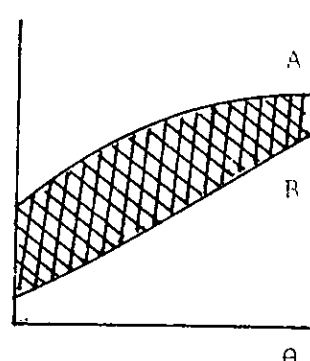
(a)

P (A)



(b)

P (A)



(c)

ภาพประกอบ 1 แสดงโค้งการตอบข้อสอบที่มีความลำเอียง 3 ลักษณะสำหรับกลุ่ม A และ B

รูป (a) แสดงถึงความลำเอียงที่คงที่ (Constant Bias) ข้อสอบมีความลำเอียงต่อกลุ่ม B อย่างสม่ำเสมอ ตลอดความต่อเนื่องของความสามารถ เป็นความลำเอียงอย่างสม่ำเสมอ (Uniform Bias)

รูป (b) แสดงว่าข้อสอบมีความยากคล้ายคลึงกันสำหรับทั้ง 2 กลุ่มแต่มีอำนาจจำแนกสำหรับกลุ่ม B น้อยกว่ากลุ่ม A ทำให้มีความลำเอียงต่อกลุ่ม B ในพวกที่มีความสามารถสูง และลำเอียงต่อกลุ่ม A ในพวกที่มีความสามารถต่ำ

รูป (c) มีจุดตัดของโค้งและแกนความเป็นไปได้ในการตอบถูกไม่เท่ากัน สำหรับ 2 กลุ่ม แต่ความยากและอำนาจจำแนกใกล้เคียงกัน มีความลำเอียงต่อกลุ่ม B ในผู้ที่มีความสามารถต่ำ แต่จะลดลงในผู้ที่มีความสามารถสูงขึ้น ลักษณะความลำเอียงเช่นนี้อาจเป็นเพราะสมาชิกในกลุ่ม A ทำข้อสอบถูกเพราะการเดาก็เป็นได้

ลินน์ และคนอื่น ๆ (Linn and Others. 1981) ได้อภิปรายดัชนีแสดงความลำเอียง 4 ตัว ซึ่ง 3 ตัวแรกที่มีพื้นฐานบนพื้นที่ระหว่างโค้งของกลุ่มเปรียบเทียบและกลุ่มอ้างอิงดังนี้

1. พื้นที่ที่โค้งของกลุ่มอ้างอิงอยู่เหนือโค้งกลุ่มเปรียบเทียบกรณีของภาพประกอบ 1 ถ้ากลุ่ม B เป็นกลุ่มฐานหรือกลุ่มอ้างอิงแล้วพื้นที่ประเภtnี้ได้แก่พื้นที่ด้านซ้ายในรูป (b)
2. พื้นที่ที่โค้งกลุ่มฐานหรือกลุ่มอ้างอิงอยู่ใต้โค้งกลุ่มศึกษา ได้แก่ พื้นที่ด้านขวาของรูป (b) และพื้นที่ในรูป (a) และ (c)
3. พื้นที่รวม ได้แก่พื้นที่ในข้อ 1 และข้อ 2 รวมกัน
4. รากที่สองของค่าเฉลี่ยความแตกต่างกำลังสองของโค้งของ 2 กลุ่ม เป็นดัชนีความลำเอียงในช่วงความสามารถ -3 ถึง 3 ที่ได้จากการคำนวณดังนี้

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

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

ดัชนี 2 ตัวแรก แสดงค่าความลำเอียงต่อกลุ่มใดกลุ่มหนึ่งโดยตรง แต่ 2 ตัวหลังเป็นการให้ภาพรวมของความลำเอียงของข้อสอบ ค่ายิ่งใกล้ 0 แสดงว่า ข้อใดข้อหนึ่งมีความใกล้เคียงมาก การคำนวณค่าดัชนีดังกล่าวจะแสดงความลำเอียง ได้ดีกว่าการพิจารณาเฉพาะความแตกต่างของค่าพารามิเตอร์ ซึ่งอาจเกิดขึ้นได้โดย พื้นที่ข้อใดข้อหนึ่งไม่ต่างกันเลยก็เป็นได้ ดังที่ ลินน์ และคนอื่น ๆ พบว่าข้อสอบข้อที่ i สำหรับกลุ่มหนึ่งมีค่า $a = 1.8$ $b = 3.5$ และ $c = .2$ อีกกลุ่มหนึ่งมีค่า $a = .5$, $b = 5.0$ และ $c = .2$ ดูแล้วค่า a และ b ต่างกันมาก แต่ค่า ความแตกต่างระหว่างข้อใด 2 ข้อ มีค่า $= .05$ ซึ่งเป็นค่าที่ยอมรับได้ในการสรุปว่า ข้อสอบไม่ลำเอียง

เชปเพิร์ด และคนอื่น ๆ (Shepard and Others. 1985 : 81) กล่าวว่า เมื่อพบว่าข้อใดลักษณะข้อสอบของกลุ่มอ้างอิงและกลุ่มเป้าหมายต่างกันมากพอที่จะ เป็นหลักฐานว่าข้อสอบที่ศึกษาอยู่ไม่ได้วัดลักษณะแฝงเดียวกันสำหรับกลุ่มทั้งสองแล้ว จะใช้ดัชนีชี้ถึงระดับความลำเอียงดังนี้

1. พื้นที่ชนิดไม่มีเครื่องหมาย (Unsigned Area) เป็นค่าสัมบูรณ์ของ พื้นที่ระหว่างข้อใดทั้งสอง

2. ผลบวกกำลังสองแบบที่ 1 (SOS 1) เป็นค่าผลบวกของความแตกต่าง ระหว่างกำลังสองทุกค่าของความสามารที่ เป็นไปได้

3. ผลบวกกำลังสองแบบที่ 2 (SOS 2) คล้าย SOS 1 แต่ให้น้ำหนัก ความแตกต่างกำลังสองด้วยส่วนกลับของความคลาดเคลื่อนของความแปรปรวนในการ ประมวลค่าความเป็นไปได้ ดังนั้นถ้ามีการประมวลข้อใดอย่างไม่ดีในส่วนใดส่วนหนึ่ง แล้ว จะทำให้ลดประโยชน์ต่อความแตกต่างระหว่างสองกลุ่ม

4. ไคสแควร์ ($IRT - \chi^2$) เป็นการทดสอบนัยสำคัญโดยการเปรียบเทียบ ความแตกต่างของค่าความยากและอำนาจจำแนกระหว่างกลุ่มพร้อม ๆ กัน โดย ลอร์ด (Lord. 1980)

5. พื้นที่ชนิดมีเครื่องหมาย เหมือนพื้นที่ชนิดไม่มีเครื่องหมายแต่มีการให้ เครื่องหมายไว้ เพื่อแสดงให้เห็นว่ากลุ่มใดเสียประโยชน์ และกลุ่มใดได้ประโยชน์ ถ้าข้อใดการตอบข้อสอบตัดกัน แสดงว่ามีความลำเอียงต่างกันสำหรับผู้มีความสามารถ ต่างกัน และค่าสัมบูรณ์ของพื้นที่ชนิดมีเครื่องหมายจะน้อยกว่าชนิดไม่มีเครื่องหมาย

6. ผลบวกกำลังสองแบบที่ 3 (SOS 3) เป็นผลบวกกำลังสองของความ แตกต่างชนิดมีเครื่องหมายของ SOS 1 โดยการใช้เครื่องหมายของความแตกต่าง แต่ละตัวที่ยังไม่ยกกำลังสอง

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

การวัดพื้นที่ของความแตกต่างระหว่างโค้งลักษณะข้อสอบของกลุ่มตัวอย่าง เมื่อพิจารณาถึงขนาดของความลำเอียงของข้อสอบนั้นมีอยู่ 2 ลักษณะคือ การวัดช่วงเปิดและช่วงปิด (Open and Closed Intervals) กรณีของการวัดช่วงปิดนั้น จะกำหนดขอบเขตของการวัดให้อยู่ในช่วงความสามารถที่กำหนด ในขณะที่การวัดช่วงเปิดนั้นวัดภายในช่วงความสามารถทั้งหมดเพื่อให้ได้พื้นที่ที่แน่นอน (Exact Area) ระหว่างโค้ง 2 โค้ง ภายใต้การวัดช่วงปิดนั้นสามารถวัดพื้นที่ได้ 2 ลักษณะคือ พื้นที่ชนิดมีเครื่องหมายและไม่มีเครื่องหมาย (Signed and Unsigned Areas) โดยมีสูตรการคำนวณทั่ว ๆ ไปดังนี้

จากฟังก์ชันการตอบข้อสอบแบบ 3 พารามิเตอร์

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

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

เมื่อ a , b และ c เป็นพารามิเตอร์ที่แสดงลักษณะของข้อสอบ และ D เป็นค่าคงที่ที่กำหนดให้เท่ากับ 1.7

พื้นที่ระหว่างความสามารถที่กำหนดช่วงไว้ 2 จุดบนสเกลความสามารถได้จาก

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

ในการศึกษาความลำเอียงของข้อสอบนั้น จะมีพารามิเตอร์ของ 2 กลุ่มคือ กลุ่มอ้างอิง ได้แก่ a_R , b_R และ c_R และกลุ่มเปรียบเทียบคือ a_F , b_F และ c_F พื้นที่ชนิดที่มีเครื่องหมาย (Closed - interval Signed Area - CSA) ได้จาก

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

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

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

ลักษณะและคอฟแมน (สุรศักดิ์ อมรรัตนศักดิ์. 2531 : 54 ; อ้างอิงมาจาก Laksana and Coffman. 1980 : 11) เสนอให้ใช้ค่า CUA > .20 เป็นเกณฑ์ ที่ความลำเอียงของข้อสอบ

การวิเคราะห์ความลำเอียงของข้อสอบด้วยทฤษฎีการตอบข้อคำถามนั้น สับโคเวียค และคนอื่น ๆ (Subkoviak and Others. 1984) พบว่าจากการ เปรียบเทียบกับวิธีแปลงค่าความยากและโคสแควร์แบบ 5 ช่วง พบว่าการวิเคราะห์ ด้วยทฤษฎีการตอบข้อคำถามแบบ 3 พารามิเตอร์ให้ความถูกต้องมากที่สุด ซึ่ง เชปเพิร์ด คามิลลี และวิลเลียมส์ (Shepard, Camilli and Williams. 1984 : 93-128) ค้นพบสอดคล้องกันในแง่ที่ว่า เป็นวิธีที่ดีที่สุดดีกว่าวิธีโคสแควร์ และค่าความยากแปลง บาเกิ และเฟอร์รารา (Baghi and Ferrara. 1990) พบว่าในขนาดตัวอย่าง 750 คน สามารถใช้วิธีแมนเทิล-แฮนส์เซล และทฤษฎีการ ตอบข้อคำถามชนิด 3 พารามิเตอร์ได้ คิมและโคเฮน (Kim and Kohen. 1991) พบว่าการเปรียบเทียบความแตกต่างของพื้นที่ใต้โค้งโดยไม่มีเครื่องหมายจะดีกว่าแบบ มีเครื่องหมาย รวมทั้งมีความคลาดเคลื่อนต่ำกว่าแบบมีเครื่องหมายและระบุค่าลำเอียง โดยที่ไม่ลำเอียงจริง น้อยกว่าวิธีอื่น ๆ คิดเป็นร้อยละ 5 ข้อ ลอเตนชเลเจอร์ และพาร์ค (Lautenschlager and Park. 1988) พบว่าวิธีของลินน์ (Linn. 1980) ดีกว่าวิธีของวอร์ม (Warm. 1978) ส่วนรัดเนอร์ เกทสัน และ ไนท์ (Rudner, Getson and Knight. 1980) พบว่าวิธีการตรวจสอบความลำเอียง ด้วยทฤษฎีการตอบข้อคำถามแบบ 3 พารามิเตอร์ไม่มีผลกระทบจากความยาวของ ข้อสอบ

การใช้โปรแกรมการถดถอยแบบโลจิสต์ ของสวามินาธาน และโรเจอร์ส (Swaminathan and Rogers. 1990 : 361-370) โดยใช้ข้อมูลจำลองพบว่า สามารถตรวจค้นความลำเอียงของข้อสอบได้ร้อยละ 50-75 แล้วแต่ความยาวของ ข้อสอบและขนาดของกลุ่มตัวอย่างในกรณีที่เป็นการลำเอียงอย่างไม่สม่ำเสมอ แต่ สามารถตรวจค้นความลำเอียงอย่างสม่ำเสมอได้ใกล้เคียงกับวิธีการของแมนเทิล - แฮนส์เซล (MH) แต่ค่าใช้จ่ายมากกว่าถึง 3-4 เท่า และโรเจอร์ส (Rogers. 1989) ได้ใช้การถดถอยแบบแบบโลจิสต์เพื่อตรวจค้นความลำเอียงของข้อสอบแบบ ไม่สม่ำเสมอ พบว่าสามารถใช้ได้ดีกับกลุ่มตัวอย่างขนาด 250 คนขึ้นไป และจะ ตรวจค้นความลำเอียงอย่างสม่ำเสมอได้เท่ากับวิธีการของแมนเทิล - แฮนส์เซล (MH) และจะตรวจค้นได้ดีกว่าในกรณีของความลำเอียงแบบไม่สม่ำเสมอ

การตรวจสอบความลำเอียงของข้อสอบตามวิธีของแมนเทล - แฮนส์เซล (MH)

เทคนิคแมนเทล - แฮนส์เซล เป็นเทคนิคที่ แมนเทล และแฮนส์เซล ได้เสนอไว้ตั้งแต่ปี 1959 ต่อมาฮอลแลนด์ และเซเยอร์ (Holland and Thayer) ได้นำเสนอเพื่อใช้ตรวจสอบความลำเอียงของข้อสอบในปี 1986 และได้เริ่มใช้ตรวจสอบความลำเอียงของข้อสอบด้วยเทคนิคแมนเทล - แฮนส์เซล โดยโดแรนส์ (Dorans) แฮมเบิลตัน โรเจอร์ส และอะราสมิธ (Hambleton, Rogers and Arrasmith) และไรท์ (Wright) (Raju, Bode and Larsen. 1989 : 2)

เทคนิคแมนเทล - แฮนส์เซล เป็นเทคนิคที่มีความคล้ายคลึงกับวิธีโคสแควร์ที่เสนอโดยชูเนแมน (Scheuneman. 1979) มาราสคูโล และสลาเตอร์ (Marascuilo and Slaughter. 1981) โดยเป็นการเปรียบเทียบผลการตอบข้อสอบของผู้สอบ 2 กลุ่ม หรือที่ฮอลแลนด์เรียกว่ากลุ่มอ้างอิงและกลุ่มเปรียบเทียบ (Reference and Focal group) โดยที่กลุ่มแรกจะใช้อ้างอิงถึงกลุ่มที่คาดว่าจะได้ประโยชน์จากข้อสอบและกลุ่มหลังเป็นกลุ่มเสียประโยชน์จากข้อสอบในกรณีที่ข้อสอบมีความลำเอียง โดยมีการตรวจสอบทุก ๆ ระดับคะแนนรวมจากการสอบ ข้อสอบใดที่ผู้สอบทั้ง 2 กลุ่ม ทำได้เท่า ๆ กัน จะเป็นข้อสอบที่ถือได้ว่าไม่มีความลำเอียงต่อกลุ่มใดกลุ่มหนึ่ง วิธีการในการตรวจสอบมีดังนี้

ถ้าให้ N_{R_j} และ N_{F_j} แทนจำนวนผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ

ที่มีคะแนนรวมในช่วงคะแนน j และให้ $N_j = N_{R_j} + N_{F_j}$ เป็นจำนวนผู้สอบรวม

ทั้ง 2 กลุ่มที่ได้คะแนนรวมจากการสอบอยู่ในช่วงคะแนน j แล้วนำมาเขียนในรูป

ตาราง 2 ทาง แสดงผลการตอบถูก (1) และตอบผิด (0) ดังตาราง 1

ตาราง 1 ข้อมูลผลการสอบของกลุ่มตัวอย่างกลุ่ม R และ F ที่มีคะแนนรวมอยู่ในช่วงคะแนน j

กลุ่มผู้สอบ	คะแนนที่ได้จากข้อสอบที่ต้องการตรวจสอบความลำเอียง			
		1	0	รวม
	R	A_j	B_j	N_{Rj}
	F	C_{Fj}	D_{Fj}	N_{Fj}
	รวม	N_{1j}	N_{0j}	N_j

เมื่อ A_j , B_j , C_j และ D_j เป็นความถี่ที่สังเกตได้ของการตอบถูก (1) และตอบผิด (0) และแสดงในรูปสัดส่วนได้ดังตาราง 2

ตาราง 2 สัดส่วนการตอบข้อสอบของกลุ่มประชากร 2 กลุ่ม

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

โดยที่ P_{Rj} คือ สัดส่วนของกลุ่มอ้างอิงที่อยู่ในช่วงความสามารถ j ที่ตอบข้อสอบถูก

P_{Fj} คือ สัดส่วนของกลุ่มสนใจที่อยู่ในช่วงความสามารถ j ที่ตอบข้อสอบถูก

Q_{Rj} คือ $1 - P_{Rj}$

Q_{Fj} คือ $1 - P_{Fj}$

จากตาราง 1 จะได้ว่าในช่วงระดับคะแนน j A_j และ B_j คือจำนวนผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบข้อที่ถูกต้องและไม่ถูกต้องตามลำดับ ในขณะที่ C_j และ D_j คือจำนวนผู้สอบกลุ่มเปรียบเทียบที่ตอบข้อสอบข้อที่ถูกต้องและไม่ถูกต้องตามลำดับเช่นกัน

ตามเทคนิควิธีแมนเทิล - แฮนส์เซล ที่เสนอโดยฮอลล์แลนด์และเฮเยอร์นั้น ข้อมูลการตอบข้อสอบแต่ละข้อจะถูกนำมาแสดงในรูปตาราง 2 ทาง เป็นจำนวนตารางเท่ากับคะแนนรวมที่มีผู้สอบทำได้บวกด้วย 1 จากนั้นจะคำนวณความเป็นไปได้ในรูปของค่าสัดส่วนการตอบข้อสอบถูก-ผิด ระหว่างกลุ่มโดยกำหนดให้เป็นค่า α ดังนี้

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

ค่า α_{MH} จะมีค่าระหว่าง 0 และ 1 โดยที่ $\alpha_{MH} = 1$ จะแสดงถึงสมมติฐานศูนย์ที่ว่าไม่มีความแตกต่างหรือไม่มีความลำเอียง ซึ่งแสดงว่ากลุ่มตัวอย่างทั้ง 2 กลุ่ม ทำข้อสอบได้ผลเช่นเดียวกันในแต่ละระดับคะแนนรวม ค่า $A_j D_j / N$ และ $B_j C_j / N$ มีค่าเท่ากันทั้งสองกลุ่มทำให้ $\alpha_{MH} = 1$

ในทางปฏิบัติแล้ว แม้ว่าจะมีการควบคุมให้ผู้สอบ 2 กลุ่ม มีความสามารถในสิ่งที่ต้องการวัดเท่ากัน แต่เป็นไปได้ที่ผลการตอบข้อสอบของผู้สอบทั้ง 2 กลุ่ม จะเป็นเช่นเดียวกัน แมนเทิล - แฮนส์เซล จึงเสนอค่าสถิติไคสแควร์เพื่อทดสอบค่าที่ได้ว่าจะมีความแตกต่างจาก 1.00 อย่างมีนัยสำคัญหรือไม่ที่ระดับชั้นแห่งความเป็นอิสระ = 1 ดังนี้

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

โดยที่ค่าผลรวมเป็นค่าที่ได้จากการรวมทุกชั้นของคะแนนรวม และค่า $E(A_j)$ มีค่าดังนี้

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

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

และสมมติฐานศูนย์แสดงว่าข้อสอบไม่ลำเอียง ได้แก่

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

สมมติฐานศูนย์นี้เป็นสมมติฐานของความเป็นอิสระอย่างมีเงื่อนไขของสมาชิกกลุ่ม และคะแนนที่ได้จากการตอบข้อสอบที่ต้องการตรวจสอบความลำเอียงและภายใต้สมมติฐานศูนย์ จะได้ค่าคาดหวังของการตอบข้อสอบดังนี้

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

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

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

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

สมมติฐานอื่นของแมนเทิล - แฮนส์เซล ได้แก่

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

เมื่อ α มีค่า = 1 ซึ่งสอดคล้องกับสมมติฐานศูนย์ จะได้ว่า

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

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

นอกจากนั้นฮอลแลนด์ และเฮเยอร์ ได้เสนอแนะให้แปลงค่า α_{MH} ให้เป็น เดลต้า (Δ) เพื่อให้เป็นค่าที่มีค่าเฉลี่ยเป็น 0 ดังนี้

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

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

เท่าที่ผ่านมามีผู้นำวิธีการวิเคราะห์ความลำเอียงของแมนเทล-แฮนส์เซล มาศึกษาเช่น คลอสเซอร์ และคนอื่น ๆ (Clauser and Others. 1991) ใช้วิธีแมนเทล-แฮนส์เซล วิเคราะห์ความลำเอียงอย่างสม่ำเสมอ พบว่าวิธีแมนเทล-แฮนส์เซล จะใช้ได้ดีในกรณีที่ข้อสอบมีค่าอำนาจจำแนกสูง แต่จะไม่สามารถวิเคราะห์ข้อสอบที่มีค่าความยากมากได้ ซึ่งสอดคล้องกับข้อค้นพบของเมเชอร์ และคนอื่น ๆ (Mazor and Others. 1991) ชูตวิคส์ และทอลแมน (Sudweeks and Tolman. 1990) ที่พบว่าข้อสอบที่ลำเอียงมักเป็นข้อสอบที่ยาก มาก และเฟอรารา (Baghi and Ferrara. 1990) พบว่าเมื่อใช้ขนาดกลุ่มตัวอย่าง 750 คนขึ้นไป วิธีแมนเทล - แฮนส์เซล ใช้แทนวิธีใช้ทฤษฎีการตอบข้อคำถาม 3 พารามิเตอร์ได้ สวามินาธานและโรเจอร์ส (Swaminathan and Rogers. 1990) ได้ศึกษาจากข้อมูลจำลองพบว่า วิธีแมนเทล-แฮนส์เซล วิเคราะห์ได้ดีกว่าการถดถอยแบบโลจิสต์ เล็กน้อย โดยตรวจค้นได้ถูกต้องร้อยละ 75 ในกรณีใช้กลุ่มตัวอย่าง 250 คน และตรวจค้นได้ถูกต้องร้อยละ 100 กรณีที่กลุ่มตัวอย่าง 500 คน กรณีที่ลำเอียงอย่างสม่ำเสมอและกรณีไม่สม่ำเสมอที่ติดกันที่ปลายข้างใดข้างหนึ่ง แต่วิธีแมนเทล-แฮนส์เซล มีค่าใช้จ่ายน้อยกว่าโลจิสต์ประมาณ 3-4 เท่า แฮมเบิลตันและคนอื่น ๆ (Hambleton and Others. 1986) พบว่าวิธีแมนเทล-แฮนส์เซล ให้ค่าใกล้เคียงกับวิธีของทฤษฎีการตอบข้อคำถาม ทั้งที่ใช้ค่าความแตกต่างของค่าเฉลี่ยกำลังสอง และการตรวจสอบความแตกต่างของพื้นที่รวมได้ไค์ แต่วิธีแมนเทล-แฮนส์เซล มีค่าใช้จ่ายต่ำกว่า และใช้เวลาน้อยกว่า ลินเคอร์ (Linacre. 1986) พบว่า วิธีแมนเทล-แฮนส์เซล ใช้ได้ดีในสถานการณ์จำลองทุกสถานการณ์พอ ๆ กับโปรแกรม PROX ของราสช์ (Rasch) แต่ควรใช้เกณฑ์ในการคำนวณความลำเอียง และความคลาดเคลื่อนหลาย ๆ เกณฑ์ และใช้ตัวประมาณค่าความคลาดเคลื่อนมาตรฐานมากกว่า 1 ตัว เปริลแมน และคนอื่น ๆ (Perlman and Others. 1988) พบว่าวิธีแมนเทล-แฮนส์เซล มีปัญหาด้านความเชื่อมั่น เมื่อจำนวนกลุ่มตัวอย่างน้อยกว่า 660 คน ธิสเซน สไตน์เบิร์ก และวายเนอร์ (Thissen, Steinberg and

Wainer. 1988) พบว่าวิธีแมนเทิล-แฮนส์เซล ให้ผลการวิเคราะห์คล้ายกับวิธี ทฤษฎีการตอบข้อคำถามแบบกระตอยเชิงเส้น (IRT-LR) และอาจใช้วิเคราะห์ ความลำเอียงก่อนใช้วิธีทฤษฎีการตอบข้อคำถามแบบกระตอยเชิงเส้น (IRT-LR)

การตรวจสอบความลำเอียงของข้อสอบวิธีการทดสอบ SIB (SIBTEST)

การทดสอบ SIB เป็นวิธีการตรวจสอบความลำเอียงของแบบทดสอบชนิด พหุมิติ (Multidimensional) โดยมีพื้นฐานจากทฤษฎีการตอบข้อคำถามแบบพหุมิติ แต่การทดสอบ SIB แตกต่างจากวิธีอิงทฤษฎีการตอบข้อคำถามตรงที่เป็นการใช้การ ทดสอบค่าสถิติไร้พารามิเตอร์ (non - parametric) นอกจากนั้นยังไม่มี การใช้ ฟังก์ชันการตอบข้อสอบหรือการประมาณค่าความสามารถที่มีอยู่ของบุคคล และมีการ คำนวณที่ง่ายไม่ซับซ้อน ใช้ได้ดีสำหรับการตรวจสอบความลำเอียงอย่างสม่ำเสมอ และมีทิศทางเดียว (Unidirectional)

การทดสอบ SIB เป็นวิธีการตรวจสอบความลำเอียงของข้อสอบที่เสนอ โดยซิลลี และสแตร์ท (Shealy and Stout. 1992 : 2 ; citing Shealy. 1989 Shealy and Stout. 1991) และรูปแบบอยู่ที่การวิเคราะห์ข้อมูลที่มี ข้อตกลงว่ามีมิติการวัด 2 มิติ มิติหนึ่งเป็นความสามารถหรือลักษณะแฝงเข้าหมายที่ ต้องการวัด และอีกมิติเป็นลักษณะแฝงสอดแทรกที่ไม่ต้องการวัด

ซิลลี และสแตร์ท (Shealy and Stout. 1992) เริ่มจากกลุ่มผู้สอบ 2 กลุ่ม คือ กลุ่มอ้างอิงและกลุ่มเปรียบเทียบ การตอบข้อสอบของผู้สอบที่สุ่มมาจาก กลุ่มผู้สอบคือ U โดยที่ $U = (U_1, \dots, U_n)$ เมื่อ U_i มีค่าเป็น 0 ถ้าตอบ ข้อสอบข้อ i ผิด และเป็น 1 เมื่อตอบถูก

สำหรับโมเดลอิงทฤษฎีการตอบข้อคำถามแล้ว โดยทั่ว ๆ ไป U จะเกิด ได้จาก 2 องค์ประกอบ คือ พารามิเตอร์ความสามารถของผู้สอบจำนวน d มิติ และ ฟังก์ชันการตอบข้อสอบแต่ละข้อ ซึ่งเป็นตัวกำหนดความน่าจะเป็นของการตอบข้อสอบ ได้ถูก

ในที่นี้ d คือ 2 เพราะเราจะพิจารณาความสามารถแทรกซ้อนเพิ่มเข้ามา จากความสามารถเข้าหมายที่ต้องการวัด เวกเตอร์ความสามารถในที่นี้คือ (θ, η) เมื่อ θ คือ ความสามารถเข้าหมาย และ η คือความสามารถสอดแทรก ส่วน ฟังก์ชันการตอบข้อสอบสำหรับข้อ i คือ $P_i(\theta, \eta)$ โดยที่ข้อสอบทุกข้อจะประกอบด้วย θ ในขณะที่บางข้อจะมี η อยู่ด้วย สำหรับข้อที่ประกอบด้วย θ อย่างเดียว

ฟังก์ชันการตอบข้อสอบคือ $P_i(\theta)$ ค่า $P_i(\theta, \eta_i)$ จะสูงขึ้นใน θ และใน η_i เมื่อข้อสอบข้อ i ต้องการความสามารถทั้ง 2 อย่าง และจะสูงขึ้นใน θ ถ้าต้องการเฉพาะความสามารถ θ

นิยามของความลำเอียงของข้อสอบในการทดสอบ SIB คือ ความลำเอียงของข้อสอบจะเกิดขึ้นต่อกลุ่มเปรียบเทียบที่ระดับความสามารถ θ ถ้าฟังก์ชันการตอบข้อสอบสำหรับความสามารถเป้าหมาย (target ability) ของกลุ่มอ้างอิง มีค่ามากกว่าฟังก์ชันการตอบข้อสอบของกลุ่มเปรียบเทียบ (Stout, 1990 : 7-8) ซึ่งแสดงได้ในรูปของ

$$\begin{aligned}
 & T_R(\theta) > T_F(\theta) \\
 \text{เมื่อ} \quad & T_R(\theta) = E_R [P(\theta, n) \mid \theta] \\
 & = \int P(\theta, n) f_R(n \mid \theta) dn \\
 & T_F(\theta) = E_F [P(\theta, n) \mid \theta] \\
 & = T_F(\theta) \\
 & \theta = (\theta, n)
 \end{aligned}$$

โดยที่ R คือ กลุ่มอ้างอิง
 F คือ กลุ่มเปรียบเทียบ
 θ คือ ความสามารถเป้าหมาย
 n คือ ความสามารถสอดแทรกซึ่งอาจเป็นพหุมิติก็ได้
 $T_R(\theta)$ คือ ฟังก์ชันการตอบข้อสอบของกลุ่มอ้างอิง
 $T_F(\theta)$ คือ ฟังก์ชันการตอบข้อสอบของกลุ่มเปรียบเทียบ
 ความลำเอียงของข้อสอบจะเกิดขึ้นเมื่อ
 $\eta_F \mid \theta$ คือ $\eta_R \mid \theta$

นิยามความลำเอียงในแง่ของแบบทดสอบทั้งฉบับ เมื่อให้ U เป็นการตอบข้อสอบแบบสุ่มของผู้สอบคนหนึ่ง $h(U)$ คือคะแนนจากแบบทดสอบ โดยที่ $h(U) = \sum U_i$ แล้ว จะได้ว่าความลำเอียงของข้อสอบจะเกิดต่อกลุ่มเปรียบเทียบในระดับความสามารถ θ ถ้า

$$\begin{aligned}
 & T_R(\theta) > T_F(\theta) \\
 \text{เมื่อ} \quad & T_R(\theta) = \text{def } E_R [h(U) \mid \theta] \\
 & T_F(\theta) = \text{def } E_F [h(U) \mid \theta]
 \end{aligned}$$

โดยที่ปริมาณของความลำเอียงที่ระดับ θ คือ $\beta(\theta) \text{ def } T_R(\theta) - T_F(\theta)$ และดัชนีของความลำเอียงคือ $\beta_U = \int \beta(\theta) f_F(\theta) d(\theta)$ ซึ่งเป็นความลำเอียงเฉลี่ยต่อผู้สอบแต่ละคนในกลุ่มที่สนใจศึกษา

สมมติฐานเพื่อการทดสอบความลำเอียงของข้อสอบหรือแบบทดสอบของ SIBTEST

$$H_0 : \beta = 0 ; H_1 : |\beta| > 0$$

การทดสอบทำได้โดยแบ่งแบบทดสอบออกเป็นแบบทดสอบย่อย 2 ฉบับ คือ V เป็นแบบทดสอบที่มีความเที่ยงตรง และ S เป็นแบบทดสอบที่ต้องการศึกษา และคะแนนของแต่ละแบบทดสอบย่อยคือ

$$Y = \frac{\sum U_i}{S}$$

$$X = \frac{\sum U_i}{V}$$

สำหรับการตรวจสอบความลำเอียงของข้อสอบรายข้อ S คือ ข้อสอบแต่ละข้อที่ตรวจสอบ และ V คือ ข้อสอบที่เหลือในแต่ละรอบของการตรวจสอบ

Y_{Rk} คือ คะแนนเฉลี่ยที่ได้จากแบบทดสอบย่อยที่ศึกษาสำหรับผู้เข้าสอบทุกคนที่มีคะแนนรวมในแบบทดสอบย่อยที่ถือว่าเป็นแบบทดสอบที่มีความเที่ยงตรง

$(x) = k$ และความลำเอียงวัดได้จาก $(\bar{Y}_{Rk} - \bar{Y}_{Fk})S$ ทั้งนี้เพราะ

$\bar{Y}_{Rk} - \bar{Y}_{Fk}$ เป็นความแตกต่างของการตอบของกลุ่มย่อยที่มีความสามารถระดับ

เดียวกัน ($= k$) บนความสามารถ เข้าหมาย และกรณีที่ข้อสอบไม่มีความลำเอียง

ค่า $\bar{Y}_{Rk} - \bar{Y}_{Fk} = 0$

สำหรับค่าสถิติที่ประมาณค่าความลำเอียง β ได้จาก

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

เมื่อ P_k = สัดส่วนของผู้สอบกลุ่มที่สนใจศึกษาที่ตอบแบบทดสอบย่อยที่ถือว่าไม่ลำเอียงได้ถูก k ข้อ

สถิติทดสอบของ SIBTEST คือ

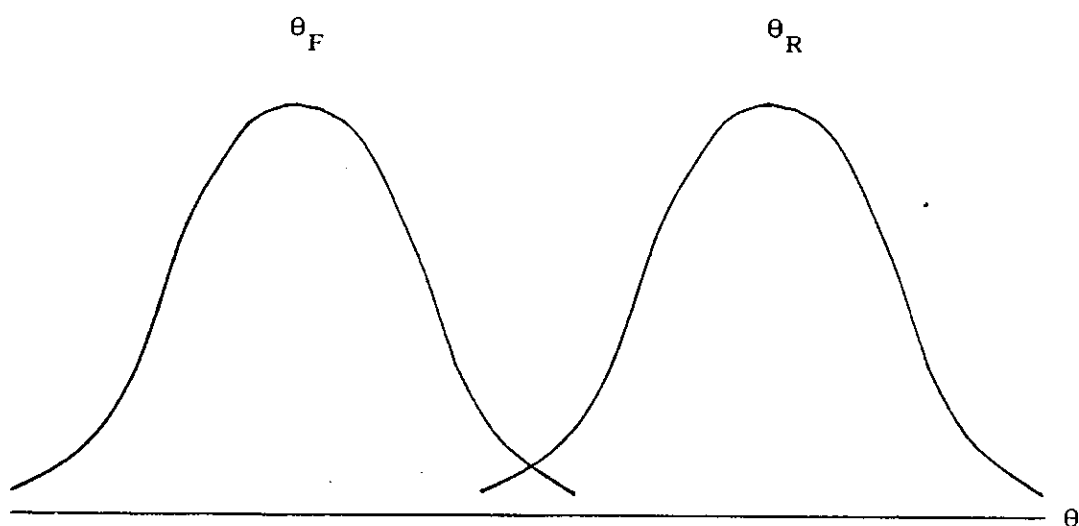
$$\beta = \frac{\hat{\beta}}{(\hat{\beta})} \text{ ซึ่ง } = N(0, 1) \text{ เมื่อ } \beta = 0$$

โดยที่

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

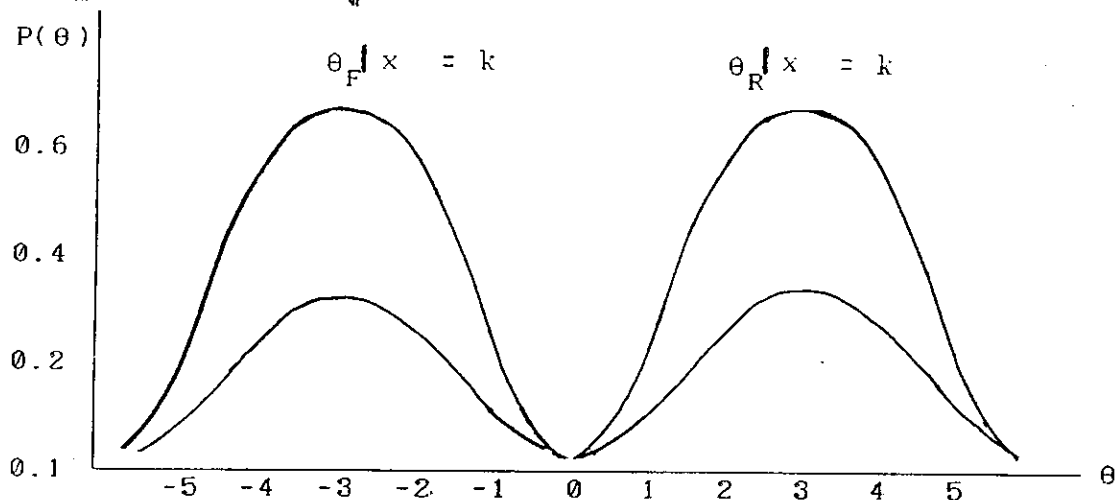
เมื่อ $\theta^2(y | k, g)$ เป็นความแปรปรวนของคะแนนของกลุ่มอ้างอิง และกลุ่มเปรียบเทียบในช่วงคะแนน k

อย่างไรก็ตามเมื่อมีความแตกต่างระหว่างความสามารถเป้าหมาย ค่าสถิติ β จะสูงกว่าปกติ ทำให้เกิดการระบุว่าข้อสอบข้อนั้นลำเอียงทั้ง ๆ ที่ไม่ลำเอียง ดังแสดงในภาพประกอบ 2



ภาพประกอบ 2 แสดงความแตกต่างระหว่างกลุ่มประชากรย่อยในความสามารถเป้าหมาย (θ)

เมื่อแบ่งตามระดับความสามารถตามแบบทดสอบย่อยที่ถือว่าไม่ลำเอียง
จะปรากฏปัญหาการมีค่าสถิติสูงกว่าปกติดังแสดงในภาพประกอบ 3



ภาพประกอบ 3 แสดงค่าสถิติที่แบ่งตามระดับความสามารถตามแบบทดสอบย่อย

นั่นคือ ที่ระดับความสามารถ k เราคาดได้ว่า แม้ว่าข้อสอบไม่ลำเอียงแล้ว
ผู้สอบในกลุ่มสนใจศึกษา จะมีความสามารถต่ำกว่าผู้สอบในกลุ่มอ้างอิง นั่นคือ

$$\bar{Y}_{Rk} - \bar{Y}_{Fk} > 0 \quad \text{ทุก } k \text{ ค่าของ } k$$

ดังนั้นจึงใช้ค่าแก้ด้วยการใช้ค่าถดถอยตามทฤษฎีการตอบข้อคำถามชนิด
3 พารามิเตอร์ เพื่อกำจัดค่าเพื่อของสถิติ โดยการแปลงค่า

$$\bar{Y}_{Rk} - \bar{Y}_{Fk} \text{ ให้เป็น } \bar{Y}_{Rk}^* - \bar{Y}_{Fk}^*$$

โดยที่ $\bar{Y}_{gk}^* = \bar{Y}_{gk} + M_{gk} [V(K) - V_g(k)]$; g คือ R หรือ F

$$\text{และ } M_{gH} = \frac{Y_{g, k+1} - Y_{g, k-1}}{V_g(k+1) - V_g(k-1)}$$

$$V(k) = \frac{1}{2} [V_R(k) + V_F(k)]$$

การทดสอบนัยสำคัญทางสถิติเมื่อ $H_0 : \beta = 0$ ได้แก่

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

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

แหล่ง/สาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบ

ชูเนแมน (Scheuneman. 1982 : 181-198) ได้ศึกษาข้อสอบหลังจากการวิเคราะห์ด้วยค่าสถิติเพื่อตรวจสอบความลำเอียงแล้ว และจากการศึกษาผลการวิเคราะห์ความลำเอียงของข้อสอบในแบบทดสอบมาตรฐานหลายฉบับ พบว่าสามารถวิเคราะห์ข้อสอบที่ลำเอียงเหล่านั้นได้ว่า น่าจะเกิดจากแหล่ง/สาเหตุใดมากที่สุด ซึ่งเขาได้สรุปแหล่ง/สาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบมากที่สุดได้ดังนี้

1. การเดา (guessing) ซึ่งอาจจะเกิดจากข้อสอบยากเกินไปหรือเวลาไม่เพียงพอ ก่อให้เกิดความไม่เท่าเทียมกันในโอกาสการตอบข้อสอบข้อนั้นถูก

2. ความเร็ว (Speed) หรือเวลาในการตอบ ซึ่งจะเป็นผลให้เกิดการเดาหรือทำไม่ทันในกรณีเวลาน้อย จะมีผลกับข้อสอบข้อหลัง ๆ โดยเฉพาะในการศึกษาความลำเอียงของข้อสอบวัดความถนัด

3. ความกำกวมหรือความไม่ชัดเจน (Unclear) ของข้อคำถาม นั่นคือข้อคำถามขาดความเป็นปรนัย หรือใช้ภาษา/คำที่ไม่สื่อเป็นภาษากลาง ซึ่งจะก่อให้เกิดความลำเอียงเฉพาะกลุ่มภาษาใดภาษาหนึ่งได้

4. ลำดับชั้นของคำถาม (Series) จะเป็นสิ่งที่อาจจะก่อให้เกิดความสับสน หรือชี้แนะคำตอบบางข้อได้

5. สถานภาพทางสังคมหรือความเกี่ยวข้องทางสังคม (Social implication) ก็จะเป็นสิ่งที่ก่อให้เกิดความแตกต่างระหว่างกลุ่มได้

6. ประสบการณ์/การฝึกของแต่ละกลุ่มย่อย เป็นสิ่งที่ก่อให้เกิดการได้เปรียบเสียเปรียบค่อนข้างจะชัดเจน

7. องค์ประกอบทางวัฒนธรรม ความเป็นอยู่ ประเพณี เชื้อชาติก็จะเอื้อให้กับบางกลุ่มย่อย

วิธีการคิดคะแนน

วิธีการคิดคะแนนผลการสอบของแบบทดสอบแบบเลือกตอบนั้น นอกจากจะให้คะแนนแบบข้อถูกให้ 1 คะแนน และข้อผิดให้ 0 คะแนนแล้ว นักทดสอบได้พยายามศึกษาวิธีการคิดคะแนนของข้อสอบอีกหลายวิธีเช่น

1. วิธีการให้น้ำหนักคะแนนตัวเลือกแตกต่างกัน ซึ่งแบ่งได้เป็น 3 รูปแบบคือ

1.1 วิธีอาศัยความเห็นของผู้เชี่ยวชาญ โดยอาศัยความเห็นของผู้เชี่ยวชาญว่าตัวเลือกแต่ละตัวมีความถูกต้องมากน้อยเพียงใด หรือใช้ค่าเฉลี่ยของลำดับที่ของความถูกต้องของตัวเลือกแต่ละตัวที่ตัดสินโดยผู้เชี่ยวชาญ

1.2 วิธีใช้ค่าความสัมพันธ์ระหว่างตัวเลือกกับสิ่งที่จะทดสอบ ซึ่งอาจจะใช้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวเลือกกับคะแนนรวม หรือใช้สเกลวัดที่คำนวณจากวิธีของกัตต์แมน (Guttman's scale Analysis) หรือใช้ค่าสเกลวัดที่คำนวณจากวิธีการเปรียบเทียบรายคู่ของเทอร์สโตน (Thurstone's paired Comparison)

1.3 วิธีอาศัยความเชื่อมั่นของผู้สอบต่อตัวเลือกของแบบทดสอบ อาจจะคิดตามความมั่นใจในการตอบ หรือโดยการหักค่าการเดา หรือโดยวิธีการกำจัดตัวเลือก

อย่างไรก็ตามได้มีผู้ศึกษาผลของการคิดคะแนนตามวิธีดังกล่าว พบว่าการให้คะแนนรายข้อโดยวิธีต่าง ๆ นี้ ยังให้ผลไม่ดีเท่าที่ควร ซึ่ง กิลฟอร์ด โลเวลล์ และวิลเลียม (จำเนียร สุธหลาย. 2530 : 4 ; อ้างอิงมาจาก Guilford, Lovell and Williams, n.d.) สรุปว่าเมื่อเปรียบเทียบกับวิธีการให้คะแนนแบบตอบถูกให้ 1 คะแนน และตอบผิดให้ 0 คะแนนแล้ว การให้คะแนนรายข้อของวิธีต่าง ๆ ได้ผลไม่คุ้มค่ากับความยุ่งยากในการคำนวณ (Wang and Stanley. 1970 : 688)

2. วิธีการคิมน้ำหนักคะแนนข้อสอบแตกต่างกัน ซึ่งแบ่งได้เป็น 3 วิธีคือ

2.1 วิธีอาศัยความเห็นของผู้เชี่ยวชาญ

2.2 วิธีใช้ค่าถดถอยพหุคูณ

2.3 วิธีคิดคะแนนโดยอาศัยรูปแบบความสามารถ

สำหรับวิธีที่ 2 นี้ โดยเฉพาะวิธีที่ 2.2 และ 2.3 เป็นความหวังที่นักวัดผลจะนำมาใช้เพื่อจะช่วยให้ผลจากการวัดของแบบทดสอบมีความเชื่อมั่นและเที่ยงตรงมากขึ้น (Wang and Stanley. 1970 : 670) ซึ่งนับว่าเป็นผลดีแก่ผู้สอบและผู้ใช้ผลการสอบด้วย ซึ่งในการศึกษาคั้งนี้ผู้วิจัยเลือกวิธีการคิดคะแนนแบบต่าง ๆ ดังนี้

1. วิธีการคิดคะแนนมาตรฐานที่-ปกติ เป็นวิธีการแปลงคะแนนดิบให้มีหน่วยวัดคงที่ และสามารถนำคะแนนที่แปลงแล้วรวมกันได้

2. วิธีการคิดคะแนนน้ำหนักความสามารถ เป็นวิธีการคิดคะแนนโดยอาศัยความสามารถแท้ตามแนวทฤษฎีการตอบข้อคำถาม (Item Response Theory) แบบ 3 พารามิเตอร์ ภายใต้ความเชื่อที่ว่า โอกาสที่ผู้สอบจะทำข้อสอบข้อใด ๆ ได้ถูกจะขึ้นอยู่กับความแตกต่างของความสามารถของผู้สอบ (θ) ความยากง่ายของข้อสอบ ค่าอำนาจจำแนกและโอกาสการเดา (Wright and Stone. 1979 : 15) ซึ่งลอร์ด (Lord. 1968 : 989) ได้ทำการวิเคราะห์ผลการสอบวัดความถนัดทางการเรียน (Scholastic Aptitude Test) โดยกำหนดค่าน้ำหนักคะแนนของผู้สอบแต่ละคน และแต่ละข้อต่างกันตามความสามารถ (θ) ซึ่งลอร์ด (Lord. 1980 : 75) ได้ให้สูตรในการคำนวณดังนี้

$$W_i(\theta) = \frac{D_{ai}(b_i - c_i)}{b_i(1 - c_i)} = \frac{D_{ai}}{1 + c_{ie}^{-DL_i}}$$

เมื่อ	W_i	คือ น้ำหนักสำคัญข้อที่ i
	θ	คือ คะแนนมาตรฐานความสามารถ
	a	คือ ค่าอำนาจจำแนกข้อที่ i
	b	คือ ค่าความยากของข้อที่ i
	c	คือ ค่าการเดาของข้อที่ i
	D	คือ 1.7
	L_i	คือ $a_i(\theta - b_i)$

งานวิจัยที่เกี่ยวข้องกับการตรวจสอบความลำเอียง

งานวิจัยที่เกี่ยวข้องกับการตรวจสอบความลำเอียงมีมากมายพอจะสรุปได้

ดังนี้

ฮัมฟรีสและทาเบอร์ (Humphreys and Taber. 1973) ใช้การวิเคราะห์ตัวประกอบศึกษาความแตกต่างของผู้ที่มีระดับเชาวน์ปัญญา และมีสถานภาพทางเศรษฐกิจต่างกัน โดยเปรียบเทียบจำนวนตัวประกอบและน้ำหนักตัวประกอบข้ามกลุ่มจากแบบทดสอบวัดผลสัมฤทธิ์และแบบทดสอบความถนัด จำนวน 21 ฉบับ พบว่าในด้านของจำนวนตัวประกอบไม่แตกต่างกันแต่น้ำหนักตัวประกอบพบว่าแตกต่างกันในกลุ่มที่มีสติปัญญาต่างกัน ซึ่งการพิจารณาน้ำหนักของตัวประกอบเป็นข้อมูลในการตรวจสอบความลำเอียงของข้อสอบ

นินเจสเตอร์ (Berk. 1982 : 166 ; Citing Nungester. 1977) ได้วิเคราะห์ความลำเอียงของข้อสอบโดยใช้วิธีแปลงค่าความยากที่เสนอโดยแองกอฟฟ์ (Angoff. 1972) วิธีโคสแควร์ (Scheuneman. 1975) และวิธีแปลงค่าความยากที่เสนอโดยฟิชเบย์ (Fishbein. 1975) จากการใช้แบบทดสอบมาตรฐานของฟลอริดา ระดับ 9 (Florida Statwise Ninth Grade) ฉบับวิชาคณิตศาสตร์และวิชาภาษาอังกฤษ โดยวิเคราะห์ข้อมูลหาสหสัมพันธ์ระหว่างวิธีการวิเคราะห์ความลำเอียงแบบต่าง ๆ ผลการวิจัยพบว่าสหสัมพันธ์ของวิธีการวิเคราะห์ความลำเอียงทั้ง 3 วิธีในแบบทดสอบวิชาคณิตศาสตร์ มีค่าระหว่าง .87-.98 และในแบบทดสอบวิชาภาษาอังกฤษมีค่าระหว่าง .55-.77 ค่าความเชื่อมั่นของแบบทดสอบเมื่อคัดข้อสอบที่มีความลำเอียงออกแล้วมีค่าลดลงเล็กน้อย

รัตเนอร์และคนอื่นๆ (Subkoviak. 1984 : 5 ; Citing Rudner and Others. 1978) วิเคราะห์ความลำเอียงของข้อสอบโดยใช้วิธีแปลงค่าความยาก วิธีโคสแควร์ วิธีโค้งลักษณะข้อสอบ และวิธีวิเคราะห์องค์ประกอบ โดยใช้ข้อมูลจากแบบทดสอบมาตรฐานของแสตนฟอร์ด (Stanford Achievement) ฉบับความเข้าใจในการอ่าน ผลการวิจัยพบว่า วิธีโค้งลักษณะข้อสอบและวิธีแปลงค่าความยากเป็นวิธีที่เหมาะสม วิธีโคสแควร์เป็นวิธีที่เหมาะสมรองลงมา

เมอร์ส และกรอสเซน (Berk. 1982 : 164 ; citing Merz and Grossen. 1979) ได้วิเคราะห์ความลำเอียงของข้อสอบโดยใช้ข้อมูลสมมุติ ด้วยวิธีแปลงค่าความยากพอยท์ไบซีเรียล โคสแควร์ การวิเคราะห์องค์ประกอบ วิธีโค้งลักษณะข้อสอบ 1 และ 3 พารามิเตอร์ จากการศึกษาพบว่า วิธีวิเคราะห์ด้วยพอยท์ไบซีเรียลเป็นวิธีที่ให้ผลการวิเคราะห์ สอดคล้องกับวิธีอื่นน้อย วิธีแปลงค่าความยากให้ผลการวิเคราะห์สอดคล้องกับวิธีอื่นมาก รวมทั้งวิธีโค้งลักษณะข้อสอบ 1 และ 3 พารามิเตอร์ ให้ผลการวิเคราะห์สอดคล้องกับวิธีอื่นปานกลาง แต่วิธีแปลงค่าความยากเป็นวิธีที่เหมาะสม เพราะไม่มีข้อจำกัดในด้านจำนวนกลุ่มตัวอย่างและการคำนวณ

ไฮรอนสัน และสับโคเวียค (Subkoviak. 1984 : 50-51 ; citing Ironson and Subkoviak. 1979) ได้ศึกษาวิธีการวิเคราะห์ความลำเอียงของข้อสอบโดยวิธีแปลงค่าความยาก วิธีพอยท์ไบซีเรียล วิธีไคสแควร์ และวิธีไคงลักษณะข้อสอบ 3 พารามิเตอร์ โดยใช้ผลการสอบของนักเรียนผิวดำและผิวขาว ผลการวิจัยพบว่าวิธีไคงลักษณะข้อสอบ 3 พารามิเตอร์ เป็นวิธีที่ให้ผลการวิเคราะห์ที่ถูกต้องที่สุด แต่มีข้อจำกัดคือต้องใช้กลุ่มตัวอย่างมาก ส่วนวิธีแปลงค่าความยาก และวิธีไคสแควร์ให้ผลที่ถูกต้องรองลงมา

ผจงจิต อินทสุวรรณ (Intasuwat. 1979 : 92) ได้ศึกษาวิธีวิเคราะห์ความลำเอียงของข้อสอบ 3 วิธีคือ วิธีไคงลักษณะข้อสอบ 3 พารามิเตอร์ ไคงลักษณะข้อสอบโมเดลของราล์ฟ และวิธีไคสแควร์ ใช้ข้อมูลจากการสอบ ไอ อี เอ (IEA : The International Association of The Evaluation of Educational Achievement) จากการศึกษาพบว่าวิธีวิเคราะห์ความลำเอียงทั้ง 3 วิธี มีความสัมพันธ์ระหว่าง .51-.98 โดยที่วิธีวิเคราะห์แบบไคสแควร์ และวิธีราล์ฟ มีค่าสหสัมพันธ์สูงถึง .98

สงขม ลักษณะ (Burrill. 1982 : 168 ; citing Laksana. 1979) ได้วิเคราะห์ความลำเอียงของข้อสอบโดยใช้วิธีไคงลักษณะข้อสอบ 3 พารามิเตอร์ และวิธีวิเคราะห์ความแปรปรวน โดยศึกษาความลำเอียงทางเพศและเชื้อชาติ ใช้ข้อมูลจากการสอบแบบทดสอบทักษะพื้นฐานของไอโอวา (Iowa Test of Basic Skills) ฉบับคำศัพท์และคณิตศาสตร์ ผลการศึกษาพบว่า วิธีวิเคราะห์ความแปรปรวนเป็นวิธีที่ใช้ได้ดีในทางปฏิบัติ แต่ให้ความหมายไม่กว้างนัก วิธีไคงลักษณะข้อสอบ 3 พารามิเตอร์ เป็นวิธีที่ให้ผลการวิเคราะห์ที่ครอบคลุม แต่มีข้อจำกัดที่จะต้องใช้กลุ่มตัวอย่างขนาดใหญ่ และต้องใช้เวลาคำนวณยุ่งยากซับซ้อน

รัดเนอร์ และคนอื่น ๆ (Subkoviak. 1984 : 51 ; citing Rudner and Others. 1980) ได้ศึกษาการวิเคราะห์ความลำเอียงของข้อสอบ 7 วิธี คือ วิธีแปลงค่าความยาก 2 วิธี วิธีไคสแควร์ 2 วิธี และวิธีไคงลักษณะข้อสอบ 3 วิธี โดยใช้ข้อมูลสมมติ (Simulate Data) ผลการศึกษาพบว่าวิธีไคงลักษณะข้อสอบ 3 พารามิเตอร์ เป็นวิธีค้นความลำเอียงของข้อสอบจากข้อมูลสมมติได้ดีที่สุด รองลงมาได้แก่ หนึ่งในสองของวิธีไคสแควร์ ส่วนวิธีแปลงค่าความยากของข้อสอบสามารถใช้ได้ในทางปฏิบัติ

รัดเนอร์ เกทสัน และไนท์ (Rudner Getson and Knight. 1980) ได้เปรียบเทียบเทคนิคการตรวจค้นความลำเอียงของข้อสอบ 3 วิธีคือ เทคนิคความ

ยากแปลงชนิดแกนสำคัญ (Transformed Item Difficulties Major Axis
 . = TID - MA) เทคนิคความยากแปลงเส้นแกน 45 องศา (Transformed Item
 Difficulties - 45 Line = TID - 45) เทคนิคทฤษฎีการตอบข้อสอบ
 3 พารามิเตอร์ (ICC-3) และ 1 พารามิเตอร์ ชนิดค่าสถิติความเหมาะสม (ICC-
 1F) ชนิดความแตกต่างในความง่ายของข้อสอบ (ICC-1E) เทคนิคไคสแควร์แบบ
 5 ช่วง (CHI-5) และเทคนิคไคสแควร์ชนิดหลายช่วง (CHI-N) ผลจากการ
 วิเคราะห์ข้อมูลพบว่าเมื่อใช้ความยาวของข้อสอบหลายขนาดไม่พบว่ามีเทคนิคใดที่มี
 ผลจากความยาวของข้อสอบเป็นพิเศษ เมื่อพิจารณาจากค่าสหสัมพันธ์ของผลการ
 ตรวจค้นกับข้อมูลที่สร้างขึ้น พบว่า วิธีที่ถูกต้องมากที่สุด ได้แก่ ICC-3 CHI-5
 TID-45 ($r = .80, .73$ และ $.68$) และที่ถูกต้องน้อยที่สุดได้แก่ ICC-1F
 ($r = .55$) และทุกวิธี ยกเว้น ICC-1F สามารถตรวจค้นความลำเอียงได้ในกรณี
 ที่มีความยากต่างกันเท่ากัน มีเพียง ICC-1F ที่มีความไวต่อความลำเอียงในอำนาจ
 จำแนกของข้อสอบด้วย เทคนิค CHI-5 มีประสิทธิภาพเท่า ๆ กับ ICC-3 และพบ
 ว่าถ้าเพิ่มจำนวนช่วงความสามารถมากขึ้น จะลดความถูกต้องในการตรวจค้นลง

ดอนลัน ฮิคส์ และวอลมาร์ค (Carton and Marco. 1982 : 289 ;
 citing Donlon Hicks and Wallmark. 1980) ได้วิเคราะห์ความลำเอียง
 ของข้อสอบด้วยวิธีกำหนดจุดเดลตา (Delta-plot Method) ใช้ผลการทดสอบ
 แบบทดสอบวัดความถนัด (Graduate Record Examination Aptitude Test)
 มี 3 ส่วน คือ คำศัพท์ที่มีความหมายตรงกันข้าม ความหมายคล้ายคลึงและแต่ง
 ประโยค ส่วนที่ 2 วัดความเข้าใจในการอ่าน ส่วนที่สามเป็นข้อสอบด้านคณิตศาสตร์
 และสถิติ การศึกษาความลำเอียงของข้อสอบต่อเพศ โดยศึกษากับกลุ่มตัวอย่าง
 นักเรียนชายผิวขาว 1,720 คน นักเรียนหญิงผิวขาว 1,735 คน จากการศึกษา
 พบว่าแบบทดสอบแต่ละส่วนมีข้อสอบที่ลำเอียง

เชปเพิร์ด คามิลลี และเอเวอร์ริลล์ (Berk. 1982 : 169 ; citing
 Shepard Camilli and Averill. 1980) ได้เปรียบเทียบวิธีวิเคราะห์ความ
 ลำเอียงของข้อสอบจากแบบทดสอบ Lorge-Thorndike แบ่งเป็นฉบับภาษา กับไม่ใช้
 ภาษา ใช้กลุ่มตัวอย่างเป็นนักเรียนผิวขาว ผิวดำ และชิคาโน (Chicano) เชื้อชาติ
 ละ 500 คน วิเคราะห์ความลำเอียงด้วยวิธีไคสแควร์ วิธีกำหนดจุดเดลตา หรือวิธี
 แปลงค่าความยากพอยท์ไบซีเรียล ได้ังลักษณะข้อสอบ 3 พารามิเตอร์ และได้ังลักษณะ
 ข้อสอบ 1 พารามิเตอร์ (Rasch Model) การวิเคราะห์ความลำเอียงใช้ทั้งเกณฑ์
 ภายนอกและเกณฑ์ภายใน โดยที่เกณฑ์ภายนอกคือ คะแนนจาก Raven's Coloured

Progressive Matrices เกณฑ์ภายในคือคะแนนรวมจากแบบทดสอบ ผลการวิจัยพบว่าวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ เป็นวิธีที่เหมาะสมในทางทฤษฎี ถูกต้องและตรงที่สุด แต่มีข้อจำกัดคือต้องใช้กลุ่มตัวอย่างขนาดใหญ่ ส่วนวิธีอื่น ๆ ก็สามารถใช้ได้

อัลเดอร์แมน และฮอลแลนด์ (Carton and Marco. 1982 : 289 ; citing Alderman and Halland. 1981) ได้วิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีกำหนดจุดค่าเดลตา (Delta - Plot Method) และวิธีวิเคราะห์โดยล็อก-ลิเนียร์ โมเดล (Log-Linear Model) เป็นการหาโมเดลที่เหมาะสมกับข้อมูล เพื่อตรวจสอบว่าโอกาสในการตอบถูกของข้อสอบขึ้นอยู่กับระดับคะแนนและกลุ่มหรือไม่ โดยใช้ไคสแควร์ตรวจสอบความเหมาะสมของโมเดลกับข้อมูล ใช้แบบทดสอบโทเฟล (TOFEL) ของเดือน พฤศจิกายน 1976 และ 1979 ศึกษาตัวแปรทางด้านภาษา 7 กลุ่ม คือ เยอรมัน (Germanic) สเปน (Spanish) อัฟริกัน (African) จีนกลุ่ม 1 (Chinese I) จีนกลุ่ม 2 (Chinese II) ญี่ปุ่น (Japanese) และอารบิก (Arabic) ขนาดกลุ่มตัวอย่างจำนวน 500 - 1,000 คน เกณฑ์การประเมินความลำเอียงใช้ภาษาจีน 2 กลุ่ม ในการตรวจสอบผลและการแปลความหมายในแต่ละกลุ่มใช้ตารางขนาดกลุ่ม $\times$ ระดับคะแนน 9 ช่วง $\times$ ผลการตอบ พบว่าในแบบโทเฟลชุดที่ 1 ค่าไคสแควร์ในแต่ละส่วนมีค่าสูงมาก ซึ่งแสดงว่าข้อสอบมีการเบี่ยงเบน ควรให้ผู้เชี่ยวชาญทางภาษาพิจารณาและทำการวิเคราะห์ซ้ำ

สตรีกเกอร์ (Carton and Marco. 1982 : 289 ; citing Stricker. 1981) ได้เปรียบเทียบดัชนีการตัดสินความลำเอียงในวิธีวิเคราะห์ความลำเอียง 3 วิธี คือ ระยะห่าง ซึ่งคำนวณโดยวิธีการกำหนดจุดค่าเดลตา ดัชนีไคสแควร์ที่ใช้ตัดสินความแตกต่างของค่าพารามิเตอร์ a , b จากวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และใช้การหาสหสัมพันธ์แบบพาร์เซียล ของคะแนนข้อสอบและกลุ่มด้วยคะแนนรวมทั้งหมดเป็นค่าคงที่ โดยใช้แบบทดสอบวัดความถนัด (Graduate Record Examination Aptitude Test) ศึกษาเชื้อชาติ (ขาว-ดำ) และเพศ ใช้กลุ่มผู้สอบชายผิวขาว 1,122 คน ผู้หญิงผิวขาว 1,471 คน ผู้ชายผิวดำ 284 คน และผู้หญิงผิวดำ 626 คน เมื่อพิจารณาตัวแปรทางด้านเชื้อชาติ วิเคราะห์ข้อมูลแยกระหว่างผู้ชายและผู้หญิง และเมื่อพิจารณาตัวแปรทางเพศ จะแยกข้อมูลระหว่างผิวดำและผิวขาว ทดสอบด้วยสัมประสิทธิ์แคปป่า และสถิติไคสแควร์เพื่อพิจารณาความสอดคล้องของดัชนีการตัดสินความลำเอียงทั้ง 3 วิธี ผลการวิจัยพบว่า

ดัชนีที่ใช้ไคสแควร์ตัดสินความแตกต่างระหว่างไคงลักษณะข้อสอบ ได้เปรียบวิธีที่ตั้ง
ตั้งเกณฑ์ เนื่องจากมีความคงที่และไม่เปลี่ยนตามเกณฑ์

ไอรอนสัน และเครก (Subkoviak. 1984 : 50 ; citing Ironson and Craig. 1982) ได้วิเคราะห์ความลำเอียงของข้อสอบ 5 วิธี คือ การแปลง
ค่าความยากวิธีไคสแควร์ 2 วิธี วิธีไคงลักษณะข้อสอบ 3 และ 1 พารามิเตอร์ ใช้
แบบทดสอบที่ผู้วิจัยสร้างให้ลำเอียงต่อเพศชาย แล้วให้ผู้สอบประมาณค่าความลำเอียง
แต่ละข้อเป็น 5 ระดับ ผลการศึกษาพบว่า การประมาณค่าความลำเอียงของข้อสอบ
กับการวิเคราะห์ความลำเอียง 5 วิธี มีความสัมพันธ์กันในระดับสูง

สับโคเวียค และคนอื่น ๆ (Subkoviak and Others. 1984 : 49-58)
ได้ศึกษาวิธีวิเคราะห์ความลำเอียงของข้อสอบ 3 วิธี คือ วิธีไคงลักษณะข้อสอบ
3 พารามิเตอร์ วิธีไคสแควร์ และวิธีแปลงค่าความยาก ใช้แบบทดสอบวัดความรู้
เกี่ยวกับคำศัพท์ 50 ข้อ ประกอบด้วยศัพท์ภาษาอังกฤษมาตรฐาน จำนวน 40 ข้อ
ศัพท์แสดงชาวผิวดำ 10 ข้อ กลุ่มตัวอย่างเป็นนักเรียนผิวขาว 1,021 คน และ
นักเรียนผิวดำ 1,008 คน พบว่าวิธีไคงลักษณะข้อสอบ 3 พารามิเตอร์ เป็นวิธีที่มี
ประสิทธิภาพมากที่สุด รองลงมาคือวิธีไคสแควร์ วิธีแปลงค่าความยากของข้อสอบ
เป็นวิธีที่มีข้อจำกัด แต่ก็สามารถนำไปใช้ในทางปฏิบัติได้ในกรณีที่มีข้อจำกัดเกี่ยวกับ
คอมพิวเตอร์

ชูเนอร์ (Shoener. 1984) ได้เปรียบเทียบการตรวจค้นความลำเอียง
ของข้อสอบด้วยวิธีทฤษฎีการตอบข้อคำถาม 3 พารามิเตอร์ และการตัดสินด้วยผู้ตัดสิน
จำนวน 14 คน โดยใช้แบบทดสอบอิงเกณฑ์กับนักเรียนชั้นมัธยมศึกษาชายหญิงจำนวน
1,064 คน ที่มีภูมิหลังทางวัฒนธรรมต่าง ๆ กัน การตัดสินใช้ผู้ตัดสินที่มีความรู้ ทาง
ด้านคณิตศาสตร์ และหลักสูตรคณิตศาสตร์เป็นอย่างดี และใช้แบบฟอร์มการให้คะแนน
เป็นช่วง (rating form) ผู้ตัดสินเป็นผิวดำ 8 คน ผิวขาว 8 คน และเชื้อสาย
สเปน 8 คน เป็นชาย และหญิงเท่า ๆ กัน ผลการวิจัยพบว่าความสอดคล้อง
ระหว่างการตัดสินโดยผู้เชี่ยวชาญต่างกลุ่มไม่นับสำคัญทั้งความลำเอียงทางเพศและ
วัฒนธรรม ตัวบ่งชี้บางตัวที่คำนวณขึ้นเพื่อรวมการให้คะแนนของผู้ตัดสินสอดคล้องกัน
อย่างมีนัยสำคัญ แต่ไม่พบความสอดคล้องระหว่างวิธีการตรวจค้นความลำเอียงทั้ง
2 วิธี การตัดข้อสอบที่พบทางสถิติว่าลำเอียงออกไปไม่ทำให้ลำดับของคนในกลุ่มย่อย
และรวมทั้งหมดเปลี่ยนไป และแบบทดสอบฉบับเดิมและแบบทดสอบที่ตัดข้อสอบที่ลำเอียง
ออกแล้วมีความสัมพันธ์อย่างมีนัยสำคัญกับแบบทดสอบคณิตศาสตร์แบบอิงกลุ่มที่เป็นแบบ
ทดสอบมาตรฐาน

สับโคเวียค และคนอื่น ๆ (Subkoviak and Others. 1984) ได้เปรียบเทียบวิธีการตรวจสอบความลำเอียงของข้อสอบ 3 วิธี คือทฤษฎีการตอบข้อคำถามชนิด 3 พารามิเตอร์ การแปลงค่าความยากและโคสแควร์ที่เสนอโดยชูเนแมน (Scheuneman. 1979) ที่พิจารณาเฉพาะตัวเลือกถูก และที่เสนอโดยคามิลลี (Camilli. 1979) ที่พิจารณาทั้งตัวเลือกถูกและผิด วิธีการแปลงค่าความยาก ทำโดยคำนวณค่าความยากของข้อสอบตามกลุ่มผู้วัดและผู้ชว จากนั้นแปลงค่าความยาก (p) เป็นค่า X ซึ่งมีค่าเท่ากับเปอร์เซ็นต์ที่ $1-P$ ในการกระจายแบบปกติ แล้วแปลงค่า X ให้เป็นค่าเดลต้าโดยที่ $= 4X + 13$ ค่าระยะทางระหว่างจุดของค่า X ทั้งสองกลุ่ม และเส้นแกนสำคัญเป็นค่าที่แสดงความลำเอียงและจะใช้เครื่องหมายบวก ถ้าเป็นการลำเอียงเข้าข้างผู้วัด และใช้ค่าเป็นลบถ้าลำเอียงเข้าข้างผู้ชว วิธีโคสแควร์ที่พิจารณาเฉพาะตัวถูกจะแบ่งคะแนนรวมเป็น 5 ช่วง คำนวณค่าความถี่คาดหวัง (E) ในการตอบถูกในแต่ละช่วงตามกลุ่ม แล้วเปรียบเทียบความถี่คาดหวังกับความถี่จากการสังเกต (O) ด้วยค่าสถิติโคสแควร์รวมทุกช่วง คะแนนค่าที่ได้จะเป็นการวัดความลำเอียงโดยไม่มีเครื่องหมาย การวัดโดยมีเครื่องหมายหรือมีทิศทางทำโดย ให้เครื่องหมายบวกกรณีลำเอียงเข้าข้างกลุ่มผู้วัด และลบเมื่อเข้าข้างกลุ่มผู้ชว โดยให้เครื่องหมายในแต่ละช่วงคะแนนวิธีโคสแควร์ที่พิจารณาทั้งตัวเลือกถูกและผิดก็เช่นเดียวกับวิธีพิจารณาตัวเลือกถูก แต่จะวิเคราะห์ตัวเลือกผิดด้วยเครื่องมือที่ใช้เป็นแบบทดสอบเลือกค่าเหมือนแบบเลือกตอบ 4 ตัวเลือกจำนวน 50 ข้อ มีอยู่ 10 ข้อ ที่เป็นข้อสอบวัดค่าสแลงสำหรับผู้วัดอีก 40 ข้อ เป็นค่าศัพท์มาตรฐาน ให้รหัส 10 ข้อแรก เป็น 1 คือข้อที่ลำเอียงอีก 40 ข้อให้รหัส 0 เป็นข้อไม่ลำเอียงผลการวิเคราะห์มีดังนี้

ค่าสหสัมพันธ์ระหว่างผลการตรวจค้นของแต่ละวิธีการ กับข้อสอบที่สร้างขึ้น ที่ให้รหัสว่าลำเอียงหรือไม่ลำเอียง พบว่าค่าสหสัมพันธ์ระหว่างวิธีพิจารณาเฉพาะตัวถูก กับวิธีพิจารณาทั้งตัวถูกและผิดมีค่าสูงสุด (ชนิดไม่มีเครื่องหมาย = .978 ชนิดมีเครื่องหมาย = .970) รองลงมาคือค่าสหสัมพันธ์ระหว่างวิธีแปลงค่าความยาก กับพิจารณาตัวถูกและผิด ถ้ามีจำนวนตัวอย่างน้อย เช่น 300 คน หรือน้อยกว่า แทนที่จะใช้ทฤษฎีการตอบข้อคำถามควรเลือกค่า Z และที่ให้ค่าถูกต้องใกล้เคียงกับ Z คือโคสแควร์ส่วนค่าเดลต้านั้น ในกรณีสุดโต่งจะมีการชี้ถึงความลำเอียงตรงข้ามกับความเป็นจริง ดัชนีเดลต้าที่เหลือ (residual delta index) ซึ่งได้ลองนำมาศึกษานั้นพบว่าใช้ได้ดีเกือบเท่า ๆ โคสแควร์ ซึ่งควรมีการแก้ไขต่อไปเพื่อให้มาใช้แทนโคสแควร์ได้

แวน เดอ ฟลายเออร์ เมลเลนเบอร์ก แอดเดอร์ และวิน (Van Der Flier, Mellenbergh, Ader and Win. 1984) ได้เปรียบเทียบวิธีการตรวจสอบความลำเอียงของข้อสอบจากข้อมูลจำลองด้วยวิธีที่มีเงื่อนไข วิธีโลจิส แบบตัดแปลงจากวิธีของ Lord ที่เสนอการใช้ทฤษฎีการตอบข้อสอบ 3 พารามิเตอร์ แบบเป็นขั้นตอนโดย

ขั้นที่ 1 ประมวลค่าพารามิเตอร์ แล้วคำนวณค่าสถิติที่แสดงความลำเอียงของข้อสอบทุกข้อ

ขั้นที่ 2 ตัดข้อสอบที่ลำเอียงออก แล้วประมวลค่าพารามิเตอร์ใหม่ จากนั้นจึงนำค่าพารามิเตอร์ที่ได้คำนวณค่าสถิติความลำเอียงของข้อสอบทุกข้อ (รวมทั้งข้อสอบที่ลำเอียงด้วย X^2-1 ข้อที่มีค่าไคสแควร์สูงสุดในรอบนั้น และมีนัยสำคัญ ถือว่าเป็นข้อสอบที่ลำเอียงข้ออื่น ๆ นอกนั้น ถือว่าเป็นข้อที่ไม่ลำเอียงแม้ว่า ค่าไคสแควร์จะมีนัยสำคัญ

ขั้นที่ 3 ทำขั้นที่ 1 และ 2 อีกครั้ง การทำซ้ำจะจบลงถ้าทำซ้ำเท่ากับจำนวนรอบที่กำหนดไว้ หรือค่าไคสแควร์สูงสุดของรอบนั้นไม่มีนัยสำคัญ

ในแต่ละรอบของการคำนวณ ค่าไคสแควร์ คำนวณจากข้อสอบทุกข้อ รวมทั้งข้อที่ตรวจค้นว่ามีความลำเอียงในรอบก่อน ๆ

การคำนวณแต่ละรอบ จำนวนข้อสอบที่ขจัดออกไปจะมากกว่าจำนวนข้อสอบในรอบที่แล้ว 1 ข้อ

ข้อสอบที่ตัดออกจากการคำนวณคะแนนรวมในการคำนวณรอบที่แล้ว ไม่จำเป็นต้องถูกตัดออกไปในรอบต่อไป ถ้าไคสแควร์ ของค่านั้นยอมรับได้ในรอบต่อมาด้วยวิธีนี้คะแนนรวมจะเป็นอิสระจากข้อสอบลำเอียง และข้อสอบทุกข้อจะได้รับการตรวจสอบโดยมีระดับความสามารถที่ได้จากแบบสอบที่ไม่ลำเอียง

ข้อดีของการทำซ้ำคือจะแน่ใจว่าได้ข้อสอบที่ลำเอียงจริง ๆ จากการตรวจสอบเพราะจะระบุเฉพาะข้อที่มีค่าไคสแควร์ที่สูงสุด และมีนัยสำคัญเท่านั้น

กรณีของการตรวจสอบความลำเอียง โดยใช้วิธีที่มีเงื่อนไขบนระดับความสามารถของผู้สอบเท่ากันโดยการใช้ค่าไคสแควร์นี้ ได้เปรียบเทียบวิธีการไม่ทำซ้ำ (คือใช้ผลจากรอบที่ 1) และการทำซ้ำโดยใช้ข้อมูลจำลองของข้อสอบ 29 ข้อ โดยให้ระดับของความลำเอียงต่างกัน (c = องค์ประกอบของความลำเอียงมีค่าตั้งแต่ 0.00 คือ ข้อสอบที่ยากกว่าและมีอำนาจจำแนกน้อยกว่า จนถึง 1.00 ซึ่งเป็นข้อสอบที่ง่ายกว่าและมีอำนาจจำแนกสูงกว่า) พบว่าวิธีการทำซ้ำให้การตรวจสอบความลำเอียงได้ถูกต้องมากกว่า ยกเว้นที่ค่า $c = 1.00$ จะตรวจสอบได้เท่า ๆ กัน

โดยมีความถูกต้องประมาณร้อยละ 93 และ 100 ที่ระดับความมีนัยสำคัญ .05 และ .01 ตามลำดับ และการทำซ้ำจะมีความถูกต้องมากกว่าเพิ่มขึ้น เมื่อค่า c ลดลง จนถึงเมื่อค่า $c = .00$ ความถูกต้องของการทำซ้ำจะคิดเป็นร้อยละ 100 ทั้งสองระดับความมีนัยสำคัญ และมีความถูกต้องคิดเป็นสองเท่าของการไม่ทำซ้ำ

กรณีของการตรวจสอบความลำเอียงของข้อสอบชนิดไม่มีเงื่อนไขบนระดับความสามารถของผู้สอบ ได้แก่ วิธีการวิเคราะห์ความแปรปรวนแบบอาร์คชายนของเพลค (Plake) และฮูเวอร์ (Hoover, 1979) และวิธีการแปลงค่าความยาก (Delta)

ในวิธีแรกนั้นค่าความยาก (p) ของข้อสอบแต่ละข้อ แต่ละกลุ่ม จะถูกแปลงด้วยวิธีอาร์คชายน แล้วทดสอบสมมติฐานศูนย์ที่ว่า ค่าพารามิเตอร์ของปฏิสัมพันธ์ = 0 ที่ระดับความมีนัยสำคัญ .05 และ .01 ส่วนวิธีที่ 2 จะใช้การแปลงค่า p ให้เป็นค่าเดลต้า แล้วคำนวณระยะทางของเส้นตั้งฉาก จากจุดเดลต้าไปยังเส้นแกนสำคัญ การทดสอบทางสถิติของเส้นดังกล่าวนี้ยังไม่มีวิธีการที่เหมาะสม จึงมีการตัดสินใจเองว่า เพราะว่าค่า c น้อยกว่า 1.00 สำหรับข้อสอบ 15 ข้อในกลุ่มที่มีข้อสอบลำเอียง เพราะฉะนั้นข้อสอบ 15 ข้อแรกที่มีค่าระยะทางนี้สูงสุด จะเป็นข้อสอบที่ตัดสินได้ว่าลำเอียง

ผลการเปรียบเทียบปรากฏว่าวิธีแปลงค่าความยาก ได้ผลพอ ๆ กับวิธีอาร์ชายนที่ระดับ .05 เมื่อค่า $c = .75$ และ .50 แต่ได้ผลถูกต้องมากกว่าในระดับ $c = .25$ และ .00 แต่วิธีอาร์คชายน ที่ระดับ .01 นั้น วิธีแปลงค่าความยากจะดีกว่าเมื่อ $c = .75$ แต่จะดีกว่าเมื่อ $c < .50$ อย่างไรก็ตามพบว่าทั้งสองวิธีให้ความถูกต้องน้อยมาก ส่วนมากไม่ถึงร้อยละ 50 โดยมีพิสัยความถูกต้องในการตรวจค้นที่ร้อยละ 41.4 - 58.6

ค็อค, เมลเลนเบอร์ก และฟลายเออร์ (Kok, Mellenbergh and Flier, 1985) ได้ใช้วิธีโลจิสแบบทำซ้ำตรวจสอบความลำเอียงข้อสอบ โดยการแบ่งคะแนนรวมจากแบบสอบเป็นประเภท หรือช่วงคะแนนที่ไม่ซ้ำซ้อนกัน และทดสอบความแตกต่างของความน่าจะเป็นของการตอบข้อสอบถูกของผู้สอบแต่ละกลุ่มที่อยู่ในช่วงความสามารถเดียวกัน โดยนำไปใช้กับข้อมูลที่สร้างขึ้นโดยให้มีความลำเอียงเข้าข้างกลุ่มใดกลุ่มหนึ่ง มีขั้นตอนในการตรวจ ดังนี้

1) แบ่งคะแนนรวมเป็น 4 ระดับ

2) ทดสอบความเหมาะสมของแบบจำลองกับข้อมูล ด้วยแบบจำลอง

โลจิส

(the root mean squared difference) และวิธีของพื้นที่รวม ซึ่ง 2 วิธีหลังนี้เป็นวิธีการที่ใช้ทฤษฎีการตอบข้อสอบ

แบบทดสอบที่ใช้ในการศึกษา เป็นแบบทดสอบความสามารถในการอ่านของคลีฟแลนด์มีข้อสอบจำนวน 75 ข้อ โดยใช้กลุ่มตัวอย่างชาย 451 คน หญิง 486 คน คะแนนจุดตัดในการแปลความหมายค่าสถิติที่แสดงความลำเอียงได้จากการจำลองข้อมูลขึ้น ผลปรากฏว่าวิธีทั้ง 4 วิธี ให้ผลการตรวจค้นข้อสอบลำเอียงใกล้เคียงกัน ปัญหาด้านวิธีวิทยาอยู่ที่ความไม่แจ่มชัดในการกำหนดจุดตัด ความคลาดเคลื่อนชนิดที่หนึ่ง และ การประมาณค่าพารามิเตอร์ที่ไม่ดี ข้อค้นพบเด่น ๆ อยู่ที่ความสำคัญของการเลือกใช้ช่วงคะแนนความสามารถในการวัดความลำเอียง และพบว่า วิธี MH เป็นวิธีเลือกใช้แทนวิธีทฤษฎีการตอบข้อสอบได้อย่างประหยัดกว่าทั้งเงินและเวลา

ดูลิทเทิล (Doolittle. 1987) ได้ศึกษาความแตกต่างอันเนื่องจากเพศในการตอบข้อสอบ 3 ประเภท ได้แก่ การคำนวณทางคณิตศาสตร์ เรขาคณิตและการใช้เหตุผลในแบบทดสอบ ACT Assessment Mathematics Usage Test ซึ่งมี 8 แบบ ให้นักเรียนที่เป็นกลุ่มตัวอย่าง 8 กลุ่ม ทำโดยจับคู่กลุ่มตัวอย่างชายและหญิง ซึ่งเป็นนักเรียนชั้นมัธยมศึกษาตอนปลายด้วย Profile ของวิชาวิทยาศาสตร์ที่ได้เรียนมาแล้ว

ออกแบบการวิเคราะห์ข้อมูลโดยวิธี Split Plot Factorial Design ชนิด $2 \times 3 \times 8 \times 8$ (เพศ $\times$ ประเภทข้อสอบ $\times$ ภูมิหลังด้านคณิตศาสตร์ $\times$ รูปแบบของแบบทดสอบ) ผลกระทบขนาดใหญ่ที่สุดได้แก่ผลกระทบหลักสำหรับภูมิหลังและเพศ จุดเน้นประการแรก ได้แก่ ปฏิสัมพันธ์ระหว่างเพศและประเภทข้อสอบซึ่งแสดงให้เห็นว่า ชายและหญิงทำข้อสอบคำนวณได้ดีพอ ๆ กัน แต่หญิงทำได้น้อยกว่าในข้อสอบเรขาคณิตและข้อสอบเชิงเหตุผลซึ่งต้องการกระบวนการวางแผนก่อนการแก้ปัญหา

ดูลิทเทิล และเคลียร์ (Doolittle and Cleary. 1987) ได้ตรวจค้นความลำเอียงระหว่างเพศของข้อสอบในแบบวัดผลสัมฤทธิ์ทางคณิตศาสตร์โดยใช้ตัวอย่าง จำนวน 8 กลุ่ม ที่ทำแบบทดสอบ ACT Assessment Mathematic Test รวม 8 รูปแบบ แต่ละกลุ่มมีความเท่าเทียมกันและมีจำนวนประมาณ 1,300 - 1,400 คน ตัวอย่างเป็นหญิงประมาณร้อยละ 55 วิเคราะห์ด้วยดัชนีของ ลินน์ และ ฮาร์นิสซ์ (Linn and Harnisch. 1981) ซึ่งเป็นโมเดลโลจิสติกแบบ 3 พารามิเตอร์ ถ้าดัชนี Z ที่ได้เป็นบวกแสดงว่าง่ายสำหรับกลุ่มเปรียบเทียบถ้าเป็นลบแสดงว่ายากสำหรับกลุ่มเปรียบเทียบ ผลการวิเคราะห์พบว่าดัชนี Z มีค่าเป็นลบ

สำหรับข้อกระทงที่วัดด้านเรขาคณิตและการให้เหตุผลเชิงพีชคณิต และเลขคณิตในทุกฟอร์ม แสดงว่าง่ายสำหรับชายมากกว่าหญิงที่เหลือนส่วนมากเป็นบวก

ลอเตนส์ชเลเจอร์และพาร์ค (Lautenschlager and Park. 1988) ได้ใช้ข้อมูลจำลองศึกษาเปรียบเทียบวิธีการเชื่อมโยงค่าพารามิเตอร์ (parameter linking method) ซึ่งมีการแปลงค่าพารามิเตอร์ที่ประมาณได้จากทฤษฎีการตอบข้อสอบ ให้เป็นค่ามาตรฐานเพื่อเปรียบเทียบกันได้ข้ามกลุ่ม โดยมีวิธีที่ ลินน์ และคนอื่น ๆ (Lautenschlager and Park. 1988 ; citing Linn. 1980) และวิธีของวอร์ม (Lautenschlager and Park. 1988 ; citing Warm. 1978)

วิธีของวอร์มง่ายที่สุดจากการใช้ความคิดรวบยอดที่ว่า ความยากของข้อสอบของทุกกลุ่มจะเหมือนกันได้ด้วยการแปลงค่าเชิงเส้นตรง โดยได้แปลงค่าพารามิเตอร์สัมพันธ์โดยใช้สูตรดังนี้

$$b_1 = \frac{SD_{b1}}{SD_{b2}} b_2 + (M_{b1} - \frac{SD_{b1}}{SD_{b2}} M_{b2})$$

$$a_1 = a_2 \frac{SD_{b2}}{SD_{b1}}$$

- | | | |
|-------------------|-----|---|
| b_1 | คือ | ค่าพารามิเตอร์ b บนมาตรฐานของกลุ่ม 1 |
| b_2 | คือ | ค่าพารามิเตอร์ b บนมาตรฐานของกลุ่ม 2 |
| a_1 | คือ | ค่าพารามิเตอร์ a บนมาตรฐานของกลุ่ม 1 |
| a_2 | คือ | ค่าพารามิเตอร์ a บนมาตรฐานของกลุ่ม 2 |
| M_{b1}, SD_{b1} | คือ | ค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานของค่า b บนมาตรฐานของกลุ่ม 1 |
| M_{b2}, SD_{b2} | คือ | ค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานของค่า b บนมาตรฐานของกลุ่ม 2 |

วิธีการของลินน์นั้น จะประมาณความถูกต้องของ b_i ด้วยค่าเฉลี่ยและความเบี่ยงเบนมาตรฐานที่ได้ให้น้ำหนักไว้ น้ำหนักของแต่ละข้อจะได้แก่ส่วนกลับของความแปรปรวนในการสุ่มของ b_i ที่ประมาณค่าได้สูงสุด ดังนั้นข้อที่มีค่าความ

แปรปรวนในการสุ่มของ b_i มากที่สุดจะมีน้ำหนักน้อยที่สุดในการพิจารณาค่าคงที่ของการเชื่อมโยงค่าพารามิเตอร์

จากการวิเคราะห์ข้อมูลจำลอง พบว่าภายใต้สถานการณ์ลำเอียงแบบทิศทางเดียว (แต่ละข้อลําเอียงเข้าข้างกลุ่มใดกลุ่มหนึ่งเพียงกลุ่มเดียว) นั้น ยากที่จะตรวจค้นจำนวนข้อสอบที่ถูกกระทบตรงข้ามกับความเป็นจริง (ข้อสอบลําเอียงแต่ตรวจพบว่า เป็นข้อสอบไม่ลําเอียงหรือข้อสอบไม่ลําเอียงแต่ตรวจพบว่า เป็นข้อที่ลําเอียง) ทำให้ยากที่จะเลือกว่าควรใช้วิธีใด แต่ในการศึกษาครั้งนี้ พอจะพูดได้ว่าวิธีของลินน์ และคนอื่น ๆ ดูจะดีกว่า และยังพบว่าข้อสอบที่มีระดับความลําเอียงน้อย มักจะถูกกระทบจากการตรวจค้นว่าเป็นข้อสอบไม่ลําเอียง

ลินาเคอร์ (Linacre, 1988) ได้ตรวจสอบความถูกต้องของวิธีแมนเทล-แฮนส์เซล และวิธี PROX ของ ราล์ฟ ในการตรวจสอบความลําเอียงของข้อสอบที่ได้จากการจำลองข้อมูลขึ้น โดยใช้ตัวประมาณค่าความคลาดเคลื่อนมาตรฐานหลายวิธี สำหรับวิธีแมนเทล-แฮนส์เซล ผลการวิเคราะห์พบว่าทั้งสองวิธีใช้ได้ดีในสถานการณ์จำลองทุกสถานการณ์ ส่วนตัวประมาณค่าความคลาดเคลื่อนมาตรฐานของแมนเทล-แฮนส์เซล มีคุณสมบัติแตกต่างกันไปและควรมีการแปลผลอย่างระมัดระวังเมื่อใช้วิธีแมนเทล-แฮนส์เซล ผู้วิเคราะห์ควรจะคำนวณขนาดของความลําเอียงของข้อสอบ และความแปรปรวนของความคลาดเคลื่อนด้วยการใช้เกณฑ์หลาย ๆ เกณฑ์ และใช้ตัวประมาณค่าความคลาดเคลื่อนมาตรฐานมากกว่า 1 ตัว สำหรับวิธี PROX จะให้การตรวจสอบการประมาณที่ได้จากวิธีแมนเทล-แฮนส์เซล

ธิสเซน สไตน์เบอร์ก และวายนอร์ (Thissen, Steinberg and Wainer, 1988) ได้ศึกษาความแตกต่างของโด่ง การตอบข้อสอบระหว่างกลุ่มด้วยทฤษฎีการตอบข้อคำถามเพื่อศึกษาถึงสภาพการณ์ในการทดสอบสมมุติฐาน ความเท่ากันของค่าพารามิเตอร์ด้วยค่าโคสแควร์แบบอัตราส่วน likelihood (IRT-LR) โดยใช้ข้อมูลจำลอง และได้ทำการเปรียบเทียบการทดสอบสมมุติฐานความเท่ากันของค่าพารามิเตอร์ ด้วย IRT-LR และวิธีของแมนเทล-แฮนส์เซล พบว่า IRT-LR สามารถตรวจสอบความลําเอียงได้ในเมื่อค่าความแตกต่างของความยากระหว่างกลุ่มมีค่า 0.3 และความแตกต่างในด้านการเดามีค่า 0.1 จำนวนตัวอย่างในกลุ่มน้อย (= 500 คน) โดยจำนวนข้อสอบร่วมจะน้อย หรือมากก็ไม่มีผลกระทบ ค่าความแตกต่างที่น้อยกว่านี้ไม่สามารถใช้ IRT-LR ตรวจสอบได้ว่าลําเอียงหรือไม่ แต่อาจทำได้ถ้าจำนวนตัวอย่างมากกว่านี้

จากแบบทดสอบ ส่วนตัวแปรจับคู่สำหรับ แม็คเนมาร์ ได้แก่ ตัวแปรด้านเศรษฐกิจสังคม หรือระดับคะแนนเฉลี่ย

ผลการวิเคราะห์พบว่าวิธีแมนเทิล-แฮนส์เซล ซึ่งใช้เกณฑ์ภายในตรวจสอบข้อสอบที่ล่าเอียงเพียงเล็กน้อย ส่วน แม็คเนมาร์ ซึ่งใช้เกณฑ์ภายนอกนั้นพบว่าเมื่อใช้ตัวแปรจับคู่เป็นสถานภาพทางเศรษฐกิจสังคม พบว่าข้อสอบมีระดับความล่าเอียงต่ำกว่าการจับคู่ด้วยระดับคะแนนเฉลี่ยแสดงว่าระดับคะแนนเฉลี่ยของคน 2 กลุ่มนี้มีระดับของความล่าเอียงไม่เท่ากัน

✓ โอชิม่า (Oshima. 1989) ได้ศึกษาผลของข้อมูลหลายมิติที่มีต่อการตรวจค้นความล่าเอียง โดยจำลองข้อมูล 42 ชุดจากข้อสอบ 40 ข้อ มีโครงสร้างความเป็นพหุมิติ 2 ชนิด คือ ชนิดที่มีลักษณะแฝงที่เด่นอยู่ลักษณะเดียว และที่มีลักษณะแฝงเด่น 2 ลักษณะ ดัชนีความไม่แปรปรวนของค่าพารามิเตอร์ของข้อสอบที่ใช้คือ ดัชนีผลบวกของความแตกต่างกำลังสองชนิดไม่มีเครื่องหมาย ใช้ตรวจสอบผลการตรวจค้นผิด คือ ระบุว่าข้อสอบล่าเอียงทั้ง ๆ ที่เป็นข้อสอบที่ไม่ล่าเอียง (False Positive) และระบุว่าข้อสอบที่ไม่ล่าเอียงทั้ง ๆ ที่เป็นข้อที่ล่าเอียง (False Negative) ผลจากการตรวจค้นพบว่าการตัดสินข้อสอบว่าล่าเอียงทั้ง ๆ ที่ไม่ใช่ข้อที่ล่าเอียงในกรณีที่มีข้อมูลได้จากการวัดลักษณะแฝงเด่น 2 ลักษณะมากกว่าที่วัดลักษณะเดียวโดยเฉพาะอย่างยิ่งถ้าความสัมพันธ์ของลักษณะ 2 ลักษณะในแต่ละกลุ่มต่างกันมาก

เมื่อค่าสหสัมพันธ์ระหว่างลักษณะเด่นทั้งสองเป็นศูนย์ในกลุ่มอ้างอิงและ 1.00 ในกลุ่มเปรียบเทียบ พบว่ามี 29 ใน 40 ข้อที่ถูกระบุอย่างผิด ๆ ว่าล่าเอียงอย่างสุดโต่ง

เมื่อแบบทดสอบเป็นเอกมิติ คือ มีลักษณะแฝงที่เด่นอยู่ลักษณะเดียว แต่มีข้อสอบมากกว่าร้อยละ 10 ที่เป็นพหุมิติแล้ว ดัชนีนี้จะไม่ไวพอที่จะตรวจค้นข้อสอบที่ล่าเอียงได้

แฮมเบิลตัน และโรเจอร์ส (Hambleton and Rogers. 1989) ได้ศึกษาเปรียบเทียบวิธีการตรวจค้นข้อสอบล่าเอียง 2 วิธี ได้แก่ วิธีแมนเทิล-แฮนส์เซล และวิธีอิงทฤษฎีการตอบข้อคำถามโดยใช้ข้อมูลการตอบแบบทดสอบ New Mexico High School Proficiency Exam (NMHSPE) จำนวน 150 ข้อ ของนักเรียนชาวอเมริกันผิวขาว 8,000 คน และชาวอเมริกันพื้นเมือง 2,600 คน จากผู้สอบทั้งหมด 23,000 คน

NMHSPE เป็นแบบทดสอบชนิดเลือกตอบ 4 ตัวเลือก วัดทักษะการดำเนินชีวิต 5 ด้าน คือ (1) ความรู้เกี่ยวกับแหล่งทรัพยากรของชุมชน (2) ความรู้ด้านผู้บริโภค

(3) ความรู้เกี่ยวกับรัฐบาลและกฎหมาย (4) ความรู้ด้านสุขภาพทางกายและจิต และ (5) ความรู้ด้านอาชีพ

กลุ่มตัวอย่างที่นำมาศึกษา ซึ่งได้แก่ ชาวอเมริกันผิวขาวและอเมริกันพื้นเมืองนั้นได้สุ่มมาศึกษาอย่างละ 2,000 คน และแบ่งเป็น 2 กลุ่ม ๆ ละ 1,000 คน ตามลำดับคู่-คี่ การศึกษาความลำเอียง 2 คู่ เช่นนี้เพื่อพิจารณาถึงความคงที่ในการระบุข้อสอบลำเอียง

ในการศึกษาความลำเอียง ผู้วิจัยได้ใช้ข้อสอบเพียง 75 ข้อ โดยตัดข้อสอบที่ง่ายมาก ($p > .70$) และค่าอำนาจจำแนกต่ำ ($r < .10$) วิธี MH วิเคราะห์โดยใช้ 76 ระดับคะแนน ส่วนวิธีทฤษฎีการตอบข้อคำถาม ใช้การวัดพื้นที่ระหว่างโค้งลักษณะข้อสอบ (ICC) พิสัยของความสามารถที่นำมาใช้ในการคำนวณพื้นที่กำหนดใช้ 2 ค่าเบี่ยงเบนมาตรฐานสูงกว่าความสามารถเฉลี่ยของชาวอเมริกันพื้นเมือง ความคงที่ของการตรวจสอบความลำเอียง ได้แก่ ร้อยละของข้อสอบที่ถูกระบุว่าร่าเียงหรือไม่ร่าเียงซ้ำกันในการวิเคราะห์ 2 ครั้ง

ผลการวิเคราะห์พบว่า วิธีแมนเทิล-แฮนส์เซล มีความสอดคล้องของการตรวจสอบร้อยละ 80 ส่วน ICC ร้อยละ 73 และพบว่า ร้อยละ 47 ของข้อสอบที่ถูกระบุว่าร่าเียงในการวิเคราะห์คู่ที่ 1 เป็นข้อสอบลำเอียงในการวิเคราะห์คู่ที่ 2 ด้วย สำหรับวิธีแมนเทิล-แฮนส์เซล และร้อยละ 61 สำหรับ ICC ในทางกลับกันพบว่า ร้อยละ 64 ของข้อสอบลำเอียงในการวิเคราะห์คู่ที่ 2 เป็นข้อสอบลำเอียงในการวิเคราะห์คู่ที่ 1 สำหรับ MH และร้อยละ 56 สำหรับ ICC ผลดังกล่าวแสดงให้เห็นว่าความคงที่ของการวิเคราะห์ทั้งวิธีแมนเทิล-แฮนส์เซล และ ICC อยู่ในระดับปานกลางซึ่งอาจเป็นเพราะจำนวนตัวอย่างค่อนข้างน้อยวิธีแมนเทิล-แฮนส์เซลมีข้อสอบที่ระบุว่าร่าเียงในการวิเคราะห์ทั้ง 2 ครั้ง 7 ข้อ ICC 14 ข้อ และพื้นที่แสดงความแตกต่างของ ICC มีค่าระหว่าง -2.7 และ 1.5)

สรุปได้ว่า ทั้งวิธีแมนเทิล-แฮนส์เซล และ ICC มีความไม่เที่ยงในระดับหนึ่งในการใช้ตรวจสอบข้อสอบลำเอียงโดยที่วิธีแมนเทิล-แฮนส์เซล มีความคงที่ประมาณร้อยละ 80 และ ICC ประมาณ 73

แม็คโกรกัน (McGrogan, 1989) ได้ศึกษาความลำเอียงของข้อสอบจากความแตกต่างด้านวัฒนธรรมจากแบบทดสอบ California Achievement Tests โดยใช้ทฤษฎีการตอบข้อสอบโมเดล 3 พารามิเตอร์ พิจารณาแบบทดสอบย่อยความเข้าใจในการอ่าน โดยใช้นักเรียนระดับ 2 จำนวน 1,500 คน แบ่งเป็นกลุ่มละ 500 คนตามกลุ่มผิวขาว ผิวดำ และฮิสแปนนิค การวิเคราะห์ใช้วิธี Likelihood

Rate χ^2 (ORT-LR) วิเคราะห์ด้วยโปรแกรมสำเร็จรูป MULTILOG (Thissen. 1984) ผลการวิเคราะห์พบว่า ไม่มีความแตกต่างอย่างมีนัยสำคัญในด้านความยากข้ามกลุ่มและข้ามแบบทดสอบ และพบว่ามีความแตกต่างของค่าเฉลี่ยของกลุ่ม โดยที่กลุ่มผิวขาวมีคะแนนเฉลี่ยสูงกว่าอีก 2 กลุ่มประมาณ 0.5-1.0 ของความเบี่ยงเบนมาตรฐาน

บากิ และเฟอรารา (Baghi and Ferrara. 1989) ได้เปรียบเทียบผลการสืบค้นความลำเอียงด้วยวิธีการแตกต่างกัน 3 วิธี ได้แก่ การใช้ทฤษฎีการตอบข้อคำถามวิธีการใช้ค่าเดลต้า (ค่าความยากแปลง) และวิธีแมนเทิล-แฮนส์เชล โดยศึกษาในแง่ของผลกระทบจากขนาดของกลุ่มตัวอย่างในผลการตรวจสอบความลำเอียง และระดับของความสัมพันธ์ระหว่างค่าสถิติจากทั้ง 3 วิธีนี้

ข้อมูลที่ใช้ในการวิเคราะห์ เป็นข้อมูลที่สุ่มจากนักเรียนระดับ 9 จำนวน 50,000 คน ที่ตอบแบบทดสอบทักษะความเป็นพลเมืองดี ประกอบด้วยข้อสอบแบบเลือกตอบ 15 ข้อ ประเมินความรู้ และทักษะ รวม 3 ด้าน ได้แก่ด้านรัฐบาล ประชาธิปไตย ด้านการเมือง และพฤติกรรมทางการเมือง และด้านหลักการ สิทธิ และความรับผิดชอบ จำนวนกลุ่มตัวอย่างที่สุ่มมาใช้มี 4 ขนาด ได้แก่ ขนาด 1,000 คน 750 คน 500 คน และ 200 คน

ผลจากการวิเคราะห์ไม่พบว่า มีข้อสอบข้อใดที่มีความลำเอียงทั้งในการเปรียบเทียบระหว่างผู้สอบพิวตาและผิวขาว และผู้สอบหญิง และชาย การแสดงค่าประมาณความยากด้วยแผนภูมิเส้นแสดงความสัมพันธ์ระหว่างค่าความยากต่างกลุ่มอยู่ในรูปเส้นตรงเกือบสมบูรณ์ ความสอดคล้องกันระหว่างวิธีการตรวจสอบความลำเอียงของราสส์ และค่าความยากแปลงมีความสัมพันธ์กันสูงมากทุกขนาดของกลุ่มตัวอย่าง จากค่าสัมประสิทธิ์สหสัมพันธ์แบบลำดับ และทั้ง 2 วิธี มีความสอดคล้องสูงสุดกับวิธีการตอบข้อสอบ 3 พารามิเตอร์ ในการตรวจสอบความลำเอียงและไม่ลำเอียงของข้อสอบ

กรีน โครน และฟอล์ค (Green, Crone and Folk. 1989) ได้วิเคราะห์ความลำเอียงของตัวलग หรือวิเคราะห์เฉพาะการตอบผิดจากข้อมูลในการตอบแบบทดสอบ SAT ฟอรั่ม 6.K ด้วยวิธีแบบจำลองล็อก ลีเนียร์ (log-linear model) จากตาราง 3 ทาง เพื่อดูปฏิสัมพันธ์ของกลุ่มตัวलग และระดับความสามารถโดยมีตัวलगจำนวน 4 ตัว แบ่งระดับความสามารถของผู้สอบจากคะแนนรวมเป็น 5 ระดับ โดยมีกลุ่มตัวอย่างที่ศึกษา 3 กลุ่ม คือ กลุ่มลาตินอเมริกา กลุ่มพิวตา และกลุ่มผิวขาว เกณฑ์ในการพิจารณาว่าตัวलगมีความลำเอียงคือ มีกลุ่มใดกลุ่มหนึ่ง

มีแนวโน้มจะเลือกตัวลงใดแตกต่างไปทางมากหรือน้อยกว่ากลุ่มอื่น หรือไม่ก็ดูที่ปฏิสัมพันธ์ระหว่างกลุ่มและตัวลง ผลจากการวิเคราะห์ข้อมูลพบว่าในจำนวนข้อสอบ 85 ข้อ นั้น ปรากฏว่า มีจำนวนข้อที่มีตัวลงลำเอียงอยู่ 14 ข้อ และมี 1 ข้อที่ไม่สามารถระบุได้ว่ามีตัวลงที่ลำเอียงหรือไม่ ในจำนวน 14 ข้อนี้ 3 ข้อเป็นข้อสอบอุปมาอุปไมย (จำนวนรวม 20 ข้อ) 5 ข้อวัดค่าตรงข้าม (รวม 25 ข้อ) และ 6 ข้อเป็นการเติมประโยคให้สมบูรณ์ (รวม 15 ข้อ) ในขณะที่ข้อสอบวัดความเข้าใจจำนวน 25 ข้อ ไม่มีความลำเอียงเลย

เรย์โนลด์ (Reynold. 1989) ได้ใช้วิธีการ 4 วิธีในการตรวจค้นความลำเอียงของข้อสอบ CTBS/S ฉบับภาษาสเปน โดยใช้ข้อมูลจากนักเรียน 895 และ 911 คน วิธีการที่ใช้ในการตรวจค้นมี 4 วิธี ได้แก่

1. ค่าความยากแปลงของแองกอฟ
2. การทดสอบไคสแควร์
3. เปรียบเทียบความแตกต่างในค่าความยาก และการทดสอบความเหมาะสมด้วยโมเดลของ ราสส์
4. ทฤษฎีการตอบข้อสอบ 3 พารามิเตอร์ (พิจารณาจากพื้นที่และความแตกต่างในค่าสถิติที่ใช้ทดสอบความเหมาะสม)

ตรวจสอบความตรงเชิงโครงสร้าง ในการวัดความลำเอียงของข้อสอบ โดยการพิจารณาค่าสหสัมพันธ์ระหว่างวิธีต่าง ๆ ร้อยละของความสอดคล้องระหว่างวิธีต่าง ๆ ที่ใช้และการวิเคราะห์ตัวประกอบ

ผลจากการวิเคราะห์พบว่าเดลต้า และไคสแควร์ มีความสัมพันธ์เกือบจะสมบูรณ์วิธีที่ 1 และ 2 กับวิธีที่ 3 และ 4 มีความสัมพันธ์กันน้อย สนับสนุนโดยผลจากการวิเคราะห์ร้อยละของความสอดคล้อง การวิเคราะห์ตัวประกอบแสดงว่า เดลต้า ไคสแควร์ และ IRT แบบ 3 พารามิเตอร์ วัดโครงสร้าง (construct) เดียวกัน คือ ความลำเอียงด้านภาษา และความสัมพันธ์ระหว่างเดลต้าและไคสแควร์นั้นเห็นได้ชัดแม้แต่ในกรณีที่มีความลำเอียงน้อย

บาภิ และเฟอรารา (Baghi and Ferrara. 1990) ได้เปรียบเทียบผลจากการสืบค้นความลำเอียงของข้อสอบด้วยทฤษฎีการตอบข้อคำถามแบบ 3 พารามิเตอร์ (IRT) และสถิติไคสแควร์แบบแมนเทิล-แฮนส์เชล (MHCS) โดยใช้ข้อมูลจากการบริหารแบบทดสอบทักษะความเป็นพลเมืองดีของรัฐแมริแลนด์ ให้แก่นักเรียนระดับ 9 จำนวน 50,000 คน ในเดือนมกราคม - กุมภาพันธ์ 1988 จำนวนตัวอย่างที่ใช้ในการวิเคราะห์ด้วย MHCS มี 4 ขนาด ได้แก่ 1,000 คน

750 คน 500 คน และ 200 คน ส่วนจำนวนตัวอย่างในการสืบค้นด้วย IRT ใช้กลุ่มละ 1,000 คน ในแต่ละกลุ่มเปรียบเทียบ การวิเคราะห์ประกอบด้วย

- (1) การพิจารณาความคงที่ของค่าสถิติ MHCS เมื่อใช้กลุ่มตัวอย่างขนาดต่าง ๆ
- (2) ความคงที่ของดัชนีความลำเอียง (MH Alpha) ในวิธี MHCS ข้ามช่วงคะแนน
- (3) ความสัมพันธ์ระหว่างดัชนีความลำเอียงแบบ IRT และค่า MH Alpha และ
- (4) ความสอดคล้องระหว่างวิธีทั้งหมดในการตรวจสอบข้อสอบที่ลำเอียง

ผลปรากฏว่าเมื่อใช้วิธี IRT สืบค้นความลำเอียงของข้อสอบระหว่างกลุ่มหญิงและชายพบข้อสอบลำเอียง 4 ข้อ และพบข้อสอบลำเอียง 3 ข้อในการเปรียบเทียบระหว่างผู้สอบผิวขาวและผิวดำ ซึ่งมีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้อง ไม่เท่ากันอย่างมีนัยสำคัญ และในขณะที่ผลจากการหาค่าสหสัมพันธ์ระหว่างดัชนีทั้งสองตัวแสดงความไม่สอดคล้องกันในการบ่งชี้ข้อสอบที่ลำเอียงนั้น ความสอดคล้องในผลรวมแสดงว่าเทคนิค MHCS เป็นวิธีที่ใช้แทน IRT แบบ 3 พารามิเตอร์ได้อย่างเพียงพอ ถ้าจำนวนตัวอย่างมีขนาดอย่างน้อย 750 คน ชนิดของแบบทดสอบไม่มีผลต่อการทำนาย MH โครงสร้างของเนื้อเรื่อง และตัวเลือกมีผลเพิ่มเติมน้อยมากต่อความแปรปรวนของการทำนายทั้งในการเปรียบเทียบระหว่างชายกับหญิงและผิวดำกับผิวขาว แต่จำนวนข้อมีความสัมพันธ์กับ MH เป็นบวกมากขึ้นในหญิง และเป็นลบสำหรับผิวดำในแบบทดสอบ SAT และคำถามเชิงเหตุผลค่อนข้างยากสำหรับหญิง

ซูดวีคส์ และทอลแมน (Sudweeks and Tolman, 1990) ได้เปรียบเทียบผลจากการใช้วิธีแมนเทิล-แฮนส์เซล และการตัดสินด้วยผู้เชี่ยวชาญในการตรวจค้นความลำเอียงของข้อสอบวิชาวิทยาศาสตร์ ระหว่างนักเรียนชายและหญิงระดับ 5 จำนวน 926 คน เครื่องมือที่เก็บรวบรวมข้อมูลเป็นแบบทดสอบอิงเกณฑ์ชนิดเลือกตอบที่สร้างขึ้นตามมาตรฐาน และวัตถุประสงค์ของหลักสูตรวิทยาศาสตร์ร่วมในรัฐยูทาห์ ผลจากการวิเคราะห์พบว่า วิธีการทั้งสองให้ผลต่างกัน และข้อสอบหลายข้อที่ถูกระบุด้วยวิธีการ 2 วิธีนี้ว่า ลำเอียง จะเป็นข้อสอบที่ยาก

ชูเนแมน (Scheuneman, 1990) ได้ศึกษาความลำเอียงเพื่อสำรวจแหล่งความยากของข้อสอบ โดยใช้ค่าสถิติ MH ที่เสนอโดยฮอลแลนด์ และเทเยอร์ (Holland and Thayer, 1986) ตรวจค้นความลำเอียงของข้อสอบในแบบทดสอบการอ่าน SAT ของนักเรียนมัธยมโดยมีจำนวนชายและหญิงเท่า ๆ กัน แต่มีผิวดำอยู่ประมาณร้อยละ 9

แบบทดสอบอีกชุดหนึ่ง ได้แก่ GRE (Graduate Record Examination) เป็นแบบวัดความถนัดสำหรับการเข้าศึกษาในระดับสูงกว่าปริญญาตรี มีจำนวนชายและหญิงเท่า ๆ กัน มีพิวค่าอยู่ประมาณร้อยละ 4

จากนั้นมีการแยกข้อสอบตามลักษณะของเนื้อหา 5 ประเภท ตามลักษณะคำถาม 5 ลักษณะ และตามประเภทของค่าที่มีอยู่ในตัวคำถาม 3 ประเภท และเพื่อประเมินผลที่เกิดขึ้นพร้อม ๆ กันของตัวแปรต่างกันจากค่าสถิติ MH ได้ใช้การวิเคราะห์การถดถอย 2 ขั้นตอน ดังนี้

1. ใช้ค่า MH เป็นตัวแปรตามโดยวิเคราะห์การถดถอยแบบเป็นขั้นแยกกัน แต่ละชุดของตัวแปรเนื้อหาเพื่อหาชุดตัวแปรที่ให้ค่าสัมประสิทธิ์การถดถอยสูงสุด

2. ใช้ Backward Stepwise วิเคราะห์ผลจากข้อ 1 กลับอีกครั้ง

ในการวิเคราะห์นี้ค่า MH ได้ถูกบังคับให้มีค่าเฉลี่ย = 0 ค่าความเบี่ยงเบนมาตรฐาน = 5 สำหรับข้อที่มีค่าสัมบูรณ์ของ MH เท่ากับ 1.2 จะไม่นำมาใช้

ผลการวิเคราะห์พบว่าเนื้อหาของข้อสอบสามารถอธิบายความแปรปรวนในค่า MH ได้มากกว่าร้อยละ 25 ทั้งในการเปรียบเทียบระหว่างหญิงและชาย และพิวค่ากับพิวชาว

ข้อความที่เป็นลักษณะบรรยายใน SAT ให้ค่า MH ที่เป็นบวกสำหรับหญิงและเป็นลบสำหรับพิวค่า อธิบายว่าเกิดจากประสบการณ์ในการอ่านเนื้อหาลักษณะบรรยายที่มีมาก่อนของหญิง

แอนเคนแมน (Ankenmann, 1990) ได้ตรวจสอบผลการตอบข้อสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ทั่วไป ใช้แบบทดสอบเลือกตอบกับนักเรียนเกรด 11 ที่วิเคราะห์ด้วยทฤษฎีการตอบข้อคำถาม ทำการสอบวิชาคณิตศาสตร์ 111/112 ระดับภูมิภาค (ฟอร์ม 8922) ของสถาบันการศึกษา New Brunswick เมื่อเดือนกรกฎาคม ปี 1989 จำนวนกลุ่มตัวอย่าง 4,894 คน ข้อสอบได้มีการตรวจสอบความเป็นมิติเดียวตามข้อตกลงของทฤษฎีการตอบข้อคำถาม ซึ่งก็ไม่ใช่ปัญหาต่อการศึกษาความลำเอียงทางเพศ ใช้ความเร็วในการทำข้อสอบเป็นตัวแปรและศึกษาความเหมาะสมของข้อมูลโดยรวมด้วย จากการศึกษาพบว่ามีข้อสอบที่มีความลำเอียงทางเพศ 14 ข้อ มีความลำเอียงอย่างเป็นระบบต่อเพศชาย 10 ข้อ และไม่มีข้อสอบที่ลำเอียงต่อเพศหญิง ลักษณะที่ลำเอียงโดยทั่ว ๆ ไป พบว่า เกิดจากเนื้อหาในตัวแบบทดสอบไม่ใช่เนื้อหาในหลักสูตร โดยเฉพาะสัดส่วนของมโนภาพของคณิตศาสตร์ รูปแบบภาษาที่ใช้ในโจทย์ และเนื้อหา ซึ่งอาจจะไม่ได้รับความสนใจ และแรงกระตุ้นให้เกิดความกระตือรือร้นจากเพศชายเท่าที่ควรจะเป็น

สวามินาธาน และโรเจอร์ส (Swaminathan and Rogers. 1990) ได้ตรวจสอบความลำเอียงของข้อสอบ ด้วยวิธีการถดถอยแบบโลจิส และวิธีแมนเทิล-แฮนส์เซล (MH) โดยมีขนาดของตัวอย่าง 2 ขนาด ได้แก่ 250 และ 500 คน และความยาวของแบบทดสอบ 3 ขนาด ได้แก่ 40, 60 และ 80 ข้อ จำลองข้อมูลการตอบข้อสอบด้วยโปรแกรม DATAGEN แบบ 3 พารามิเตอร์ ข้อมูลการจำลองความลำเอียงใช้ 2 พารามิเตอร์ โดยให้อำนาจจำแนกของ 2 กลุ่มเท่ากัน ส่วนค่าความยากแปรผันไปเพื่อให้เกิดระดับความลำเอียงต่าง ๆ กันตามต้องการ แต่ในลักษณะการลำเอียงแบบไม่สม่ำเสมอ (Nonuniform) นั้น ให้ค่าความยากของ 2 กลุ่มเท่ากัน แต่ค่าอำนาจจำแนกแปรไป ผลการวิเคราะห์พบว่า การตรวจสอบความลำเอียงของข้อสอบ 2 วิธี ให้ผลใกล้เคียงกันโดย MH ดีกว่าเล็กน้อยโดยมีการตรวจสอบได้ถูกต้องร้อยละ 75 กรณีใช้ตัวอย่าง 250 คน และตรวจสอบได้ร้อยละ 100 ในตัวอย่าง 500 คน ในทุกความยาวแบบทดสอบและในกรณีความลำเอียงแบบสม่ำเสมอ (Uniform) ในกรณีของการลำเอียงแบบไม่สม่ำเสมอนั้น MH ไม่สามารถตรวจสอบได้ ส่วนการถดถอยแบบโลจิสตรวจสอบได้ถูกต้องประมาณร้อยละ 50 ในกรณีตัวอย่างน้อยและข้อสอบสั้น และถูกต้องประมาณร้อยละ 75 ในแบบทดสอบยาว และตัวอย่างขนาดใหญ่แต่กรณีมีความลำเอียงแบบไม่สม่ำเสมอ โดยที่โค้งตัดกันที่ปลายสุดของโค้งไม่ว่าจะเป็นด้านความสามารถสูงหรือต่ำก็ตาม MH จะดีกว่า

ในการตรวจค้นผิดคือระบุว่าลำเอียงทั้ง ๆ ที่ไม่ลำเอียงนั้น MH ดีกว่า โดยตรวจค้นผิดประมาณร้อยละ 1 ส่วนการถดถอยแบบโลจิสผิดประมาณร้อยละ 1-6

วิธีการถดถอยแบบโลจิสแม้จะยืดหยุ่นกว่า และมีความเป็นทั่วไปมากกว่า แต่ค่าใช้จ่ายมากกว่าประมาณ 3-4 เท่า

ประโยชน์ของการถดถอยแบบโลจิส อยู่ตรงที่เป็นวิธีการที่มีพื้นฐานเป็นแบบจำลองที่สามารถเพิ่มตัวแปรความสามารถเข้าไปในแบบจำลองได้ เช่นเดียวกับองค์ประกอบอื่น ๆ ที่มีความเกี่ยวข้องและสำคัญซึ่งทำให้เข้าใจในธรรมชาติของความลำเอียงได้ดีกว่า

ชมิทท์ และโดแรนส์ (Schmitt and Dorans. 1990) ได้ตรวจสอบความลำเอียงของข้อสอบในแบบทดสอบวัดความถนัดทางวิชาการ (Scholastic Aptitude Test) จากผู้สอบชาวอเมริกันเชื้อสายฮิสแปนิก (เม็กซิกันและเปอร์โตริกัน) เอเชียและชนผิวดำ ด้วยวิธีมาตรฐาน และใช้เกณฑ์ตัดสิน 2 จุด คือ ที่ขนาดความลำเอียง 0.5 และ .10 ตรวจค้นความลำเอียงทั้งในตัวเลือกถูกและผิด นอกจากนั้นได้วิเคราะห์ความสัมพันธ์ระหว่างความลำเอียงของข้อสอบกับ

เนื้อหาของข้อสอบที่มีลักษณะต่าง ๆ ได้แก่เนื้อหาที่มีความน่าสนใจเป็นพิเศษ มีคำศัพท์ที่มีตัวสะกดเหมือนกันในภาษาต่าง ๆ แต่มีความหมายต่างกัน มีกลุ่มคำหรือคำเดียวกันอยู่ในทั้งตัวคำถามและตัวเลือก โดยได้ตัดข้อสอบที่มีผู้ตอบไม่ถึง¹ ออกเพื่อกำจัดอิทธิพลของความเร็วออกจากการวิเคราะห์ ผลจากการวิเคราะห์พบว่า ข้อสอบที่มีเนื้อหาที่น่าสนใจจะง่ายสำหรับชนกลุ่มน้อย คือ ฮิสแปนิกและชาวผิวดำและพบส่วนมากในข้อสอบเติมข้อความให้ได้ใจความสมบูรณ์และความเข้าใจในการอ่าน ข้อสอบที่มีคำที่มีตัวสะกดเหมือนแต่ความหมายต่างมีความสัมพันธ์กับความลำเอียงในทางลบ ข้อสอบพวกนี้จะยากสำหรับชนกลุ่มน้อยมากกว่าผิวขาว ยิ่งมีคำพวกนี้มากก็ยิ่งยาก พบมากในข้อสอบอุปมาอุปไมยและคำตรงข้าม

การวิเคราะห์ความสัมพันธ์ของตัวลง กับเนื้อหายังให้เหตุผลได้ไม่ชัดเจนว่า เป็นเพราะอะไร และพบว่าในชนกลุ่มน้อยและชนผิวขาวที่มีความสามารถเท่ากันนั้น ชนกลุ่มน้อยทำถึงข้อท้าย ๆ ได้น้อยกว่าพวกผิวขาว ดังนั้นความเร็วในการทำข้อสอบต่างกันอาจมีส่วนทำให้เกิดความลำเอียงที่มีเครื่องหมายต่างกันคือเป็นบวกในข้อยาก ๆ และเป็นลบในข้อง่าย ๆ ก็เป็นไปได้

พาร์ค และลอเตนส์ชเลเจอร์ (Park and Lautenschlager, 1990) ได้ตรวจสอบความลำเอียงของข้อสอบโดยใช้ทฤษฎีการตอบข้อคำถามที่ได้ดัดแปลงจากวิธีของลอร์ด (Lord, 1980) หรือที่เรียกว่า วิธี "ปลอดความลำเอียง" (Purification) ซึ่งมีวิธีการดังนี้

1) วิเคราะห์แบบทดสอบทั้งฉบับ โดยการประมาณค่าพารามิเตอร์รวมแล้วแปลงค่าความยาก (b) ให้เป็นค่ามาตรฐาน จากนั้นกำหนดค่าการเดา (c_i) ตามที่ได้จากการวิเคราะห์แล้วประมาณค่าความยากและอำนาจจำแนก (a_i) ใหม่อีกครั้ง โดยแยกตามกลุ่มที่ศึกษาแล้วแปลงค่า b_i เป็นค่ามาตรฐานอีกครั้ง จากนั้นเปรียบเทียบฟังก์ชันการตอบข้อสอบเพื่อดูความลำเอียง

2) ประมาณค่าระดับความสามารถโดยรวมทุกกลุ่มเข้าด้วยกันใหม่ใช้แต่เฉพาะข้อสอบที่ไม่ลำเอียง จะได้ข้อสอบที่มีความเป็นเอกพันธ์

3) กำหนดค่า θ เท่ากับที่ประมาณได้ในข้อ 2 แล้วประมาณค่า a_i และ b_i ของข้อสอบทุกข้อรวมข้อที่มีความลำเอียงไว้ด้วย

4) เปรียบเทียบฟังก์ชันการตอบข้อสอบ หรือค่าพารามิเตอร์อีกครั้งด้วยค่าไคสแควร์

วิธีนี้เรียกว่าปลอดจากลำเอียง เพราะพยายามที่จะประมาณค่าพารามิเตอร์โดยตัดข้อสอบลำเอียงออกจากแบบทดสอบก่อน

วิธีที่นำมาใช้ในการศึกษาครั้งนี้ เป็นวิธีที่ปรับปรุงจากวิธีของลอร์ดได้แก่

1. วิธีดัดแปลงจาก ดราสโกว์ (Drasgow) ที่ได้เสนอไว้ เป็นการตรวจค้นความลำเอียงโดยการทำซ้ำ ๆ หลายรอบ จนกว่าจะได้ชุดข้อสอบที่ลำเอียงที่ซ้ำกันในการทำซ้ำ 2 รอบสุดท้าย วิธีนี้ใช้ค่าโคสแควร์ต่ำสุดของไดฟ์จิ (Divgi. 1985) แทนวิธีของสต็อกกิงและลอร์ด (Stocking and Lord. 1983) วิธีการนี้เรียกว่า M-DIL (Modified-Drasgow's Iterative Linking Procedure) มีวิธีดังนี้

1) ใช้วิธีการเชื่อมโยงมาตรฐานของ สต็อกกิงและลอร์ด ประมาณค่าพารามิเตอร์บนมาตรฐานที่เท่ากันโดยประมาณ

2) คำนวณค่าความลำเอียง

3) เชื่อมโยงค่าพารามิเตอร์ที่ประมาณได้อีกครั้ง โดยใช้แต่ข้อที่ไม่ลำเอียง

4) คำนวณค่าความลำเอียงของทุกข้ออีกครั้ง

5) ทำซ้ำ ๆ จนกว่าจะได้ข้อสอบลำเอียงชุดเดียวกันใน 2 รอบสุดท้าย

2. วิธีดัดแปลงจากวิธี "ปลอดลำเอียง" ของ ลอร์ด (Lord. 1980) ซึ่ง พาร์ค (Park. 1988) ได้เสนอขึ้นเรียกว่า M-LTP (Modified-Lord Test Purification) ซึ่งมีขั้นตอนดังนี้

1) คำนวณค่า θ จากทุกกลุ่มรวมกัน

2) จากค่า θ ของแต่ละคนที่คำนวณได้ใน 1) นำมาคำนวณค่าพารามิเตอร์ของข้อสอบ แล้วคำนวณค่าสถิติความลำเอียง

3) คำนวณค่า θ ใหม่อีกครั้ง โดยใช้เฉพาะข้อสอบที่ไม่ลำเอียง

4) คำนวณค่าพารามิเตอร์ของข้อสอบอีกครั้งโดยใช้ค่า θ ที่คำนวณได้จาก 3)

5) คำนวณค่าสถิติความลำเอียง

6) ทำซ้ำจนกว่าจะได้ข้อสอบซ้ำกัน ในการทำซ้ำ 2 รอบสุดท้าย

3. ใช้วิธีผสมระหว่าง M-DIL และวิธีปลอดลำเอียงเรียกว่า ILAP (Iterative Parameter Linking and θ Scale Purification) ซึ่งจะเริ่มด้วย M-DIL จนถึงรอบสุดท้ายแล้วเริ่มประมาณค่า θ ใหม่อีกครั้งแยกตามกลุ่มเฉพาะข้อสอบไม่ลำเอียง จากนั้นจึงคำนวณค่าพารามิเตอร์ข้อสอบแยกตามกลุ่มแล้วใช้วิธีการของไดฟ์จิ (Divgi) ประมาณค่าคงที่ของการเชื่อมโยง โดยใช้เฉพาะข้อที่ไม่ลำเอียง แล้วคำนวณค่าสถิติความลำเอียงใหม่ แล้วตามด้วย M-DIL ต่อ

ข้อมูลที่ศึกษาเป็นข้อมูลจำลองด้วยวิธีการจำลองของ ซิมป์สัน (Simpson. 1978) ภายใต้สภาพการณ์ลำเอียงชนิดทิศทางเดียว และทิศทางผสม พบว่าการวิเคราะห์ 2 วิธีแรก ให้ผลเท่ากัน แต่ M-DIL ต้องทำซ้ำในจำนวนรอบที่มากกว่า M-LTP และถ้าใช้วิธีสุดท้ายแล้วจะสามารถลดจำนวนข้อสอบที่ลำเอียง แต่ถูกระบุจากการตรวจค้นว่าไม่ลำเอียงลงได้มากกว่า 2 วิธีแรก

ฟรีเดิล และคอสติน (Freedle and Kostin. 1990) ได้ศึกษาการใช้ค่าความยากของข้อสอบเป็นตัวทำนายความลำเอียงของข้อสอบกลุ่มผู้สอบพิวตาและพิวชาวที่มีความสามารถเท่าเทียมกัน โดยใช้คะแนนรวมจากแบบทดสอบเป็นตัววัดความสามารถ แบบทดสอบที่นำมาวิเคราะห์ ได้แก่ แบบทดสอบ Graduate Record Examination (GRE) และ Scholastic Aptitude Test (SAT) โดยวัดความสามารถทางภาษา 4 ด้าน ได้แก่ อุปมาอุปไมย คำตรงข้าม เต็มคำในประโยค ให้ได้ความสมบูรณ์และความเข้าใจในการอ่านวิเคราะห์ด้วยการหาค่าความยากของข้อสอบ โดยใช้ค่าเดลต้าที่เทียบมาตรฐานแล้วโดยแปลงค่าความยากของข้อสอบให้มีค่าเฉลี่ย 13 และค่าความเบี่ยงเบนมาตรฐาน 4 ด้วยการแปลงค่าเช่นนี้ ค่าเดลต้าที่มากขึ้นจะแสดงว่าข้อสอบยากขึ้น

สำหรับขนาดของความลำเอียงนั้นใช้วิธีที่เรียกว่าวิธีการทำมาตรฐาน (Standardization Approach) ซึ่งเสนอโดย โดแรนส์ และคูลิค (Dorans and Kulick. 1986) ซึ่งมีวิธีการดังนี้

1. ในแต่ละระดับของคะแนนรวม หาค่าความแตกต่างของร้อยละของผู้ตอบ ถูกระหว่างกลุ่มผู้สอบพิวตาและพิวชาว
 2. ให้นำบันทึกความแตกต่างในร้อยละของการทำถูกในแต่ละระดับคะแนนด้วยความถี่สัมพัทธ์ในระดับคะแนนนั้น ในกลุ่มพิวตา
- โดยแน่ใจว่าความแตกต่างที่สัมพันธ์กับคนส่วนมากในกลุ่มนี้จะเป็นตัวแสดงถึงค่าความลำเอียงของข้อสอบแต่ละข้อ

3. คำนวณค่าขนาดความลำเอียงของข้อสอบแต่ละข้อ โดยใช้ค่าเฉลี่ยของความแตกต่างที่ให้นำหนักไว้แล้วข้ามระดับคะแนน เรียกว่าค่าความแตกต่างมาตรฐานในค่าร้อยละของความถูกต้อง ซึ่งมีสูตรดังนี้ (Freedle and Kostin. 1990 : 330)

$$DIF_{STD} = \frac{\sum_{S=1}^S Nf_s (P_{fs} - P_{bs})}{\sum_{S=1}^S Nf_s}$$

เมื่อ	DIF_{STD}	คือ ขนาดความลำเอียง
	S	คือ จำนวนระดับคะแนน
	Nf_s	คือ จำนวนคนที่ได้คะแนนระดับ S ในกลุ่มเปรียบเทียบ (ในที่นี้คือกลุ่มผิวดำ)
	Pf_s	คือ ร้อยละของผู้ตอบในกลุ่มผิวดำในระดับคะแนน S ที่ตอบข้อสอบถูก
	Pb_s	คือ ร้อยละของผู้ตอบในกลุ่มผิวขาว (กลุ่มอ้างอิง) ในระดับคะแนน S ที่ตอบข้อสอบถูก

ถ้าค่าที่ได้เป็นลบแสดงว่าข้อสอบนั้นลำเอียงเข้าข้างกลุ่มผิวดำ แต่ถ้าเป็นบวกแสดงว่าข้อสอบลำเอียงเข้าข้างกลุ่มผิวขาว

จากนั้นคำนวณค่าสหสัมพันธ์ระหว่างค่าความยากของข้อสอบและขนาดของความลำเอียง

นอกจากนั้น เนื่องจากมีข้ออภิปรายว่าการคำนวณค่าขนาดของความลำเอียง อาจมีความไวต่ออัตราการไม่ตอบข้อสอบของผู้สอบต่างกลุ่มที่อาจแตกต่างกันอยู่ จึงได้มีการใช้ค่าแก๊ซซิงโดแรนส์ และคูลิค (Dorans and Kulic, 1983) ได้เสนอการคำนวณร้อยละที่ทำถูกไว้ 3 วิธี (Freedle, 1990 : 337)

$$1. R/(NR + 0 + R + W)$$

$$2. R/(0 + R + W)$$

$$3. R/(R + W)$$

เมื่อ	R	ได้แก่ จำนวนข้อที่ตอบถูก
	NR	ได้แก่ จำนวนข้อที่ทำไม่ทัน
	W	ได้แก่ จำนวนข้อที่ตอบผิด
	0	ได้แก่ จำนวนข้อที่ไม่ตอบ

จากนั้นคำนวณหาขนาดความลำเอียงใหม่อีกครั้งและคำนวณค่าสหสัมพันธ์ระหว่างค่าแก๊ซซิงโดแรนส์และระหว่างค่าความลำเอียงที่ใช้ค่าแก๊ซซิงโดแรนส์กับค่าความยากของข้อสอบปรากฏผลการศึกษาดังนี้

1. ค่าความยากของข้อสอบสามารถนำมาใช้ทำนายขนาดความลำเอียงได้ทั้งใน GRE และ SAT

2. ผู้สอบพิจารณาข้อสอบยากได้ดีกว่าผู้พิจารณาข้อสอบง่ายได้ดีน้อยกว่าสำหรับผู้อยู่ในระดับความสามารถเดียวกัน

3. คำที่มีความหมายได้หลายอย่างเป็นตัวทำนายทางอ้อมของความสัมพันธ์ระหว่างความยากและความลำเอียงของข้อสอบ เพราะว่าสามารถใช้อธิบายความแตกต่างในการแปลความหมายระหว่างกลุ่มต่างกันได้ และโดยที่คำเหล่านี้จะถูกนำมาใช้ในข้อง่าย ๆ มากกว่า ผู้สอบพิจารณาจึงทำได้ดีกว่า

4. การเทียบบริบทด้านภาษาให้มากขึ้นจะลดความคลุมเครือของความหมายของคำที่มีความถี่ในการใช้สูงหรือคำที่มีความหมายได้หลายอย่าง

5. การใช้คำแก้ในการคำนวณค่าความลำเอียง ไม่ก่อให้เกิดผลแตกต่างไปจากค่าที่คำนวณโดยไม่ใช้คำแก้

6. การใช้คำแก้ 3 วิธีต่างกัน ให้ผลต่อค่าความลำเอียงไม่แตกต่างกัน วายเนอร์ เชอเรซี และทิสเซน (Wainer, Sireci and Thissen. 1991) ได้ใช้แบบจำลองของ บ็อค (Bock. 1972) ซึ่งดัดแปลงจากทฤษฎีการตอบข้อสอบมาใช้ในการค้นหาความลำเอียงของแบบทดสอบ โดยศึกษาจากแบบทดสอบทั้งฉบับ ด้วยเกณฑ์ภายในและเกณฑ์ภายนอกเปรียบเทียบกัน เริ่มด้วยการวิเคราะห์ตัวประกอบของแบบทดสอบทั้งฉบับ พบว่ามีตัวประกอบอยู่ 4 ตัว จึงแบ่งแบบทดสอบออกเป็นแบบทดสอบย่อย 4 ฉบับ ตามจำนวนตัวประกอบ จากนั้นจึงวิเคราะห์หาความลำเอียงของแบบทดสอบตามเพศ ในแต่ละวิธี ดังนี้

การใช้เกณฑ์ภายใน จับคู่ความสามารถด้วยคะแนนรวมของแบบทดสอบ 4 ฉบับ ตามรูปแบบการได้คะแนนจากแต่ละแบบทดสอบเป็นตัวแบ่ง เพราะว่าแต่ละแบบทดสอบย่อย มีระดับคะแนนที่เป็นไปได้ 10 ระดับ คือ 0 ถึง 9 เพราะฉะนั้นจะมีจำนวนรูปแบบการตอบ 10^4 รูปแบบ คูณด้วยจำนวนกลุ่ม จะได้เซลล์ในตารางรวม $2 \times 10^4 = 20,000$ เซลล์

การใช้เกณฑ์ภายนอก จับคู่ความสามารถด้วยคะแนนจากภายนอก ซึ่งอาจเป็นข้อสอบตั้งแต่ 3 ข้อ หรือมากกว่า การศึกษาครั้งนี้ใช้ข้อสอบ 6 ข้อ ซึ่งได้เลือกมาเป็นเกณฑ์ โดยใช้เกณฑ์การเลือก ดังนี้

1. มีความสัมพันธ์กับลักษณะแฝงที่แบบทดสอบย่อยนั้นวัด
2. มีความชันสูง
3. มีค่าความยากคลุมช่วงความสามารถของผู้สอบ
4. ไม่มีความลำเอียง

พบจากค่าสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันว่า การวัดพื้นที่ โดยใช้ BILOG B และ BILOG C มีความสัมพันธ์กันสูงมาก ทั้งในดัชนีแบบมีเครื่องหมาย และไม่มีเครื่องหมาย ส่วนระหว่าง LOGIST และ BILOG ก็สูงเช่นกัน แต่ความคงที่ ในรูปแบบของความคลายคลึงจะน้อยกว่า ส่วนความคลายคลึงกันระหว่าง LOGIST กับ BILOG B และ BILOG C ลดลงกว่าเล็กน้อย เป็นผลที่สอดคล้องกับการศึกษา ที่คนอื่น ๆ เคยพบ ในการวัดพื้นที่ของความเบี่ยงเบนของข้อสอบ

BILOG B และ C จะดีกว่าวิธีอื่น ๆ และมีความไวในการตรวจค้นมาก เมื่อค่า b อยู่ในช่วง $+5$ มากกว่าในช่วง $+3$ ยกเว้น BILOG B ที่เท่า ๆ กันทั้ง 2 ช่วง ส่วน MMLE จะดีกว่า JMLE มาก ในการตรวจค้นความเบี่ยงเบน ในกรณี ที่ค่าสัมบูรณ์ของ b อยู่ในช่วง $+3$ และ ดีกว่าเล็กน้อยในกรณีค่าสัมบูรณ์ของ b น้อย กว่า หรือเท่ากับ 5 ในการวัดแบบมีเครื่องหมาย จำนวนตัวอย่างที่ใช้ก็ไม่มีอิทธิพล มากนักต่อการตรวจค้นความเบี่ยงเบนของข้อสอบสรุปแล้ว การใช้ BILOG แบบ B และ C ในการตรวจค้นความลำเอียงของข้อสอบที่สร้างขึ้นมาโดยทราบค่า หรือ กำหนดค่าความลำเอียงไว้ก่อนแล้วดีกว่าวิธีอื่น

คิม และโคเฮน (Kim and Cohen, 1991) ได้ศึกษาความลำเอียงของ ข้อสอบที่สร้างขึ้นอย่างจงใจให้ลำเอียง โดยการเปรียบเทียบผลการคำนวณพื้นที่ความ แตกต่างระหว่างโค้งที่ได้จากทฤษฎีการตอบข้อสอบของกลุ่มตัวอย่างผิวขาว และผิว ดำ 2 วิธี ได้แก่ การคำนวณพื้นที่ภายใต้ความสามารถที่มีขอบเขตจำกัด และที่ไม่มี ขอบเขตจำกัด

จำนวนกลุ่มตัวอย่างนี้ใช้กลุ่มละ 1,000 คน สุ่มมาจากเดิมที่มีผิวขาว 1021 คน และผิวดำ 1008 คน เพราะขีดจำกัด ในการใช้โปรแกรม BILOG ด้วย ไมโครคอมพิวเตอร์เครื่องมือเก็บรวบรวมข้อมูล คือ แบบทดสอบคำศัพท์แบบเลือกตอบ ชนิด 4 ตัวเลือก จำนวน 50 ข้อ เป็นคำศัพท์ภาษาอังกฤษมาตรฐาน 40 ข้อ อีก 10 ข้อ เป็นคำศัพท์ที่มีความง่ายสำหรับกลุ่มผิวดำ มากกว่ากลุ่มผิวขาว

การวิเคราะห์ความเป็นมิติเดียว (Unidimension) ด้วยการวิเคราะห์ ตัวประกอบสำคัญ (principal components analysis) พบว่า ตัวประกอบแรก ที่ได้ โดยไม่มีการหมุนแกน และใช้ค่าสหสัมพันธ์แบบเด็ดตัวแสดงถึงความเป็น เอกมิติได้พอเพียงในการใช้แบบจำลองทฤษฎีการตอบข้อสอบ สำหรับกลุ่มผิวขาวโดย อธิบายความแปรปรวนได้ร้อยละ 25 ในขณะที่ค่อนข้างน้อยสำหรับผิวดำ คือ อธิบาย ได้ร้อยละ 16

ใช้โมเดลการตอบข้อสอบแบบ 2 พารามิเตอร์ และแบบ 3 พารามิเตอร์
ชนิดบังคับให้ $C_R = C_F = C = .23$

ใช้วิธีการวัดพื้นที่ 4 วิธี จากพารามิเตอร์ข้อสอบที่เทียบมาตรฐานแล้ว คือ ESA ซึ่งได้แก่ พื้นที่แบบปิดมีเครื่องหมาย (Exact Signed Area) ได้จากการคำนวณพื้นที่ตลอดช่วงความสามารถโดยไม่จำกัดขอบเขต และ EUA คือพื้นที่แบบปิดไม่มีเครื่องหมาย (Exact Unsigned Area) กรณีที่เป็นพื้นที่ที่เครื่องหมาย กลุ่มอ้างอิง (R) คือ กลุ่มผิดค่า ดังนั้นค่าที่เป็นบวกของ CSA หรือ ESA จะแสดงว่าข้อสอบข้อนั้นลำเอียงเข้าข้างกลุ่มผิดค่า

ข้อสอบที่สร้างเพื่อให้เกิดความลำเอียงเข้าข้างผิดค่า 10 ข้อ ให้ค่าเป็น 1 และอีก 40 ข้อที่เหลือเป็น 0 แล้วหาค่าสหสัมพันธ์ แบบพอยท์ไบซีเรียล ระหว่างค่าดังกล่าวกับการวัด 4 วิธี เพื่อดูผลการตรวจค้นแต่ละวิธี และหาค่าสหสัมพันธ์ระหว่างวิธีต่าง ๆ เพื่อดูความคล้ายคลึงกัน

การประมาณค่าอัตราความคลาดเคลื่อนของการตรวจสอบความลำเอียงกระทำโดยการมีข้อตกลงว่า การวัดพื้นที่แต่ละวิธีมีการกระจายเป็นปกติ ค่าเฉลี่ยและค่าความเบี่ยงเบนมาตรฐานของการวัดแต่ละวิธี ประมาณจากข้อสอบ 40 ข้อ กำหนดค่าวิกฤตแบบทางเดียวเป็น 2 ค่า คือ .05 และ .01 ข้อที่มีพื้นที่ใหญ่กว่าค่าวิกฤตจะเป็นข้อสอบที่ลำเอียง การจัดประเภทข้อสอบเช่นนี้ทำให้เกิดความคลาดเคลื่อน 2 ชนิด ได้แก่ ความคลาดเคลื่อนชนิดระบุว่าไม่ลำเอียง ทั้ง ๆ ที่ลำเอียง (FN = False Negatives) และระบุว่าลำเอียง ทั้ง ๆ ที่ไม่ลำเอียง (FP = False Positives) และเมื่อใช้ $p = .05$ พบว่ามี FP มากกว่าที่ .01

ผลจากการวิเคราะห์ข้อมูลพบว่า เกือบไม่มีความแตกต่างระหว่างการคำนวณพื้นที่แบบช่วงเปิด และแบบช่วงปิดในการตรวจค้นความลำเอียงของข้อสอบ ส่วนการใช้พื้นที่แบบไม่มีเครื่องหมายไม่ว่าจะเป็นแบบช่วงเปิดหรือปิด มีความไวต่อการตรวจค้นความลำเอียงของข้อสอบมากกว่าแบบมีเครื่องหมาย นอกจากนั้นพบว่าการคำนวณพื้นที่ช่วงเปิดนั้นพื้นที่ชนิดมีเครื่องหมาย และไม่มีเครื่องหมาย มีความสัมพันธ์กันมากกว่าการคำนวณพื้นที่ช่วงปิด

ในด้านของความคลาดเคลื่อนในการตรวจค้นความลำเอียง พบว่าวิธี 3 พารามิเตอร์ โดยใช้พื้นที่แบบมีเครื่องหมาย มีอัตราความคลาดเคลื่อนต่ำกว่าวิธีอื่น ๆ ไม่มีความคลาดเคลื่อนเลยในระดับ .01 และมีความคลาดเคลื่อนในระบบข้อสอบว่าลำเอียง โดยไม่ลำเอียงจริง จำนวน 2 ข้อ ที่ระดับ .05 ส่วนวิธี

3 พารามิเตอร์ แบบไม่มีเครื่องหมายมีอัตราความคลาดเคลื่อนสูงกว่าวิธีอื่น ๆ เล็กน้อย

คลอเซอร์ และคณะ (Clauser and Others. 1991) ได้ศึกษาถึงอิทธิพลต่าง ๆ ที่มีผลต่อวิธีการตรวจค้นความลำเอียงของข้อสอบด้วย MH ข้อมูลที่ใช้เป็นข้อมูลจำลองขึ้นเพื่อใช้ในการศึกษาการตรวจค้นความลำเอียงแบบสมำเสมอ ข้อมูลที่จำลองขึ้นเป็นแบบทดสอบจำนวน 75 ข้อ มีความลำเอียงอยู่ 16 ข้อ มีระดับของความแตกต่างในค่าความยากหลายระดับระหว่างกลุ่มอ้างอิงและกลุ่มเป้าหมาย ซึ่งมีจำนวนกลุ่มละ 1,500 คน มีการวิเคราะห์ระดับความยากที่ต่างกันหลายระดับกับระดับความยากของแบบทดสอบ 5 ระดับ อำนาจจำแนก 4 ระดับ และการกระจายของความสามารรถ 2 ลักษณะ ผลการวิเคราะห์แสดงอิทธิพลจากความแตกต่างในระดับความยากรวมทั้งค่าอำนาจจำแนกและค่าความยาก ซึ่งทำให้ได้ข้อเสนอแนะว่าค่าสถิติ Mantel Haenszel นี้ใช้ได้สำหรับข้อสอบที่มีค่าอำนาจจำแนกสูง และมีแนวโน้มจะไม่สามารถตรวจค้นความลำเอียงในข้อสอบที่มีค่าความยากสูง

เมเซอร์ และคณะ (Mazor and Others. 1991) ได้ใช้วิธีการตรวจค้นความลำเอียงแบบ MH ศึกษาข้อมูลที่จำลองขึ้นโดยใช้กลุ่มตัวอย่างขนาด 2,000 1,000 500 200 และ 100 คน เพื่อเปรียบเทียบอัตราการตรวจค้นโดยใช้ข้อสอบที่จำลองขึ้น 5 ชุด ชุดละ 75 ข้อ ผลการศึกษาพบว่าเมื่อใช้จำนวนผู้สอบ 2,000 คน วิธีนี้ตรวจค้นความลำเอียงได้ผิดพลาดร้อยละ 30 และเมื่อใช้จำนวนผู้สอบ 500 คนลงไป พบว่าตรวจค้นความลำเอียงผิดพลาดร้อยละ 50 ข้อสอบที่ไม่สามารถตรวจค้นความลำเอียงได้ ได้แก่ ข้อสอบที่ยากมาก ข้อสอบที่มีค่าความยากต่างกันเล็กน้อยสำหรับคน 2 กลุ่ม และเป็นข้อสอบที่มีค่าอำนาจจำแนกต่ำ

ไรอัน (Ryan. 1991) ได้ตรวจค้นความลำเอียงของข้อสอบด้วยวิธีการที่เสนอโดยแมนเทล และแฮนส์เซล (1959) โดยใช้ตัวอย่างกลุ่มอ้างอิง 500 คน และกลุ่มเปรียบเทียบ 100 คน เพื่อตรวจสอบความคงที่ของค่าประมาณความลำเอียงด้วย MH ข้ามกลุ่มตัวอย่างและขนาดตัวอย่างต่าง ๆ กัน และเพื่อศึกษาว่า MH มีความแข็งแกร่งต่อรบกวนของข้อสอบเพียงไร โดยมีข้อสอบรวมจำนวน 40 ข้อ และข้อสอบที่ใช้สลับกันอีกจำนวน 35 ข้อ กลุ่มตัวอย่างเป็นชาวผิวขาวและผิวดำจำนวน 5015 คน และ 670 คน ตามลำดับ โดยมีเกณฑ์การแยกประเภทข้อสอบจากผลการวิเคราะห์เป็น 3 ประเภท

ประเภท A ข้อสอบที่ทำหน้าที่ได้ตรง มีค่า MH < 1.00 และมีค่าไม่ต่างจาก 0 อย่างมีนัยสำคัญ

ประเภท B ข้อสอบที่พอจะนำมาใช้ได้ แต่ควรเลือกข้อที่มีค่า MH ต่ำคือ $1 = MH < 1.5$ และไม่ต่างจาก 1 อย่างมีนัยสำคัญ

ประเภท C เลือกมาใช้ถ้าตรงตามวัตถุประสงค์เฉพาะที่ระบุไว้อย่างมีนัยสำคัญว่าต้องการข้อสอบลักษณะนี้ได้แก่ข้อที่มีค่า $MH > 1.5$ และมากกว่า 1 อย่างมีนัยสำคัญ

ขั้นตอนในการวิเคราะห์ขั้นแรกมีการวิเคราะห์ 5 แบบในการสร้างฐาน เรียกว่า ชุด 1 ทำโดยการสุ่มตัวอย่างผิวขาว 670 คนเป็นกลุ่มเป้าหมาย ที่เหลือเป็นกลุ่มอ้างอิงแล้วสุ่มกลุ่มผิวขาวมาอีก 4 กลุ่มเป็นกลุ่มอ้างอิงได้แก่ W_1 W_2 W_3 W_4 และวิเคราะห์ข้อสอบรวม 40 ข้อด้วย MH จับคู่การวิเคราะห์เป็น W/W W/W_1 W/W_2 W/W_3 W/W_4 คะแนนที่ใช้แบ่งระดับความสามารถคือ คะแนนรวมจากข้อสอบ 40 ข้อ

การวิเคราะห์ชุด 2 ทำในลักษณะเดียวกับ ชุด 1 แต่ใช้กลุ่มเป้าหมายเป็นผิวดำวิเคราะห์ด้วย MH 5 คู่ ได้แก่ BW BW_1 BW_2 BW_3 BW_4 เป็นการวิเคราะห์เพื่อดูความคงที่ของค่า MH ข้ามกลุ่มและขนาดตัวอย่าง

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

และชุด 4 เป็นการวิเคราะห์เช่นเดียวกับชุด 3 แต่ใช้ข้อสอบ 4 ฟอร์มที่ไม่มีข้อสอบรวม (มีข้อสอบ 35 ข้อ) และใช้คะแนนรวมจาก 35 ข้อ เป็นเกณฑ์แบ่งระดับความสามารถ

ผลการวิเคราะห์พบว่า ความลำเอียงสำหรับกลุ่มผิวขาวด้วยกัน 5 คู่ นั้นมีความสัมพันธ์กันต่ำมาก ไม่มีความแตกต่างอันเนื่องมาจากความลำเอียง แต่มีความแตกต่างจากความคลาดเคลื่อนในการสุ่มตัวอย่างที่มีความซ้ำซ้อนกันมากกว่า จึงอาจเป็นการลำเอียงแบบไม่แท้ก็ได้

เกณฑ์การแบ่งช่วงระดับความสามารถ ซึ่งเป็นบริบทที่ศึกษาไม่มีผลต่อค่า MH ไม่ว่าจะใช้เกณฑ์คะแนนรวม 35 ข้อ หรือ 75 ข้อ ก็ได้ค่าสหสัมพันธ์ของการเปรียบเทียบแต่ละคู่เท่า ๆ กัน

ผลจากการจำแนกประเภทข้อสอบมีเพียง 1 ข้อ ที่มีความสอดคล้องกับทุกการวิเคราะห์ คือ เป็นประเภท B ทุกครั้ง นอกนั้นมีความแปรปรวนไปตามการวิเคราะห์

เชน (CHEN, 1991) ได้พัฒนาการใช้วิธีการเปรียบเทียบรายคู่ (P.C) ของการตอบตามแนวของราชส์โมเดล โดยการคำนวณแบบรายคู่อิสระกับรายคู่แบบ

ไม่อิสระที่ใช้วิธีการกระดถอยภายใต้น้ำหนักกำลังสองที่น้อยที่สุด และกระดถอยโลจิส การทดสอบการกะประมาณค่าพารามิเตอร์ที่ต่างกัน ที่เกี่ยวกับความไม่แปรเปลี่ยน ของพารามิเตอร์ในราชส์โมเดล ใช้ค่าพารามิเตอร์ของข้อสอบเป็นค่าประเมินความลำเอียงของข้อสอบในกลุ่มที่ต่างกัน และใช้ค่าพารามิเตอร์ของแต่ละคนทดสอบความเป็นมิติเดียวกันในกลุ่มย่อยของข้อสอบ ศึกษาโดยวิธีมอนติคาร์โล วิเคราะห์ด้วยราชส์โมเดลของไรท์ และสโตน (Wright and Stone, 1979) UML และ Prox พบว่าในข้อสอบที่ยาก วิธีการกะประมาณความลำเอียงค่าพารามิเตอร์ วิธีการกระดถอยโลจิสติก จะกะประมาณภายใต้รายคู่ที่ไม่อิสระได้คงที่กว่าภายใต้รายคู่ที่อิสระ มีความลำเอียงน้อย และค่าความสามารถลุ่มก็แปรเปลี่ยนความลำเอียงและความแปรเปลี่ยนของการลุ่มจะน้อยลงเมื่อขนาดกลุ่มตัวอย่าง หรือความยาวข้อสอบเพิ่มขึ้น รายคู่ที่อิสระจะประมาณค่าได้ใกล้เคียงกับค่าพารามิเตอร์จริงมากกว่า UML และ Prox เมื่อขนาดกลุ่มตัวอย่างใหญ่ ($N = 500$) ข้อมูลจำลองที่ใช้เปรียบเทียบการแจกแจงของค่าสถิติภายใต้สมมติฐานศูนย์ที่กล่าวถึงความคงที่ของค่าพารามิเตอร์ระหว่างกลุ่มที่มีการแจกแจงแบบ t และ χ^2 ผลการศึกษาพบว่า การวัดความเหมาะสมของข้อมูล มีการแจกแจงใกล้เคียงแบบ t มากกว่าในทางทฤษฎีเมื่อใช้การเปรียบเทียบรายคู่อิสระ

มาจุมดาร์ (Majumdar, 1991) ได้ศึกษาการกะประมาณขนาดกลุ่มตัวอย่างด้วยการเปรียบเทียบ 3 วิธีคือ วิธีแบบแมนเทิล-แฮนส์เซล (MH) วิธีของจอห์นสัน-เนย์แมน (JN) และสหสัมพันธ์ส่วนร่วม Partical Correlation (PC) ใช้วิธีมอนติคาร์โล จัดข้อสอบจำนวน 25, 50 ข้อ ขนาดกลุ่มตัวอย่าง 50, 100 คน ใช้ข้อสอบลำเอียง 3 รูปแบบ คือ ลำเอียงอย่างมีระบบ (โค้ง ICC ตัดกัน) ลำเอียงอย่างไม่เป็นระบบ (ลาดชัน) และทั้ง 2 อย่าง และระดับความสูงต่ำของความลำเอียง ข้อสอบทั้งหมด 28 ข้อ ที่ตัดมาจากคอมพิวเตอร์ 1,000 ครั้ง และวิเคราะห์ข้อมูลจริงจากกลุ่มตัวอย่าง ผลการศึกษาพบว่าภายใต้สมมติฐานศูนย์วิธีของจอห์นสัน-เนย์แมนจะคงที่ และมีศักยภาพมากกว่าวิธีสหสัมพันธ์ส่วนร่วม และวิธีแมนเทิล-แฮนส์เซล ในข้อสอบที่ลำเอียงอย่างไม่มีรูปแบบ ส่วนกรณีข้อสอบลำเอียงอย่างมีรูปแบบนั้น วิธีจอห์นสัน-เนย์แมน มีศักยภาพพอ ๆ กับวิธีสหสัมพันธ์ส่วนร่วม และวิธีแมนเทิล-แฮนส์เซล และเมื่อรวมทั้ง 2 อย่าง วิธีของจอห์นสัน-เนย์แมนก็ยังคงดีกว่า และคาดว่าเทคนิคของจอห์นสัน-เนย์แมนมีข้อมูลในการทดสอบนัยสำคัญทางสถิติในมิติของความสามารถดีกว่า

โอนัล (O'Neal. 1991) ได้เปรียบเทียบวิธีการตรวจสอบความลำเอียง 6 วิธี คือ การวิเคราะห์ความแปรปรวน (Distractor analysis) วิธีแมนเทล-แฮนส์เซล, วิธีการหนดจุดเดลต้าของแองกอฟฟ์ วิธี modified เดลต้าของแองกอฟฟ์ ดัชนีสหสัมพันธ์ร่วม และค่าความยาก โดยพิจารณาความสามารถในการตรวจสอบทั้ง 3 ขั้นตอนของการพัฒนาข้อสอบ ใช้ผลจากการทดลองข้อสอบเมื่อ Spring 1989 จำนวน 1,000 คน ข้อสอบ 36 ชุด Fall 1989 จำนวน 460 คน ข้อสอบ 3 ชุด และ Spring 1990 จำนวน 46,000 คน เป็นข้อสอบการอ่าน, คณิตศาสตร์ และ ภาษาของนักเรียนเกรด 9 ที่วัดความรู้พื้นฐานของรัฐ Alabama ศึกษาเปรียบเทียบใน 4 กลุ่มย่อย คือ ผู้หญิงผิวขาวกับผิวดำ ผู้ชายผิวขาวกับผิวดำ ผู้หญิงผิวดำกับผู้ชายผิวดำ และผู้หญิงผิวขาวกับผู้ชายผิวขาว วิธีการตรวจสอบที่คงที่มากที่สุดคือ วิธีใช้ค่าความยากโดยตรง ตามด้วยการวิเคราะห์องค์ประกอบ วิธีแมนเทล-แฮนส์เซล และวิธีสหสัมพันธ์ส่วนร่วมคงที่น้อย ส่วนวิธีการของแองกอฟฟ์ทั้ง 2 จะคงที่น้อยที่สุด ซึ่งวิธีการของแองกอฟฟ์ไม่ใช่เป็นวิธีการที่ผู้วิจัยต้องการ เพราะแยกข้อสอบได้ไม่เหมาะสม ผลการศึกษาพบว่ามีความเป็นไปได้ที่จะใช้มากกว่า 2 ดัชนีในตัวผู้สอบในการทำข้อสอบ (60 ข้อ)

มิลเลอร์ และโอชิม่า (Miller and Oshima. 1992) ได้ใช้วิธีการกะประมาณความลำเอียงแบบ 2 ขั้นตอน กับ 6 ดัชนี คือ ขนาดกลุ่มตัวอย่างจำนวน ข้อสอบที่ลำเอียง และขนาดของความลำเอียง โดยใช้วิธีการตรวจสอบความลำเอียงของแมนเทล-แฮนส์เซล ใช้ 2 ขั้นตอนเพื่อพิจารณาข้อสอบที่ตรวจพบว่าไม่ลำเอียงแต่ดูเหมือนจะลำเอียง และใช้ในการพิจารณาขนาดของความลำเอียงมีขนาดใหญ่ และจำนวนข้อสอบที่ลำเอียงมีมาก (20 % หรือ 40 %) น้ำหนักของค่าดัชนีจะสอดคล้องทางบวกมากกว่าน้ำหนักดัชนีในทฤษฎีการตอบข้อสอบ และวิธีแมนเทล-แฮนส์เซลผิดพลาดทางบวก แต่ไม่เหมือนกับความลำเอียง เช่นเดียวกับทฤษฎีการตอบข้อคำถาม

✓ โอชิม่า และมิลเลอร์ (Oshima and Miller. 1992) ได้ตรวจสอบมิติของแบบทดสอบ และตรวจสอบความลำเอียงของแบบทดสอบโดยใช้ทฤษฎีการตอบข้อคำถาม ผลจากการพิสูจน์ในหลายมิติจากข้อมูลจำลองพบว่า ดัชนี ทฤษฎีการตอบข้อคำถามสามารถแยกข้อสอบหลายมิติที่ลำเอียงได้ แต่ไม่สามารถแยกข้อสอบที่มีหลายมิติแต่ไม่ลำเอียง แต่การศึกษานี้ไม่ได้พิจารณาผสมกันระหว่างกลุ่มต่าง ๆ

✓ อัคเคอร์แมน (Ackerman. 1992) ได้ตรวจสอบความลำเอียงของข้อสอบ ผลกระทบและความเที่ยงตรงจากหลายมิติ ใช้วิธีทฤษฎีการตอบข้อคำถาม และวิธี

แมนเทล-แฮนส์เซล โดยการจำลองข้อสอบที่ลำเอียงในทุกลำดับ โดยใช้วิธีมอนติคาร์โล พบว่าความแตกต่างของความลำเอียงของข้อสอบจะมีผลกระทบต่อข้อสอบและเกี่ยวข้องกับความเที่ยงตรงของข้อสอบ

ออดตาโร (UTTARO. 1992) ได้ศึกษาวิธีวิเคราะห์ DIF ด้วยวิธีแมนเทล-แฮนส์เซล โดยเปรียบเทียบกับวิธีทฤษฎีการตอบข้อคำถาม (IRT) ใช้ข้อมูลจำลอง ประเมินความคงที่ของ λ^2 โดยประมาณค่าความคลาดเคลื่อนแบบที่ 1 และ 2 ความแปรปรวนและการจำแนก DIF ของ ETS วิธีการที่ศึกษาทำให้ไม่ตรงกันกับการเปรียบเทียบใช้ความยากของราชส์โมเดลิ องค์ประกอบที่มีผลในการกะประมาณความแปรปรวนของ α_{MH} คือผู้สอบ ซึ่งจะไวต่อโค้งลักษณะเฉพาะของข้อสอบ (ICC) และให้ค่าความสามารถ (θ) ไม่ตรงกัน ค่าความเชื่อมั่นของคะแนนรวม ทำให้ทราบความสามารถในการกะประมาณและเป็นจุดที่น่าสนใจของวิธีการแมนเทล-แฮนส์เซล

อิบราฮิม (IBRAHIM. 1992) ศึกษาผลการตรวจสอบความลำเอียงด้วยวิธีแปลงค่าความยากไคสแควร์ แมนเทล-แฮนส์เซล, แมนเทล-แฮนส์เซลแบบเดลต้า กระถกถอยโลจิส SOS2 SOS4 และไคสแควร์ ของลอร์ด ใช้ขนาดกลุ่มตัวอย่าง 3 ขนาด ความยาวข้อสอบ 2 ขนาด 4 กรณีของค่าอำนาจจำแนก และค่าความยาก 3 ระดับ โดยศึกษา 2 ส่วน ส่วนแรกแจกแจงผลการสอบที่ไม่ลำเอียง และส่วนที่สองสมมติให้มีความลำเอียง ความสอดคล้องของแต่ละวิธีได้ผลดังนี้ วิธีไคสแควร์ของลอร์ด ทำได้แน่นอนกว่าตรวจสอบได้ดี และลำดับที่ไม่เป็นผลจากค่าอำนาจจำแนกและความยาวของข้อสอบ ถ้ากลุ่มตัวอย่างน้อย ใช้ค่าจากตารางจะทำได้ดีพอ ๆ กับการตรวจสอบความลำเอียง ค่าดัชนีของวิธีแมนเทล-แฮนส์เซล และกระถกถอยโลจิสก็ได้เช่นเดียวกับไคสแควร์ของคามิลลี และวิธีแปลงค่าความยากตรวจสอบความลำเอียงได้อย่างไม่มีนัยสำคัญ สำหรับความลำเอียงที่เป็นระบบทั้งวิธีแมนเทล-แฮนส์เซล ไคสแควร์ และกระถกถอยโลจิสต้องปรับปรุง ในขณะที่ข้อสอบที่มีความลำเอียงไม่เป็นระบบ วิธีกระถกถอยโลจิสตรวจสอบได้เหมาะสม และวิธีไคสแควร์ ของลอร์ด มีประสิทธิภาพในการตรวจสอบได้ทุกรูปแบบ ขนาดของกลุ่มตัวอย่างจะสัมพันธ์กับค่าไคสแควร์ ทุก ๆ ดัชนีไคสแควร์ค่าที่สังเกตได้จะต่ำกว่าค่าในตาราง อาจจะเป็นเพราะว่าเกิดจากการสุ่มกลุ่มตัวอย่าง ซึ่งจะต้องมีมาก จึงจะค้นหาความลำเอียงของข้อสอบได้ อย่างไรก็ตามสัดส่วนของข้อสอบที่ลำเอียงจะพบมากขึ้นถ้ากลุ่มตัวอย่างใหญ่ สำหรับวิธีการตรวจสอบด้วยไคสแควร์ของคามิลลีวิธีแมนเทล-แฮนส์เซล และกระถกถอยโลจิส

✓ อัคเคอร์แมน และอีแวน (Ackerman and Evans. 1992) ได้ศึกษาความสัมพันธ์ระหว่างระดับของความเชื่อมั่นและศักยภาพของวิธีการตรวจสอบความลำเอียง 2 วิธีคือ วิธีแมนเทล-แฮนส์เซล ของฮอลแลนด์ และเธอเยอร์ (Holland and Thayer. 1988) และ SIB ของเชอร์เลย์ และสแตท์ (Shealy and Stout. 1991) ใช้คะแนนดิบจากกลุ่มที่สนใจ ผลปรากฏว่าคะแนนจากการสอบได้แสดงถึงความสามารถของผู้สอบ ถ้าผู้สอบทำคะแนนสอบผิดพลาด (แบบทดสอบมีความเชื่อมั่นต่ำ) วิธีการตรวจสอบก็ไม่สามารถตรวจสอบให้แน่นอนได้ด้วย จากการศึกษามอนติคาร์โล ข้อสอบ 40 ข้อ 720 เงื่อนไขพบว่า ข้อสอบที่มีความยาวจำกัด ศักยภาพของค่าสถิติทั้งสองจะสูงพอสมควร เช่นเดียวกันกับค่าความเชื่อมั่น เมื่อกลุ่มตัวอย่างเพิ่มขึ้น อย่างไรก็ตาม การจับคู่กลุ่มตัวอย่างเล็ก ๆ และมีความเชื่อมั่นสูง ศักยภาพจะลดลงสำหรับเงื่อนไขที่สมมติทั้งวิธีแมนเทล-แฮนส์เซล และ SIB

สเคค และ ลิสลิทซ์ (Skaggs and Lissitz. 1992) ได้ศึกษาความคงที่ของการตรวจสอบความลำเอียงของข้อสอบระหว่างการจัดการข้อสอบที่ต่างกัน ของข้อสอบเดียวกัน โดยใช้ข้อมูลจากการสอบวิชาคณิตศาสตร์ของนักเรียนเกรด 8 เกือบ 6,600 คน พบว่าวิธีแมนเทล-แฮนส์เซล และทฤษฎีการตอบข้อคำถามที่ใช้ผลรวมยกกำลังสองจะคงที่มากที่สุด

แฮร์ริส และคาร์ลตัน (Harris and Carlton. 1993) ได้ศึกษาลักษณะกว้าง ๆ ของข้อสอบเพื่อชี้ให้เห็นจุดเด่นและจุดด้อย สำหรับผู้สอบหญิงและชายโดยแบ่งลักษณะที่ศึกษาเป็นประเด็นที่วัด รูปแบบและเนื้อหาของข้อสอบ

ข้อสอบที่นำมาศึกษาเป็นข้อมูลการตอบข้อสอบในแบบทดสอบ SAT-M จำนวน 6 ฟอรัม ซึ่งเป็นตอนที่ทดสอบเฉพาะคณิตศาสตร์ ประกอบด้วยข้อสอบ 60 ข้อ เป็นการแก้ปัญหาทางคณิตศาสตร์ทั่วไป 40 ข้อ และการเปรียบเทียบปริมาณตัวเลข 20 ข้อ ผู้สอบเป็นนักเรียนชั้นมัธยมทั้งต้นและปลาย เป็นชาย 181,228 คน หญิง 198,668 คน ทุกคนรายงานด้วยตนเองว่า ภาษาอังกฤษเป็นภาษาที่แต่ละคนใช้ได้ดีที่สุด จำนวนกลุ่มตัวอย่างที่ใช้ใน SAT-M ฟอรัม 1-6 เป็นชาย 6,329 - 74,283 คน หญิง 6,712 - 83,945 คน

ผู้วิจัยได้ระบุลักษณะเดิมของข้อสอบโดยใช้ข้อมูลจากการศึกษาในครั้งก่อน ๆ ของผู้อื่น เช่น Bleistein, 1986 ; Donlon, Ekstrom และ Lockheed, 1979 ; Doolittle และ Cleary, 1987 ; McPeck และ Wild, 1987 ; Schmitt และ Dorans. 1987 ; Wendler และ Carlton, 1987 เป็นต้น สิ่งที่จะได้แก่ ประเด็นที่วัด (นั่นคือ เนื้อหาหลักได้แก่ เลขคณิต พีชคณิต เรขาคณิต

หรือทั่ว ๆ ไป) รูปแบบข้อสอบ (มีการใช้ตัวเลือกที่ว่า "ไม่สามารถตัดสินได้หรือไม่") และเนื้อหาของข้อสอบ (มีการอ้างอิงด้านเพศหรือไม่)

ตรวจสอบด้วยวิธีแมนเทิล-แฮนส์เซล และพิจารณาความแตกต่างของค่าเฉลี่ย ซึ่งใช้แสดงถึงความยากของข้อสอบใน SAT อยู่แล้ว โดยที่ค่าเฉลี่ย = 1.00 แสดงว่าข้อสอบมีความยาก สำหรับกลุ่มหนึ่งมากกว่าอีกกลุ่มเท่ากับ 1 เฉลี่ย หรือเท่ากับความแตกต่างร้อยละ 10 โดยประมาณ

ค่าเฉลี่ยที่เป็นลบแสดงว่ายากสำหรับหญิงมากกว่าชาย ค่าบวกแสดงว่ายากสำหรับชายมากกว่าหญิง

ผลการวิเคราะห์โดยจับคู่ชายและหญิงในด้านความสามารถ ซึ่งได้แก่คะแนนรวมจากแบบทดสอบแล้ว ได้ค่าเฉลี่ยมีค่า -1.86 ถึง 1.27 ค่าเฉลี่ย -.01 และค่าความเบี่ยงเบนมาตรฐาน = .51 และใช้การวิเคราะห์ความแปรปรวนทดสอบนัยสำคัญของความแตกต่างเฉลี่ยระหว่างข้อที่มีและไม่มีลักษณะดังที่ระบุไว้ พบว่าโดยทั่ว ๆ ไป ถ้าเนื้อหาหลักเป็นเรขาคณิตหรือเลขคณิตแล้วชายจะทำได้ดีกว่าหญิง แต่กรณีเป็นเนื้อหาทั่ว ๆ ไป (เช่น โครงสร้างของระบบจำนวน เซต ฯลฯ) หญิงจะทำได้ดีกว่าชาย ถ้าเป็นเลขคณิตและพีชคณิต หญิงจะทำได้ดีกว่าชาย แต่ถ้าเป็นเลขคณิตและเรขาคณิตแล้ว ชายจะทำได้ดีกว่าหญิง การพบว่าหญิงทำได้ดีกว่าในกรณีของพีชคณิตและชายจะทำได้ดีกว่าในเรขาคณิต (โดยที่หญิงและชายมีความสามารถรวมเท่ากัน) นั้นสอดคล้องกับการค้นพบของ ดูลิทเทิล และเคลียร์ (Doolittle and Clearly, 1987)

นอกจากนั้นพบว่า ชายจะทำได้ดีกว่าในข้อสอบที่ต้องการกระบวนการทางสมองในระดับสูงกว่า ในข้อที่เป็นการประยุกต์นำมาใช้ในชีวิตประจำวัน ส่วนหญิงจะทำได้ดีกว่ากรณีที่มีตัวแปร (เช่น x , a , b) ในปัญหาและในตัวเลือก กรณีเป็นปัญหาตรงไปตรงมาจากหลักสูตรหรือหนังสือเรียนโดยไม่มีการนำมาประยุกต์ และในกรณีมีการอ้างอิงถึงบุคคลโดยไม่มีการแสดงถึงเพศหรือสถานภาพทางเชื้อชาติ

ราชู และคนอื่น ๆ (Raju and Others, 1993) ได้ศึกษาเปรียบเทียบการทดสอบของลอร์ดและ Z ของราชู (1990) ในการวัดพื้นที่ความแตกต่างระหว่างโดเมนของข้อสอบ และวิธีแมนเทิล-แฮนส์เซล ในการตรวจสอบข้อสอบลำเอียงโดยใช้ข้อมูลจากการทดสอบใช้แบบทดสอบ Gater-Mac Ginitie Reading Tests (GMRT) เฉพาะแบบทดสอบย่อยที่วัดคำศัพท์ 45 ข้อ เป็นแบบเลือกตอบ 5 ตัวเลือก กับนักเรียนระดับ 10 และ 12 ในปี 1987 จำนวน 839 คน ในจำนวนนี้มีนักเรียนผิวดำ 245 คน ผิวขาว 436 คน เป็นหญิง 440 และชาย 399 คน

ผลจากการศึกษา ในกรณีการตรวจสอบความลำเอียงระหว่างนักเรียนผิวดำ และผิวขาว พบว่า มีข้อสอบลำเอียง 1 ข้อ สำหรับการทดสอบด้วย X^2 และ Z ทั้งพื้นที่ชนิดมีเครื่องหมายและไม่มีเครื่องหมาย ได้แก่ ข้อ 41 ทั้ง 3 วิธี ส่วนวิธี แมนเทิล-แฮนส์เซล พบข้อสอบลำเอียง 2 ข้อ ได้แก่ ข้อ 14 และ 27 ทั้ง 3 ข้อ เป็นข้อที่มีความยากต่างกันมากสำหรับผู้สอบผิวดำและผิวขาว

ส่วนผลการตรวจสอบความลำเอียงระหว่างชายและหญิง 3 วิธีแรก พบ ข้อสอบลำเอียงซ้ำกัน 4 ข้อ ส่วนวิธีแมนเทิล-แฮนส์เซล พบ 5 ข้อ ซ้ำกับ 3 วิธีแรก 4 ข้อ (ร้อยละ 80) ได้แก่ ข้อ 2, 18, 23, 33 และข้อที่ไม่ซ้ำกับ 3 วิธีแรกคือ ข้อ 41 ทุกข้อมีค่าความยากสูงกว่าข้อที่ไม่ลำเอียง

การพบว่าวิธีแมนเทิล-แฮนส์เซล ค้นพบข้อสอบลำเอียงได้มากกว่าวิธีที่ใช้ ทฤษฎีการตอบข้อสอบเช่นนี้ค้านกับการค้นพบของ สวามินนาธาน และโรเจอร์ส (Swaminathan and Rogers. 1990) ที่พบว่าวิธีที่ใช้ทฤษฎีการตอบข้อสอบค้นพบ ข้อสอบลำเอียงได้มากกว่าวิธีแมนเทิล-แฮนส์เซล

ข้อควรสังเกต คือ การทดสอบ Z ของราชูอาจให้ผลถูกต้องกว่าการ กำหนดเกณฑ์ตัดสินตามอำเภอใจ (arbitrary) สำหรับพื้นที่ความแตกต่างระหว่าง ไค้ เพราะค่าความคลาดเคลื่อนมาตรฐานในการวัดเป็นตัวหนึ่งที่เกี่ยวข้อง ทำให้ข้อสอบข้อที่มีความแตกต่างของพื้นที่น้อยกว่าเป็นข้อสอบลำเอียง และข้อสอบที่มี ค่าความแตกต่างของพื้นที่มากกว่าเป็นข้อสอบไม่ลำเอียงได้

รูสโซส์ และสตาท์ (Roussos and Stout. 1993) ได้ศึกษาผลของ การใช้จำนวนตัวอย่างขนาดเล็ก ความแตกต่างกันของค่าพารามิเตอร์ข้อสอบรวมทั้งความแตกต่างในค่าเฉลี่ยของความสามารถของกลุ่มอ้างอิง และกลุ่มเปรียบเทียบ เพื่อศึกษาถึงผลที่มีต่อความคลาดเคลื่อนชนิดที่ 1 (Type I error) ซึ่งเกิดจากการปฏิเสธสมมติฐานศูนย์ที่เป็นจริง หรือในกรณีนี้ คือ การยอมรับว่าข้อสอบลำเอียง ทั้ง ๆ ที่ข้อสอบไม่ลำเอียง

การศึกษาที่ 1 ใช้กลุ่มตัวอย่าง และกลุ่มเปรียบเทียบกลุ่มละ 100 200 500 และ 1000 คน เลือกข้อสอบ 1 ข้อ จากแบบทดสอบ ASVAB เลือกข้อที่มีค่า พารามิเตอร์ของข้อสอบคือ a b และ c ใกล้เคียงกับค่าเฉลี่ยของแบบทดสอบมากที่สุด คือ 1.32 0.03 และ 0.25 (ค่าเฉลี่ยคือ 1.22 0.00 และ 0.02) และเป็น ข้อสอบที่ไม่ลำเอียงความสามารถ (θ) ของกลุ่มตัวอย่างทั้ง 2 กลุ่มถูกจำลองให้มีการกระจายเป็นปกติมีค่าความแปรปรวนเป็น 1.00 ความแตกต่างระหว่างค่าเฉลี่ย ของการแจกแจงของ 2 กลุ่ม ถูกจำลองให้มีค่า 0.0 0.5 และ 1.0 แล้วคำนวณ

ค่าความสามารถเฉลี่ยของแต่ละกลุ่มให้จุดกลางของค่าเฉลี่ยอยู่บนระดับความยากเฉลี่ยของแบบทดสอบ จำลองการตอบ 400 ครั้ง และวิเคราะห์ 400 ครั้ง ในแต่ละสถานการณ์

ผลการใช้ SIBTEST และวิธีแมนเทิล-แฮนส์เซล ตรวจสอบความลำเอียงของข้อสอบที่เลือกขึ้นม้างกล่าว พบว่านอกจากกรณีกลุ่มตัวอย่างขนาด 100 คน และค่าความแตกต่างของความสามารถ (d_r) = 0.0 แล้ว อัตราการปฏิเสธสมมติฐานศูนย์ หรืออัตราความคลาดเคลื่อนประเภทที่ 1 ของวิธีทั้งสองแตกต่างกันอย่างไม่มีนัยสำคัญ ทั้งยังมีความแตกต่างจากค่าระดับนัยสำคัญที่กำหนด คือ 0.05 อย่างไม่มีนัยสำคัญเช่นกัน สำหรับความแตกต่างที่พบในจำนวนตัวอย่าง 100 คน และ $d_r = 0.0$ นั้น อัตราความคลาดเคลื่อนประเภทที่ 1 ของทั้งสองวิธีต่ำกว่า 0.05 อย่างไรก็ตาม SIBTEST มีอัตราความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธีแมนเทิล-แฮนส์เซลเล็กน้อย สำหรับการวิเคราะห์ความลำเอียงโดยใช้กลุ่มตัวอย่างไม่เกิน 1,000 คน และมีแนวโน้มว่าอัตราความคลาดเคลื่อนจะสูงขึ้นเมื่อกลุ่มตัวอย่างมากขึ้นทั้ง 2 วิธี

การศึกษาที่ 2 ใช้แบบทดสอบเดิม แต่เลือกค่า a b และ c สำหรับข้อสอบที่นำมาวิเคราะห์ให้แตกต่างกันโดย

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

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

$$c = 0.20$$

จำลองการกระจายของความสามารถเป้าหมายให้เป็นปกติ d_r มีค่า 0.0 และ 1.0 นอกจากนั้นการจำลองทุกอย่างเป็นเช่นเดียวกับการศึกษาที่ 1 แต่ใช้จำนวนผู้สอบขนาดใหญ่ คือ 500 1,000 และ 3,000 คน และแต่ละสถานการณ์จำลองการตอบ 100 ครั้ง และตรวจสอบโดยใช้ SIBTEST และวิธีแมนเทิล-แฮนส์เซล พบว่าเมื่อค่า a สูงขึ้นและ b ต่ำ อัตราการปฏิเสธสมมติฐานจะสูงกว่าระดับที่กำหนดไว้คือ 0.05 ทั้ง 2 วิธี และวิธีแมนเทิล-แฮนส์เซล จะมีอัตราการปฏิเสธสูงกว่า SIBTEST และยิ่งจำนวนกลุ่มตัวอย่างใหญ่ขึ้น อัตราการปฏิเสธจะยิ่งสูงขึ้น ในขนาดกลุ่มตัวอย่าง 3,000 คน ที่ค่า $a = 2.5$ และ $b = -1.5$ อัตราการปฏิเสธของวิธีแมนเทิล-แฮนส์เซลสูงถึง 1.00 และ SIBTEST = 0.91 ในกรณีที่ $d_r = 1.0$ และ $c = 0.20$ ในภาพรวมแล้ว SIBTEST มีอัตราการปฏิเสธต่ำกว่า MH แต่เมื่อใช้ค่า $d_r = 0.0$ ไม่มีค่า c ผลปรากฏว่า อัตราการ

ปฏิเสธของทั้ง 2 วิธีอยู่ในระดับใกล้เคียงกับ 0.05 แม้ว่า SIBTEST จะสูงกว่าวิธีแมนเทล-แฮนส์เซล แต่ไม่มีนัยสำคัญ

สรุปว่าเมื่อใช้จำนวนตัวอย่างน้อย ไม่เกิน 1,000 คน วิธีแมนเทล-แฮนส์เซล และ SIBTEST ให้ผลในด้านความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกัน และอยู่ภายในระดับ 0.05 ในขณะที่พบว่า SIBTEST มีอัตราความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธีแมนเทล-แฮนส์เซล ในกรณีที่จำนวนตัวอย่างมีขนาดใหญ่ (มากกว่า 1,000 คน) และค่าเฉลี่ยของการแจกแจงของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบมีความแตกต่างกัน รวมทั้งข้อสอบมีความยากต่ำ และค่าอำนาจจำแนกสูง นอกจากผลทางด้านอัตราการใช้ปฏิเสธแล้ว ยังพบว่าตัวประมาณความลำเอียงคือ β และ Δ ที่มีค่าต่ำกว่าเกณฑ์หรือมีค่าแสดงความลำเอียงเพียงเล็กน้อยก็อาจมีนัยสำคัญได้ ($\Delta = 1.5$ และ 1.0 $\beta = .10$ หรือ $.05$)

สมิท (Smith. 1993) ได้ศึกษาความลำเอียงของข้อสอบประเมินผลการศึกษาส่วนกลาง โดยใช้วิธีการตรวจสอบความลำเอียง 4 วิธี คือ ไคสแควร์ของซุเนแมน ค่าความยากของรัตเนอร์และคอนเวย์ 1 พารามิเตอร์ ของราซส์ โมเดล และแมนเทล-แฮนส์เซล ใช้วิชาภาษาฝรั่งเศสเป็นตัวตรวจสอบมาตรฐานส่วนกลางใน 6 วิชา กับนักเรียนเกรด 4, 8, และ 11 ให้เฉพาะข้อมูลของนักเรียนเกรด 8 ที่สอบในปี ค.ศ. 1988 - 1989 เป็นวิชาการอ่านทั่วไปกับคณิตศาสตร์ แบ่งนักเรียนเป็น 2 กลุ่ม กลุ่มที่ 1 ใช้ภาษาฝรั่งเศสเป็นภาษาที่ 2 และสามารถพูดภาษาอังกฤษได้คล่อง จำนวน 336 คน และอีกกลุ่ม 336 คน จะพูดภาษาอังกฤษอย่างเดียว พบว่าข้อสอบวิชาคณิตศาสตร์ 8 ข้อใน 50 ข้อ และข้อสอบการอ่าน 9 ข้อใน 50 ข้อ แสดงถึงความแตกต่างระหว่างกลุ่มอย่างชัดเจน แต่ไม่มีนัยสำคัญ แต่ลงความเห็นว่าจะแนบสอบที่ต่ำกว่าคะแนนเฉลี่ยของกลุ่มนักเรียนที่ใช้ทั้ง 2 ภาษา ไม่ใช่มาจากความลำเอียงของข้อสอบ แต่จะมาจากพื้นฐานก่อนการเรียน ซึ่งจะเห็นเช่นเดียวกับที่พบในกรณีของความขัดแย้งของคะแนน

โคเฮน และคิม (Cohen and Kim. 1993) ศึกษาเปรียบเทียบการใช้ทฤษฎีการตอบข้อสอบ 2 พารามิเตอร์ โดยใช้สถิติทดสอบ χ^2 ของลอร์ด และสถิติทดสอบพื้นที่ความแตกต่างระหว่างโค้งลักษณะข้อสอบชนิดมีเครื่องหมายและไม่มีเครื่องหมาย อันได้แก่ สถิติ Z ของราซส์ โดยใช้ข้อมูลจำลองตามทฤษฎีการตอบข้อสอบ 2 พารามิเตอร์ จำลองการตอบข้อสอบ 2 ขนาด คือ 20 และ 60 ข้อ จำนวนผู้สอบ 100 และ 500 คน และมีจำนวนข้อสอบลำเอียงในแบบทดสอบจำนวนร้อยละ 0 10 และ 20

ผลการวิเคราะห์พบว่า

1. แบบทดสอบที่ยาวการระบุข้อสอบลำเอียงผิดพลาด (FP) จะมากขึ้น และถ้ามีจำนวนข้อสอบลำเอียงอยู่ในแบบทดสอบน้อย จำนวนข้อสอบที่ถูกระบุว่าลำเอียงผิดพลาดจะยิ่งมากขึ้น

องค์ประกอบที่มีอิทธิพลต่อการระบุว่าข้อสอบลำเอียงผิดพลาด (FP) อีกประการ คือ ระดับนัยสำคัญ (α) ค่า α ยิ่งมาก FP จะเกิดมาก (ที่ .05 จะเกิด FP มากกว่าที่ .01) ผลดังนี้เกิดในทุกสถานการณ์ที่ศึกษา สำหรับ Z แต่เกิดในบางสถานการณ์สำหรับ X^2

2. แบบทดสอบที่ยาวกว่า มีร้อยละของข้อสอบลำเอียงมากกว่า จะมีข้อสอบที่ถูกระบุว่าไม่ลำเอียงผิดพลาด (FN) มากกว่า แต่จะเกิดขึ้นกับการทดสอบด้วย Z มากกว่า

กรณีที่จำนวนผู้สอบเท่ากับ 100 คน พบว่าความแตกต่างในการระบุว่าข้อสอบไม่ลำเอียงผิดพลาด สำหรับ Z และ X^2 ไม่แตกต่างกันมากเท่ากรณีใช้ตัวอย่าง 500 คน ซึ่งมีการระบุว่าไม่ลำเอียงผิดพลาด สำหรับ X^2 น้อยกว่า Z โดยเฉพาะอย่างยิ่งในกรณีที่ข้อสอบลำเอียงในแบบทดสอบมากที่สุดคือ ร้อยละ 20

โคเฮน และคิม วิเคราะห์ว่า การที่ค่าสถิติทดสอบ X^2 ให้ผลการตรวจสอบข้อสอบลำเอียงดีกว่า Z ของพื้นที่ที่ติดเครื่องหมายอาจเป็นเพราะ Z (ESA) ใช้การเปรียบเทียบเฉพาะค่าความยาก (b) ในขณะที่ X^2 ใช้ทั้งค่าความยากและอำนาจจำแนก (b และ a) แต่ไม่สามารถอธิบายความแตกต่างได้ในกรณีใช้ Z (H) ซึ่งทดสอบพื้นที่ที่ไม่ติดเครื่องหมายเพราะ Z (H) ก็ใช้การทดสอบค่าพารามิเตอร์ทั้ง 2 ตัวเช่นกัน

✓ เชลีย์ และสแตท (Shealy and stout. 1993) ได้พัฒนาวิธีการที่จะตรวจสอบความลำเอียงของข้อสอบจากกลุ่มตัวอย่างที่มีความสามารถต่างกัน และตรวจสอบได้ดี โดยพัฒนาจากหลายมุมมองของทฤษฎีการตอบข้อคำถามที่ใช้ตรวจสอบความลำเอียงและกะประมาณ DIF ไปพร้อม ๆ กันหลาย ๆ ข้อ ซึ่งจะใช้วิธีจำลองข้อมูลด้วยคอมพิวเตอร์คือ โมเดล SIBTEST

เวลช์ และฮูเวอร์ (Welch and Hoover. 1993) ได้พัฒนาวิธีการ 2 วิธีคือ HW1 และ HW3 ที่ใช้ตรวจสอบความลำเอียงของข้อสอบที่คะแนนสอบเป็นโพลีโทมัส (Polytomously) ได้ทดลองใช้กับข้อมูลจำลองได้ดี

คลัสเซอร์ (Clauser. 1993) ได้ศึกษาวิธีการตรวจสอบความลำเอียงด้วยวิธีแมนเทิล-แฮนส์เซลเพิ่มเติมใน 5 สถานการณ์ คือ

ครั้งที่ 1 ศึกษาตามการแนะนำของ ฮอลแลนด์ และเธเยอร์ (Holland and Thayer) เป็นวิธีการ 2 ขั้นตอน พบว่า 2 ขั้นตอนสัมพันธ์กับการลดความคลาดเคลื่อนแบบที่ 2 (Type II)

ครั้งที่ 2 ทดสอบความสามารถของวิธีแมนเทล-แฮนส์เซล ที่จะพิสูจน์ความลำเอียงอย่างเป็นระบบกับความแปรเปลี่ยนของค่าความยากและอำนาจจำแนก และกับค่าความยากที่แตกต่างระหว่างกลุ่ม พบว่าเมื่อความแตกต่างของค่าความยากคงที่ค่าอำนาจจำแนกต่ำ เช่นเดียวกับข้อสอบที่ยาก

ครั้งที่ 3 ศึกษาขนาดของกลุ่มตัวอย่าง พบว่าถ้ากลุ่มตัวอย่างต่ำกว่า 200 คน จะไม่เหมาะที่จะใช้วิธีแมนเทล-แฮนส์เซล ควรใช้กลุ่มตัวอย่างใหญ่ขึ้น

ครั้งที่ 4 ใช้ผลของคะแนนกลุ่มพบว่าถ้าใช้น้อยกว่า NT1 จะทำให้ค่าความคลาดเคลื่อนแบบที่ 1 (Type I) เพิ่มขึ้นทันที

ครั้งที่ 5 ทดสอบกับความลำเอียงที่ไม่เป็นระบบ พบว่าวิธีแมนเทล-แฮนส์เซล ไม่สามารถตรวจสอบรูปแบบของข้อสอบลำเอียงอย่างไม่เป็นระบบได้

เฟอร์กอน (Furgon, 1993) ได้ตรวจสอบความเป็นไปได้ของการใช้การถดถอย flexible logistic (Flod) ในการประเมิน DIF โดยตรวจสอบความแน่นอนของวิธีการกับมาตรฐานการถดถอยโลจิสต์ และวิธีแมนเทล-แฮนส์เซล และศึกษาผลของเปอร์เซ็นต์ของ DIF ที่อยู่ในแบบทดสอบ (0 % 20 % 40 % และ 60 %) แล้วพิจารณาความคงที่ในการตรวจสอบและเปอร์เซ็นต์ของความสอดคล้อง ผลการศึกษาพบว่าการถดถอย Flod ดีกว่ามาตรฐานการถดถอยโลจิสต์ และวิธีแมนเทล-แฮนส์เซล ในการคิด DIF ที่ไม่เป็นระบบ โดยเฉพาะค่าอำนาจจำแนกของ 2 กลุ่มต่างกัน การถดถอย Flod จะทำได้ดีกว่าแมนเทล-แฮนส์เซล เมื่อค่าความยากและค่าอำนาจจำแนกต่างกันทั้ง 2 กลุ่ม ในข้อสอบที่เบี่ยงเบนและไม่เบี่ยงเบน ในทางปฏิบัติแล้วการถดถอย Flod ตรวจสอบในเปอร์เซ็นต์ของความสอดคล้องได้ต่ำคล้าย ๆ กับมาตรฐานการถดถอยโลจิสต์ แต่ดีกว่าวิธีแมนเทล-แฮนส์เซล สรุปโดยรวมแล้ว ทุกวิธีสามารถทดสอบเปอร์เซ็นต์ของความสอดคล้องดีกว่าอัตราการตรวจสอบ ไม่มีแบบใดดีกว่ากัน แต่จะขึ้นอยู่กับส่วนและรูปแบบของ DIF

เวสเตอร์ส (Westers, 1993) ได้เสนอตัวอย่างของ DIF ที่นำไปสู่การใช้ Loglinear ของราซส์ คือ โมเดล SERE ซึ่งจะแสดง DIF เมื่อความยากต่างกัน ใช้เฉพาะกลุ่มตัวอย่างน้อย ๆ จากข้อมูลจำลองพบว่าสามารถตรวจสอบ

ความแตกต่างระหว่างกลุ่ม (DIF) ได้ดี จะนำไปสู่การพัฒนาคะแนนหลายระดับที่อธิบายคุณลักษณะหลายมิติ และช่วยนำไปสู่การปรับปรุงโมเดล SERE ต่อไป

นารายานัน และ สวามินาธาน (Narayanan and Swaminathan, 1993) ได้ศึกษาโดยการเปรียบเทียบความคลาดเคลื่อนชนิดที่ 1 (Type I) และ ศักยภาพ (power) ของวิธีตรวจสอบความลำเอียงวิธีแมนเทิล-แฮนส์เชล และ SIB โดยพิจารณาสัดส่วนการแจกแจงของ SIB และแมนเทิล-แฮนส์เชล ใช้ข้อมูลจำลองความสนใจขนาดของประชากร การแจกแจงของความสามารถในต่างกลุ่ม เพอร์เซนต์ของ DIF รูปแบบของข้อสอบและขนาดของ DIF การตรวจสอบการกระจายของ SIB และแมนเทิล-แฮนส์เชลพบว่า SIB มีการกระจายเป็นไปตามทฤษฎี เมื่อขนาดกลุ่มตัวอย่างเกินกว่า 200 ในขณะที่วิธีแมนเทิล-แฮนส์เชลไม่สามารถกระจายให้เป็นไปตามทฤษฎี วิธีการแมนเทิล-แฮนส์เชล และ SIB มีศักยภาพในการตรวจสอบ DIF ได้พอ ๆ กัน และ SIB มีศักยภาพสำหรับการแจกแจงความสามารถที่ไม่เท่ากันมากกว่าวิธีแมนเทิล-แฮนส์เชล อัตราส่วนของค่าความคลาดเคลื่อนรูปแบบที่ 1 ในวิธีแมนเทิล-แฮนส์เชลค่อนข้างจำกัด แต่ก็สูงกว่าวิธี SIB

เชียง (Chang, 1993 : 1330 A) ได้ตรวจสอบผลการใช้วิธีตรวจสอบความลำเอียง บทบาทของตัวแปรต่าง ๆ ที่ผู้สอบมาจากภูมิหลังต่างกัน ปัญหาทั่วไปในการกะประมาณค่าดัชนีความลำเอียงของข้อสอบ และเปรียบเทียบความสัมพันธ์ของวิธีการตรวจสอบความลำเอียง 5 วิธีคือ การแปลงค่าความยาก ไคสแควร์ของคามิลลี Loglinear

โมเดล วิธีแมนเทิล-แฮนส์เชล และโลจิสติกกรีเกรสชั่น โดยใช้กลุ่มตัวอย่างเป็นนักเรียนเกรด 8 จำนวน 330 คน ในโรงเรียนกึ่งชานเมืองโดยให้เรียนรวมกันทั้งกลุ่มทดลองและกลุ่มควบคุม โดยกลุ่มควบคุมได้รับการสอนปรูพิวติยา 4 สัปดาห์ก่อนทำการทดสอบด้วยข้อสอบ 44 ข้อ ขณะที่กลุ่มทดลองได้รับการสอบเฉพาะส่วนแรกของข้อสอบก่อนที่จะทดสอบทั้งหมด ข้อสอบส่วนแรกจะเป็นข้อสอบที่มั่นใจว่าไม่ลำเอียง และส่วนที่สองจะเป็นข้อสอบที่ลำเอียง ทุกวิธีการตรวจสอบความลำเอียงจะทำการเปรียบเทียบใน 8 กลุ่ม (ภูมิหลังการสอบ 5 กลุ่ม กับภายใต้จริยธรรม 3 กลุ่ม) และ 2 ระดับ คือความลำเอียงและความไม่ลำเอียง พบว่า คำสั่งเป็นส่วนสำคัญต่อจำนวนข้อสอบที่ไม่ลำเอียงและลำเอียงที่แท้จริง ส่วนของข้อสอบจะแสดงความลำเอียง และวิธีการตรวจสอบของข้อสอบที่ลำเอียง กลุ่มทดลองมีบทบาทในการนับจำนวนข้อสอบที่ลำเอียง กระบวนการกะประมาณความสามารถแสดงถึงความลำเอียงของข้อสอบที่ลำเอียงและไม่ลำเอียงได้ชัดเจน วิธีไคสแควร์มีความสัมพันธ์

กันมากกว่าการแปลงค่าความยากในการแยกกลุ่มข้อสอบลำเอียง และไม่ลำเอียง และสอดคล้องกับวิธีอื่น ๆ และองค์ประกอบเช่นขนาดของกลุ่มตัวอย่าง ความรู้ของนักเรียนก่อนจะได้รับคำสั่งมีผลต่อความลำเอียงของข้อสอบ

มูน (Moon, 1994) ได้ศึกษาวิธีตรวจสอบความลำเอียงโดยวิธีแมนเทล-แฮนส์เซลล์ กับวิธีประยุกต์ของแมนเทล-แฮนส์เซลล์ ใช้ตรวจสอบ DIF ในข้อสอบวัดการเขียนกับกลุ่มย่อย 3 ระดับ คือ ยึดเกณฑ์ภายนอก (MC) ปฏิบัติจากการเรียนรู้ข้อสอบ และทั้ง 2 อย่างรวมกัน ใช้ส่วนหนึ่งของโปรแกรมการสอบ State-mandated ระหว่างปี ค.ศ.1991 และ 1993 จากนักเรียนเกรด 6 ที่ได้รับการกระตุ้นให้เขียนในปี ค.ศ.1993 และเรียนเกรด 4 และสอบ ITBS ในปี 1991 พบว่าเมื่อใช้ตารางปฏิบัติซึ่งใช้เกณฑ์ของแมนเทล (Mantel, 1963) จะได้ค่าดีกว่าการประยุกต์วิธีแมนเทล-แฮนส์เซลล์ (1995) ในการตรวจสอบ DIF ในทางเลือกของการจัดคู่ระดับเป็นการตัดสินใจ และต้องการให้สัมพันธ์กันสูงสุดกับการศึกษาครั้งนี้

มัลเลอร์ (Maller, 1994) ได้พัฒนาแบบทดสอบวัดเชาวน์ปัญญาสำหรับเด็กของเวสเลอร์ที่ปรับปรุงครั้งที่ 3 (WISC-III) ซึ่งเดิมสร้างไว้สำหรับเด็กที่สามารถพูดฟังภาษาอังกฤษได้ ทำให้เกิดปัญหาว่าถ้าจะใช้กับเด็กหูตึงจะมีความเที่ยงตรงเชิงโครงสร้างเพียงพอหรือไม่ ข้อสอบจะทำหน้าที่เบี่ยงเบนไปหรือไม่ และธรรมชาติของเนื้อหาข้อสอบแต่ละข้อจะทำหน้าที่ลำเอียงหรือไม่ ซึ่งแบบทดสอบถูกแปลเป็นภาษาสัญลักษณ์ (ภาษามือ) และทดลองสอบกับเด็กนักเรียนหูตึง 110 คน จาก 3 แหล่ง อายุระหว่าง 8-11 ปี ($M = 13.25$, $SD = 2.37$) ที่หูตึงโดยกำเนิด ใช้ภาษามือและมี IQ ใกล้เคียงกัน ตรวจสอบความเที่ยงตรงเชิงโครงสร้างด้วย LISREL วิเคราะห์ความแปรปรวนร่วม ตรวจสอบความลำเอียงโดยราล์สโมเดลในกลุ่มย่อยของแบบทดสอบ ภาพสำเร็จ, การสื่อสาร, ความคล้ายเรขาคณิต ศัพท์ และบทสนทนา ในทุกแบบทดสอบย่อยจะมี DIF และ DIF จะเพิ่มความแตกต่างในข้อสอบที่ยาก ยกเว้นคำศัพท์ ความลำเอียงของข้อสอบอาจจะมาจากการแปลความและหลักสูตร

งานวิจัยที่เกี่ยวข้องกับการตรวจสอบความลำเอียงภายในประเทศไทย ได้มีการศึกษาตามลำดับ ดังนี้

ชัชชัย เผ่าพงศ์ (2527 : 73-76) ศึกษาความลำเอียงของข้อสอบด้วยวิธีโค้งลักษณะข้อสอบ -3 พารามิเตอร์ ตัวแปรที่ศึกษาคือเพศ แบบทดสอบที่ใช้คือแบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์ และภาษาในระดับมัธยมต้น

ซึ่งแบบทดสอบฉบับนี้พัฒนาโดยสำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชายและหญิงชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2524 ทั่วทุกภาคของประเทศ (ภาคเหนือ ภาคกลาง ภาคใต้ ภาคตะวันออก และภาคตะวันออกเฉียงเหนือ) กลุ่มซึ่งทำการวิเคราะห์แบบทดสอบด้านคณิตศาสตร์เป็นนักเรียนชาย 1,610 คน นักเรียนหญิง 1,337 คน อีกกลุ่มหนึ่งทำการวิเคราะห์แบบทดสอบด้านภาษาเป็นนักเรียนชาย 1,316 คน นักเรียนหญิง 985 คน ผลการศึกษาพบว่าแบบทดสอบด้านคณิตศาสตร์ จำนวน 30 ข้อ มีความลำเอียงต่อกลุ่มใดกลุ่มหนึ่ง โดยเฉพาะ 9 ข้อ คือลำเอียงต่อกลุ่มนักเรียนชาย 7 ข้อ และลำเอียงต่อกลุ่มนักเรียนหญิง 2 ข้อ ข้อสอบที่ลำเอียงต่อกลุ่มนักเรียนชายและหญิงในระดับปานกลางขึ้นไปมีจำนวน 5 ข้อ ซึ่งวัดในเรื่อง ร้อยละ การหาปริมาตร และการหาความยาวของด้านรูปสามเหลี่ยม จำนวนเรื่องละ 1 ข้อ อีกจำนวน 2 ข้อ เป็นเรื่องเกี่ยวกับโจทย์ปัญหาและจากข้อสอบจำนวน 5 ข้อนี้เป็นข้อสอบที่มีความลำเอียงต่อกลุ่มนักเรียนหญิงในช่วงความสามารถแรก และลำเอียงต่อกลุ่มนักเรียนชายในช่วงความสามารถต่อมา จำนวน 1 ข้อ ส่วนแบบทดสอบด้านภาษาเกี่ยวกับความเข้าใจในการอ่าน จำนวน 40 ข้อ พบว่ามีความลำเอียงต่อกลุ่มนักเรียนชายโดยเฉพาะ 3 ข้อ และลำเอียงต่อกลุ่มนักเรียนหญิงโดยเฉพาะมี 8 ข้อ ข้อสอบที่มีความลำเอียงในระดับปานกลางขึ้นไปมีจำนวน 9 ข้อ ซึ่งวัดความเข้าใจเกี่ยวกับการอ่านคำประพันธ์, บทร้อยกรอง อย่างละ 1 ข้อ และวัดความเข้าใจเกี่ยวกับการอ่านข้อความ จำนวน 7 ข้อ และจากข้อสอบจำนวน 9 ข้อนี้เป็นข้อที่มีความลำเอียงต่อกลุ่มนักเรียนชายโดยเฉพาะ 1 ข้อ เป็นข้อที่มีความลำเอียงต่อกลุ่มนักเรียนหญิงโดยเฉพาะ 6 ข้อ และมีข้อสอบ 2 ข้อที่มีความลำเอียงต่อกลุ่มนักเรียนชายในช่วงความสามารถแรก, ช่วงความสามารถต่อมาลำเอียงต่อกลุ่มนักเรียนหญิง ผลการวิจัยครั้งนี้เขาสรุปว่า แบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์และด้านภาษาฉบับนี้มีความลำเอียงต่อเพศ เขาเสนอว่าในการที่จะนำแบบทดสอบฉบับนี้ไปใช้กับกลุ่มผู้สอบเพศชายและเพศหญิง ควรจะพิจารณาข้อที่มีค่าอำนาจจำแนกของทั้ง 2 กลุ่มที่มีความแตกต่างกันไม่มาก ไม่ว่าข้อนั้นจะมีค่าอำนาจจำแนกสูงหรือต่ำก็ตาม

ทัศนีย์ พิรมนตรี (2530 : 76-78) ศึกษาวิเคราะห์ความลำเอียงของข้อสอบ 3 วิธีคือ วิธีกำหนดจุดค่าเฉลี่ย วิธีทดสอบความแตกต่างระหว่างกลุ่มด้วยสถิติไคสแควร์ในโมเดลลอกลีเนียร์ คือโมเดลที่ไม่มีพารามิเตอร์ผลรวมระหว่างระดับคะแนนกับกลุ่ม และโมเดลที่ไม่มีพารามิเตอร์ของผลหลักที่เกิดจากกลุ่ม และวิธีโค้ง

ลักษณะข้อสอบ -3 พารามิเตอร์ แบบทดสอบที่ใช้ศึกษา คือแบบทดสอบวิชาคณิตศาสตร์ โครงการตรวจสอบคุณภาพการศึกษาของสำนักงานทดสอบทางการศึกษา กรมวิชาการ กระทรวงศึกษาธิการ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 ทั่วทุกภาคของประเทศ (ภาคกลาง ภาคใต้ ภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคตะวันออกเฉียงเหนือ) และกรุงเทพมหานคร จำนวนทั้งหมด 7,035 คน

ผู้วิจัยเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างนักเรียน

กรุงเทพมหานคร กับกลุ่มนักเรียนภาคภูมิศาสตร์ทั้ง 5 ภาคดังกล่าว ผลการวิจัยพบว่า การวิเคราะห์ความลำเอียงด้วยวิธีโค้งลักษณะข้อสอบ -3 พารามิเตอร์ พบข้อสอบที่มีความลำเอียงมากที่สุด และในแต่ละวิธีวิเคราะห์พบว่ามีข้อสอบที่ลำเอียงซ้ำกันระหว่างกลุ่มนักเรียนกรุงเทพมหานครกับกลุ่มนักเรียนทุกภาค

สรุปได้ อมรรัตน์ศักดิ์ (2531 : 200-207) ได้ศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบ 4 วิธีคือ วิธีวิเคราะห์ความแปรปรวน วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ -1 พารามิเตอร์ และวิธีโค้งลักษณะข้อสอบ -3 พารามิเตอร์ แบบทดสอบที่ใช้ศึกษาคือแบบทดสอบที่ใช้สอบแข่งขันเพื่อบรรจุเข้ารับราชการครู ปี พ.ศ. 2529 ซึ่งแบบทดสอบชุดนี้มี 4 ฉบับ คือ ฉบับที่ 1 แบบทดสอบวิชาเอก (ความรู้เกี่ยวกับการประถมศึกษา) ฉบับที่ 2 แบบทดสอบวิชาการศึกษาและกฎหมาย ฉบับที่ 3 แบบทดสอบภาษาไทย ฉบับที่ 4 แบบทดสอบความสามารถทั่วไป กลุ่มตัวอย่างคือ ผู้เข้าสอบแข่งขันเพื่อบรรจุเข้ารับราชการเป็นข้าราชการครูในปี พ.ศ. 2529 จำนวน 2,340 คน เป็นชาย 1,170 คน หญิง 1,170 คน ผู้วิจัยวิเคราะห์หาดัชนีความลำเอียงของข้อสอบแล้วหาสัมประสิทธิ์สหสัมพันธ์ระหว่างวิธีการวิเคราะห์ทั้ง 4 วิธี และเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อน และหลังการศึกษาความลำเอียงของข้อสอบ ตามวิธีการคิดคะแนนรวมที่แตกต่างกัน 6 วิธี ในด้านจำนวนผู้ที่ได้รับการคัดเลือก, สัดส่วนของชาย : หญิงที่ได้รับการคัดเลือก และความเชื่อมั่นของแบบทดสอบ ผลการศึกษาพบว่า การวิเคราะห์ความลำเอียงด้วยวิธีโค้งลักษณะข้อสอบ -3 พารามิเตอร์ พบจำนวนข้อสอบที่ลำเอียงมากที่สุด รองลงมาคือวิธีวิเคราะห์ความแปรปรวน และวิธีแปลงค่าความยากของข้อสอบ พบจำนวนข้อที่ลำเอียงน้อยที่สุด ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียง ระหว่างวิธีวิเคราะห์ความลำเอียง 4 วิธี พบว่ามีค่าอยู่ระหว่าง .7535 ถึง .9921 ในด้านจำนวนผู้ที่ได้รับคัดเลือกเมื่อเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อนและหลังการศึกษาความลำเอียง ของข้อสอบตามวิธีคิดคะแนนดิบกับวิธีรวมคะแนนแบบอื่น ๆ อีก 5 วิธี มีจำนวนผู้ที่ได้รับการ

คัดเลือกแตกต่างกันประมาณ 4 ถึง 24 เปอร์เซนต์ ส่วนวิธีแปลงคะแนนดิบเป็นคะแนนมาตรฐานที่ปกติกับวิธีรวมคะแนนแบบอื่น ๆ อีก 4 วิธี มีจำนวนผู้ที่ได้รับการคัดเลือกแตกต่างกันประมาณ 5 ถึง 23 เปอร์เซนต์ และค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดข้อสอบที่ลำเอียงออกแล้ว มีค่าลดลงกว่าแบบทดสอบเดิมเล็กน้อย

พัชรีย์ ปิยภัณฑ (2531 : 65-67) ได้ทำการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแปลงค่าความยาก วิธีโคลสแควร์ และวิธีโค้งลักษณะข้อสอบ -1 พารามิเตอร์จากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์จำนวน 45 ข้อ ซึ่งผู้วิจัยสร้างขึ้นเอง ตัวแปรที่ศึกษาคือเพศ และภาคภูมิศาสตร์ กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักเรียนชายหญิงชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2530 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ ในกรุงเทพมหานคร และจังหวัดสมุทรสาคร จำนวน 1,948 คน ผู้วิจัยวิเคราะห์หาดัชนีความลำเอียงของข้อสอบ แล้วหาสัมประสิทธิ์สหสัมพันธ์ของวิธีวิเคราะห์ 3 วิธี และศึกษาเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบก่อน และหลังการคัดเลือกข้อสอบที่ลำเอียงออก ผลการวิจัยพบว่า การวิเคราะห์ความลำเอียงของข้อสอบ วิธีโค้งลักษณะข้อสอบพบจำนวนข้อสอบที่ลำเอียงน้อยที่สุด ค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงต่อเพศระหว่างวิธีวิเคราะห์ความลำเอียง 3 วิธี มีค่าระหว่าง .1713 ถึง .5618 และค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงต่อกลุ่มนักเรียนกรุงเทพมหานครกับกลุ่มนักเรียนสมุทรสาคร มีค่าระหว่าง .1868 ถึง .6009 ส่วนค่าความเชื่อมั่นของแบบทดสอบก่อน และหลังคัดข้อสอบที่ลำเอียงแล้วแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

สุพัฒน์ สุกมลสันต์ (2534) ได้วิเคราะห์ความลำเอียงของข้อสอบภาษาอังกฤษเพื่อคัดเลือกเข้าศึกษาในมหาวิทยาลัย โดยใช้ข้อมูลการตอบแบบทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยชุด กข และชุด กขค ปี 2531 - 2533 ซึ่งมีข้อสอบชุดละ 100 ข้อ ตัวแปรที่ศึกษาความลำเอียงคือ เพศ และภาคภูมิศาสตร์ของผู้สอบ ซึ่งแยกออกเป็น 5 ภาคตามภูมิลำเนาของผู้สอบ ได้แก่ ภาคกลาง ภาคตะวันออก ภาคตะวันออกเฉียงเหนือ ภาคเหนือ และภาคใต้ เทคนิควิธีที่นำมาวิเคราะห์ความลำเอียง มี 3 วิธี ได้แก่ วิธีกำหนดจุดค่าเดลต้า (Delta-Plot Method) วิธีโคลสแควร์ชนิดที่แบ่งความสามารถของผู้สอบเป็น 3 ระดับ ได้แก่ กลุ่มความสามารถระดับต่ำ (ผู้ได้คะแนนรวม 0-40 คะแนน) กลุ่มความสามารถระดับปานกลาง (ผู้ได้คะแนนรวม 41-70 คะแนน) และระดับสูง (ผู้ได้คะแนนรวม 71-100 คะแนน)

วิธีที่ 3 ที่นำมาใช้วิเคราะห์ความลำเอียงของข้อสอบได้แก่ วิธีการวัดพื้นที่ความแตกต่างระหว่างโค้งลักษณะข้อสอบที่วิเคราะห์ตามทฤษฎีการตอบข้อสอบแบบ 3 พารามิเตอร์

เกณฑ์การตัดสินความลำเอียงของข้อสอบสำหรับวิธีต่าง ๆ ที่ สัพพันธ์ สุกมลสันต์ กำหนดขึ้น คือ

1. วิธีการหาค่าจุดค่าเฉลี่ย ข้อสอบลำเอียงคือ ข้อที่มีระยะตั้งฉากจากค่าคู่อันดับเฉลี่ยไปยังเส้นแกนหลัก (d) มากกว่า ± 1.964 SD

2. วิธีไคสแควร์ ข้อสอบลำเอียงคือ ข้อสอบที่ผู้สอบต่างกลุ่มที่อยู่ในระดับคะแนนเดียวกัน มีสัดส่วนการตอบถูกหรือผิดแตกต่างกันอย่างมีนัยสำคัญที่ระดับ $\alpha = 0.017$

3. วิธีการวัดพื้นที่ความแตกต่างระหว่างโค้งลักษณะของข้อสอบ ข้อสอบลำเอียงคือ ข้อที่มีพื้นที่ความแตกต่างมากกว่า 0.40 โดยแบ่งเป็น 3 ระดับ

3.1 พื้นที่ > 0.70 แสดงว่าลำเอียงมาก

3.2 พื้นที่ระหว่าง 0.40-0.70 แสดงว่าลำเอียงปานกลาง

3.3 พื้นที่ที่มีค่ามากกว่า 0.00 แต่น้อยกว่า 0.40 แสดงว่าลำเอียงน้อย

นิรมล ชัยชวลิต (2537 : 102) ได้เปรียบเทียบผลการวิเคราะห์ความลำเอียงของแบบทดสอบความเข้าใจในการอ่านภาษาไทยตามทฤษฎีคลาสสิกอลที่ใช้วิธีวิเคราะห์ต่างกัน 3 วิธีคือ วิธีแปลงค่าความยาก วิธีไคสแควร์และวิธีวิเคราะห์ความแปรปรวน โดยเปรียบเทียบจำนวนข้อที่มีความลำเอียงออก ค่าความเชื่อมั่นที่ใช้สูตรคำนวณแบบแบ่งครึ่งฉบับ ของแบบทดสอบหลังคิดเลือกที่มีความลำเอียงออก กลุ่มตัวอย่างที่ใช้ในการศึกษาครั้งนี้เป็นนักเรียนชายและนักเรียนหญิง ชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2536 ของโรงเรียนในสังกัดกรมสามัญศึกษา เขตพระโขนง กรุงเทพมหานคร จำนวน 1,066 คน พบว่าการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยาก กับวิธีไคสแควร์และวิธีแปลงค่าความยากกับวิธีวิเคราะห์ความแปรปรวน มีจำนวนข้อที่มีความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีไคสแควร์กับวิธีวิเคราะห์ความแปรปรวนพบจำนวนข้อที่มีความลำเอียงไม่แตกต่างกัน ส่วนค่าความเชื่อมั่นของแบบทดสอบความเข้าใจในการอ่านภาษาไทยที่ใช้สูตรคำนวณทั้งแบบ KR-20 และแบบแบ่งครึ่งฉบับ หลังคิดเลือกข้อที่มีความลำเอียงจากการวิเคราะห์ด้วยวิธีแปลงค่าความยาก วิธีไคสแควร์ และวิธีวิเคราะห์ความแปรปรวน มีค่าไม่แตกต่างกัน

กาญจนา วัฒนสุนทร (2537 : ง) ได้พัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ โดยใช้ข้อมูลเชิงประจักษ์ สำหรับดัชนี 4 ตัวคือ พื้นที่ระหว่างโค้งการตอบข้อสอบชนิดที่มีเครื่องหมาย (SA) และไม่มีเครื่องหมาย (VA) จากวิธีทฤษฎีการตอบข้อสอบโมเดล 2 พารามิเตอร์ α_{MH} จากวิธีแมนเทิล-แฮนส์เชล และ β จากวิธี SIBTEST โดยใช้ข้อมูลการสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของทบวงมหาวิทยาลัย ปีการศึกษา 2535 จัดความยาวแบบทดสอบ 20 30 และ 40 ข้อ สำหรับวิชาคณิตศาสตร์ และ 50 60 70 และ 80 ข้อ สำหรับวิชาภาษาอังกฤษ กลุ่มตัวอย่าง 6 ขนาดคือ 100 200 400 600 800 และ 1,000 คน การพัฒนาเกณฑ์ กระทำโดยคำนวณค่าดัชนีทั้ง 4 ตัว จากข้อมูลการตอบข้อสอบของผู้สอบเพศเดียวกัน เพศละ 50 ค่า สำหรับแต่ละความยาวแบบทดสอบ และขนาดของผู้สอบ จากนั้นนำค่าดัชนีที่ได้ทั้งหมดมาวิเคราะห์ค่าเฉลี่ย และกำหนดเกณฑ์จากค่าเฉลี่ย 2 ลักษณะ คือ เกณฑ์ที่กำหนด โดยพิจารณาจากค่าเฉลี่ยรวมจากค่าดัชนีทุกข้อ โดยไม่พิจารณาความแตกต่างในด้านความยาวแบบทดสอบและขนาดของกลุ่มผู้สอบและเกณฑ์ที่พัฒนาจากค่าเฉลี่ย ซึ่งพิจารณาถึงความยาวแบบทดสอบและขนาดผู้สอบด้วย จากการนำเกณฑ์ที่กำหนดไว้ไปตัดสินดัชนีที่ได้จากการวิเคราะห์ระหว่างผู้สอบเพศหญิงและชายพบว่า ไม่ว่าใช้เกณฑ์ลักษณะใด ผลด้านจำนวนข้อสอบในรูปของร้อยละ ของข้อสอบที่ลำเอียง และความสอดคล้องของการตัดสินใจ มีความไม่คงที่ข้ามความยาวของแบบทดสอบและขนาดผู้สอบโดยที่เกณฑ์ในลักษณะหลังจะลดความไม่คงที่ลงได้ระดับหนึ่ง และพบว่าที่ความยาวแบบทดสอบ 40 ข้อ สำหรับวิชาคณิตศาสตร์ และ 80 ข้อ สำหรับวิชาภาษาอังกฤษ มีจำนวนข้อสอบที่ได้รับการตัดสินมากกว่าในความยาวอื่นมาก และเกณฑ์ที่กำหนดได้จากการศึกษาครั้งนี้มีค่าสูงกว่าเกณฑ์ที่มีผู้ใช้กันอยู่ในกรณีของค่า UA (0) ในขณะที่ค่า α_{MH} มีค่าใกล้เคียงกับเกณฑ์ขั้นต่ำที่ควรใช้ในการตัดสินข้อสอบลำเอียงที่พัฒนาได้ครั้งนี้คือ $ISA1 > .40$ $VA > .50$ เพื่อให้ความยาวแบบทดสอบ 50 ข้อขึ้นไป $ISA1 > .80$ $VA > 1.20$ กรณีความยาวแบบทดสอบสั้นกว่า 50 ข้อ ค่า $.60 > \alpha_{MH} > 1.40$ และ $ISIB1 > .06$ และควรปรับใช้เกณฑ์สูงขึ้น โดยพิจารณาถึงวัตถุประสงค์ที่ต้องการใช้แบบทดสอบ แต่ไม่ควรต่ำกว่านี้ เพื่อลดปัญหาของการเกิดความคลาดเคลื่อนประเภท 2 และควรศึกษาเพิ่มเติม เพื่อหาความยาวแบบทดสอบ และจำนวนผู้สอบที่ให้ค่าดัชนีที่มีความคงที่เพิ่มขึ้น

งานวิจัยที่เกี่ยวข้องกับวิธีการคิดคะแนน

งานวิจัยที่เกี่ยวข้องกับวิธีการคิดคะแนนนั้น มีผู้ศึกษาต่าง ๆ กัน ดังนี้

นีเดิลสกี (Wang and Stanley. 1970 : 671 ; citing Nedelsky. 1954) ได้ใช้ผลการสอบรวบยอดของนักศึกษา 306 คน ใช้การคิดคะแนนโดยผู้เชี่ยวชาญ ปรากฏว่าทำให้แบบทดสอบมีค่าความเชื่อมั่นสูงกว่าการให้คะแนนแบบ 0-1 จากนั้น จาคอบส์ และแวนเดอร์ แวนเดอร์ (Wang and Stanley. 1970 : 693 ; citing Jacobs and Vandervanter. 1968) ได้ใช้วิธีเดียวกัน พบว่า การคิดคะแนนวิธีดังกล่าว แบบทดสอบมีค่าความเชื่อมั่นแบบสอบซ้ำ ค่าความเที่ยงตรงตามสภาพ (Concurrent Validity) มีค่าสูงกว่าการคิดคะแนนแบบ 0-1

แฮมเบิลตัน โรเบิร์ตและทราับ (Wang and Stanley. 1970 : 693 ; citing Hambleton, Roberts and Traub. 1970) ใช้การคิดคะแนนโดยใช้ค่าเฉลี่ยของลำดับที่ของความถูกต้องของตัวเลือกแต่ละตัวที่ตัดสินโดยผู้เชี่ยวชาญ พบว่าวิธีดังกล่าวทำให้แบบทดสอบเลือกตอบมีค่าความเชื่อมั่นและความเที่ยงตรงเชิงพยากรณ์สูงขึ้นกว่าความเชื่อมั่นและความเที่ยงตรงที่คำนวณจากคะแนนที่ใช้แบบเดิม แต่ไม่มีนัยสำคัญทางสถิติ

แรฟฟีลด์ (Wang and Stanley. 1970 : 693 ; citing Raffeld. 1975 : 285) ได้ศึกษาการคิดคะแนนตามวิธีของกัตแมน (Guttman's Scale Analysis) โดยใช้นักเรียนเกรด 9 จำนวน 1,354 คน ที่สอบวิชาการอ่าน จับใจความ ผัฟท์ ไวยากรณ์ การสะกดคำ และคณิตศาสตร์ พบว่า ความเชื่อมั่นและความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบที่คิดคะแนนตามวิธีของกัตแมนมีค่าสูงกว่าที่คำนวณจากการคิดคะแนนแบบวิธีดั้งเดิมอย่างมีนัยสำคัญ เมื่อใช้คะแนนปลายปีเป็นเกณฑ์สำหรับการพยากรณ์

แพทเนียค และทราับ (Patnaik and Traub. 1973 : 281) ได้ศึกษาผลการคิดคะแนนต่อความเชื่อมั่นและความเที่ยงตรงเชิงพยากรณ์ ของนักศึกษาระดับบัณฑิตศึกษา จำนวน 61 คน โดยใช้แบบทดสอบวัดเชาวน์ปัญญาแบบเลือกตอบ 5 ตัวเลือก และใช้การเปรียบเทียบรายคู่ ของเทอร์สโตนเป็นคะแนนตัวเลือกแต่ละตัว พบว่าการคิดคะแนนดังกล่าวทำให้ค่าความเชื่อมั่นแบบแบ่งครึ่งสูงกว่าค่าความเชื่อมั่นที่คำนวณจากคะแนนแบบเดิมอย่างมีนัยสำคัญทางสถิติที่ .05 แต่มีค่าความเที่ยงตรง

เชิงพยากรณ์ที่ใช้ผลสัมฤทธิ์ทางการเรียนในวิชาต่าง ๆ เป็นเกณฑ์ต่ำกว่าค่าความเที่ยงตรงดังกล่าวอย่างมีนัยสำคัญทางสถิติที่ .05

โดมอน (Diamond, 1975 : 129) ศึกษาการคิดคะแนน 3 วิธีกับกลุ่มตัวอย่าง 84 คน ในระดับบัณฑิตศึกษาด้วยแบบทดสอบวิชาวัดผลที่ค่อนข้างง่าย และมีอำนาจจำแนกต่ำ พบว่า แบบทดสอบที่คิดคะแนนตามความมั่นใจในการตอบ การคิดหักค่าการเดา และคิดคะแนนโดยการกำจัดตัวเลือก ทั้ง 3 วิธี มีค่าความเชื่อมั่นแบบฟลานาแกน (Flanagan) สูงกว่าค่าความเชื่อมั่นที่คำนวณจากคะแนนที่ใช้แบบดั้งเดิม แต่ค่าความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบที่ใช้ผลการสอบประจำภาคเป็นเกณฑ์ ไม่แตกต่างจากค่าความเที่ยงตรงเชิงพยากรณ์ที่คำนวณจากคะแนนแบบเดิม ซึ่งสอดคล้องกับการศึกษาของ ฮันนา (Hanna, 1975 : 145) ที่คิดคะแนนตัวเลือกของแบบทดสอบเลือกตอบ 4 ตัวเลือก 2 ฉบับ มี 10 และ 30 ข้อ กับกลุ่มตัวอย่าง 85 คน โดยวิธีคิดคะแนนตามความมั่นใจในการตอบ

ฮากสไตน์ และคานสับ (Hakstain and Kansup, 1975b : 231) ศึกษาผลการคิดคะแนนโดยวิธีคิดคะแนนตามความมั่นใจในการตอบและวิธีคิดคะแนนโดยวิธีการกำจัดตัวเลือก โดยให้นักเรียนระดับเกรด 9 จำนวน 1,023 คน พบว่าสอดคล้องกับการศึกษาของ โดมอนและฮันนา คือการคิดคะแนนทั้ง 2 แบบ ทำให้แบบทดสอบมีค่าความเชื่อมั่นสูงขึ้นเล็กน้อย แต่ให้ค่าความเที่ยงตรงเชิงพยากรณ์ต่ำกว่าค่าความเที่ยงตรงที่คำนวณจากคะแนนที่ใช้แบบเดิม และฮันนา (Hanna, 1977 : 643) ได้ศึกษาโดยใช้แบบทดสอบเลือกตอบ 18 ข้อ 2 ฉบับ วิชาวิทยาศาสตร์และคณิตศาสตร์ กับนักเรียนเกรด 5 และเกรด 6 จำนวน 1,391 คน พบว่าแบบทดสอบมีค่าความเชื่อมั่นสูงกว่าความเชื่อมั่นที่คิดคะแนนแบบเดิม แต่ความเที่ยงตรงไม่แตกต่างกัน

ลอร์ด (Lord, 1968 : 989) ได้คิดคะแนนข้อสอบวิชา Scholastic Aptitude Test โดยอาศัยโมเดลโลจิสต์ 3 พารามิเตอร์ โดยการกำหนดค่าน้ำหนักคะแนนของผู้สอบแต่ละคน แต่ละข้อตามความสามารถจากโด่งลักษณะของข้อสอบ (ICC) พบว่า กรณีที่ผู้สอบมีความสามารถสูงมาก การคิดคะแนนข้อสอบรายข้อแบบเดิมดีกว่า ถ้าความสามารถในระดับต่ำหรือปานกลาง การคิดคะแนนตามความสามารถจะถูกต้องมากกว่า ต่อมาในปี 1975 ลอร์ด (Lord, 1980 : 74) ได้เปรียบเทียบผลของการคิดคะแนนต่างกันอีก โดยใช้แบบทดสอบทางภาษา SCAT II Form 2A มี 4 ตัวเลือก 40 ข้อ พบว่าการคิดคะแนนข้อสอบแต่ละข้อโดยวิธีใช้น้ำหนักที่เหมาะสม (W) ให้ผลการสอบดีที่สุด

ชู และชูศักดิ์ (ชูศักดิ์ ชัมภลลิขิต. 2525 : 33-44 ; อ้างอิงจาก Hsu and C. Khampalikit. 1982 : 19-23) ได้เปรียบเทียบผลการคิดคะแนน โดยวิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสม (W) กับการคิดคะแนนแบบเดิมในข้อสอบวิชา ภาษาจีน และวิชาประวัติศาสตร์ พบว่าการคิดคะแนนวิธีใช้ค่าน้ำหนักคะแนนที่เหมาะสม ทำให้แบบทดสอบมีค่าอำนาจจำแนก และค่าความเชื่อมั่นสูงกว่าที่คำนวณจากคะแนนที่ให้โดยวิธีเดิม

จำเนียร สุขหลาย (จำเนียร สุขหลาย. 2530 : ง-ช) ได้ศึกษาวิธีการคิดคะแนนสอบคัดเลือกเข้าศึกษาในวิทยาลัยครู โดยใช้วิธีคิดคะแนน 4 วิธีคือ ใช้ความสามารถแท้ (O) การคิดคะแนนแบบถ่วงน้ำหนัก (W) การคิดคะแนนแบบ ทิ-ปกติ (T) และการคิดคะแนนดิบ (S) โดยศึกษาผลของอันดับที่ของผู้เข้าสอบ จากการคิดคะแนนแต่ละวิธี ความสัมพันธ์ระหว่างคะแนนแต่ละวิธี และระหว่าง คะแนนแต่ละวิธีกับผลสัมฤทธิ์ทางการเรียน และประสิทธิภาพของการสอบคัดเลือก ที่คิดคะแนนแต่ละวิธีในการทำนายผลสัมฤทธิ์ทางการเรียน ใช้ผลการสอบเข้าศึกษา ในวิทยาลัยครูเทพสตรี และวิทยาลัยราไพพรรณี จำนวน 2,496 และ 1,429 คน ในวิชาความรู้ทั่วไปคือ คณิตศาสตร์ วิทยาศาสตร์ ภาษาไทย ภาษาอังกฤษ และ สังคมศึกษา และวิชาเฉพาะวิชาเอก ผลการศึกษาพบว่า การคิดคะแนนแบบความสามารถแท้ (O) ทำให้อันดับที่ของผู้เข้าสอบแปรเปลี่ยนไปจากการคิดแบบคะแนนดิบ มากที่สุด คะแนนการสอบทั้ง 4 วิธี มีความสัมพันธ์ในทางบวกอย่างมีนัยสำคัญที่ .01 คะแนนการสอบสัมพันธ์กับผลสัมฤทธิ์ทางการเรียนอย่างมีนัยสำคัญที่ .05 และ .01 และประสิทธิภาพการทำนายผลสัมฤทธิ์ทางการเรียนได้ดีที่สุด

จากการศึกษางานวิจัยที่เกี่ยวข้องกับการตรวจสอบความลำเอียง และการคิดคะแนนของแบบทดสอบ ทั้งในและต่างประเทศ พบว่ายังไม่มียงานวิจัยใดที่สามารถสรุปได้ว่าวิธีการตรวจสอบความลำเอียง และการคิดคะแนนของแบบทดสอบวิธีใดให้ผลได้เที่ยงตรง สอดคล้อง และยุติธรรมแก่ผู้สอบในกลุ่มย่อยต่าง ๆ ได้มากที่สุด ดังนั้นเพื่อเป็นแนวทางในการตัดสินใจที่จะนำวิธีการตรวจสอบความลำเอียง วิธีการคิดคะแนนและการศึกษาสาเหตุของความลำเอียงของข้อสอบดังกล่าว ผู้วิจัยจึงศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกโดยการตรวจสอบความลำเอียง ด้วยวิธีการใช้ทฤษฎีการตอบข้อคำถาม วิธีแมนเทิล-แฮนส์เซล และวิธี SIBTEST โดยศึกษาจากกลุ่มเพศชายและหญิง และการใช้วิธีการคิดคะแนนแบบคะแนนมาตรฐาน ทิ-ปกติ และคะแนนน้ำหนักความสามารถ

บทที่ 3

วิธีดำเนินการวิจัย

การวิจัยเพื่อศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่ใช้
วิธีการตรวจสอบความลำเอียง การคิดคะแนนต่างกันและสาเหตุของความลำเอียง
มีขั้นตอนการวิจัยดังนี้

แหล่งข้อมูลในการวิจัย

1. ดัชนีความลำเอียงและความสอดคล้องของการตัดสินผลการสอบ ศึกษา
จากการสอบคัดเลือกเข้าเป็นนักศึกษาชั้นปีที่ 1 ประเภทรับตรง ปีการศึกษา 2538
ของมหาวิทยาลัยสงขลานครินทร์ โดยมีรายละเอียดดังนี้

ตาราง 3 แสดงจำนวนข้อมูลที่ใช้ในการศึกษาดัชนีความลำเอียงและการตัดสินผล
การสอบ

วิชา	ฉบับที่	จำนวนข้อสอบ		จำนวนผู้สอบ (คน)		
		ฉบับ เดิม	ใช้ใน การวิจัย	ชาย	หญิง	รวม
ภาษาไทย ก	1	50	42	1345	2687	4032
	2	50	42	1377	2718	4095
สังคมศึกษา ก	1	80	64	1361	2688	4049
	2	80	64	1358	2714	4072
ภาษาอังกฤษ กข	1	40	36	1388	2689	4077
	2	40	36	1327	2712	4039
ภาษาไทย กข	1	50	37	704	1983	2687
	2	50	37	749	1974	2723

ตาราง 3 (ต่อ)

วิชา	ฉบับที่	จำนวนข้อสอบ		จำนวนผู้สอบ (คน)		
		ฉบับเดิม	ใช้ในการวิจัย	ชาย	หญิง	รวม
สังคมศึกษา กข	1	80	59	715	1970	2685
	2	80	59	739	1991	2730
ภาษาอังกฤษ กขค	1	40	37	722	1997	2719
	2	40	37	736	1965	2701

2. ความเที่ยงตรงเชิงพยากรณ์ ศึกษาจากนักศึกษาที่ได้รับการคัดเลือกเข้าศึกษาในชั้นปีที่ 1 ปีการศึกษา 2538 ประเภทรับตรง ของมหาวิทยาลัยสงขลานครินทร์ รวมทั้งสิ้น 1,051 คน เป็นสายวิทยาศาสตร์ 765 คน และสายศิลปศาสตร์ 286 คน

3. สาเหตุของความลำเอียงของข้อสอบศึกษาจากการตอบแบบสำรวจความลำเอียงของข้อสอบ โดยมีผู้ระบุสาเหตุของความลำเอียงประกอบด้วยบุคคลต่าง ๆ ดังนี้

3.1 นักวัดผลการศึกษา และอาจารย์ผู้สอนในรายวิชาต่าง ๆ รายวิชาละ 50 คน

3.2 นิสิตนักศึกษาที่เรียนในสาขาวิชาที่เกี่ยวข้อง รายวิชาละ 50 คน

วิธีดำเนินการวิจัย

1. ติดต่อกับฝ่ายวิชาการของมหาวิทยาลัยสงขลานครินทร์ เพื่อขอใช้ข้อมูลการรับสมัครผลการตอบข้อสอบ ผลการสอบคัดเลือก และระดับคะแนนผลการเรียนในรายวิชาต่าง ๆ ในภาคเรียนที่ 1 ของปีการศึกษา 2538

2. นำผลการตอบข้อสอบคัดเลือกทั้งหมดวิเคราะห์ด้วยโปรแกรม SPSS-PC⁺ และวิเคราะห์องค์ประกอบเพื่อตรวจสอบความเป็นมิติเดียวของแบบทดสอบแต่ละฉบับ

3. วิเคราะห์หาคุณภาพของข้อสอบเป็นรายข้อในทุกวิชาโดย

3.1 วิเคราะห์หาคุณภาพรายข้อตามทฤษฎีการตอบข้อคำถาม ด้วยโปรแกรมสำเร็จรูป BILOG

3.2 วิเคราะห์ความลำเอียงของข้อสอบในแต่ละข้อ แต่ละวิชา ตามทฤษฎีการตอบข้อคำถาม วิธีแมนเทิล-แฮนส์เซล และวิธี SIBTEST ด้วยโปรแกรมสำเร็จรูป IRIDIF และ SIBTEST

3.3 ทดสอบความแตกต่างของจำนวนข้อสอบที่ลำเอียงในแต่ละฉบับแต่ละวิชา ทั้ง 3 วิธีของแบบทดสอบ ด้วยสถิติอนุพารามेटริก แบบ Cochran Q Test

4. นำข้อสอบที่ตรวจพบว่ามีควมลำเอียงทั้ง 3 วิธี ไปให้ผู้ระบุสาเหตุระบุสาเหตุของความลำเอียงต่อกลุ่มเพศชายและหญิง แล้วสรุปสาเหตุของความลำเอียงข้อสอบแต่ละข้อ

5. นำคะแนนจากการสอบของข้อสอบทั้งหมดกับคะแนนจากผลการสอบที่คิดเฉพาะข้อสอบที่ปราศจากความลำเอียง มาหาความสัมพันธ์ในรูปของคะแนนมาตรฐานที่-ปกติ และคะแนนน้ำหนักความสามารถในแต่ละวิชา และหาความสัมพันธ์ลำดับที่ (Rank Correlation) ในคะแนนรวม

6. คำนวณค่าการถดถอยพหุคูณ (Multiple Regression Analysis) ของคะแนนสอบคัดเลือกกับระดับคะแนนที่เรียนในชั้นปีที่ 1 ของนักเรียนที่ได้รับการคัดเลือก โดยการใช้คะแนนแบบมาตรฐานที่-ปกติ และคะแนนน้ำหนักความสามารถ

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้เป็นแบบทดสอบวิชา ภาษาไทย ก สังคมศึกษา ก ภาษาอังกฤษ กข สังคมศึกษา กข ภาษาไทย กข และ ภาษาอังกฤษ กขค ที่ใช้สอบคัดเลือกบุคคลเข้าเป็นนักศึกษาชั้นปีที่ 1 ประเภทรับตรง ปีการศึกษา 2538 ของมหาวิทยาลัยสงขลานครินทร์ และแบบประเมินสาเหตุของความลำเอียงที่สร้างโดยผู้วิจัย (รายละเอียดในภาคผนวก)

สถิติที่ใช้ในการวิจัย

การวิจัยครั้งนี้ผู้วิจัยใช้โปรแกรมสำเร็จรูปเพื่อคำนวณค่าสถิติต่าง ๆ ดังนี้

1. โปรแกรมสำเร็จรูป SPSS วิเคราะห์องค์ประกอบ (Factor Analysis) ค่าสถิติพื้นฐาน ค่าสหสัมพันธ์ และค่าสหสัมพันธ์พหุคูณ
2. โปรแกรมสำเร็จรูป BILOG เพื่อคำนวณค่าพารามิเตอร์ตามทฤษฎีการตอบข้อคำถาม
3. โปรแกรม IRTDIF คำนวณค่าพื้นที่ของความแตกต่างระหว่างโค้งของข้อสอบจากกลุ่มย่อย 2 กลุ่ม
4. โปรแกรม SIBTEST คำนวณค่าเบต้า (β) ของวิธี SIBTEST และค่าแอลฟา (α) ค่าไคสแควร์ของวิธีแมนเทิล-แฮนส์เซล

ผลการวิเคราะห์ข้อมูล

สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล

ตรงกัน	ในการเสนอผลการวิเคราะห์ข้อมูล เพื่อความสะดวกและความเข้าใจ ผู้วิจัยได้กำหนดสัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูลดังนี้	
N	แทน	จำนวนคน
a	แทน	ค่าอำนาจจำแนกของข้อสอบ
b	แทน	ค่าความยากง่ายของข้อสอบ
c	แทน	ค่าสัมประสิทธิ์การเดาของข้อสอบ
Max	แทน	คะแนนสูงสุด
Min	แทน	คะแนนต่ำสุด
$\bar{X}$	แทน	ค่าเฉลี่ย
SD	แทน	ค่าส่วนเบี่ยงเบนมาตรฐาน
IRT	แทน	วิธีวิเคราะห์ความลำเอียงวิธีใช้ทฤษฎีการตอบข้อคำถาม
SIB	แทน	วิธีวิเคราะห์ความลำเอียงวิธี SIBTEST
MH	แทน	วิธีวิเคราะห์ความลำเอียงวิธีแมนเทิล-แฮนส์เซล
Q	แทน	ค่าการทดสอบ Cochran Q Test
χ^2	แทน	การทดสอบไคสแควร์
T	แทน	การคิดคะแนนโดยการแปลงคะแนนดิบเป็นคะแนน มาตรฐานที-ปกติ
θ	แทน	การคิดคะแนนโดยใช้คะแนนน้ำหนักความสามารถ
b	แทน	สัมประสิทธิ์ของตัวทำนายในรูปคะแนนดิบ
β	แทน	สัมประสิทธิ์ของตัวทำนายในรูปคะแนนมาตรฐาน
A	แทน	ค่าคงที่ของสมการทำนาย
SE_b	แทน	ความคลาดเคลื่อนมาตรฐานของการทำนาย
SE_{est}	แทน	ความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ตัวทำนาย
R	แทน	สัมประสิทธิ์สหสัมพันธ์หาคู่
R^2	แทน	สัดส่วนความแปรปรวนที่ทำนายได้

F	แทน	การทดสอบ F
X_j	แทน	ตัวทำนายที่ j ($j = 1, 2, 3, 4, 5, 6$)
Y	แทน	ตัวเกณฑ์
X	แทน	ตัวเกณฑ์ ซึ่งได้จากการทำนายในรูปคะแนนมาตรฐาน
X_1	แทน	คะแนนวิชาภาษาไทย ก
X_2	แทน	คะแนนวิชาสังคมศึกษา ก
X_3	แทน	คะแนนวิชาภาษาอังกฤษ กข
X_4	แทน	คะแนนวิชาภาษาไทย กข
X_5	แทน	คะแนนวิชาสังคมศึกษา กข
X_6	แทน	คะแนนวิชาภาษาอังกฤษ กขค
θ -UB	แทน	ผลการสอบที่คิดคะแนนน้ำหนักความสามารถที่ใช้เฉพาะ ข้อสอบที่ปราศจากความลำเอียง
θ -ALL	แทน	ผลการสอบที่คิดคะแนนน้ำหนักความสามารถและใช้ข้อสอบ ทั้งหมด
T-UB	แทน	ผลการสอบที่คิดคะแนนที่-ปกติที่ใช้เฉพาะข้อสอบที่ ปราศจากความลำเอียง
T-ALL	แทน	ผลการสอบที่คิดคะแนนที่-ปกติ และใช้ข้อสอบทั้งหมด

การนำเสนอผลการวิเคราะห์ข้อมูล

ผู้วิจัยวิเคราะห์ข้อมูลตามขั้นตอนของการดำเนินการวิจัยตามลำดับดังนี้

1. ตรวจสอบความเป็นมิติเดียวกันของแบบทดสอบแต่ละวิชา
2. ค่าสถิติพื้นฐานของผลการสอบและคุณภาพของแบบทดสอบแต่ละวิชา
3. วิเคราะห์ความลำเอียงของข้อสอบแต่ละวิชาทั้ง 3 วิธี
4. ทดสอบความแตกต่างของจำนวนข้อสอบที่ลำเอียงจากการวิเคราะห์

ด้วย 3 วิธีของแบบทดสอบแต่ละวิชา

5. สาเหตุความลำเอียงของข้อสอบแต่ละวิชา

6. ความสัมพันธ์ของผลการสอบที่คิดคะแนนมาตรฐานที่-ปกติ และคิดคะแนน
น้ำหนักความสามารถที่ใช้ข้อสอบทั้งหมดกับใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง

7. ประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการเรียนของการคิดคะแนน
และการใช้จำนวนข้อสอบ

1. การตรวจสอบความเป็นมิติเดียวกันของแบบทดสอบแต่ละวิชา

จากผลการตอบข้อสอบจำนวน 6 วิชา 12 ฉบับ นำมาตรวจสอบหาคุณภาพเบื้องต้นโดยวิเคราะห์ค่าสถิติพื้นฐานและค่าพารามิเตอร์ พบว่าข้อสอบของแบบทดสอบแต่ละฉบับ บางข้อจะต้องตัดทิ้งไปเนื่องจากค่าความยากของข้อสอบมีค่าต่ำหรือสูงมาก คือมีค่ามากกว่าหรือน้อยกว่า 4.00 จึงเหลือจำนวนข้อสอบที่ใช้ในการวิจัย วิชาภาษาไทย ก ฉบับที่ 1, 2 ฉบับละ 42 ข้อ วิชาสังคมศึกษา ก ฉบับที่ 1, 2 ฉบับละ 64 ข้อ วิชาภาษาอังกฤษ กข ฉบับที่ 1, 2 ฉบับละ 36 ข้อ วิชาภาษาไทย กข ฉบับที่ 1, 2 ฉบับละ 37 ข้อ วิชาสังคมศึกษา กข ฉบับที่ 1, 2 ฉบับละ 59 ข้อ และ วิชาภาษาอังกฤษ กขค ฉบับที่ 1, 2 ฉบับละ 37 ข้อ จากนั้นนำผลการตอบข้อสอบและวิชาแต่ละฉบับวิเคราะห์หองค์ประกอบด้วยโปรแกรมสำเร็จรูป SPSS-PC⁺ ซึ่งได้ค่าไอเกนและร้อยละของความแปรปรวน และกราฟแสดงค่าไอเกนขององค์ประกอบที่ 1-4 ของแบบทดสอบแต่ละฉบับ ดังตาราง 4 และภาพประกอบ 4-9

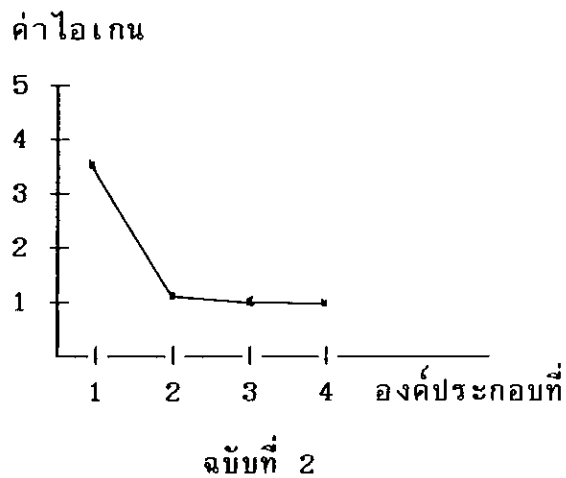
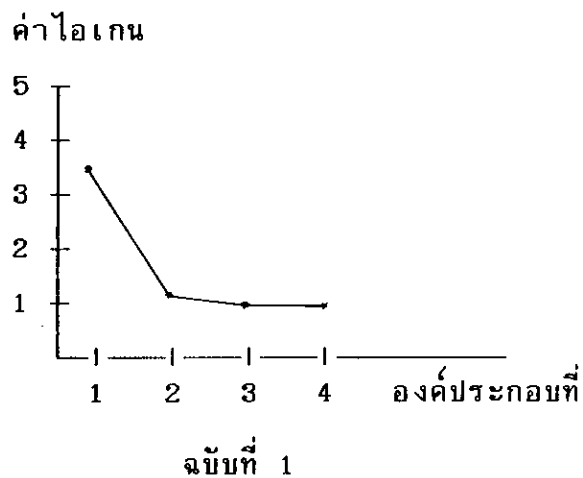
ตาราง 4 ค่าไอเกนและร้อยละของความแปรปรวนขององค์ประกอบที่ 1-4 ของ
แบบทดสอบวิชาภาษาไทย ก สังคมศึกษา ก ภาษาอังกฤษ กข ภาษาไทย กข
สังคมศึกษา กข และภาษาอังกฤษ กขค ทั้ง 12 ฉบับ

แบบทดสอบวิชา	ฉบับที่	ค่าไอเกนขององค์ประกอบที่			
		1	2	3	4
ภาษาไทย ก	1	3.17 (6.3)	1.33 (2.7)	1.18 (2.4)	1.16 (2.3)
	2	3.19 (6.4)	1.33 (2.7)	1.18 (2.4)	1.14 (2.3)
สังคมศึกษา ก	1	4.58 (5.7)	1.49 (1.9)	1.33 (1.7)	1.26 (1.6)
	2	4.78 (6.0)	1.51 (1.9)	1.38 (1.7)	1.30 (1.6)
ภาษาอังกฤษ กข	1	2.39 (6.0)	1.36 (3.4)	1.20 (3.0)	1.15 (2.9)
	2	2.47 (6.2)	1.33 (3.3)	1.24 (3.1)	1.14 (2.9)
ภาษาไทย กข	1	2.38 (4.8)	1.28 (2.6)	1.22 (2.5)	1.20 (2.4)
	2	2.49 (5.0)	1.30 (2.6)	1.24 (2.5)	1.23 (2.5)
สังคมศึกษา กข	1	3.40 (4.2)	1.57 (2.0)	1.38 (1.7)	1.32 (1.7)
	2	3.34 (4.2)	1.65 (2.1)	1.33 (1.7)	1.32 (1.6)
ภาษาอังกฤษ กขค	1	2.81 (5.6)	1.46 (2.9)	1.30 (2.6)	1.24 (2.5)
	2	2.75 (5.5)	1.46 (2.9)	1.33 (2.7)	1.23 (2.5)

* ในวงเล็บคือค่าร้อยละของความแปรปรวน

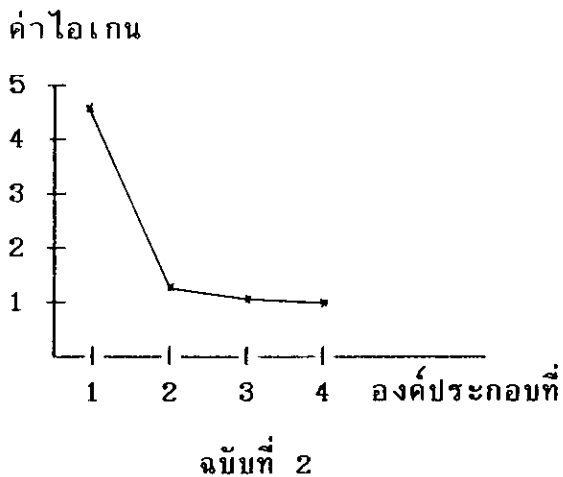
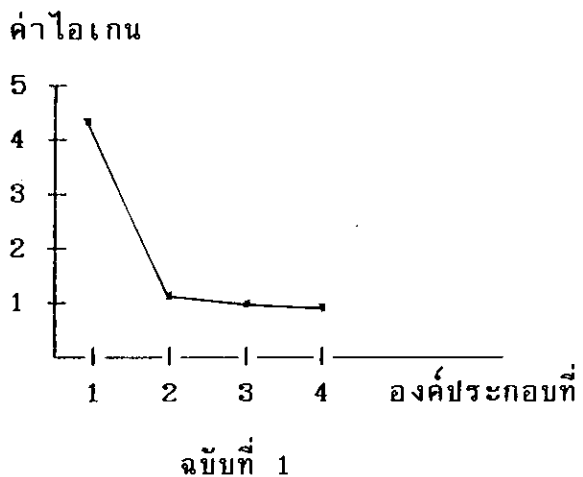
จากตาราง 4 ผลของการตรวจสอบความเป็นมิติเดียวกันของแบบทดสอบทั้ง 6 วิชา 12 ฉบับ พบว่าแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 1 และ 2 ค่าไอเกนขององค์ประกอบที่ 1 มีค่าระหว่าง 3-4 เท่าขององค์ประกอบที่ 2, 3 และ 4 นอกจากนั้นมีค่าระหว่าง 2-3 เท่า ซึ่งเมื่อพิจารณาโดยรวมแล้ว แบบทดสอบทั้ง 6 วิชา 12 ฉบับ ค่าไอเกนขององค์ประกอบที่ 1 มีค่าเป็น 3 เท่าขององค์ประกอบที่ 2, 3 และ 4 พอจะสรุปได้ว่าแบบทดสอบทั้ง 6 วิชา 12 ฉบับ มีความเป็นมิติเดียว ซึ่งแสดงค่าไอเกนของ 4 องค์ประกอบในแต่ละวิชาแต่ละฉบับ ดังแสดงในภาพประกอบ 4-9

1.1 แบบทดสอบวิชาภาษาไทย ก



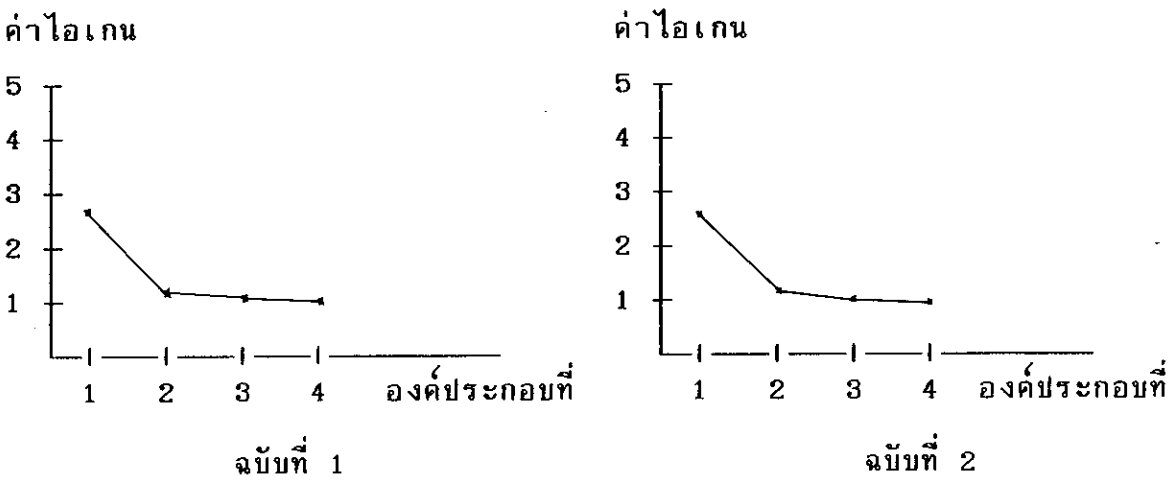
ภาพประกอบ 4 แสดงค่าไอเคนของ 4 องค์ประกอบของแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 1 และฉบับที่ 2

1.2 แบบทดสอบวิชาสังคมศึกษา ก



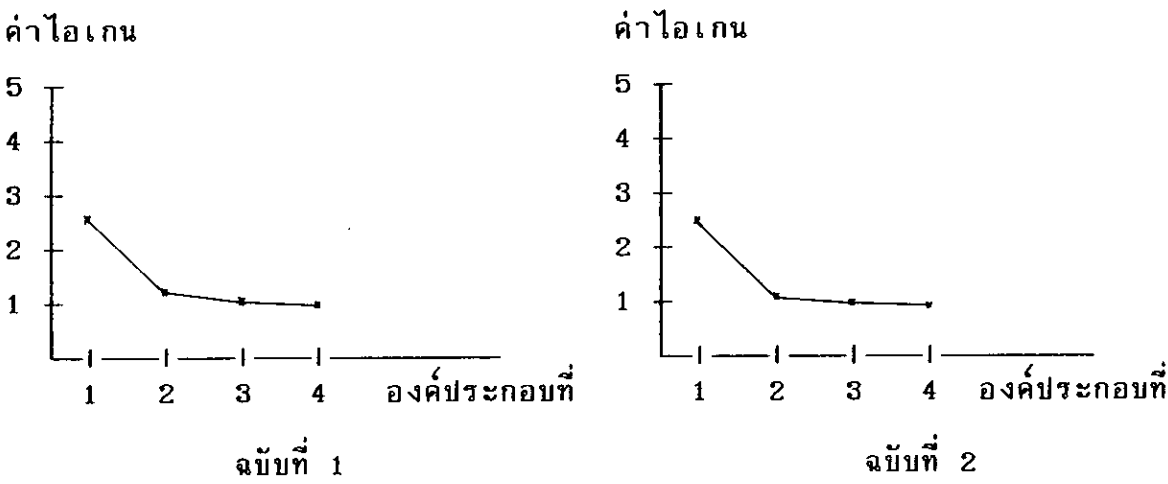
ภาพประกอบ 5 แสดงค่าไอเคนของ 4 องค์ประกอบของแบบทดสอบวิชาสังคม-ศึกษา ก ฉบับที่ 1 และฉบับที่ 2

1.3 แบบทดสอบวิชาภาษาอังกฤษ กข



ภาพประกอบ 6 แสดงค่าไอเกนของ 4 องค์ประกอบของแบบทดสอบวิชาภาษาอังกฤษ กข ฉบับที่ 1 และฉบับที่ 2

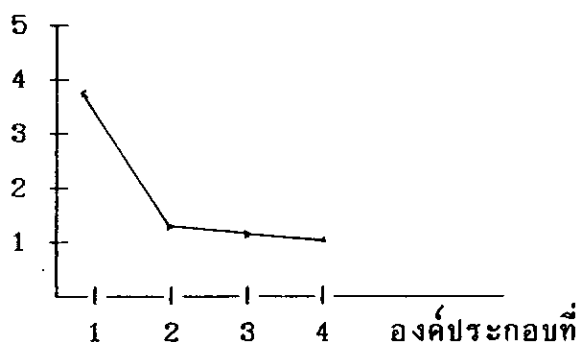
1.4 แบบทดสอบวิชาภาษาไทย กข



ภาพประกอบ 7 แสดงค่าไอเกนของ 4 องค์ประกอบของแบบทดสอบวิชาภาษาไทย กข ฉบับที่ 1 และฉบับที่ 2

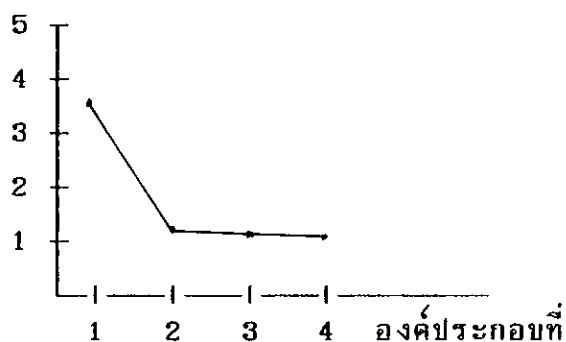
1.5 แบบทดสอบวิชาสังคมศึกษา กข

ค่าไอเคน



ฉบับที่ 1

ค่าไอเคน

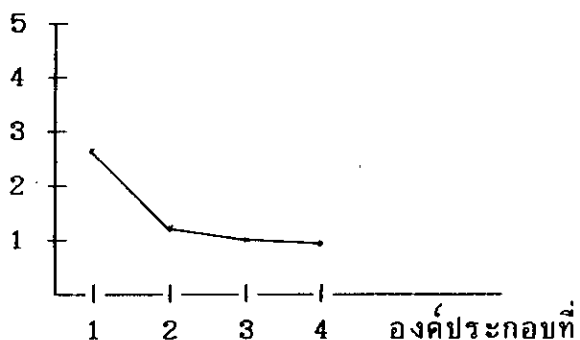


ฉบับที่ 2

ภาพประกอบ 8 แสดงค่าไอเคนของ 4 องค์ประกอบของแบบทดสอบวิชาสังคมศึกษา กข ฉบับที่ 1 และฉบับที่ 2

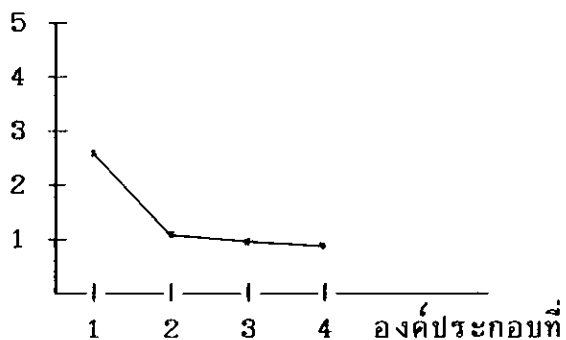
1.6 แบบทดสอบวิชาภาษาอังกฤษ กขค

ค่าไอเคน



ฉบับที่ 1

ค่าไอเคน



ฉบับที่ 2

ภาพประกอบ 9 แสดงค่าไอเคนของ 4 องค์ประกอบของแบบทดสอบวิชาภาษาอังกฤษ กขค ฉบับที่ 1 และฉบับที่ 2

2. ค่าสถิติพื้นฐานของผลการสอบและคุณภาพของแบบทดสอบแต่ละวิชา

2.1 ผลการวิเคราะห์ค่าสถิติพื้นฐานของแบบทดสอบ 6 วิชา 12 ฉบับ ดังแสดงในตาราง 5

ตาราง 5 ค่าสถิติพื้นฐานของผลการสอบของแบบทดสอบแต่ละวิชา

แบบทดสอบวิชา	ฉบับที่	จำนวน ข้อสอบ	รวม				เพศชาย				เพศหญิง			
			n	max	min	$\bar{X}$	n	max	min	$\bar{X}$	n	max	min	$\bar{X}$
ภาษาไทย ก	1	42	4032	26	4	13.03	1345	26	5	13.17	2687	26	4	12.95
	2	42	4095	36	5	20.10	1377	35	7	19.48	2718	36	5	20.41
สังคมศึกษา ก	1	64	4049	57	14	34.29	1316	56	15	34.76	2688	57	14	34.05
	2	64	4072	48	4	16.92	1358	48	7	16.69	2714	44	4	17.04
ภาษาอังกฤษ กข	1	36	4077	33	2	11.60	1388	33	3	11.71	2689	29	2	11.55
	2	36	4039	22	2	10.51	1327	20	2	10.33	2712	22	2	10.60
ภาษาไทย กข	1	37	2687	37	7	18.55	704	35	8	18.25	1983	37	7	18.65
	2	37	2723	21	4	12.07	749	20	4	12.05	1974	21	4	12.08
สังคมศึกษา กข	1	59	2685	52	13	27.04	715	52	13	26.69	1970	48	13	27.17
	2	59	2730	35	11	21.75	739	34	11	21.85	1991	35	11	21.71
ภาษาอังกฤษ กขค	1	37	2719	36	4	17.74	722	32	4	14.29	1977	36	4	14.90
	2	37	2701	27	3	13.32	736	24	3	13.08	1965	27	3	13.41

จากตาราง 5 ค่าสถิติพื้นฐานของแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 1 จำนวน 42 ข้อ จากผู้สอบทั้งหมด 4,032 คน เป็นเพศชาย 1,345 คน เพศหญิง 2,687 คน ได้คะแนนสูงสุด 26 คะแนนทั้ง 3 กลุ่ม คะแนนต่ำสุด 4 คะแนนในกลุ่มผู้สอบทั้งหมดและเพศหญิง ส่วนเพศชายคะแนนต่ำสุด 5 คะแนน คะแนนเฉลี่ยของกลุ่มผู้สอบทั้งหมดเท่ากับ 13.03 เพศชายเท่ากับ 13.17 และเพศหญิงเท่ากับ 12.95 วิชาภาษาไทย ก ฉบับที่ 2 จำนวน 42 ข้อ จากผู้สอบทั้งหมด 4,095 คน เป็นเพศชาย 1,377 คน เพศหญิง 2,718 คน ได้คะแนนสูงสุด และคะแนนต่ำสุดของกลุ่มผู้สอบทั้งหมดและเพศหญิงเป็น 36 และ 5 คะแนนตามลำดับ ส่วนเพศชายคะแนนสูงสุดและต่ำสุด 35 และ 7 คะแนนตามลำดับ คะแนนเฉลี่ยในกลุ่มผู้สอบทั้งหมดเท่ากับ 20.10 เพศชายเท่ากับ 19.48 และเพศหญิงเท่ากับ 20.41 วิชาสังคมศึกษา ก ฉบับที่ 1 64 ข้อ จากผู้สอบทั้งหมด 4,049 คน เป็นเพศชาย 1,316 คน เพศหญิง 2,688 คน คะแนนสูงสุดและต่ำสุดของกลุ่มผู้สอบทั้งหมดและเพศหญิงเป็น 57 และ 14 คะแนนตามลำดับ ส่วนเพศชายคะแนนสูงสุด และต่ำสุด 56 และ 15 คะแนนตามลำดับ คะแนนเฉลี่ยในกลุ่มผู้สอบทั้งหมดเท่ากับ 34.29 เพศชายเท่ากับ 34.76 และเพศหญิงเท่ากับ 34.05 วิชาสังคมศึกษา ก ฉบับที่ 2 จำนวน 64 ข้อจากผู้สอบทั้งหมด 4,072 คน เป็นเพศชาย 1,358 คน เพศหญิง 2,714 คน คะแนนสูงสุดของกลุ่มผู้สอบทั้งหมด และเพศชาย 48 คะแนน เพศหญิงคะแนนสูงสุด 44 คะแนน คะแนนต่ำสุดของกลุ่มผู้สอบทั้งหมด และเพศหญิง 4 คะแนน เพศชายคะแนนต่ำสุด 7 คะแนน คะแนนเฉลี่ยในกลุ่มผู้สอบทั้งหมดเท่ากับ 16.92 เพศชายเท่ากับ 16.69 และเพศหญิงเท่ากับ 17.04 วิชาภาษาอังกฤษ กข ฉบับที่ 1 36 ข้อ จากผู้สอบทั้งหมด 4,077 คน เพศชาย 1,388 คน เพศหญิง 2,689 คน คะแนนสูงสุดของกลุ่มผู้สอบทั้งหมด และเพศชาย 33 คะแนน เพศหญิงคะแนนสูงสุด 29 คะแนน คะแนนต่ำสุดกลุ่มผู้สอบทั้งหมด และเพศหญิง 2 คะแนน เพศชายคะแนนต่ำสุด 3 คะแนน คะแนนเฉลี่ยในกลุ่มผู้สอบทั้งหมดเท่ากับ 11.60 เพศชายเท่ากับ 11.71 และเพศหญิงเท่ากับ 11.55 วิชาภาษาอังกฤษ กข ฉบับที่ 2 36 ข้อ จากผู้สอบทั้งหมด 4,039 คน เพศชาย 1,327 คน เพศหญิง 2,712 คน คะแนนสูงสุดของผู้สอบทั้งหมด และเพศหญิง 22 คะแนน คะแนนสูงสุดของเพศชาย 20 คะแนน คะแนนต่ำสุดทั้ง 3 กลุ่ม 2 คะแนน คะแนนเฉลี่ยในกลุ่มผู้สอบทั้งหมดเท่ากับ 10.51 เพศชายเท่ากับ 10.33 และเพศหญิงเท่ากับ 10.60 วิชาภาษาไทย กข ฉบับที่ 1 37 ข้อ จากผู้สอบทั้งหมด 2,687 คน เพศชาย 704 คน เพศหญิง 1,983 คน คะแนนสูงสุดและต่ำสุดในกลุ่มผู้สอบทั้งหมด และเพศหญิง 37 และ

7 คะแนน ส่วนเพศชายคะแนนสูงสุดและต่ำสุด 35 และ 8 คะแนน คะแนนเฉลี่ยในกลุ่มผู้สอบทั้งหมดเท่ากับ 18.55 เพศชายเท่ากับ 18.25 และเพศหญิงเท่ากับ 18.25 และเพศหญิงเท่ากับ 18.65 วิชาภาษาไทย กข ฉบับที่ 2 37 ข้อ จากผู้สอบทั้งหมด 2,723 คน เพศชาย 749 คน เพศหญิง 1,974 คน คะแนนสูงสุดกลุ่มผู้สอบทั้งหมด และเพศหญิง 21 คะแนน เพศชายคะแนนสูงสุด 20 คะแนน ส่วนคะแนนต่ำสุดทั้ง 3 กลุ่ม 4 คะแนน คะแนนเฉลี่ยของกลุ่มผู้สอบทั้งหมดเท่ากับ 12.07 เพศชายเท่ากับ 12.05 และเพศหญิงเท่ากับ 12.08 วิชาสังคมศึกษา กข ฉบับที่ 1 59 ข้อ จากผู้สอบทั้งหมด 2,685 คน เพศชาย 715 คน เพศหญิง 1,970 คน คะแนนสูงสุดของผู้สอบทั้งหมดและเพศชาย 52 คะแนน เพศหญิงคะแนนสูงสุด 48 คะแนน ส่วนคะแนนต่ำสุดทั้ง 3 กลุ่ม 13 คะแนน เท่ากับคะแนนเฉลี่ยของผู้สอบทั้งหมดเท่ากับ 27.04 เพศชายเท่ากับ 26.69 เพศหญิงเท่ากับ 27.17 วิชาสังคมศึกษา กข ฉบับที่ 2 59 ข้อ จากกลุ่มผู้สอบทั้งหมด 2,730 คน เพศชาย 739 คน เพศหญิง 1,991 คน คะแนนสูงสุดของผู้สอบทั้งหมดและเพศหญิง 35 คะแนน คะแนนสูงสุดของเพศชาย 34 คะแนน ส่วนคะแนนต่ำสุดทั้ง 3 กลุ่ม 11 คะแนน เท่ากับคะแนนเฉลี่ยของกลุ่มผู้สอบทั้งหมดเท่ากับ 21.75 เพศชายเท่ากับ 21.85 และเพศหญิงเท่ากับ 21.71 วิชาภาษาอังกฤษ กขค ฉบับที่ 1 37 ข้อ จากผู้สอบทั้งหมด 2,719 คน เพศชาย 722 คน เพศหญิง 1,997 คน คะแนนสูงสุดของผู้สอบทั้งหมด และเพศหญิง 36 คะแนน คะแนนสูงสุดของเพศชาย 32 คะแนน คะแนนต่ำสุดทั้ง 3 กลุ่ม 4 คะแนน เท่ากัน คะแนนเฉลี่ยผู้สอบทั้งหมดเท่ากับ 14.74 เพศชายเท่ากับ 14.29 และเพศหญิงเท่ากับ 14.90 และวิชาภาษาอังกฤษ กขค ฉบับที่ 2 37 ข้อ จากผู้สอบทั้งหมด 2,701 คน เพศชาย 736 คน เพศหญิง 1,965 คน คะแนนสูงสุดของผู้สอบทั้งหมด และเพศหญิง 27 คะแนน คะแนนสูงสุดของเพศชาย 24 คะแนน คะแนนต่ำสุดทั้ง 3 กลุ่ม 3 คะแนน เท่ากัน คะแนนเฉลี่ยผู้สอบทั้งหมดเท่ากับ 13.32 เพศชายเท่ากับ 13.08 และเพศหญิงเท่ากับ 13.41

2.2 ผลการประมาณค่าพารามิเตอร์แบบ 3 พารามิเตอร์รายข้อ และค่าฟังก์ชันสหสัมพันธ์ของแบบทดสอบแต่ละฉบับ ดังแสดงในตาราง 6-17 และภาพประกอบ 10-33

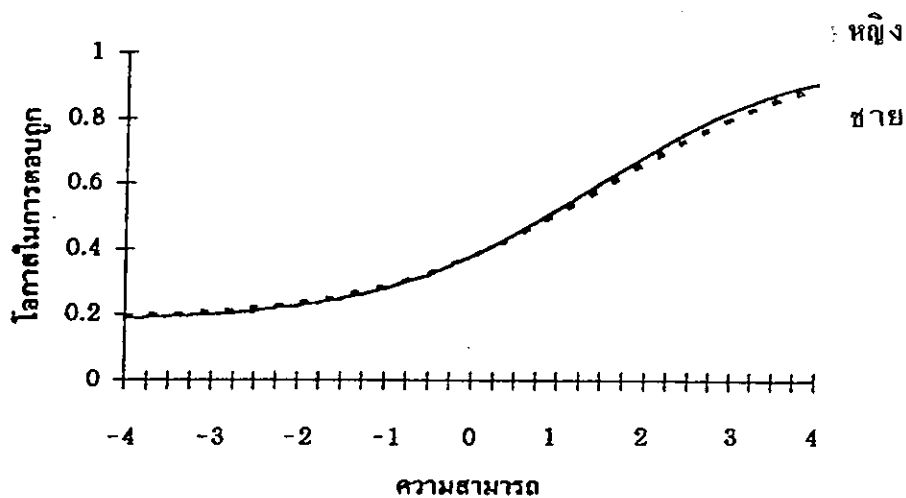
ตาราง 6 ค่าพารามิเตอร์ของแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 1

ข้อที่	รวม (N = 4032)			ชาย (N = 1345)			หญิง (N = 2687)		
	a	b	c	a	b	c	a	b	c
1	.968	1.508	.181	.824	1.938	.205	.870	1.372	.152
2	1.096	1.857	.273	.564	2.270	.246	1.046	1.814	.261
3	.675	-.340	.218	.587	-.271	.177	.692	-.422	.227
4	1.072	1.339	.070	.829	1.719	.082	1.084	1.224	.059
5	.482	.948	.177	.624	.748	.193	.453	1.073	.182
6	.529	4.836	.103	.371	4.764	.164	.484	4.905	.106
7	.684	2.202	.212	.411	3.920	.197	.669	1.934	.207
8	.714	2.220	.085	.551	2.929	.100	.715	2.072	.073
9	.425	.928	.114	.393	1.272	.151	.462	.871	.124
10	.956	1.735	.064	1.069	1.620	.063	.950	1.777	.071
11	.967	2.856	.225	.417	5.740	.215	.946	2.744	.227
12	.901	2.263	.182	.544	3.439	.173	.879	2.118	.178
13	.379	4.129	.276	.393	5.468	.297	.359	3.741	.259
14	.536	2.282	.252	.419	3.083	.265	.480	2.063	.217
15	.178	-.200	.247	.239	.869	.292	.163	-.824	.233
16	.587	2.406	.370	.269	3.121	.291	.662	2.185	.373
17	.422	1.087	.086	.517	1.354	.136	.412	1.043	.095
18	1.109	-.215	.210	.981	-.114	.193	1.107	-.336	.189
19	.361	-2.787	.196	.349	-2.454	.208	.375	-3.044	.192
20	.495	.550	.079	.729	.873	.185	.467	.494	.075

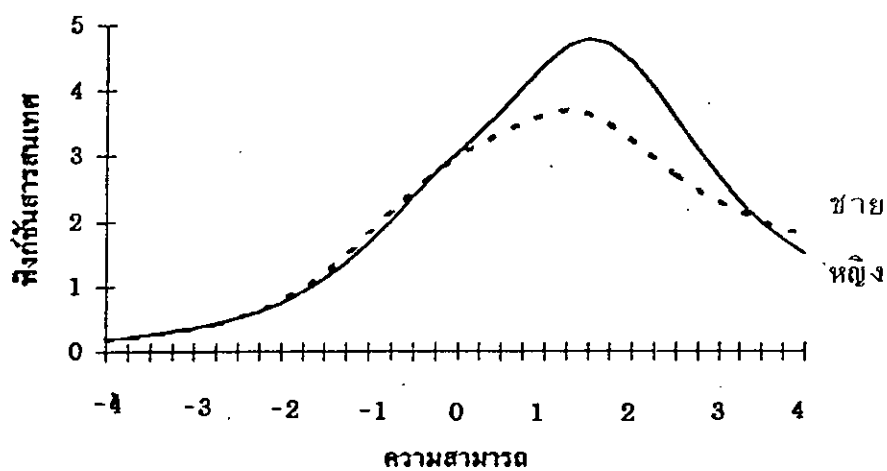
ตาราง 6 (ต่อ)

ข้อที่	รวม (N = 4032)			ชาย (N = 1345)			หญิง (N = 2687)		
	a	b	c	a	b	c	a	b	c
21	.356	-.944	.164	.366	-.479	.206	.354	-1.100	.168
22	.361	1.572	.107	.424	1.514	.142	.364	1.646	.117
23	.184	-4.161	.195	.231	-3.131	.205	.177	-4.415	.201
24	.400	-.112	.104	.431	.371	.170	.400	-.213	.112
25	.237	1.061	.178	.268	1.151	.189	.244	1.243	.214
26	.413	6.488	.303	.370	5.645	.287	.423	7.890	.308
27	.390	4.297	.097	.432	4.479	.107	.416	4.061	.102
28	.157	.864	.212	.209	.314	.205	.160	1.383	.243
29	.305	-.569	.168	.281	-.834	.204	.348	-.361	.168
30	.635	.120	.127	.681	.376	.152	.621	.033	.135
31	.611	-.460	.088	.629	-.155	.143	.614	-.525	.102
32	.312	1.721	.152	.300	3.125	.191	.332	1.433	.176
33	.417	.437	.122	.522	.602	.134	.389	.499	.160
34	.379	2.544	.304	.385	1.857	.249	.374	2.818	.319
35	.487	2.764	.204	.426	3.019	.206	.504	2.725	.257
36	.355	-1.336	.141	.300	-.813	.191	.382	-1.472	.151
37	.593	3.033	.238	.629	2.698	.229	.519	3.310	.235
38	.335	4.552	.378	.399	6.733	.374	.287	4.231	.372
39	.308	.604	.152	.383	.657	.164	.296	.749	.185
40	.309	5.030	.361	.311	3.389	.324	.334	5.323	.364
41	.304	2.207	.136	.343	3.079	.199	.314	1.938	.133
42	.479	-.349	.144	.540	.205	.226	.463	-.551	.133
$\bar{X}$	.520	1.403	.183	.481	1.811	.198	.514	1.367	.187
SD	.259	2.076	.082	.205	2.169	.063	.251	2.206	.081

จากตาราง 6 ค่าพารามิเตอร์ของแบบทดสอบวิชา ภาษาไทย ก ฉบับที่ 1 จำนวน 42 ข้อ จากผู้สอบจำนวน 4,032 คน เป็นชาย 1,345 คน หญิง 2,687 คน เมื่อคำนวณจากจำนวนผู้สอบทั้งหมด ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .157 - 1.109 ซึ่งมีค่าเฉลี่ยเท่ากับ .520 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .259 ค่าความยากง่ายอยู่ระหว่าง -4.161 - 6.488 มีค่าเฉลี่ยเท่ากับ 1.403 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.076 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .070 - .378 มีค่าเฉลี่ยเท่ากับ .183 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .082 ส่วนกลุ่มชาย ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .209 - 1.069 มีค่าเฉลี่ยเท่ากับ .481 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .205 ค่าความยากง่ายอยู่ระหว่าง -3.131 - 6.733 มีค่าเฉลี่ยเท่ากับ 1.811 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.169 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .063 - .374 มีค่าเฉลี่ยเท่ากับ .198 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .063 และในกลุ่มหญิง ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .160 - 1.107 มีค่าเฉลี่ยเท่ากับ .514 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .251 ค่าความยากง่ายอยู่ระหว่าง -4.415 - 7.890 ค่าเฉลี่ยเท่ากับ 1.367 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.206 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .059 - .373 มีค่าเฉลี่ยเท่ากับ .187 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .081 เพื่อพิจารณาค่าเฉลี่ยของความยากง่ายระหว่างเพศ พบว่าชายมีค่าความยากเฉลี่ยสูงกว่าหญิง ซึ่งสามารถพิจารณาได้จากโค้งลักษณะของแบบทดสอบและโค้งสารสนเทศ ที่แสดงในภาพประกอบ 10-11



ภาพประกอบ 10 ได้ังลักษณะของแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 1

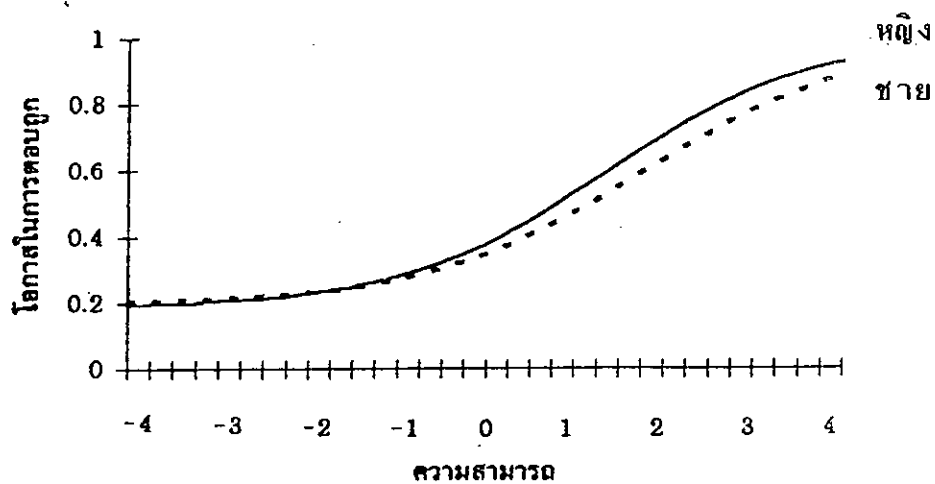


ภาพประกอบ 11 ได้ังสารสนเทศของแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 1

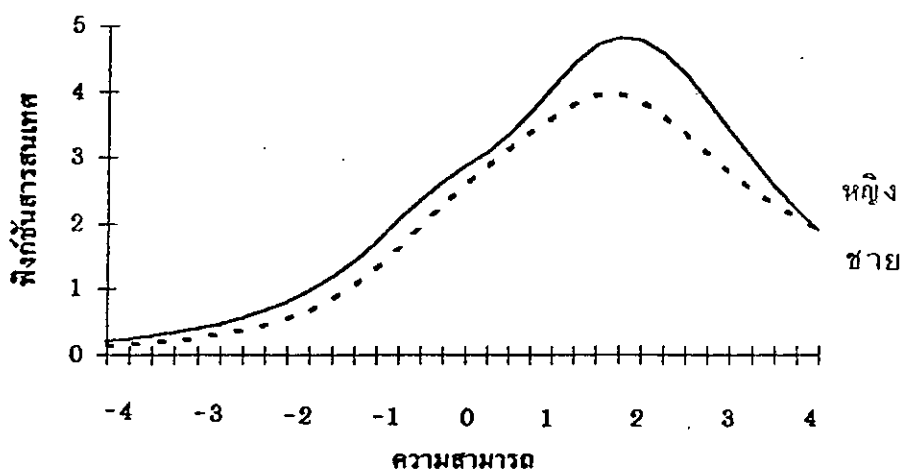
ตาราง 7 ค่าพารามิเตอร์ของแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 2

ข้อที่	รวม (N = 4095)			ชาย (N = 1377)			หญิง (N = 2718)		
	a	b	c	a	b	c	a	b	c
1	.794	2.224	.283	.350	3.368	.254	.804	2.029	.271
2	.901	1.571	.167	.629	1.963	.190	.925	1.393	.145
3	1.032	1.379	.052	.880	1.400	.069	1.104	1.285	.047
4	.659	-.474	.178	.648	-.598	.172	.655	-.546	.176
5	.540	1.174	.239	.407	.749	.178	.641	1.181	.257
6	.723	2.074	.080	.474	2.826	.080	.814	1.850	.081
7	.589	2.121	.200	.434	2.382	.152	.601	2.020	.216
8	1.008	1.733	.055	1.120	1.382	.065	1.008	1.758	.053
9	.449	1.153	.188	.387	.665	.138	.636	1.315	.261
10	.426	3.020	.141	.359	4.081	.167	.470	2.671	.139
11	.480	3.634	.199	.416	4.243	.185	.442	3.686	.204
12	.556	2.237	.263	.426	3.158	.299	.499	1.927	.214
13	.294	4.944	.273	.338	4.188	.273	.312	4.770	.277
14	.415	2.798	.297	.368	3.016	.283	.354	2.758	.279
15	.179	-.233	.246	.209	.454	.266	.174	-.804	.230
16	1.093	-.170	.230	.972	-.361	.196	1.119	-.231	.228
17	.454	.740	.093	.476	.768	.149	.478	.697	.095
18	.580	.408	.071	.558	.380	.106	.618	.384	.082
19	.390	-2.204	.197	.397	-1.977	.198	.374	-2.552	.192
20	.288	-1.195	.158	.361	-.994	.180	.265	-1.290	.180
21	.457	1.559	.155	.445	1.509	.189	.473	1.478	.142
22	.348	-.452	.124	.299	-.231	.190	.387	-.493	.125
23	.202	-3.333	.182	.302	-2.507	.194	.177	-3.714	.192

จากตาราง 7 ค่าพารามิเตอร์ของแบบทดสอบวิชา ภาษาไทย ก ฉบับที่ 2 จำนวน 42 ข้อ จากผู้สอบจำนวน 4,095 คน เป็นชาย 1,377 คน หญิง 2,718 คน เมื่อคำนวณจากจำนวนผู้สอบทั้งหมด ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .151 - 1.093 ซึ่งมีค่าเฉลี่ยเท่ากับ .471 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .233 ค่าความยากง่ายอยู่ระหว่าง -3.333 - 6.320 มีค่าเฉลี่ยเท่ากับ 1.522 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.063 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .052 - .330 มีค่าเฉลี่ยเท่ากับ .174 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .070 ส่วนกลุ่มชาย ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .206 - 1.120 มีค่าเฉลี่ยเท่ากับ .456 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .202 ค่าความยากง่ายอยู่ระหว่าง -2.507 - 5.383 มีค่าเฉลี่ยเท่ากับ 1.522 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.950 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .065 - .366 มีค่าเฉลี่ยเท่ากับ .185 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .061 และในกลุ่มหญิง ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .174 - 1.119 มีค่าเฉลี่ยเท่ากับ .483 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .239 ค่าความยากง่ายอยู่ระหว่าง -4.28 - 7.715 ค่าเฉลี่ยเท่ากับ 1.414 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.434 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .047 - .370 มีค่าเฉลี่ยเท่ากับ .180 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .072 เมื่อพิจารณาค่าเฉลี่ยของความยากง่ายระหว่างเพศ พบว่าชายมีค่าความยากง่ายเฉลี่ยสูงกว่าหญิง ซึ่งสามารถพิจารณาได้จากโค้งลักษณะของแบบทดสอบ และได้สังเกตเห็นเพศที่แสดงในภาพประกอบ 12-13



ภาพประกอบ 12 ได้ังลักษณะของแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 2



ภาพประกอบ 13 ได้ังสารสนเทศของแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 2

ตาราง 8 ค่าพารามิเตอร์ของแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 1

ข้อที่	รวม (N = 4049)			ชาย (N = 1361)			หญิง (N = 2688)		
	a	b	c	a	b	c	a	b	c
1	.375	6.955	.156	.696	-.583	.174	.780	-.244	.273
2	.429	2.390	.091	.395	.418	.138	.373	.253	.154
3	.328	1.822	.189	.310	-1.858	.201	.378	-1.603	.198
4	.650	-2.545	.196	.564	-.742	.211	.487	-1.232	.248
5	.858	3.827	.192	.683	3.758	.044	.583	3.725	.040
6	.134	4.399	.231	.371	5.696	.082	.475	5.656	.090
7	.222	.982	.402	.247	1.072	.189	.275	.760	.154
8	.205	4.730	.500	.320	2.821	.266	.254	4.095	.321
9	.628	2.524	.089	.286	4.568	.293	.287	3.823	.164
10	.833	1.120	.085	.491	-2.872	.203	.303	-3.982	.219
11	.341	-3.061	.203	.618	.254	.128	.583	.043	.095
12	.816	3.263	.062	.265	1.955	.248	.407	2.860	.386
13	.243	3.601	.242	.341	5.780	.234	.654	3.107	.183
14	.674	2.990	.145	.396	3.065	.118	.761	2.544	.193
15	.225	-1.539	.222	.649	-1.710	.169	.850	-1.266	.128
16	.489	6.351	.099	.510	-.183	.160	.602	.242	.161
17	.225	3.065	.033	.181	2.463	.270	.313	6.023	.449
18	.515	4.671	.412	.497	3.013	.112	.618	2.805	.104
19	.846	1.034	.228	.354	-3.780	.196	.257	-5.391	.214
20	.300	.654	.276	.557	.397	.145	.611	.726	.153
21	.931	1.643	.099	.651	.350	.073	.257	-5.391	.214
22	.383	.510	.400	.435	2.440	.067	.611	.726	.153
23	.600	3.742	.121	.348	4.356	.249	.936	.220	.054

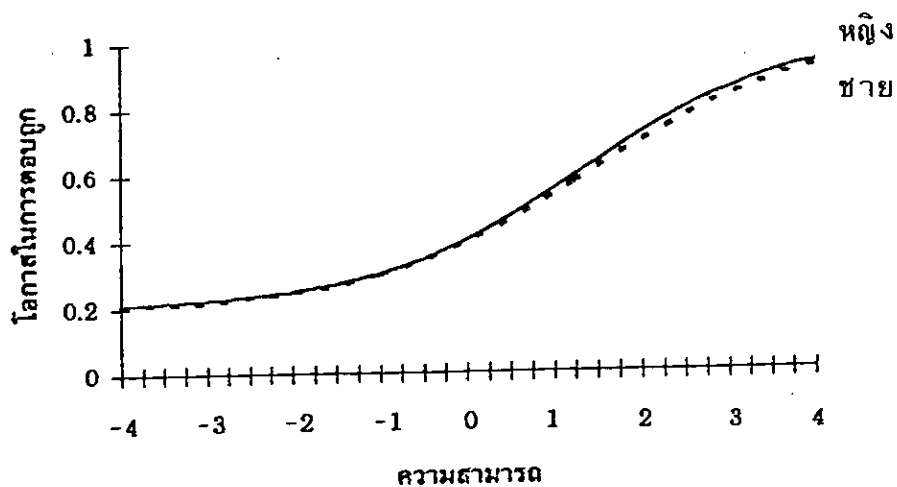
ตาราง 8 (ต่อ)

ข้อที่	รวม (N = 4049)			ชาย (N = 1316)			หญิง (N = 2688)		
	a	b	c	a	b	c	a	b	c
24	.404	6.476	.188	.345	.960	.180	.467	2.773	.068
25	.247	4.532	.430	.623	-.074	.190	.248	3.223	.265
26	.305	.719	.267	1.049	.800	.269	.647	.046	.117
27	.647	1.430	.185	.386	-.637	.170	.933	.921	.259
28	.635	3.536	.400	.310	.112	.181	.325	1.618	.279
29	.501	.056	.109	.289	4.384	.454	.376	-.735	.159
30	.290	2.884	.159	.253	3.205	.217	.371	.233	.101
31	.507	5.862	.051	.384	2.608	.100	.142	6.160	.500
32	.116	-3.272	.236	.354	.980	.174	.540	2.469	.217
33	.262	5.668	.500	.614	1.325	.188	.524	2.544	.122
34	.446	7.087	.087	.719	.133	.118	.328	1.875	.307
35	.246	1.984	.291	.681	-2.557	.192	.862	1.019	.284
36	.403	.809	.237	.421	1.030	.299	.652	.516	.104
37	.390	3.981	.343	.898	-.993	.176	.619	-2.353	.212
38	.857	3.889	.025	.504	-1.811	.185	.450	2.513	.261
39	.743	1.365	.173	.344	1.199	.301	.775	-.875	.152
40	.273	3.940	.314	.396	5.651	.286	.398	-1.511	.208
41	.398	.863	.115	.128	-.515	.271	.606	2.098	.498
42	.289	2.565	.261	.618	2.169	.402	.642	2.149	.397
43	.573	3.265	.234	.720	3.007	.047	.809	2.916	.059
44	.278	4.050	.195	.337	3.940	.152	.345	3.419	.140
45	.349	-2.236	.188	.584	-1.657	.179	.603	-1.256	.142

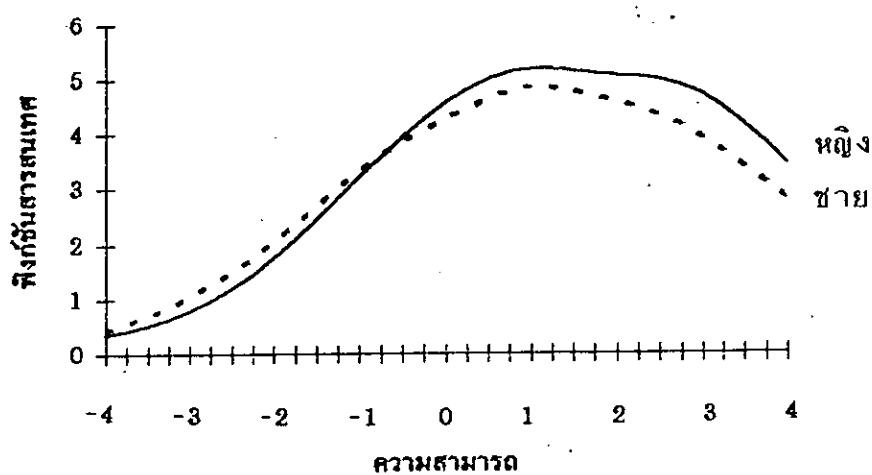
ตาราง 8 (ต่อ)

ข้อที่	รวม (N = 4049)			ชาย (N = 1316)			หญิง (N = 2688)		
	a	b	c	a	b	c	a	b	c
46	.214	.957	.240	.321	-2.062	.199	.370	-1.305	.158
47	1.060	2.507	.216	.571	-.748	.234	.501	-.793	.247
48	.316	.965	.164	.723	1.019	.221	.664	.959	.220
49	.392	5.128	.229	.426	4.640	.166	.419	4.484	.133
50	.522	4.147	.139	.513	2.466	.154	.663	2.792	.187
51	.907	3.042	.090	.493	.099	.137	.511	.767	.143
52	.246	-.425	.150	.583	-.448	.174	.571	-.402	.150
53	1.081	2.262	.040	.543	.219	.244	.510	.438	.247
54	.360	1.548	.183	.333	6.250	.423	.343	3.909	.364
55	.322	.617	.193	.439	-1.465	.187	.383	-1.660	.167
56	1.169	.918	.112	.245	4.681	.301	.267	4.103	.336
57	.639	1.902	.152	.342	3.490	.183	.475	2.913	.218
58	.492	4.153	.258	.374	2.807	.134	.451	2.746	.082
59	.997	1.241	.121	.354	4.370	.331	.325	4.376	.353
60	.828	3.008	.115	.711	2.481	.208	.726	2.470	.234
61	.694	1.336	.163	.356	1.227	.161	.304	1.306	.210
62	.505	5.403	.120	.949	2.561	.062	.632	3.621	.065
63	.349	5.162	.163	.934	1.267	.201	.623	1.494	.145
64	.340	6.398	.450	.468	3.230	.152	.439	3.659	.165
$\bar{X}$	.497	2.587	.202	.481	1.406	.196	.507	1.314	.203
SD	.258	2.291	.113	.191	2.349	.082	.185	2.451	.103

จากตาราง 8 ค่าพารามิเตอร์ของแบบทดสอบวิชา สังคมศึกษา ก ฉบับที่ 1 จำนวน 64 ข้อ จากผู้สอบจำนวน 4,049 คน เป็นชาย 1,361 คน หญิง 2,688 คน เมื่อคำนวณจากจำนวนผู้สอบทั้งหมด ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .116 - 1.169 ซึ่งมีค่าเฉลี่ยเท่ากับ .497 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .258 ค่าความยากง่ายอยู่ระหว่าง -3.272 - 7.087 มีค่าเฉลี่ยเท่ากับ 2.587 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.291 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .025 - .500 มีค่าเฉลี่ยเท่ากับ .202 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .113 ส่วนกลุ่มชาย ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .128 - 1.049 มีค่าเฉลี่ยเท่ากับ .481 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .191 ค่าความยากง่ายอยู่ระหว่าง -3.780 - 6.250 มีค่าเฉลี่ยเท่ากับ 1.406 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.349 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .044 - .459 มีค่าเฉลี่ยเท่ากับ .196 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .082 และในกลุ่มหญิง ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .142 - .936 มีค่าเฉลี่ยเท่ากับ .507 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .185 ค่าความยากง่ายอยู่ระหว่าง -5.391 - 6.160 ค่าเฉลี่ยเท่ากับ 1.314 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.451 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .040 - .500 มีค่าเฉลี่ยเท่ากับ .203 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .103 เมื่อพิจารณาค่าเฉลี่ยของความยากง่ายระหว่างเพศ พบว่าชายมีค่าความยากเฉลี่ยสูงกว่าหญิง ซึ่งสามารถพิจารณาได้จากโค้งลักษณะของแบบทดสอบ และโค้งสารสนเทศที่แสดงในภาพประกอบ 14-15



ภาพประกอบ 14 ได้ังลักษณะของแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 1



ภาพประกอบ 15 ได้ังสารสนเทศของแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 1

ตาราง 9 ค่าพารามิเตอร์ของแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 2

ข้อที่	รวม (N = 4072)			ชาย (N = 1358)			หญิง (N = 2714)		
	a	b	c	a	b	c	a	b	c
1	.713	3.182	.193	.545	3.171	.169	.632	3.662	.198
2	.575	1.240	.138	.663	1.502	.228	.564	1.123	.101
3	.740	3.336	.108	.514	4.182	.096	.784	3.199	.112
4	.288	1.502	.176	.287	1.773	.187	.319	1.469	.201
5	.620	3.420	.325	.392	5.095	.300	.591	3.214	.328
6	.306	5.195	.393	.279	4.505	.323	.317	5.206	.414
7	.402	2.770	.081	.463	2.198	.111	.416	3.098	.088
8	.526	2.461	.163	.418	2.499	.149	.597	2.461	.172
9	.193	3.964	.249	.359	7.106	.363	.219	2.390	.194
10	.453	-1.682	.135	.495	-1.785	.159	.430	-1.564	.161
11	.235	3.763	.289	.248	4.726	.340	.290	3.130	.290
12	.558	.032	.258	.450	.440	.193	.582	.127	.257
13	.546	-.772	.145	.537	-.868	.147	.565	-.634	.181
14	.515	.360	.078	.548	.224	.115	.521	.528	.094
15	.497	2.404	.152	.652	1.772	.162	.415	2.922	.148
16	.401	4.419	.154	.466	3.068	.131	.379	5.762	.172
17	.498	.817	.158	.428	.978	.140	.551	.783	.180
18	.576	-.198	.159	.667	-.186	.165	.538	-.109	.161
19	.290	-1.648	.164	.308	-1.887	.189	.280	-1.449	.176
20	.699	-1.419	.136	.789	-1.400	.180	.667	-1.391	.140
21	.265	3.074	.078	.325	2.816	.095	.272	3.251	.104
22	.694	3.030	.054	.612	3.441	.060	.710	2.956	.053
23	.700	1.792	.394	.305	1.740	.262	.801	1.701	.402

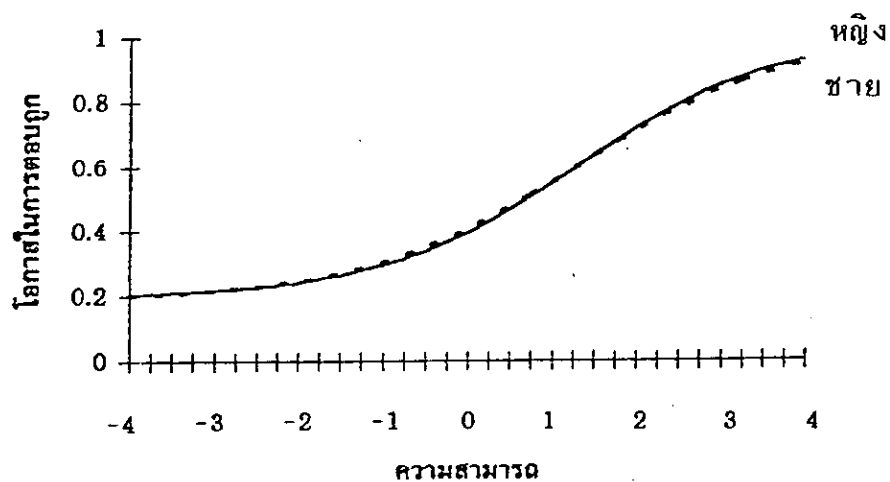
ตาราง 9 (ต่อ)

ข้อที่	รวม (N = 4072)			ชาย (N = 1358)			หญิง (N = 2714)		
	a	b	c	a	b	c	a	b	c
24	.159	.617	.257	.216	.173	.214	.153	1.055	.292
25	.143	-2.379	.211	.155	-1.731	.219	.163	-2.280	.213
26	.681	2.812	.479	.265	2.448	.319	.758	1.997	.499
27	.509	-.878	.118	.606	-1.163	.156	.483	-.617	.141
28	.695	-1.250	.177	.752	-1.447	.162	.670	-1.134	.197
29	.307	2.210	.209	.277	.875	.195	.376	2.803	.231
30	.601	-2.359	.156	.527	-2.870	.177	.624	-2.161	.174
31	.694	.441	.110	.683	.221	.112	.714	.584	.124
32	.685	.862	.154	.730	1.229	.162	.677	.619	.135
33	.271	.952	.169	.275	1.183	.188	.288	.939	.183
34	.410	2.598	.116	.327	4.059	.184	.521	2.190	.113
35	.243	3.309	.183	.230	3.475	.209	.311	3.148	.210
36	.700	2.879	.463	.370	2.438	.328	.526	3.911	.487
37	.293	.049	.109	.280	.200	.147	.314	.166	.132
38	.342	-1.070	.151	.296	-1.053	.174	.376	-.962	.170
39	.317	1.635	.327	.216	1.221	.264	.333	1.482	.302
40	.926	.872	.382	.694	.572	.302	.940	.941	.392
41	.573	-.018	.120	.486	-.230	.159	.631	.140	.131
42	1.090	2.503	.311	.563	3.544	.298	1.066	2.467	.310
43	.206	4.671	.346	.184	3.258	.300	.258	4.718	.365
44	.543	2.774	.080	.585	2.512	.099	.515	3.021	.077
45	.821	.138	.051	.680	.158	.138	.925	.188	.040
46	.684	.867	.080	.688	.813	.123	.698	.936	.076

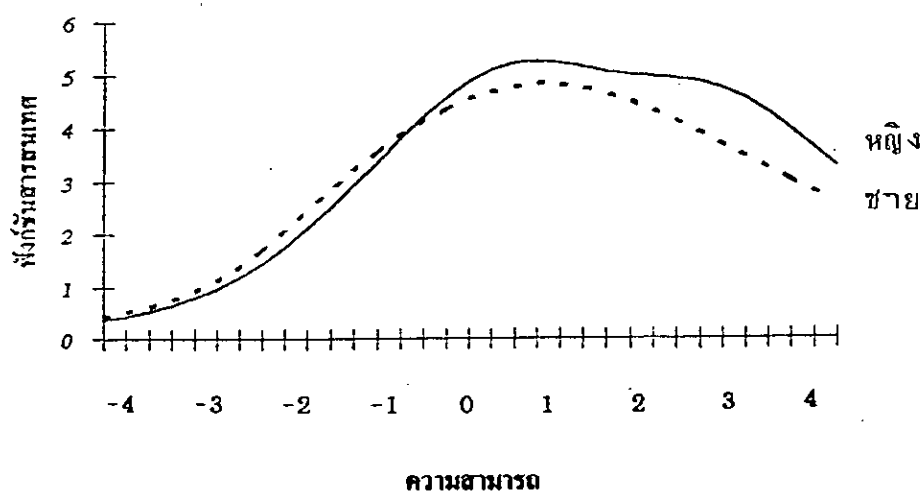
ตาราง 9 (ต่อ)

ข้อที่	รวม (N = 4072)			ชาย (N = 1358)			หญิง (N = 2714)		
	a	b	c	a	b	c	a	b	c
47	.343	-4.099	.181	.362	-3.813	.186	.336	-4.206	.192
48	.434	2.895	.090	.458	2.692	.105	.451	3.006	.097
49	.151	4.114	.270	.216	2.214	.256	.196	5.748	.363
50	.654	.277	.174	.609	-.007	.155	.693	.431	.191
51	.857	-1.025	.092	.848	-1.103	.119	.862	-.944	.118
52	.695	2.409	.133	.783	2.056	.128	.598	2.734	.133
53	.581	3.378	.181	.448	3.796	.178	.552	3.501	.176
54	.308	2.221	.335	.316	1.416	.233	.251	2.473	.331
55	.679	.420	.108	.784	.607	.159	.664	.353	.099
56	.327	-3.861	.183	.386	-3.473	.185	.300	-4.078	.194
57	.239	3.172	.147	.378	2.600	.243	.233	3.563	.147
58	.240	4.009	.262	.290	4.657	.312	.266	3.409	.253
59	.343	.252	.119	.374	.523	.148	.351	.225	.136
60	.565	3.866	.063	.427	6.510	.086	.583	3.484	.052
61	.515	-.563	.232	.652	-.067	.256	.476	-.971	.179
62	.328	-2.010	.163	.354	-1.577	.173	.330	-2.114	.179
63	.472	1.204	.133	.524	1.170	.107	.475	1.222	.158
64	.762	-.415	.210	.731	-.529	.157	.802	-.335	.246
$\bar{X}$	.500	1.219	.187	.465	1.277	.188	.502	1.304	.196
SD	.205	2.140	.099	.180	2.298	.071	.208	2.196	.102

จากตาราง 9 ค่าพารามิเตอร์ของแบบทดสอบวิชา สังคมศึกษา ก ฉบับที่ 2 จำนวน 64 ข้อ จากผู้สอบจำนวน 4,072 คน เป็นชาย 1,358 คน หญิง 2,714 คน เมื่อคำนวณจากจำนวนผู้สอบทั้งหมด ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .151 - 1.090 ซึ่งมีค่าเฉลี่ยเท่ากับ .500 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .205 ค่าความยากง่ายอยู่ระหว่าง -4.009 - 4.114 มีค่าเฉลี่ยเท่ากับ 1.219 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.140 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .051 - .394 มีค่าเฉลี่ยเท่ากับ .187 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .099 ส่วนกลุ่มชาย ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .155 - .848 มีค่าเฉลี่ยเท่ากับ .465 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .180 ค่าความยากง่ายอยู่ระหว่าง -3.813 - 5.095 มีค่าเฉลี่ยเท่ากับ 1.277 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.298 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .086 - .363 มีค่าเฉลี่ยเท่ากับ .188 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .071 และในกลุ่มหญิง ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .153 - 1.066 มีค่าเฉลี่ยเท่ากับ .502 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .208 ค่าความยากง่ายอยู่ระหว่าง -4.078 - 5.762 ค่าเฉลี่ยเท่ากับ 1.304 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.196 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .052 - .499 มีค่าเฉลี่ยเท่ากับ .196 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .102 เมื่อพิจารณาค่าเฉลี่ยของความยากง่ายระหว่างผู้เข้าสอบแต่ละเพศ พบว่าค่าความยากง่ายเฉลี่ยของข้อสอบในกลุ่มหญิงมีค่าสูงกว่าค่าความยากง่ายเฉลี่ยของข้อสอบในกลุ่มชาย ซึ่งสามารถพิจารณาได้จาก ไค้ดงลักษณะของแบบทดสอบและไค้ดงสารสนเทศที่แสดงในภาพประกอบ 16-17



ภาพประกอบ 16 ได้ังลักษณะของแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 2

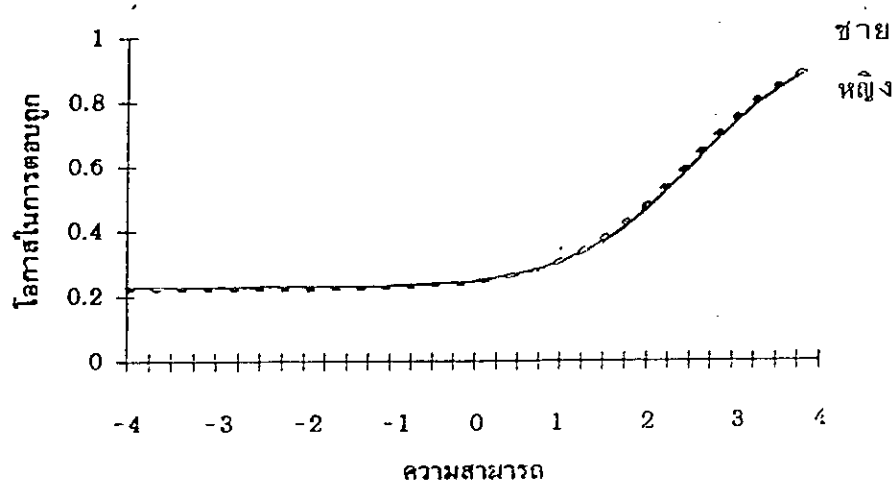


ภาพประกอบ 17 ได้ังสารสนเทศของแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 2

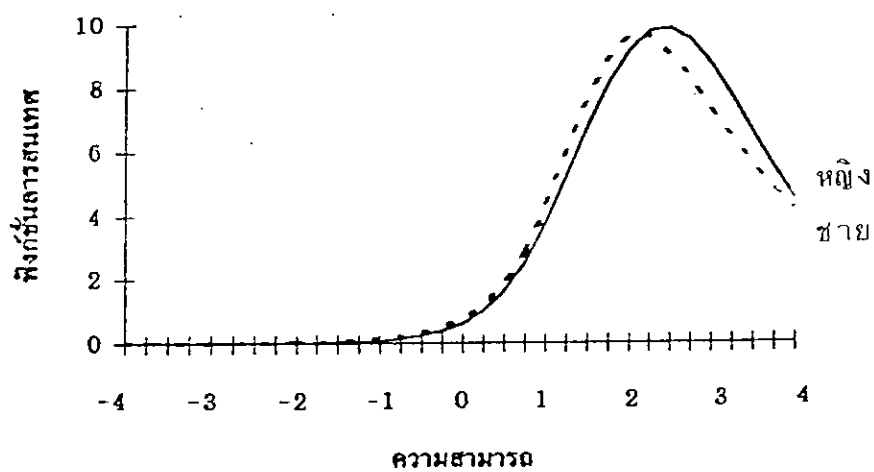
ตาราง 10 (ต่อ)

ข้อที่	รวม (N = 4077)			ชาย (N = 1388)			หญิง (N = 2689)		
	a	b	c	a	b	c	a	b	c
24	.834	3.654	.338	.820	3.215	.336	.732	4.127	.336
25	.906	2.754	.150	.640	2.925	.134	1.152	2.637	.155
26	1.019	2.464	.053	1.152	2.374	.078	.974	2.534	.043
27	1.008	2.447	.213	.805	2.695	.208	1.033	2.399	.213
28	1.218	2.595	.190	.993	2.933	.199	1.108	2.567	.181
29	1.366	2.953	.247	.792	4.685	.250	1.207	2.918	.244
30	.852	2.570	.199	1.352	2.154	.206	.654	2.846	.190
31	.591	2.610	.391	.626	1.725	.311	.662	2.995	.424
32	.610	3.688	.261	.644	3.049	.274	.655	3.847	.255
33	.877	1.793	.315	.586	1.930	.261	.991	1.703	.326
34	.927	1.781	.146	.845	1.648	.124	.985	1.854	.157
35	.695	1.642	.166	.846	1.781	.179	.672	1.534	.163
36	.879	1.490	.322	1.364	1.374	.375	.702	1.525	.283
$\bar{X}$	.851	2.533	.226	.821	2.499	.220	.832	2.589	.228
SD	.253	0.901	.110	.268	1.008	.094	.221	0.952	.113

จากตาราง 10 ค่าพารามิเตอร์ของแบบทดสอบวิชา ภาษาอังกฤษ กข
 ฉับที่ 1 จำนวน 36 ข้อ จากผู้สอบจำนวน 4,077 คน เป็นชาย 1,388 คน
 หญิง 2,689 คน เมื่อคำนวณจากจำนวนผู้สอบทั้งหมด ข้อสอบมีค่าอำนาจจำแนกอยู่
 ระหว่าง .482 - 1.379 ซึ่งมีค่าเฉลี่ยเท่ากับ .851 ส่วนเบี่ยงเบนมาตรฐาน
 เท่ากับ .253 ค่าความยากง่ายอยู่ระหว่าง 1.480 - 5.316 มีค่าเฉลี่ยเท่ากับ
 2.533 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .091 และค่าสัมประสิทธิ์การเดาอยู่
 ระหว่าง .053 - .491 มีค่าเฉลี่ยเท่ากับ .226 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ
 .110 ส่วนกลุ่มชาย ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .345 - 1.352 มีค่า
 เฉลี่ยเท่ากับ .821 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .268 ค่าความยากง่ายอยู่
 ระหว่าง .857 - 5.218 มีค่าเฉลี่ยเท่ากับ 2.499 ส่วนเบี่ยงเบนมาตรฐาน
 เท่ากับ 1.008 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .078 - .457 มีค่าเฉลี่ย
 เท่ากับ .220 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .094 และในกลุ่มหญิง ข้อสอบมี
 ค่าอำนาจจำแนกอยู่ระหว่าง .347 - 1.213 มีค่าเฉลี่ยเท่ากับ .832 ส่วน
 เบี่ยงเบนมาตรฐานเท่ากับ .221 ค่าความยากง่ายอยู่ระหว่าง 1.432 - 5.572
 ค่าเฉลี่ยเท่ากับ 2.589 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .952 และค่าสัมประสิทธิ์
 การเดาอยู่ระหว่าง .043 - .500 มีค่าเฉลี่ยเท่ากับ .228 ส่วนเบี่ยงเบน
 มาตรฐานเท่ากับ .113 เมื่อพิจารณาค่าเฉลี่ยของความยากง่ายระหว่างผู้เข้าสอบ
 แต่ละเพศพบว่าค่าความยากง่ายเฉลี่ยของข้อสอบในกลุ่มหญิงมีค่าสูงกว่าค่าความยาก
 ง่ายเฉลี่ยของข้อสอบในกลุ่มชาย ซึ่งสามารถพิจารณาได้จากโค้งลักษณะของแบบ
 ทดสอบและโค้งสารสนเทศที่แสดงในภาพประกอบ 18-19



ภาพประกอบ 18 ได้ังลักษณะของแบบทดสอบวิชาภาษาอังกฤษ กข ฉบับที่ 1



ภาพประกอบ 19 ได้ังสารสนเทศของแบบทดสอบวิชาภาษาอังกฤษ กข ฉบับที่ 1

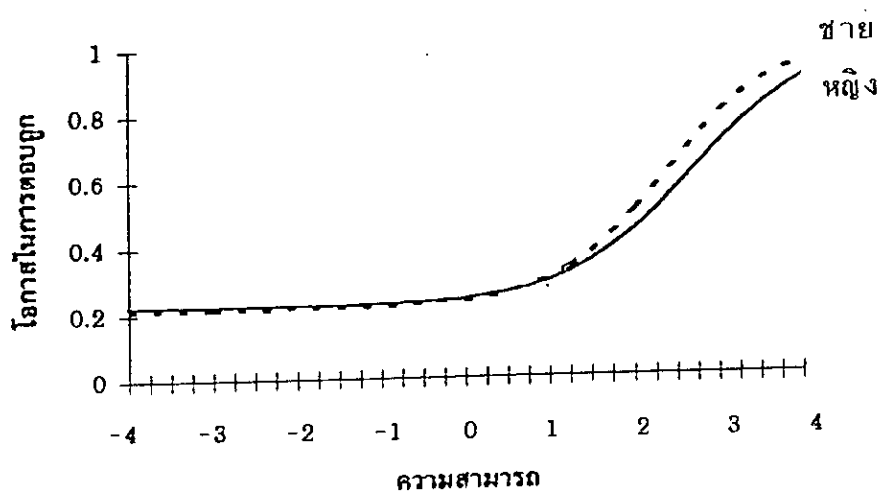
ตาราง 11 ค่าพารามิเตอร์ของแบบทดสอบวิชา ภาษาอังกฤษ กข ฉบับที่ 2

ข้อที่	รวม (N = 4039)			ชาย (N = 1327)			หญิง (N = 2712)		
	a	b	c	a	b	c	a	b	c
1	.626	3.000	.130	.919	2.583	.134	.579	3.199	.136
2	1.357	1.778	.095	1.510	1.757	.098	1.226	1.822	.092
3	.510	2.531	.199	.517	2.470	.189	.589	2.498	.220
4	.814	2.991	.147	.811	3.704	.164	.789	2.830	.138
5	1.671	2.432	.268	1.811	2.377	.236	1.467	2.493	.282
6	.450	2.806	.311	.491	2.611	.297	.479	2.813	.325
7	.784	1.913	.134	1.293	1.792	.179	.704	1.969	.122
8	.658	1.478	.160	.627	1.694	.154	.720	1.433	.180
9	.602	2.868	.160	.636	3.191	.177	.677	2.684	.165
10	1.394	2.229	.493	1.103	1.914	.457	1.174	2.540	.500
11	.670	1.619	.245	.611	1.671	.259	.699	1.580	.235
12	1.475	2.793	.228	1.325	2.858	.236	1.112	3.078	.223
13	1.176	1.785	.161	1.390	1.680	.173	1.051	1.878	.155
14	1.173	1.275	.127	1.179	1.325	.115	1.197	1.287	.142
15	1.103	1.820	.148	1.191	1.676	.126	1.046	1.939	.160
16	.832	2.052	.327	.791	1.782	.288	.844	2.207	.344
17	.947	3.342	.153	.899	2.863	.152	.831	3.946	.151
18	1.574	2.856	.297	1.209	2.719	.285	1.093	3.441	.299
19	.346	5.690	.500	.744	3.831	.462	.242	5.994	.500
20	.586	1.185	.307	.468	1.448	.288	.635	1.038	.303
21	.903	2.158	.153	.914	2.184	.147	.913	2.166	.159
22	1.040	2.449	.062	1.110	2.297	.080	1.033	2.528	.056
23	.687	2.669	.192	.722	2.847	.195	.654	2.631	.189

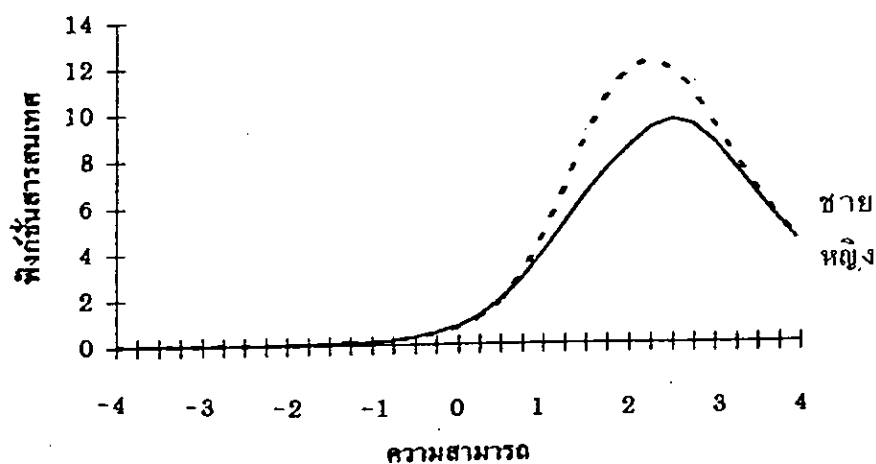
ตาราง 11 (ต่อ)

ข้อที่	รวม (N = 4039)			ชาย (N = 1327)			หญิง (N = 2712)		
	a	b	c	a	b	c	a	b	c
24	1.460	2.590	.216	1.039	2.807	.200	1.458	2.619	.221
25	1.087	2.983	.222	.921	3.502	.227	1.008	2.954	.217
26	.496	2.707	.151	.613	2.316	.154	.511	2.889	.167
27	.326	1.533	.315	.453	1.265	.269	.289	1.665	.338
28	.754	4.849	.268	.777	4.824	.275	.696	5.114	.263
29	1.002	1.716	.084	1.265	1.936	.119	1.013	1.628	.078
30	1.316	3.172	.349	.917	3.118	.349	.696	8.140	.348
31	.797	3.671	.312	.853	3.619	.317	.692	3.775	.306
32	.617	2.898	.214	.684	2.624	.131	.638	2.993	.128
33	.862	1.533	.120	.883	1.527	.102	.878	1.562	.137
34	.780	1.746	.264	1.000	1.689	.255	.690	1.817	.269
35	.602	1.462	.136	.875	1.556	.156	.554	1.447	.145
36	.686	1.467	.283	.687	1.200	.251	.629	1.558	.279
$\bar{X}$	0.893	2.446	0.220	0.923	2.368	0.214	0.820	2.671	0.221
SD	0.356	0.956	0.103	0.314	0.845	0.092	0.289	1.395	0.105

จากตาราง 11 ค่าพารามิเตอร์ของแบบทดสอบวิชา ภาษาอังกฤษ กข
 ฉัยที่ 2 จำนวน 36 ข้อ จากผู้สอบจำนวน 4,039 คน เป็นชาย 1,327 คน
 หญิง 2,712 คน เมื่อคำนวณจากจำนวนผู้สอบทั้งหมด ข้อสอบมีค่าอำนาจจำแนกอยู่
 ระหว่าง .326 - 1.671 ซึ่งมีค่าเฉลี่ยเท่ากับ .893 ส่วนเบี่ยงเบนมาตรฐาน
 เท่ากับ .356 ค่าความยากง่ายอยู่ระหว่าง 1.275 - 5.690 มีค่าเฉลี่ยเท่ากับ
 2.446 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .956 และค่าสัมประสิทธิ์การเดาอยู่
 ระหว่าง .084 - .500 มีค่าเฉลี่ยเท่ากับ .220 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ
 .103 ส่วนกลุ่มชาย ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .453 - 1.811 มีค่า
 เฉลี่ยเท่ากับ .923 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .314 ค่าความยากง่ายอยู่
 ระหว่าง 1.200 - 4.824 มีค่าเฉลี่ยเท่ากับ 2.368 ส่วนเบี่ยงเบนมาตรฐาน
 เท่ากับ .845 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .080 - .462 มีค่าเฉลี่ย
 เท่ากับ .214 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .092 และในกลุ่มหญิง ข้อสอบมีค่า
 อำนาจจำแนกอยู่ระหว่าง 1.467 - .242 มีค่าเฉลี่ยเท่ากับ .820 ส่วนเบี่ยงเบน
 มาตรฐานเท่ากับ .289 ค่าความยากง่ายอยู่ระหว่าง 1.038 - 8.140 ค่าเฉลี่ย
 เท่ากับ 2.671 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.395 และค่าสัมประสิทธิ์การ
 เดาอยู่ระหว่าง .056 - .500 มีค่าเฉลี่ยเท่ากับ .221 ส่วนเบี่ยงเบนมาตรฐาน
 เท่ากับ .105 เมื่อพิจารณาค่าเฉลี่ยของความยากง่ายระหว่างผู้เข้าสอบแต่ละเพศ
 พบว่า ค่าความยากง่ายเฉลี่ยของข้อสอบในกลุ่มหญิงมีค่าสูงกว่าค่าความยากง่ายเฉลี่ย
 ของข้อสอบในกลุ่มชาย ซึ่งสามารถพิจารณาได้จากโค้งลักษณะของแบบทดสอบ และ
 โค้งสารสนเทศที่แสดงในภาพประกอบ 20-21



ภาพประกอบ 20 ได้ังลักษณะของแบบทดสอบวิชาภาษาอังกฤษ กข ฉบับที่ 2



ภาพประกอบ 21 ได้ังสารสนเทศของแบบทดสอบวิชาภาษาอังกฤษ กข ฉบับที่ 2

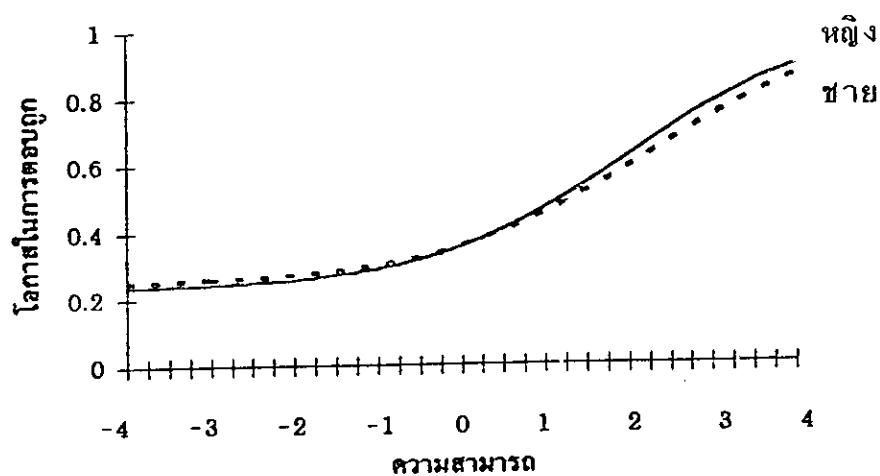
ตาราง 12 ค่าพารามิเตอร์ของแบบทดสอบวิชา ภาษาไทย กข ฉบับที่ 1

ข้อที่	รวม (N = 2687)			ชาย (N = 704)			หญิง (N = 1983)		
	a	b	c	a	b	c	a	b	c
1	.304	.599	.237	.266	1.553	.273	.319	.452	.244
2	.622	4.161	.101	.467	6.366	.126	.604	4.003	.091
3	.273	1.999	.268	.286	1.875	.273	.295	2.097	.288
4	.619	2.379	.237	.805	1.682	.223	.597	2.597	.241
5	.947	1.955	.088	.905	2.152	.076	.947	1.952	.098
6	.524	1.909	.226	.460	2.478	.222	.552	1.857	.241
7	.336	2.700	.163	.402	3.935	.210	.334	2.501	.167
8	.293	2.685	.193	.425	2.722	.257	.288	2.643	.189
9	.506	-.449	.161	.584	-.164	.219	.491	-.461	.172
10	.443	2.530	.257	.588	2.644	.300	.406	2.398	.232
11	.479	4.101	.176	.476	3.924	.174	.483	4.260	.180
12	.366	.069	.263	.411	.210	.250	.347	.038	.272
13	.403	5.214	.460	.401	4.675	.450	.356	4.814	.444
14	.163	-.934	.271	.242	-1.399	.269	.178	-.539	.269
15	.321	.477	.242	.463	.467	.264	.305	.533	.247
16	.395	6.265	.383	.360	4.416	.387	.391	6.175	.367
17	.417	1.333	.167	.478	1.758	.217	.417	1.267	.169
18	.381	2.629	.276	.398	2.781	.286	.374	2.578	.271
19	.230	3.266	.337	.319	2.558	.362	.231	3.013	.315
20	.320	-.875	.221	.332	-.289	.246	.335	-.993	.224
21	.402	5.449	.328	.409	1.551	.302	.401	5.694	.334
22	.432	1.124	.223	.706	1.259	.241	.369	1.093	.225
23	.368	-1.886	.214	.402	-2.002	.246	.388	-1.686	.215

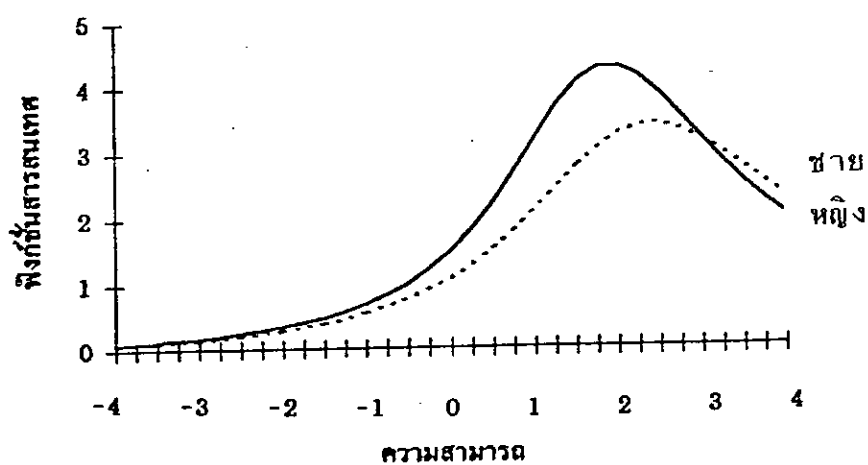
ตาราง 12 (ต่อ)

ข้อที่	รวม (N = 2687)			ชาย (N = 704)			หญิง (N = 1983)		
	a	b	c	a	b	c	a	b	c
24	.346	1.607	.158	.338	2.470	.244	.374	1.476	.156
25	.328	4.301	.282	.348	2.502	.234	.408	4.154	.290
26	.787	1.302	.332	.584	1.469	.304	.692	1.202	.305
27	.732	1.930	.247	.521	3.114	.245	.700	1.804	.243
28	.601	1.841	.147	.442	3.211	.184	.641	1.643	.142
29	.909	1.577	.147	.653	2.657	.158	.915	1.420	.143
30	.683	.255	.191	.491	.755	.221	.739	.167	.192
31	.452	2.571	.422	.301	2.564	.358	.515	2.312	.423
32	.714	1.626	.237	.593	2.312	.237	.653	1.511	.225
33	.593	2.812	.224	.416	3.742	.231	.643	2.672	.223
34	.862	1.581	.083	.629	2.464	.094	.924	1.470	.091
35	.702	2.629	.311	.443	4.521	.300	.615	2.545	.304
36	.893	1.907	.131	.673	2.644	.146	.878	1.823	.125
37	.407	2.783	.203	.400	3.206	.268	.432	2.649	.189
$\bar{X}$	0.506	2.038	0.233	0.470	2.237	0.246	0.501	1.977	0.231
SD	0.204	1.760	0.086	0.149	1.675	0.075	0.202	1.722	0.083

จากตาราง 12 ค่าพารามิเตอร์ของแบบทดสอบวิชา ภาษาไทย กข ฉบับ
ที่ 1 จำนวน 37 ข้อ จากผู้สอบจำนวน 2,687 คน เป็นชาย 704 คน หญิง
1,983 คน เมื่อคำนวณจากจำนวนผู้สอบทั้งหมด ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง
.163 - .947 ซึ่งมีค่าเฉลี่ยเท่ากับ .506 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .204
ค่าความยากง่ายอยู่ระหว่าง -1.886 - 6.265 มีค่าเฉลี่ยเท่ากับ 2.038 ส่วน
เบี่ยงเบนมาตรฐานเท่ากับ 1.760 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .083 -
.460 มีค่าเฉลี่ยเท่ากับ .233 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .086 ส่วนกลุ่ม
ชาย ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .242 - .905 มีค่าเฉลี่ยเท่ากับ .470
ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .149 ค่าความยากง่ายอยู่ระหว่าง -2.002 -
6.366 มีค่าเฉลี่ยเท่ากับ 2.237 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.675 และ
ค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .076 - .450 มีค่าเฉลี่ยเท่ากับ .246 ส่วน
เบี่ยงเบนมาตรฐานเท่ากับ .075 และในกลุ่มหญิง ข้อสอบมีค่าอำนาจจำแนกอยู่
ระหว่าง .178 - .947 มีค่าเฉลี่ยเท่ากับ .501 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ
.202 ค่าความยากง่ายอยู่ระหว่าง -1.686 - 6.175 ค่าเฉลี่ยเท่ากับ 1.977
ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.722 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง
.091 - .444 มีค่าเฉลี่ยเท่ากับ .231 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .083
เมื่อพิจารณาค่าเฉลี่ยของความยากง่ายระหว่างผู้เข้าสอบแต่ละเพศ พบว่า ค่าความ
ยากง่ายเฉลี่ยของข้อสอบในกลุ่มชาย มีค่าสูงกว่าค่าความยากง่ายเฉลี่ยของข้อสอบ
ในกลุ่มหญิง ซึ่งสามารถพิจารณาได้จากโค้งลักษณะของแบบทดสอบ และโค้งสาร
สนเทศที่แสดงในภาพประกอบ 22-23



ภาพประกอบ 22 ได้ังลักษณะของแบบทดสอบวิชาภาษาไทย กข ฉบับที่ 1



ภาพประกอบ 23 ได้ังสารสนเทศของแบบทดสอบวิชาภาษาไทย กข ฉบับที่ 1

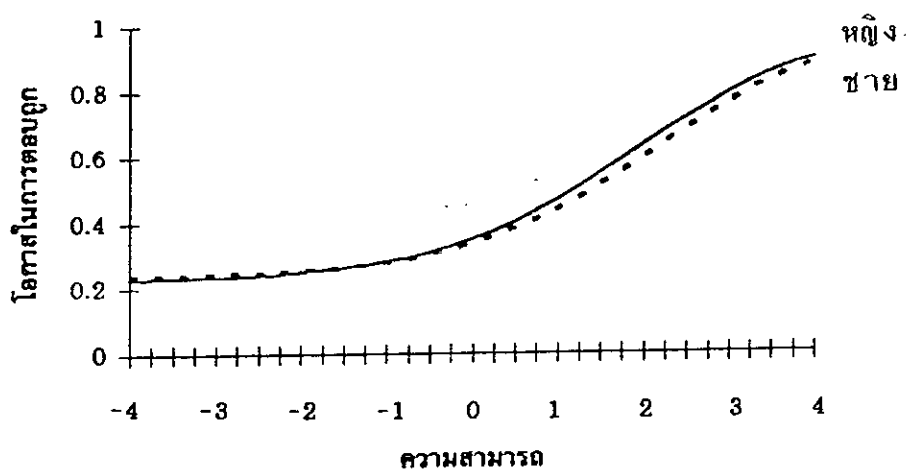
ตาราง 13 ค่าพารามิเตอร์ของแบบทดสอบวิชา ภาษาไทย กข ฉบับที่ 2

ข้อที่	รวม (N = 2723)			ชาย (N = 749)			หญิง (N = 1974)		
	a	b	c	a	b	c	a	b	c
1	.351	1.576	.157	.301	3.214	.270	.417	1.311	.154
2	.381	5.280	.352	.427	4.846	.373	.384	5.187	.343
3	.559	1.379	.358	.369	1.294	.270	.562	1.277	.351
4	.887	1.836	.233	.655	2.442	.208	.851	1.747	.236
5	.608	1.940	.154	.548	2.180	.147	.641	1.913	.168
6	.621	1.653	.103	.781	2.240	.132	.597	1.517	.107
7	.824	.531	.303	.777	.540	.269	.744	.424	.282
8	.799	2.082	.465	.308	2.140	.376	.691	2.035	.441
9	.743	1.469	.201	.724	1.765	.198	.676	1.381	.191
10	.828	2.614	.276	.493	3.213	.285	.756	2.641	.258
11	1.005	1.674	.087	1.313	1.716	.098	.855	1.715	.083
12	.390	3.669	.306	.398	4.180	.321	.393	3.437	.298
13	.985	1.728	.148	1.133	1.638	.145	.850	1.807	.143
14	.383	3.569	.247	.440	3.893	.304	.369	3.212	.210
15	.518	2.806	.151	.525	2.956	.203	.502	2.723	.125
16	.445	.068	.282	.327	.280	.287	.473	-.108	.249
17	.210	-1.398	.257	.296	-1.631	.251	.219	-1.060	.259
18	.409	.031	.172	.499	.105	.189	.405	.132	.201
19	.462	1.199	.161	.542	1.620	.197	.456	1.142	.173
20	.435	6.777	.295	.434	4.723	.282	.435	7.586	.295
21	.277	3.051	.277	.313	3.132	.268	.317	2.983	.311
22	.256	1.990	.281	.354	2.033	.319	.242	1.962	.278
23	.897	3.065	.091	.590	3.842	.118	.896	3.003	.077

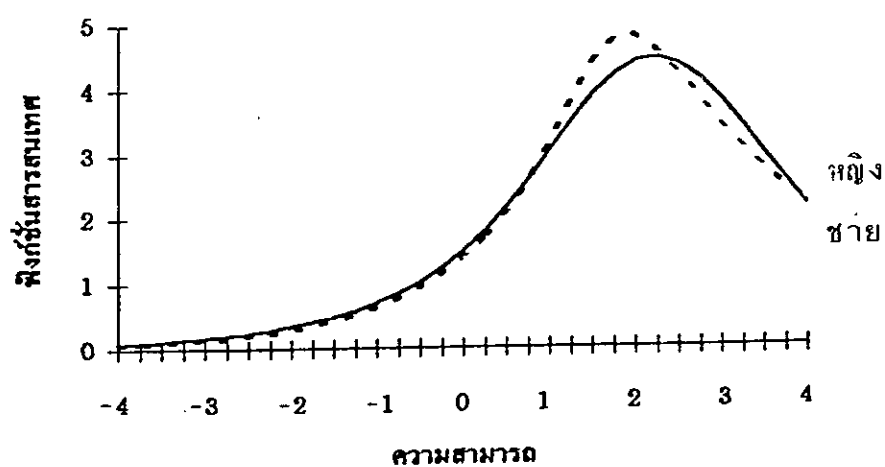
ตาราง 13 (ต่อ)

ข้อที่	รวม (N = 2723)			ชาย (N = 749)			หญิง (N = 1974)		
	a	b	c	a	b	c	a	b	c
24	.313	-1.201	.226	.332	-.523	.235	.309	-1.409	.240
25	.413	5.552	.330	.396	4.637	.309	.421	5.783	.333
26	.374	.841	.173	.731	1.203	.218	.305	.811	.196
27	.331	-2.036	.219	.428	-1.458	.231	.310	-2.205	.229
28	.346	.629	.177	.578	.597	.179	.292	.902	.223
29	.264	2.763	.284	.317	1.901	.259	.290	2.950	.308
30	.797	2.803	.281	.372	3.537	.304	.839	2.598	.249
31	.832	1.931	.077	.784	2.513	.089	.806	1.842	.077
32	.558	2.105	.213	.649	1.898	.190	.526	2.293	.232
33	.304	2.844	.152	.363	3.724	.198	.322	2.708	.172
34	.411	1.749	.169	.547	2.140	.226	.412	1.645	.171
35	.546	-.141	.156	.654	.239	.193	.517	-.227	.170
36	.304	2.080	.239	.333	2.732	.316	.332	1.826	.225
37	.557	2.709	.259	.472	3.343	.262	.580	2.550	.257
$\bar{X}$	.538	1.924	.224	.527	2.130	.235	.520	1.892	.224
SD	.224	1.783	.085	.226	1.614	.071	.197	1.850	.081

จากตาราง 13 ค่าพารามิเตอร์ของแบบทดสอบวิชา ภาษาไทย กข ฉบับ
 ที่ 2 จำนวน 37 ข้อ จากผู้สอบจำนวน 2,723 คน เป็นชาย 749 คน หญิง
 1,974 คน เมื่อคำนวณจากจำนวนผู้สอบทั้งหมด ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง
 .210 - 1.005 ซึ่งมีค่าเฉลี่ยเท่ากับ .538 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .224
 ค่าความยากง่ายอยู่ระหว่าง -2.036 - 5.280 มีค่าเฉลี่ยเท่ากับ 1.924 ส่วน
 เบี่ยงเบนมาตรฐานเท่ากับ 1.783 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .077 -
 .465 มีค่าเฉลี่ยเท่ากับ .224 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .085 ส่วนกลุ่มชาย
 ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .296 - 1.313 มีค่าเฉลี่ยเท่ากับ .527 ส่วน
 เบี่ยงเบนมาตรฐานเท่ากับ .226 ค่าความยากง่ายอยู่ระหว่าง -1.631 - 4.846
 มีค่าเฉลี่ยเท่ากับ 2.130 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.614 และค่าสัมประสิทธิ์
 การเดาอยู่ระหว่าง .089 - .376 มีค่าเฉลี่ยเท่ากับ .235 ส่วนเบี่ยงเบนมาตรฐาน
 เท่ากับ .071 และในกลุ่มหญิง ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .242 - .855
 มีค่าเฉลี่ยเท่ากับ .520 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .197 ค่าความยากง่ายอยู่
 ระหว่าง -2.205 - 7.586 ค่าเฉลี่ยเท่ากับ 1.892 ส่วนเบี่ยงเบนมาตรฐาน
 เท่ากับ 1.850 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .077 - .441 มีค่าเฉลี่ย
 เท่ากับ .224 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .081 เมื่อพิจารณาค่าเฉลี่ยของ
 ความยากง่ายระหว่างผู้เข้าสอบแต่ละเพศ พบว่า ค่าความยากง่ายเฉลี่ยของข้อสอบ
 ในกลุ่มชายมีค่าสูงกว่าค่าความยากง่ายเฉลี่ยของข้อสอบในกลุ่มหญิง ซึ่งสามารถ
 พิจารณาได้จากโค้งลักษณะของแบบทดสอบ และโค้งสารสนเทศที่แสดงในภาพประกอบ
 24-25



ภาพประกอบ 24 ได้ังลักษณะของแบบทดสอบวิชาภาษาไทย กข ฉบับที่ 2



ภาพประกอบ 25 ได้ังสารสนเทศของแบบทดสอบวิชาภาษาไทย กข ฉบับที่ 2

ตาราง 14 ค่าพารามิเตอร์ของแบบทดสอบวิชา สังคมศึกษา กข ฉบับที่ 1

ข้อที่	รวม (N = 2685)			ชาย (N = 715)			หญิง (N = 1970)		
	a	b	c	a	b	c	a	b	c
1	.209	.868	.188	.411	.628	.252	.608	1.442	.469
2	.363	.220	.125	.484	2.955	.166	.448	2.199	.078
3	.352	-1.723	.188	.346	1.706	.194	.334	1.872	.197
4	.517	-.910	.283	.726	-2.338	.196	.597	-2.719	.195
5	.393	8.949	.052	.117	-.868	.253	.071	-2.179	.287
6	.401	5.931	.085	.195	-3.624	.205	.138	-4.100	.229
7	.246	3.772	.292	.198	2.718	.341	.201	4.141	.500
8	.242	4.079	.177	.503	3.499	.121	.672	2.323	.082
9	.358	-3.529	.211	1.033	1.330	.100	.774	1.053	.084
10	.575	.064	.087	.249	-4.860	.203	.392	-2.576	.197
11	.489	3.775	.193	.362	3.320	.245	.231	3.547	.250
12	.568	2.780	.168	.674	3.265	.116	.598	3.101	.154
13	.560	.053	.141	.422	7.716	.113	.482	6.269	.094
14	.569	2.834	.103	.340	3.946	.362	.496	4.851	.409
15	.281	4.864	.202	.719	.881	.251	.268	1.933	.020
16	.811	.122	.039	1.069	1.855	.104	.869	1.575	.096
17	.379	3.814	.319	.446	3.817	.132	.644	3.807	.115
18	.262	2.062	.212	.426	6.054	.141	.396	7.243	.209
19	1.033	2.048	.345	.598	1.756	.178	.503	1.984	.123
20	.635	-.031	.127	.255	2.615	.358	.212	4.435	.398
21	.185	4.179	.291	.509	3.911	.136	.543	4.443	.139
22	.373	-.789	.139	.674	1.437	.162	.597	1.435	.186
23	.347	.074	.098	.366	2.747	.363	.473	4.650	.384

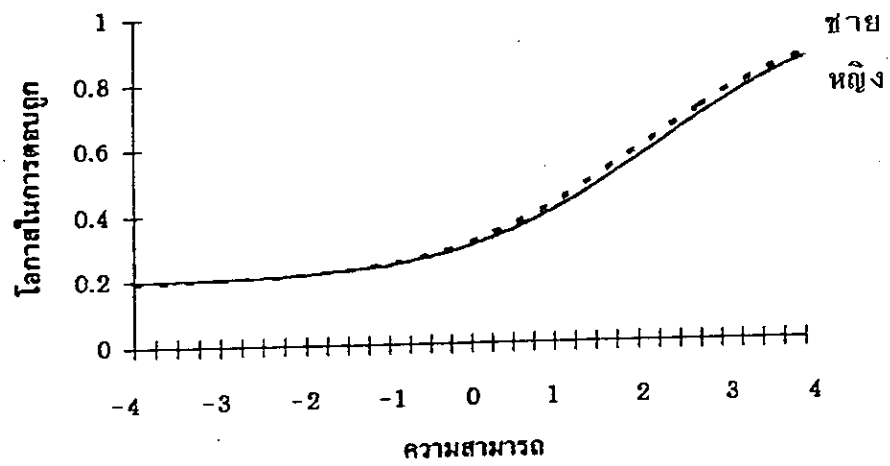
ตาราง 14 (ต่อ)

ข้อที่	รวม (N = 2685)			ชาย (N = 715)			หญิง (N = 1970)		
	a	b	c	a	b	c	a	b	c
24	.148	7.295	.500	.516	.442	.179	.505	-.005	.109
25	.414	2.695	.209	.316	3.050	.200	.319	2.761	.166
26	.423	2.606	.098	.508	5.097	.058	.454	7.422	.051
27	.520	1.530	.256	.170	-1.739	.223	.122	-3.476	.232
28	.776	1.152	.261	.195	1.457	.260	.253	5.996	.500
29	.651	-2.393	.198	.206	1.473	.246	.291	1.792	.292
30	.823	-.918	.152	.472	2.802	.285	.341	4.652	.358
31	.434	-1.638	.188	.517	5.616	.040	.854	3.946	.021
32	.744	1.896	.492	.828	1.442	.175	.685	1.319	.165
33	.084	-1.646	.278	.482	.685	.146	.388	1.020	.128
34	.258	3.484	.473	.411	3.302	.267	.340	2.071	.198
35	.799	2.041	.408	.308	4.082	.239	.316	2.155	.208
36	.765	2.938	.052	.619	2.719	.229	.430	3.956	.224
37	.222	-1.036	.298	.444	-.794	.196	.531	-1.309	.184
38	.306	3.690	.128	.346	3.478	.215	.283	4.176	.202
39	.602	-1.409	.132	.443	-1.673	.193	.326	-2.443	.191
40	.358	-1.561	.154	.289	.843	.232	.197	.924	.239
41	.551	-.631	.285	.553	3.535	.194	1.036	2.491	.218
42	.698	1.001	.228	.335	1.607	.214	.321	.787	.160
43	.425	4.410	.142	.403	3.524	.202	.405	5.863	.234
44	.225	3.631	.146	.321	3.788	.185	.305	3.569	.178
45	.629	2.637	.180	1.174	2.636	.144	.383	5.150	.136
46	.596	.469	.120	.622	3.384	.106	.926	3.107	.083

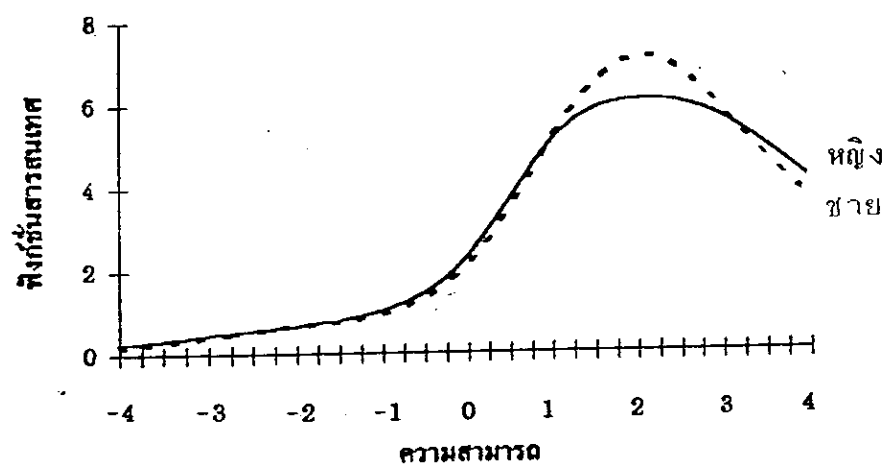
ตาราง 14 (ต่อ)

ข้อที่	รวม (N = 2685)			ชาย (N = 715)			หญิง (N = 1970)		
	a	b	c	a	b	c	a	b	c
47	.159	1.204	.208	.793	1.450	.120	.835	.987	.097
48	.162	2.041	.256	.483	1.963	.200	.516	3.121	.325
49	.556	.438	.268	1.204	2.211	.050	.952	2.358	.036
50	.296	4.745	.378	.359	1.644	.178	.373	1.514	.193
51	.396	-1.657	.154	.350	.611	.212	.372	.599	.187
52	.223	4.773	.319	1.035	.989	.136	1.182	.891	.103
53	.291	1.414	.302	.668	1.417	.246	.872	1.407	.195
54	.398	3.190	.202	.527	2.159	.196	.599	1.856	.126
55	.406	2.693	.084	.350	5.136	.208	.551	3.869	.275
56	.343	4.546	.355	.923	1.307	.110	1.001	1.234	.128
57	.383	3.953	.081	.400	4.273	.178	.475	3.340	.147
58	.820	2.376	.232	.652	3.028	.093	.830	3.100	.121
59	.303	1.236	.178	.796	1.051	.171	.619	1.437	.149
$\bar{X}$	.436	1.774	.211	.505	2.075	.191	.503	2.243	.194
SD	.207	2.493	.107	.252	2.211	.071	.248	2.480	.110

จากตาราง 14 ค่าพารามิเตอร์ของแบบทดสอบวิชา สังคมศึกษา กข ฉบับที่ 1 จำนวน 59 ข้อ จากผู้สอบจำนวน 2,685 คน เป็นชาย 715 คน หญิง 1,970 คน เมื่อคำนวณจากจำนวนผู้สอบทั้งหมด ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .084 - 1.033 ซึ่งมีค่าเฉลี่ยเท่ากับ .436 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .207 ค่าความยากง่ายอยู่ระหว่าง -3.529 - 8.949 มีค่าเฉลี่ยเท่ากับ 1.774 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.493 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .039 - .500 มีค่าเฉลี่ยเท่ากับ .211 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .107 ส่วนกลุ่มชาย ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .117 - 1.033 มีค่าเฉลี่ยเท่ากับ .505 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .252 ค่าความยากง่ายอยู่ระหว่าง -4.860 - 7.716 มีค่าเฉลี่ยเท่ากับ 2.075 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.211 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .040 - .363 มีค่าเฉลี่ยเท่ากับ .191 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .071 และในกลุ่มหญิง ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .071 - 1.182 มีค่าเฉลี่ยเท่ากับ .503 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .248 ค่าความยากง่ายอยู่ระหว่าง -4.100 - 7.422 ค่าเฉลี่ยเท่ากับ 2.243 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.480 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .020 - .500 มีค่าเฉลี่ยเท่ากับ .194 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .110 เมื่อพิจารณาค่าเฉลี่ยของความยากง่ายระหว่างผู้เข้าสอบแต่ละเพศ พบว่าค่าความยากง่ายเฉลี่ยของข้อสอบในกลุ่มหญิงมีค่าสูงกว่าค่าความยากง่ายเฉลี่ยของข้อสอบในกลุ่มชาย ซึ่งสามารถพิจารณาได้จากโค้งลักษณะของแบบทดสอบ และโค้งสารสนเทศที่แสดงในภาพประกอบ 26-27



ภาพประกอบ 26 ได้ังลักษณะของแบบทดสอบวิชาสังคมศึกษา กข ฉบับที่ 1



ภาพประกอบ 27 ได้ังสารสนเทศของแบบทดสอบวิชาสังคมศึกษา กข ฉบับที่ 1

ตาราง 15 ค่าพารามิเตอร์ของแบบทดสอบวิชา สังคมศึกษา กข ฉบับที่ 2

ข้อที่	รวม (N = 2730)			ชาย (N = 739)			หญิง (N = 1991)		
	a	b	c	a	b	c	a	b	c
1	.482	4.903	.378	.386	3.780	.355	.395	6.687	.375
2	.556	.116	.099	.587	.280	.117	.555	.113	.118
3	.336	3.110	.192	.339	3.595	.247	.363	2.867	.181
4	.564	5.432	.041	.515	5.426	.053	.511	6.223	.039
5	.168	-2.404	.208	.193	-1.806	.212	.179	-2.373	.211
6	.290	5.100	.491	.255	3.437	.397	.176	4.230	.407
7	.506	5.117	.086	.452	6.910	.116	.474	5.203	.077
8	.237	1.208	.244	.230	.908	.233	.257	1.149	.239
9	.799	1.117	.157	.679	1.381	.135	.848	1.038	.169
10	.424	1.113	.168	.551	.662	.178	.377	1.395	.173
11	.378	2.261	.200	.494	1.870	.186	.341	2.425	.205
12	.356	3.705	.324	.406	6.500	.344	.332	2.987	.292
13	.539	3.261	.224	.685	2.385	.236	.410	4.094	.211
14	.497	1.109	.168	.558	-.673	.165	.480	-1.268	.186
15	.493	3.142	.245	.482	2.740	.229	.422	3.552	.240
16	.460	-1.611	.184	.524	-1.507	.178	.437	-1.663	.190
17	.232	.730	.226	.300	.749	.217	.224	.632	.226
18	1.255	2.054	.203	.533	2.617	.158	1.632	1.945	.208
19	.413	.983	.188	.366	1.289	.218	.436	.843	.173
20	.479	3.741	.212	.382	4.345	.240	.519	3.617	.204
21	.324	2.689	.118	.339	2.601	.163	.347	2.806	.129
22	.478	4.165	.129	.407	4.638	.115	.525	4.024	.139
23	.956	3.229	.080	.815	2.901	.090	.687	4.398	.075

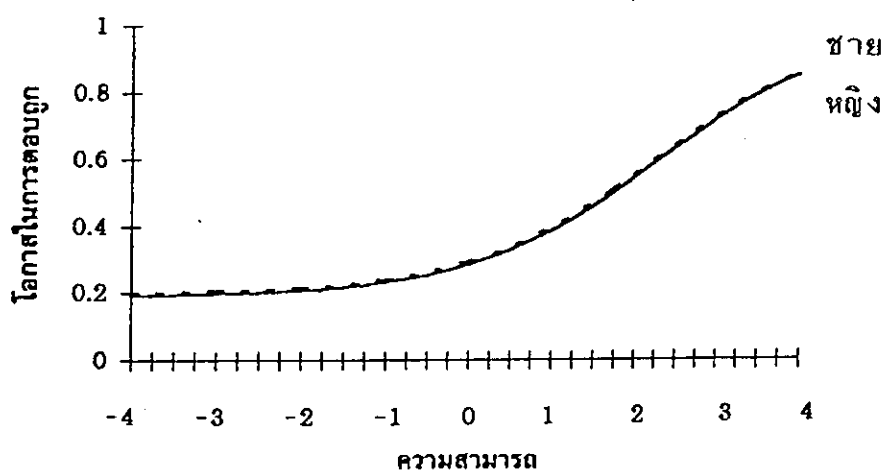
ตาราง 15 (ต่อ)

ข้อที่	รวม (N = 2730)			ชาย (N = 739)			หญิง (N = 1991)		
	a	b	c	a	b	c	a	b	c
24	.826	1.235	.118	.924	1.221	.143	.765	1.235	.105
25	.554	2.808	.308	.567	2.780	.269	.404	3.032	.294
26	.896	2.656	.030	1.036	2.483	.036	.800	2.840	.030
27	.281	1.764	.171	.324	1.976	.246	.284	1.708	.157
28	.348	.300	.168	.289	.420	.223	.375	.296	.157
29	.910	.833	.067	.974	.588	.091	.918	.950	.073
30	.750	1.456	.247	.521	1.601	.268	.732	1.390	.215
31	.440	2.099	.123	.349	2.606	.157	.512	2.020	.134
32	.635	3.203	.227	.865	2.588	.195	.404	4.020	.220
33	.892	1.368	.100	.616	1.709	.109	1.049	1.292	.105
34	.400	4.198	.165	.368	4.383	.164	.446	3.962	.170
35	.574	4.562	.121	.504	4.650	.130	.463	5.576	.119
36	.913	1.315	.197	.630	.934	.109	1.155	1.415	.226
37	.401	7.498	.196	.432	8.111	.248	.379	8.137	.179
38	.204	3.945	.357	.385	4.355	.379	.176	2.281	.272
39	.664	1.589	.463	.418	.553	.258	.483	1.794	.448
40	.494	2.586	.147	.523	3.013	.181	.462	2.487	.130
41	.338	2.056	.216	.361	1.711	.208	.334	2.185	.221
42	.764	-2.183	.171	.975	-2.047	.185	.685	-2.284	.175
43	.107	5.080	.500	.132	.910	.269	.084	-1.008	.271
44	.119	-3.790	.230	.204	-2.300	.214	.116	-3.952	.228
45	.218	5.454	.500	.401	5.241	.483	.227	4.612	.500
46	.842	2.218	.108	.695	2.656	.117	.807	2.178	.102

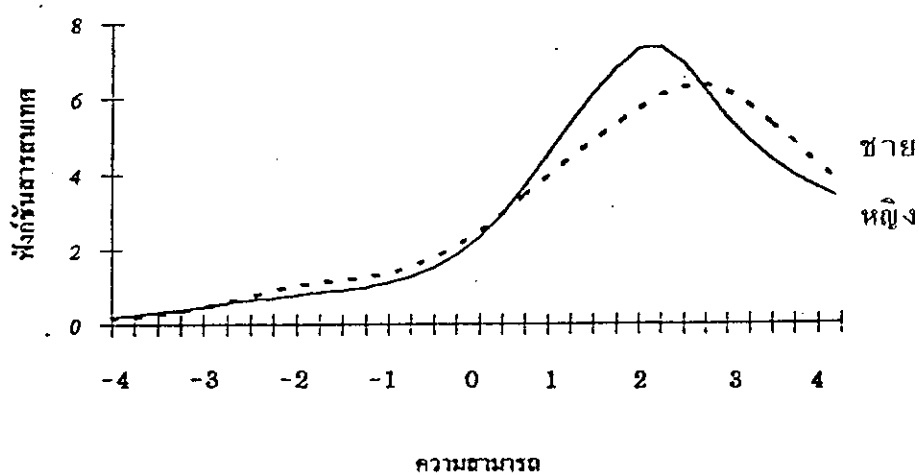
ตาราง 15 (ต่อ)

ข้อที่	รวม (N = 2730)			ชาย (N = 739)			หญิง (N = 1991)		
	a	b	c	a	b	c	a	b	c
47	.835	1.462	.108	.866	1.847	.131	.838	1.320	.099
48	.434	-2.161	.182	.420	-2.549	.195	.442	-2.024	.184
49	.512	6.290	.090	.452	6.596	.106	.381	9.357	.089
50	.269	2.865	.182	.402	2.227	.953	.246	3.207	.210
51	1.000	2.406	.149	.991	2.535	.124	.849	2.484	.153
52	.524	3.182	.387	.425	3.552	.414	.410	3.325	.351
53	.846	1.051	.233	.799	.836	.251	.776	1.086	.206
54	.782	1.836	.101	.634	2.224	.099	.829	1.758	.106
55	.621	3.660	.128	.590	3.333	.143	.528	4.238	.119
56	.615	2.222	.153	.560	1.751	.124	.696	2.356	.169
57	.252	1.359	.186	.238	2.139	.246	.282	1.181	.186
58	.460	4.741	.143	.566	3.523	.117	.412	5.599	.154
59	.608	1.430	.157	.514	1.774	.173	.648	1.323	.154
$\bar{X}$	0.534	2.319	0.198	0.515	2.287	0.195	0.513	2.310	0.189
SD	0.248	2.153	0.108	0.214	2.197	0.089	0.276	2.469	0.091

จากตาราง 15 ค่าพารามิเตอร์ของแบบทดสอบวิชา สังคมศึกษา กข ฉบับที่ 2 จำนวน 59 ข้อ จากผู้สอบจำนวน 2,730 คน เป็นชาย 739 คน หญิง 1,991 คน เมื่อคำนวณจากจำนวนผู้สอบทั้งหมด ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .107 - 1.255 ซึ่งมีค่าเฉลี่ยเท่ากับ .534 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .248 ค่าความยากง่ายอยู่ระหว่าง -3.790 - 7.498 มีค่าเฉลี่ยเท่ากับ 2.319 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.153 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .030 - .500 มีค่าเฉลี่ยเท่ากับ .198 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .108 ส่วนกลุ่มชาย ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .132 - 1.036 มีค่าเฉลี่ยเท่ากับ .515 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .214 ค่าความยากง่ายอยู่ระหว่าง -2.549 - 8.111 มีค่าเฉลี่ยเท่ากับ 2.287 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.197 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .036 - .483 มีค่าเฉลี่ยเท่ากับ .195 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .089 และในกลุ่มหญิง ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .084 - 1.155 มีค่าเฉลี่ยเท่ากับ .513 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .276 ค่าความยากง่ายอยู่ระหว่าง -3.952 - 6.687 ค่าเฉลี่ยเท่ากับ 2.310 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 2.469 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .030 - .500 มีค่าเฉลี่ยเท่ากับ .189 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .091 เมื่อพิจารณาค่าเฉลี่ยของความยากง่ายระหว่างผู้เข้าสอบแต่ละเพศ พบว่าค่าความยากง่ายเฉลี่ยของข้อสอบในกลุ่มหญิงมีค่าสูงกว่าค่าความยากง่ายเฉลี่ยของข้อสอบในกลุ่มชาย ซึ่งสามารถพิจารณาได้จากโค้งลักษณะของแบบทดสอบและโค้งสารสนเทศที่แสดงในภาพประกอบ 28-29



ภาพประกอบ 28 โค้งลักษณะของแบบทดสอบวิชาสังคมศึกษา กข ฉบับที่ 2

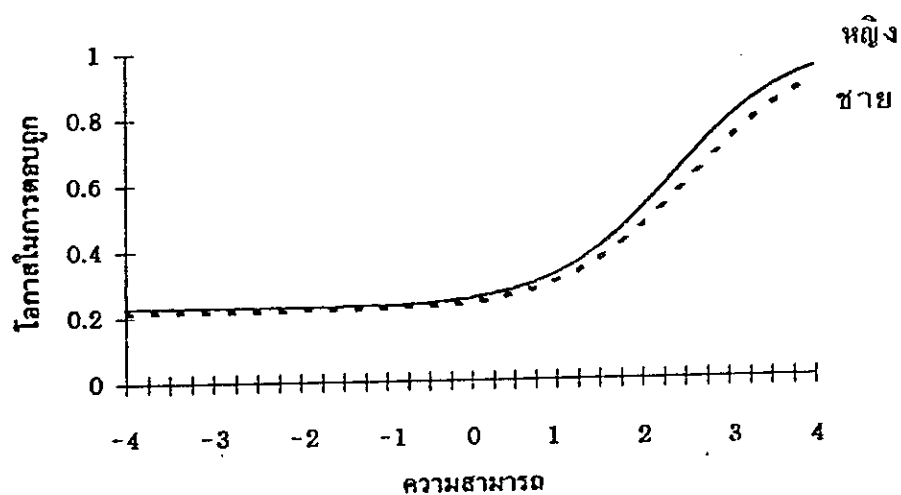


ภาพประกอบ 29 โค้งสารสนเทศของแบบทดสอบวิชาสังคมศึกษา กข ฉบับที่ 2

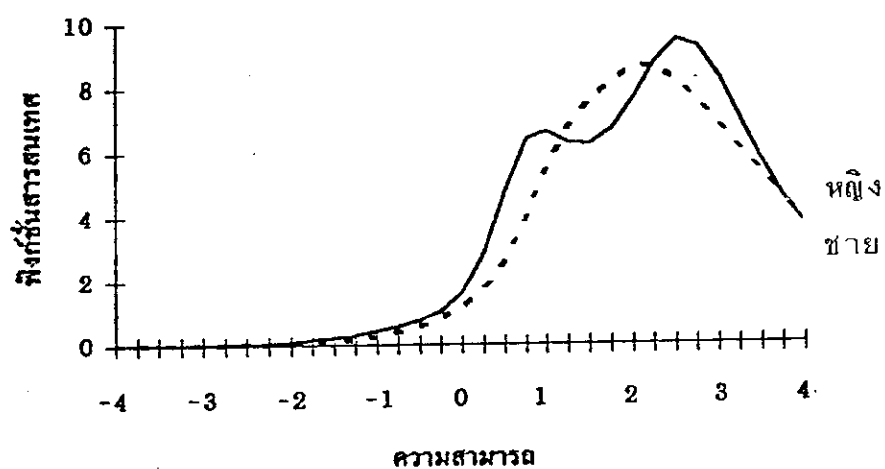
ตาราง 16 ค่าพารามิเตอร์ของแบบทดสอบวิชา ภาษาอังกฤษ กชค ฉบับที่ 1

ข้อที่	รวม (N = 2719)			ชาย (N = 722)			หญิง (N = 1997)		
	a	b	c	a	b	c	a	b	c
1	.678	1.577	.123	.692	1.901	.144	.681	1.474	.117
2	.498	1.822	.276	.469	2.064	.290	.473	1.598	.241
3	2.233	.822	.090	1.812	1.125	.096	2.268	.727	.085
4	1.042	1.800	.262	.937	2.028	.253	1.003	1.746	.261
5	.681	2.842	.293	.630	3.016	.257	.671	2.837	.304
6	.821	2.763	.186	.860	3.230	.215	.707	2.703	.166
7	.975	1.457	.168	.854	1.547	.131	.987	1.441	.182
8	.947	2.875	.227	.743	2.873	.213	1.015	2.801	.228
9	.776	2.210	.248	.735	3.158	.282	.737	1.966	.224
10	1.564	2.855	.244	.624	3.658	.227	1.362	2.993	.244
11	.840	4.285	.218	.517	7.505	.238	.782	4.300	.211
12	.765	4.753	.129	.630	6.254	.135	.694	5.028	.128
13	.485	2.734	.185	.511	3.143	.167	.555	2.635	.219
14	.776	2.839	.224	.883	2.896	.226	.684	2.869	.218
15	.437	1.920	.204	.583	1.642	.189	.421	2.117	.231
16	.539	4.785	.366	.477	4.466	.328	.615	4.705	.378
17	.605	2.246	.097	1.021	2.050	.098	.537	2.390	.111
18	.547	3.170	.298	.541	2.250	.208	.828	3.117	.338
19	.950	2.273	.247	.611	4.040	.284	1.031	2.026	.234
20	1.217	1.113	.185	.927	1.407	.167	1.286	1.021	.189
21	.602	1.747	.211	.586	2.147	.219	.556	1.623	.195
22	.633	.991	.207	.591	1.114	.195	.685	1.013	.234
23	.747	.062	.156	.840	.413	.156	.701	-.042	.177

จากตาราง 16 ค่าพารามิเตอร์ของแบบทดสอบวิชา ภาษาอังกฤษ กชค
ฉบับที่ 1 จำนวน 37 ข้อ จากผู้สอบจำนวน 2,719 คน เป็นชาย 722 คน หญิง
1,997 คน เมื่อดำเนินการจากจำนวนผู้สอบทั้งหมด ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง
.221 - 2.233 ซึ่งมีค่าเฉลี่ยเท่ากับ .877 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .409
ค่าความยากง่ายอยู่ระหว่าง .020 - 4.785 มีค่าเฉลี่ยเท่ากับ 2.336 ส่วน
เบี่ยงเบนมาตรฐานเท่ากับ 1.072 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .089
- .344 มีค่าเฉลี่ยเท่ากับ .461 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .283 ส่วนกลุ่ม
ชาย ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .407 - 1.812 มีค่าเฉลี่ยเท่ากับ .795
ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .337 ค่าความยากง่ายอยู่ระหว่าง .413 - 7.505
มีค่าเฉลี่ยเท่ากับ 2.582 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.396 และค่าสัมประสิทธิ์
การเดาอยู่ระหว่าง .096 - .298 มีค่าเฉลี่ยเท่ากับ .213 ส่วนเบี่ยงเบน
มาตรฐานเท่ากับ .057 และในกลุ่มหญิง ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .410
- 1.568 มีค่าเฉลี่ยเท่ากับ .856 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .396 ค่าความ
ยากง่ายอยู่ระหว่าง -.171 - 5.028 ค่าเฉลี่ยเท่ากับ 2.357 ส่วนเบี่ยงเบน
มาตรฐานเท่ากับ 1.176 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .080 - .470
มีค่าเฉลี่ยเท่ากับ .224 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .081 เมื่อพิจารณาค่า
เฉลี่ยของความยากง่ายระหว่างผู้เข้าสอบแต่ละเพศ พบว่าค่าความยากง่ายเฉลี่ย
ของข้อสอบในกลุ่มชายมีค่าสูงกว่าค่าความยากง่ายเฉลี่ยของข้อสอบในกลุ่มหญิง ซึ่ง
สามารถพิจารณาได้จากโค้งลักษณะของแบบทดสอบ และโค้งสารสนเทศที่แสดงใน
ภาพประกอบ 30-31



ภาพประกอบ 30 ได้ังลักษณะของแบบทดสอบวิชาภาษาอังกฤษ กชค ฉบับที่ 1



ภาพประกอบ 31 ได้ังสารสนเทศของแบบทดสอบวิชาภาษาอังกฤษ กชค ฉบับที่ 1

ตาราง 17 ค่าพารามิเตอร์ของแบบทดสอบวิชา ภาษาอังกฤษ กชค ฉบับที่ 2

ข้อที่	รวม (N = 2701)			ชาย (N = 736)			หญิง (N = 1965)		
	a	b	c	a	b	c	a	b	c
1	.624	3.356	.207	.600	4.094	.234	.548	3.348	.197
2	.948	1.472	.135	1.126	1.560	.131	.908	1.389	.158
3	.637	1.410	.104	.602	2.232	.130	.633	1.216	.135
4	.583	1.070	.229	.708	1.341	.254	.496	.906	.236
5	1.651	.825	.076	1.705	1.218	.078	1.594	.608	.110
6	1.021	1.940	.268	.740	2.351	.233	.910	1.872	.279
7	1.519	2.193	.321	.638	3.956	.306	1.471	2.033	.325
8	.750	3.416	.220	.778	3.105	.207	.712	3.581	.225
9	.628	3.605	.205	.612	4.788	.189	.604	3.563	.219
10	.786	2.330	.255	.511	2.745	.233	.792	2.209	.257
11	.862	2.923	.256	.684	3.982	.277	.777	2.793	.245
12	.438	1.932	.183	.540	2.331	.195	.374	1.979	.209
13	.541	2.604	.193	.696	3.005	.229	.546	2.368	.197
14	.691	4.533	.377	.489	3.356	.332	.640	3.532	.233
15	.986	2.037	.137	.980	2.319	.152	.856	1.988	.134
16	.367	3.183	.253	.543	1.618	.176	.518	3.471	.318
17	1.113	2.085	.253	.787	2.661	.255	.980	1.949	.246
18	1.042	1.069	.153	.867	1.418	.153	1.030	.863	.166
19	.530	1.221	.113	.675	1.385	.127	.467	1.166	.145
20	.731	5.682	.191	.616	6.311	.205	.659	7.008	.186
21	.598	.759	.146	.907	.904	.165	.502	.608	.173
22	.734	-.037	.155	.624	.710	.210	.671	-.434	.194
23	.856	2.408	.196	.822	2.400	.194	.772	2.428	.196

ตาราง 17 (ต่อ)

ข้อที่	รวม (N = 2701)			ชาย (N = 736)			หญิง (N = 1965)		
	a	b	c	a	b	c	a	b	c
24	.245	2.527	.318	.410	3.114	.380	.208	2.787	.342
25	2.153	2.395	.091	.803	4.435	.091	1.986	2.261	.092
26	.461	1.702	.263	.548	2.689	.340	.412	1.469	.256
27	.673	-.166	.163	.763	.479	.187	.536	-.631	.211
28	.786	3.511	.282	.690	4.078	.303	.705	3.270	.267
29	1.436	2.535	.221	.630	3.701	.190	1.501	2.480	.230
30	1.447	2.606	.241	.655	5.511	.248	1.177	2.545	.237
31	1.154	2.026	.147	.890	2.583	.125	1.049	1.968	.161
32	.617	2.790	.333	.598	4.253	.351	.506	2.569	.315
33	1.230	1.522	.228	1.192	1.651	.238	1.155	1.385	.240
34	1.682	2.385	.257	.813	2.812	.240	1.778	2.324	.258
35	1.115	2.995	.228	.617	7.015	.251	.900	3.039	.219
36	1.374	2.078	.300	.534	3.179	.248	1.455	1.868	.309
37	.533	2.514	.302	.543	2.816	.286	.511	2.520	.322
$\bar{X}$	0.906	2.254	0.216	0.728	2.922	0.220	0.847	2.170	0.223
SD	0.424	1.147	0.072	0.236	1.496	0.072	0.418	1.317	0.062

จากตาราง 17 ค่าพารามิเตอร์ของแบบทดสอบวิชา ภาษาอังกฤษ กชค
ฉบับที่ 2 จำนวน 37 ข้อ จากผู้สอบจำนวน 2,701 คน เป็นชาย 736 คน หญิง
1,965 คน เมื่อคำนวณจากจำนวนผู้สอบทั้งหมด ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง
.245 - 2.153 ซึ่งมีค่าเฉลี่ยเท่ากับ .906 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ
.424 ค่าความยากง่ายอยู่ระหว่าง -.166 - 5.682 มีค่าเฉลี่ยเท่ากับ 2.254
ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.147 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .076
- .377 มีค่าเฉลี่ยเท่ากับ .216 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .072 ส่วนกลุ่ม
ชาย ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .410 - 1.705 มีค่าเฉลี่ยเท่ากับ .728
ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .236 ค่าความยากง่ายอยู่ระหว่าง .479 - 7.015
มีค่าเฉลี่ยเท่ากับ 2.922 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ 1.496 และค่าสัมประสิทธิ์
การเดาอยู่ระหว่าง .078 - .351 มีค่าเฉลี่ยเท่ากับ .220 ส่วนเบี่ยงเบน
มาตรฐานเท่ากับ .072 และในกลุ่มหญิง ข้อสอบมีค่าอำนาจจำแนกอยู่ระหว่าง .208
- 1.778 มีค่าเฉลี่ยเท่ากับ .847 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .418 ค่าความ
ยากง่ายอยู่ระหว่าง -.631 - 7.008 ค่าเฉลี่ยเท่ากับ 2.170 ส่วนเบี่ยงเบน
มาตรฐานเท่ากับ 1.317 และค่าสัมประสิทธิ์การเดาอยู่ระหว่าง .092 - .342 มี
ค่าเฉลี่ยเท่ากับ .223 ส่วนเบี่ยงเบนมาตรฐานเท่ากับ .062 เมื่อพิจารณาค่าเฉลี่ย
ของความยากง่ายระหว่างผู้เข้าสอบแต่ละเพศ พบว่าค่าความยากง่ายของข้อสอบใน
กลุ่มชาย มีค่าสูงกว่าค่าความยากง่ายเฉลี่ยของข้อสอบในกลุ่มหญิง ซึ่งสามารถ
พิจารณาได้จากโค้งลักษณะของแบบทดสอบ และได้สังสรสนเทศที่แสดงในภาพ
ประกอบ 32-33

ตาราง 18 (ต่อ)

ข้อที่	ฉบับที่ 1			ฉบับที่ 2		
	IRT (SA)	SIB (β)	MH (α)	IRT (SA)	SIB (β)	MH (α)
19	-2.027	-0.026	0.88	-0.276	-0.041	0.74
20	-0.615	0.040	1.20	-0.010	-0.010	0.97
21	-2.660	-0.009	0.99	0.017	0.032	1.20
22	-1.720	-0.018	0.92	-0.195	-0.014	0.97
23	-0.950	0.001	1.04	0.082	0.012	1.09
24	0.478	-0.006	0.98	-0.503	0.018	1.16
25	-1.003	0.043	1.23	0.356	-0.011	0.97
26	-1.569	0.073	1.44	0.233	-0.003	0.98
27	-1.470	-0.011	0.93	0.952	0.087	1.43
28	-1.857	-0.009	0.95	0.300	0.064	1.36
29	-0.877	-0.084	0.71	0.136	0.027	1.09
30	-1.210	-0.014	0.93	-0.316	-0.067	0.72
31	0.147	0.025	1.12	-0.763	-0.069	0.77
32	0.211	-0.026	0.91	-0.163	0.001	0.99
33	-2.280	-0.055	0.76	0.021	-0.030	0.89
34	0.574	0.013	1.10	0.460	0.044	1.23
35	0.392	-0.062	0.80	-0.635	-0.016	0.95
36	0.842	0.025	1.24	-0.288	-0.055	0.78
37	1.527	0.003	1.06	-1.482	-0.072	0.77
38	-1.104	0.008	1.03	-0.287	-0.038	0.87
39	1.067	0.034	1.20	1.235	0.017	1.09
40	-0.644	-0.030	0.90	-0.190	-0.029	0.89
41	-1.829	-0.034	0.86	-0.485	-0.026	0.90
42	1.400	-0.025	0.87	0.727	0.034	1.35

จากตาราง 18 ดัชนีความลำเอียงของข้อสอบวิชา ภาษาไทย ก ฉบับที่ 1 เมื่อวิเคราะห์ความลำเอียงด้วยวิธีใช้ทฤษฎีการตอบข้อคำถามได้ค่าดัชนีอยู่ระหว่าง .069 - 2.660 วิธี SIBTEST ได้ค่าดัชนีอยู่ระหว่าง .001 - .065 และวิธี แมนเทิล-แฮนส์เซล ได้ค่าดัชนีอยู่ระหว่าง .76 - 1.44 ส่วนฉบับที่ 2 เมื่อวิเคราะห์ความลำเอียงด้วยวิธีใช้ทฤษฎีการตอบข้อคำถามได้ค่าดัชนีอยู่ระหว่าง .010 - 1.482 วิธี SIBTEST ได้ค่าดัชนีอยู่ระหว่าง .001 - .087 และวิธีแมนเทิล-แฮนส์เซล ได้ค่าดัชนีอยู่ระหว่าง .74 - 1.40

ตาราง 19 ค่าดัชนีความลำเอียงของทั้ง 3 วิธีของแบบทดสอบวิชา สังคมศึกษา ก

ข้อที่	ฉบับที่ 1			ฉบับที่ 2		
	IRT (SA)	SIB (β)	MH (α)	IRT (SA)	SIB (β)	MH (α)
1	0.290	0.007	1.03	0.433	0.000	0.98
2	-0.115	0.048	1.23	-0.376	0.012	1.07
3	-0.000	-0.016	1.09	-0.110	-0.014	0.89
4	-0.314	0.082	1.59	-0.085	-0.029	0.87
5	-0.178	0.016	1.30	-0.378	-0.040	0.83
6	0.309	-0.004	0.95	0.446	-0.053	0.80
7	-0.107	0.006	1.05	0.426	0.065	1.49
8	0.269	0.026	1.12	0.305	0.020	1.12
9	-0.292	-0.081	0.70	-2.293	-0.061	0.78
10	0.033	-0.005	0.90	0.246	0.039	1.27
11	-0.178	0.048	1.24	-0.434	-0.004	0.99
12	0.874	0.004	1.04	0.419	0.032	1.17
13	-0.382	-0.037	0.83	0.170	0.005	1.03
14	0.289	0.026	1.15	0.231	0.045	1.22
15	0.262	-0.023	0.85	0.379	0.034	1.18
16	0.342	-0.051	0.80	0.894	0.015	1.11
17	1.932	-0.026	0.91	-0.049	-0.046	0.82
18	0.085	-0.017	0.90	0.004	-0.012	0.95
19	-0.165	0.016	1.26	0.287	0.048	1.29
20	0.287	-0.031	0.85	0.075	0.016	1.15
21	0.179	-0.031	0.86	-0.003	-0.017	0.90
22	0.271	-0.025	0.85	-0.164	-0.003	0.96
23	0.848	-0.054	0.81	0.599	-0.016	0.94

ตาราง 19 (ต่อ)

ข้อที่	ฉบับที่ 1			ฉบับที่ 2		
	IRT (SA)	SIB (β)	MH (α)	IRT (SA)	SIB (β)	MH (α)
24	-0.264	0.002	1.01	0.241	-0.011	0.95
25	0.125	-0.035	0.84	-0.192	-0.029	0.86
26	0.099	0.002	1.00	0.592	-0.064	0.76
27	0.400	-0.036	0.87	0.502	0.095	1.71
28	-0.053	0.025	1.11	0.309	0.037	1.32
29	0.085	-0.048	0.83	1.250	0.160	1.99
30	-0.476	0.076	1.35	-0.333	0.024	1.36
31	0.249	-0.037	0.85	0.324	0.049	1.25
32	0.268	-0.014	0.92	-0.548	-0.113	0.59
33	0.462	0.038	1.16	-0.102	-0.022	0.91
34	-0.149	0.108	1.62	-0.485	0.012	1.07
35	0.325	-0.054	0.77	0.216	0.023	1.10
36	0.221	-0.016	0.77	1.193	-0.056	0.79
37	1.006	-0.154	0.52	-0.008	-0.003	0.97
38	0.138	-0.021	0.85	-0.052	-0.022	0.90
39	0.363	-0.046	0.72	0.356	0.026	1.10
40	0.976	0.038	1.17	0.388	-0.001	0.99
41	-0.252	0.041	1.22	0.291	0.062	1.32
42	1.064	0.003	1.01	-0.249	-0.007	0.96
43	0.012	-0.004	0.99	0.929	0.044	1.19
44	0.026	0.011	1.17	0.202	0.028	1.24
45	-0.208	0.020	1.13	0.040	0.048	1.25
46	0.295	-0.041	0.73	0.110	0.039	1.21

ตาราง 19 (ต่อ)

ข้อที่	ฉบับที่ 1			ฉบับที่ 2		
	IRT (SA)	SIB (β)	MH (α)	IRT (SA)	SIB (β)	MH (α)
47	0.327	-0.054	0.74	-0.071	-0.008	0.90
48	0.004	0.016	1.09	0.174	0.018	1.12
49	-0.077	0.040	1.02	1.075	0.033	1.13
50	-0.087	-0.020	0.89	0.385	0.047	1.24
51	0.445	-0.014	0.92	0.141	0.014	1.10
52	0.513	-0.087	0.70	0.304	0.003	1.02
53	0.041	0.003	1.00	0.089	0.012	1.05
54	0.162	-0.008	0.96	0.336	-0.032	0.87
55	-0.815	0.010	1.04	-0.248	-0.042	0.83
56	-0.034	0.042	1.20	0.058	0.005	1.12
57	-0.096	0.025	1.13	-0.105	0.046	1.23
58	0.078	-0.059	0.71	-0.613	-0.011	0.94
59	0.165	0.036	1.17	-0.194	-0.041	0.83
60	-0.117	0.035	1.20	-0.782	-0.001	0.99
61	0.005	0.053	1.26	-0.656	-0.105	0.60
62	-0.042	0.033	1.22	-0.244	-0.046	0.75
63	0.392	0.004	1.05	-0.010	-0.056	0.79
64	0.045	-0.008	0.96	0.165	-0.022	0.90

จากตาราง 19 ดัชนีความลำเอียงของข้อสอบวิชา สังคมศึกษา ก ฉบับที่ 1 เมื่อวิเคราะห์ความลำเอียงด้วยวิธีใช้ทฤษฎีการตอบข้อคำถามได้ค่าดัชนีอยู่ระหว่าง .000 - 1.932 วิธี SIBTEST ได้ค่าดัชนีอยู่ระหว่าง .002 - .154 และวิธี แมนเทล-แฮนส์เซล ได้ค่าดัชนีอยู่ระหว่าง .52 - 1.59 ส่วนฉบับที่ 2 เมื่อ

วิเคราะห์ความลำเอียงด้วยวิธีใช้ทฤษฎีการตอบข้อคำถามได้ค่าดัชนีอยู่ระหว่าง .003 - 2.293 วิธี SIBTEST ได้ค่าดัชนีอยู่ระหว่าง .000 - .160 และวิธีแมนเทิล-แฮนส์เซล ได้ค่าดัชนีอยู่ระหว่าง .59 - 1.71

ตาราง 20 ค่าดัชนีความลำเอียงของทั้ง 3 วิธีของแบบทดสอบวิชา ภาษาอังกฤษ กข

ข้อที่	ฉบับที่ 1			ฉบับที่ 2		
	IRT (SA)	SIB (β)	MH (α)	IRT (SA)	SIB (β)	MH (α)
1	-0.030	-0.017	0.84	0.060	-0.018	0.93
2	-0.041	0.016	1.14	0.032	-0.004	1.03
3	-0.199	-0.019	0.91	0.133	0.003	1.02
4	0.306	0.048	1.27	-0.578	-0.008	1.04
5	-0.826	-0.041	0.82	0.074	-0.055	0.81
6	-0.231	0.016	1.08	0.100	-0.011	0.99
7	-0.164	-0.011	0.96	-0.064	0.008	1.11
8	0.020	-0.030	0.86	-0.143	-0.037	0.89
9	-0.186	0.006	1.04	-0.278	-0.012	1.00
10	-0.088	-0.031	0.82	0.570	-0.006	0.99
11	0.000	-0.020	0.86	-0.002	0.029	1.21
12	0.264	0.027	1.09	0.094	-0.005	1.09
13	0.021	0.018	1.11	0.130	0.017	1.11
14	1.160	0.046	1.16	-0.035	-0.037	0.90
15	0.635	0.037	1.18	0.210	-0.018	0.93
16	0.240	0.007	1.04	0.384	-0.008	1.00
17	0.417	0.030	1.21	0.616	0.021	1.20
18	-0.770	-0.059	0.78	0.484	-0.006	1.00
19	0.772	0.013	1.04	-0.620	-0.080	0.80
20	-0.071	0.007	1.10	-0.166	-0.030	0.95
21	0.347	-0.036	0.87	-0.015	-0.019	0.95
22	0.146	0.020	1.12	0.166	0.032	1.47
23	0.250	0.044	1.17	-0.230	-0.012	0.98

ตาราง 20 (ต่อ)

ข้อที่	ฉบับที่ 1			ฉบับที่ 2		
	IRT (SA)	SIB (β)	MH (α)	IRT (SA)	SIB (β)	MH (α)
24	0.416	0.019	1.03	-0.030	-0.013	0.96
25	0.166	0.010	1.05	-0.312	-0.013	1.05
26	0.064	0.022	1.32	0.215	0.003	1.08
27	-0.096	-0.003	0.98	-0.042	-0.066	0.81
28	-0.234	0.007	1.04	0.001	0.007	1.10
29	-0.703	-0.007	0.94	-0.334	-0.017	0.93
30	0.178	-0.017	0.96	1.204	0.013	1.12
31	0.959	0.014	0.89	-0.095	0.000	1.08
32	0.479	0.045	1.21	0.190	0.016	1.16
33	0.070	-0.024	0.89	0.029	-0.021	0.92
34	0.235	0.005	1.02	-0.045	-0.044	0.89
35	-0.311	-0.062	0.77	-0.296	-0.062	0.78
36	-0.031	0.031	1.16	0.253	0.003	1.09

จากตาราง 20 ดัชนีความลำเอียงของข้อสอบวิชา ภาษาอังกฤษ กข ฉบับที่ 1 เมื่อวิเคราะห์ความลำเอียงด้วยวิธีใช้ทฤษฎีการตอบข้อคำถามได้ค่าดัชนีอยู่ระหว่าง .000 - 1.160 วิธี SIBTEST ได้ค่าดัชนีอยู่ระหว่าง .003 - .062 และวิธีแมนเทิล-ฮันส์เซล ได้ค่าดัชนีอยู่ระหว่าง .77 - 1.32 ส่วนฉบับที่ 2 เมื่อวิเคราะห์ความลำเอียงด้วยวิธีใช้ทฤษฎีการตอบข้อคำถามได้ค่าดัชนีอยู่ระหว่าง .001 - 1.204 วิธี SIBTEST ได้ค่าดัชนีอยู่ระหว่าง .003 - .080 และวิธีแมนเทิล-ฮันส์เซล ได้ค่าดัชนีอยู่ระหว่าง .78 - 1.47

ตาราง 21 (ต่อ)

ข้อที่	ฉบับที่ 1			ฉบับที่ 2		
	IRT (SA)	SIB (β)	MH (α)	IRT (SA)	SIB (β)	MH (α)
24	-0.477	0.002	1.03	-0.463	-0.067	0.71
25	1.016	0.056	1.29	0.489	0.001	1.03
26	-0.140	0.002	1.00	-0.627	-0.112	0.63
27	-0.672	-0.052	0.83	-0.187	-0.009	0.92
28	-0.722	-0.027	0.87	-0.014	-0.029	0.89
29	-0.814	-0.061	0.74	0.430	0.022	1.12
30	-0.415	-0.042	0.83	0.228	0.098	1.61
31	0.372	0.003	1.02	-0.522	-0.042	0.76
32	-0.546	-0.044	0.82	0.130	-0.025	0.90
33	-0.154	0.016	1.10	-0.635	-0.049	0.81
34	-0.600	-0.043	0.74	-0.547	-0.034	0.86
35	-0.715	-0.035	0.88	-0.382	-0.050	0.81
36	-0.487	-0.010	0.95	-0.480	0.010	1.09
37	-0.225	0.053	1.30	-0.273	-0.006	0.99

จากตาราง 21 ดัชนีความลำเอียงของข้อสอบวิชา ภาษาไทย กข ฉบับที่ 1 เมื่อวิเคราะห์ความลำเอียงด้วยวิธีใช้ทฤษฎีการตอบข้อคำถามได้ค่าดัชนีอยู่ระหว่าง .050 - 1.016 วิธี SIBTEST ได้ค่าดัชนีอยู่ระหว่าง .002 - .071 และวิธีแมนเทิล-แฮนส์เซล ได้ค่าดัชนีอยู่ระหว่าง .74 - 1.41 ส่วนฉบับที่ 2 เมื่อวิเคราะห์ความลำเอียงด้วยวิธีใช้ทฤษฎีการตอบข้อคำถามได้ค่าดัชนีอยู่ระหว่าง .009 - .788 วิธี SIBTEST ได้ค่าดัชนีอยู่ระหว่าง .001 - .098 และวิธีแมนเทิล-แฮนส์เซล ได้ค่าดัชนีอยู่ระหว่าง .63 - 1.49

ตาราง 22 ค่าดัชนีความลำเอียงของทั้ง 3 วิธี ของแบบทดสอบวิชาสังคมศึกษา กข

ข้อที่	ฉบับที่ 1			ฉบับที่ 2		
	IRT (SA)	SIB (β)	MH (α)	IRT (SA)	SIB (β)	MH (α)
1	0.762	-0.039	0.88	1.067	0.034	1.13
2	-0.540	0.007	1.07	-0.140	-0.038	0.86
3	0.064	0.007	1.08	-0.277	0.021	1.17
4	-0.118	0.005	1.00	0.211	0.017	1.51
5	-0.106	-0.035	0.86	-0.142	-0.023	0.89
6	0.224	0.036	1.21	-0.167	-0.023	0.90
7	0.505	-0.066	0.77	-0.408	0.025	1.30
8	-0.472	0.008	1.06	0.150	0.028	1.10
9	-0.321	-0.057	0.73	-0.226	-0.053	0.77
10	0.435	0.040	1.59	0.348	0.053	1.26
11	-0.445	-0.067	0.78	0.014	-0.021	0.89
12	-0.213	-0.061	0.69	-1.626	-0.058	0.78
13	-0.153	0.017	1.20	0.455	0.037	1.26
14	0.883	-0.003	1.04	-0.390	-0.079	0.66
15	0.139	0.086	1.51	0.301	0.003	1.07
16	-0.314	-0.045	0.77	0.009	0.006	1.02
17	0.414	0.050	1.47	-0.135	-0.032	0.88
18	0.197	-0.068	0.65	-0.013	0.007	0.99
19	0.043	0.050	1.28	-0.207	0.002	1.00
20	0.048	0.020	1.10	0.044	0.047	1.25
21	0.324	0.008	1.10	0.131	0.051	1.25
22	-0.060	-0.035	0.87	0.053	-0.011	0.88
23	1.239	0.067	1.36	0.695	0.034	1.50

ตาราง 22 (ต่อ)

ข้อที่	ฉบับที่ 1			ฉบับที่ 2		
	IRT (SA)	SIB (β)	MH (α)	IRT (SA)	SIB (β)	MH (α)
24	-0.345	-0.029	0.91	-0.039	0.010	1.30
25	-0.129	0.000	1.09	-0.205	-0.053	0.80
26	0.462	0.017	1.31	0.127	0.001	1.00
27	-0.253	-0.047	0.82	-0.237	0.028	1.12
28	1.895	0.012	1.06	-0.030	0.033	1.14
29	0.351	0.029	1.14	0.333	0.090	1.55
30	0.492	-0.037	0.87	0.008	0.064	1.32
31	-0.153	0.022	1.74	-0.003	0.035	1.17
32	-0.177	-0.023	0.90	0.096	-0.051	0.78
33	0.148	0.039	1.21	-0.161	0.013	1.01
34	-0.864	-0.047	0.84	0.068	0.006	1.01
35	-0.876	-0.072	0.72	0.225	0.019	1.16
36	0.301	0.011	1.10	0.585	0.044	1.20
37	-0.452	-0.075	0.68	-0.102	0.067	1.52
38	0.036	0.019	1.09	-1.713	-0.096	0.69
39	-0.192	-0.016	0.89	0.925	-0.006	0.99
40	-0.092	-0.018	0.95	-0.446	-0.009	0.95
41	-0.217	-0.016	0.91	0.200	0.019	1.06
42	-0.481	-0.048	0.84	-0.003	0.014	1.24
43	0.946	0.031	1.18	-0.402	-0.061	0.77
44	0.013	0.016	1.09	-0.013	-0.010	0.97
45	0.298	-0.010	0.98	-1.049	-0.064	0.75
46	0.152	0.041	1.52	-0.263	-0.001	0.97

ตาราง 23 ค่าดัชนีความลำเอียงของทั้ง 3 วิธี ของแบบทดสอบวิชาภาษาอังกฤษ
กชค

ข้อที่	ฉบับที่ 1			ฉบับที่ 2		
	IRT (SA)	SIB (β)	MH (α)	IRT (SA)	SIB (β)	MH (α)
1	-0.353	-0.027	0.92	-0.491	0.027	1.19
2	-0.308	0.006	1.06	-0.207	-0.009	0.96
3	-0.395	-0.064	0.74	-0.755	-0.069	0.76
4	-0.225	-0.015	0.95	-0.490	-0.015	0.95
5	-0.061	-0.035	0.84	-0.607	-0.059	0.66
6	-0.509	0.011	1.14	-0.284	-0.038	0.91
7	-0.053	-0.032	0.86	-0.905	-0.030	0.96
8	0.149	-0.003	1.06	0.217	0.004	1.05
9	-0.869	-0.008	1.03	-0.586	0.043	0.81
10	0.104	0.009	1.06	-0.034	0.016	1.14
11	3.390	0.020	1.19	-0.608	0.034	1.21
12	-0.231	0.014	1.06	-0.538	-0.066	0.79
13	-0.230	-0.060	0.75	-0.639	-0.007	1.03
14	-0.208	0.001	0.98	-6.290	0.008	1.08
15	0.084	-0.015	0.92	-0.338	0.022	1.15
16	0.388	-0.013	0.90	1.162	0.008	1.10
17	-0.132	-0.059	0.72	-0.463	0.015	1.11
18	0.939	-0.035	0.89	-0.480	-0.018	0.96
19	-0.989	0.002	1.10	-0.344	-0.028	0.88
20	-0.314	-0.040	0.85	0.151	0.018	1.18
21	-0.421	-0.029	0.90	-0.389	-0.018	0.89
22	-0.029	-0.028	0.96	-0.977	-0.120	0.67

ตาราง 23 (ต่อ)

ข้อที่	ฉบับที่ 1			ฉบับที่ 2		
	IRT (SA)	SIB (β)	MH (α)	IRT (SA)	SIB (β)	MH (α)
23	-0.422	-0.101	0.67	-0.016	0.026	1.18
24	0.324	-0.005	1.05	-0.880	-0.049	0.86
25	1.539	-0.031	0.90	-1.091	-0.010	0.91
26	-0.148	0.038	1.55	-1.031	-0.015	0.97
27	-0.020	-0.002	0.99	-0.932	-0.132	0.59
28	-0.017	-0.024	0.91	-0.436	0.040	1.26
29	0.025	-0.017	0.93	-0.339	-0.023	0.91
30	-0.445	0.017	1.11	-1.115	-0.009	1.07
31	0.134	0.005	1.18	-0.442	-0.038	0.83
32	-0.049	-0.007	1.01	-1.116	-0.013	1.01
33	0.573	0.039	1.15	-0.256	0.029	1.12
34	-0.625	-0.107	0.67	-0.123	0.012	1.09
35	-0.109	0.000	1.08	-1.116	0.023	1.19
36	0.262	0.023	1.15	-0.449	-0.017	0.92
37	-0.154	-0.011	1.05	-0.247	-0.038	0.92

จากตาราง 23 ดัชนีความลำเอียงของข้อสอบวิชา ภาษาอังกฤษ กขค ฉบับที่ 1 เมื่อวิเคราะห์ความลำเอียงด้วยวิธีใช้ทฤษฎีการตอบข้อคำถามได้ค่าดัชนีอยู่ระหว่าง .017 - 3.390 วิธี SIBTEST ได้ค่าดัชนีอยู่ระหว่าง .000 - .107 และวิธีแมนเทิล-แฮนส์เซล ได้ค่าดัชนีอยู่ระหว่าง .67 - 1.55 ส่วนฉบับที่ 2 เมื่อวิเคราะห์ความลำเอียงด้วยวิธีใช้ทฤษฎีการตอบข้อคำถามได้ค่าดัชนีอยู่ระหว่าง .034 - 1.162 วิธี SIBTEST ได้ค่าดัชนีอยู่ระหว่าง .007 - .132 และวิธีแมนเทิล-แฮนส์เซล ได้ค่าดัชนีอยู่ระหว่าง .59 - 1.26

จากดัชนีความลำเอียงของข้อสอบที่ใช้วิธีการวิเคราะห์ทั้ง 3 วิธี ของแบบทดสอบในแต่ละวิชาจากตาราง 18-23 เมื่อพิจารณาถึงจำนวนข้อสอบที่ลำเอียงโดยยึดถือเกณฑ์ของแต่ละวิธี ดังนี้

1. วิธีใช้ทฤษฎีการตอบข้อคำถาม ใช้ค่าพื้นที่ความแตกต่างมีขนาด (ISA) มากกว่า .20 ถือว่าเป็นข้อสอบลำเอียง

2. วิธีแมนเทิล-แฮนส์เซล ใช้ค่า α_{MH} มากกว่า 1.00 และการทดสอบไคสแควร์ของแมนเทิล-แฮนส์เซล

3. วิธี SIBTEST ใช้การทดสอบนัยสำคัญของค่า Z-test

เมื่อเทียบค่าดัชนีความลำเอียงของข้อสอบที่วิเคราะห์ด้วยวิธีการทั้ง 3 วิธี กับเกณฑ์ดังกล่าวแล้วสรุปผลจำนวนข้อสอบที่ลำเอียงและไม่ลำเอียงในแต่ละวิชาดังตาราง 24-35

จากตาราง 24 ในแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 1 พบว่า วิธีใช้
 ทฤษฎีการตอบข้อความตัดสินจำนวนข้อสอบที่ลำเอียงมากกว่าวิธีแมนเทิล-แฮนส์เซล
 และวิธี SIBTEST โดยพบจำนวนข้อสอบที่ลำเอียงเป็น 35, 19 และ 9 ข้อ ตาม
 ลำดับ จำนวนข้อสอบที่ลำเอียงตรงกันทั้ง 3 วิธี มีจำนวน 3 ข้อ คือ 5, 25 และ
 26 ข้อสอบที่ลำเอียงตรงกัน 2 วิธี มีจำนวน 16 ข้อ คือ 1, 4, 8, 13, 14,
 18, 20, 23, 29, 33, 34, 35, 36, 37, 38 และ 39 ข้อสอบที่ลำเอียง
 เฉพาะวิธีใดวิธีหนึ่ง มีจำนวน 22 ข้อ คือ 2, 3, 6, 7, 9, 11, 12, 15, 16,
 17, 19, 21, 22, 24, 27, 28, 30, 31, 32, 40, 41 และ 42 และ
 ข้อสอบที่ไม่ลำเอียงทั้ง 3 วิธี มีจำนวน 1 ข้อ คือ 10

จากตาราง 25 ในแบบทดสอบวิชาภาษาไทย ก ฉบับที่ 2 พบว่า วิธีใช้
 ทฤษฎีการตอบข้อคำถามตัดสินจำนวนข้อสอบที่ลำเอียงมากกว่าวิธีแมนเทิล-แฮนส์เซล
 และวิธี SIBTEST โดยพบจำนวนข้อสอบที่ลำเอียงเป็น 24, 19 และ 13 ข้อตาม
 ลำดับ จำนวนข้อสอบที่ลำเอียงตรงกันทั้ง 3 วิธี มีจำนวน 6 ข้อ คือ 8, 9, 27,
 28, 34 และ 42 ข้อสอบที่ลำเอียงตรงกัน 2 วิธี มีจำนวน 11 ข้อ คือ 2, 5,
 12, 15, 19, 24, 30, 31, 36, 37 และ 39 ข้อสอบที่ลำเอียงเฉพาะวิธีใด
 วิธีหนึ่ง มีจำนวน 16 ข้อ คือ 1, 3, 6, 7, 10, 13, 17, 18, 21, 23, 25,
 26, 29, 35, 38, และ 41 และข้อสอบที่ไม่ลำเอียงทั้ง 3 วิธี มีจำนวน 9 ข้อ
 คือ 4, 11, 14, 16, 20, 22, 32, 33 และ 40

จากตาราง 26 ในแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 1 พบว่า วิธีใช้
 ทฤษฎีการตอบข้อคำถามตัดสินจำนวนข้อสอบที่ลำเอียงมากกว่าวิธีแมนเทิล-แฮนส์เซล
 และ SIBTEST โดยพบจำนวนข้อสอบที่ลำเอียงเป็น 34, 27 และ 20 ตามลำดับ
 จำนวนข้อสอบที่ลำเอียงตรงกันทั้ง 3 วิธี มีจำนวน 3 ข้อ คือ 4, 30 และ 41
 ข้อสอบที่ลำเอียงตรงกัน 2 วิธี มีจำนวน 26 ข้อ คือ 1, 8, 9, 12, 13, 14,
 16, 23, 24, 33, 34, 35, 35, 37, 39, 40, 42, 45, 46, 47, 49,
 52, 55, 56, 61, 62, และ 63 ข้อสอบที่ลำเอียงเฉพาะวิธีใดวิธีหนึ่ง มีจำนวน
 20 ข้อ คือ 2, 3, 5, 6, 7, 11, 15, 17, 19, 20, 22, 26, 27, 28,
 29, 31, 32, 36, 51 และ 58 และข้อสอบที่ไม่ลำเอียงทั้ง 3 วิธี มีจำนวน
 12 ข้อ คือ 10, 18, 21, 25, 38, 43, 50, 53, 54, 57, 59 และ 64

จากตาราง 27 ในแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 2 พบว่าวิธีใช้
 ทฤษฎีการตอบข้อคำถามตัดสินจำนวนข้อสอบที่ลำเอียงมากกว่าวิธีแมนเทิล-แฮนส์เซล
 และ SIBTEST โดยพบจำนวนข้อสอบที่ลำเอียงเป็น 39, 26 และ 25 ข้อ ตาม
 ลำดับ จำนวนข้อสอบที่ลำเอียงตรงกันทั้ง 3 วิธี มีจำนวน 12 ข้อ คือ 7, 10, 14,
 19, 27, 28, 29, 30, 31, 41, 43 และ 50 ข้อสอบที่ลำเอียงตรงกัน 2 วิธี
 มีจำนวน 19 ข้อ คือ 2, 6, 8, 9, 12, 15, 16, 26, 32, 34, 35, 36,
 39, 45, 46, 52, 57, 61, และ 62 ข้อสอบที่ลำเอียงเฉพาะวิธีใดวิธีหนึ่ง มี
 จำนวน 16 ข้อ คือ 1, 5, 13, 17, 20, 23, 24, 40, 42, 44, 54, 55,
 58, 59, 60 และ 63 และข้อสอบที่ไม่ลำเอียงทั้ง 3 วิธี มีจำนวน 17 ข้อ คือ 3,
 4, 11, 18, 21, 22, 25, 33, 37, 38, 47, 48, 49, 51, 53, 56
 และ 64

จากตาราง 28 ในแบบทดสอบวิชาภาษาอังกฤษ กข ฉบับที่ 1 พบว่า วิธีแมนเทิล - แชนส์เซล ตัดสินจำนวนข้อสอบที่ลำเอียงมากกว่าวิธีใช้ทฤษฎีการตอบข้อคำถาม และวิธี SIBTEST โดยพบจำนวนข้อสอบที่ลำเอียงเป็น 24, 19 และ 6 ข้อ ตามลำดับ จำนวนข้อสอบที่ลำเอียงตรงกันทั้ง 3 วิธี มีจำนวน 6 ข้อ คือ 4, 5, 14, 18, 32 และ 35 ข้อสอบที่ลำเอียงตรงกัน 2 วิธี มีจำนวน 10 ข้อ คือ 6, 12, 15, 16, 17, 19, 23, 24, 28 และ 34 ข้อสอบที่ลำเอียงเฉพาะวิธีใดวิธีหนึ่ง มีจำนวน 11 ข้อ คือ 2, 9, 13, 20, 21, 22, 25, 26, 29, 31 และ 36 และข้อสอบที่ไม่ลำเอียงทั้ง 3 วิธี มีจำนวน 9 ข้อ คือ 1, 3, 7, 8, 10, 11, 27, 30 และ 33

จากตาราง 29 ในแบบทดสอบวิชาภาษาอังกฤษ กข ฉบับที่ 2 พบว่า วิธีแมนเทิล - แชนส์เซลล์ ตัดสินจำนวนข้อสอบที่ลำเอียงมากกว่าวิธีใช้ทฤษฎีการตอบข้อคำถาม และวิธี SIBTEST โดยพบจำนวนข้อสอบที่ลำเอียงเป็น 19, 14 และ 7 ข้อตามลำดับ จำนวนข้อสอบที่ลำเอียงตรงกันทั้ง 3 วิธี มีจำนวน 2 ข้อ คือ 19 และ 35 ข้อสอบที่ลำเอียงตรงกัน 2 วิธี มีจำนวน 8 ข้อ คือ 4, 16, 17, 18, 22, 25, 26 และ 30 ข้อสอบที่ลำเอียงเฉพาะวิธีใดวิธีหนึ่ง มีจำนวน 19 ข้อ คือ 2, 3, 5, 7, 9, 10, 11, 12, 13, 14, 15, 23, 27, 28, 29, 31, 32, 34 และ 36 และข้อสอบที่ไม่ลำเอียงทั้ง 3 วิธี มีจำนวน 6 ข้อ คือ 1, 6, 20, 21, 24 และ 33

จากตาราง 30 ในแบบทดสอบวิชาภาษาไทย กข ฉบับที่ 1 พบว่า วิธีใช้
 ทฤษฎีการตอบข้อคำถามตัดสินจำนวนข้อสอบที่ลำเอียงมากกว่าวิธีแมนเทิล-แฮนส์เซล
 และวิธี SIBTEST โดยพบจำนวนข้อสอบที่ลำเอียงเป็น 24, 19 และ 9 ข้อ ตาม
 ลำดับ จำนวนข้อสอบที่ลำเอียงตรงกันทั้ง 3 วิธี มีจำนวน 2 ข้อ คือ 14 และ 23
 ข้อสอบที่ลำเอียงตรงกัน 2 วิธี มีจำนวน 15 ข้อ คือ 1, 2, 4, 7, 8, 10, 16,
 22, 24, 25, 27, 29, 31, 34 และ 37 ข้อสอบที่ลำเอียงเฉพาะวิธีใดวิธีหนึ่ง
 มีจำนวน 18 ข้อ คือ 3, 6, 9, 11, 13, 15, 17, 18, 19, 20, 21, 26,
 28, 30, 32, 33, 35 และ 36 และข้อสอบที่ไม่ลำเอียงทั้ง 3 วิธี มีจำนวน 4
 ข้อ คือ 5, 12, 35 และ 36

จากตาราง 31 ในแบบทดสอบวิชาภาษาไทย กข ฉบับที่ 2 พบว่า วิธีใช้
 ทฤษฎีการตอบข้อคำถามตัดสินจำนวนข้อสอบที่ลำเอียงมากกว่าวิธีแมนเทิล-แฮนส์เซล
 และวิธี SIBTEST โดยพบจำนวนข้อสอบที่ลำเอียงเป็น 24, 16 และ 10 ข้อ ตาม
 ลำดับ จำนวนข้อสอบที่ลำเอียงตรงกันทั้ง 3 วิธี มีจำนวน 2 ข้อ คือ 17 และ 30
 ข้อสอบที่ลำเอียงตรงกัน 2 วิธี มีจำนวน 16 ข้อ คือ 1, 6, 8, 12, 14, 15,
 19, 20, 23, 24, 25, 26, 29, 31, 34 และ 36 ข้อสอบที่ลำเอียงเฉพาะ
 วิธีใดวิธีหนึ่ง มีจำนวน 12 ข้อ คือ 2, 3, 4, 9, 10, 13, 16, 18, 21, 33,
 35 และ 37 และข้อสอบที่ไม่ลำเอียงทั้ง 3 วิธี มีจำนวน 7 ข้อ คือ 5, 7, 11,
 21, 27, 28 และ 32

จากตาราง 32 ในแบบทดสอบวิชาสังคมศึกษา กข ฉบับที่ 1 พบว่า วิธีใช้
 ทฤษฎีการตอบข้อคำถาม ตัดสินจำนวนข้อสอบที่ลำเอียงมากกว่าวิธีแมนเทิล-ฮันส์เชล
 และวิธี SIBTEST โดยพบว่าจำนวนข้อสอบที่ลำเอียงเป็น 34, 32 และ 17 ข้อ
 ตามลำดับ จำนวนข้อสอบที่ลำเอียงตรงกันทั้ง 3 วิธี มีจำนวน 4 ข้อ คือ 10, 17,
 23 และ 59 ข้อสอบที่ลำเอียงตรงกัน 2 วิธี มีจำนวน 23 ข้อ คือ 2, 6, 7, 8,
 9, 11, 12, 14, 15, 16, 20, 21, 26, 28, 29, 36, 37, 43, 46,
 47, 53, 54 และ 57 ข้อสอบที่ลำเอียงเฉพาะวิธีใดวิธีหนึ่ง มีจำนวน 25 ข้อ
 คือ 1, 3, 4, 13, 18, 19, 24, 25, 27, 30, 31, 33, 34, 35, 38,
 41, 42, 44, 45, 48, 49, 51, 52, 55 และ 58 และข้อสอบที่ไม่ลำเอียง
 ทั้ง 3 วิธี มีจำนวน 7 ข้อ คือ 5, 22, 32, 39, 40, 50 และ 56

จากตาราง 33 ในแบบทดสอบวิชาสังคมศึกษา กข. ฉบับที่ 2 พบว่า วิธีใช้
 ทฤษฎีการตอบข้อคำถาม ตัดสินจำนวนข้อสอบที่ลำเอียงมากกว่าวิธีแมนเทิล-แฮนส์เซล
 และวิธี SIBTEST โดยพบจำนวนข้อสอบที่ลำเอียงเป็น 34, 33 และ 19 ข้อ ตาม
 ลำดับ จำนวนข้อสอบที่ลำเอียงตรงกันทั้ง 3 วิธี มีจำนวน 5 ข้อ คือ 10, 23, 29,
 48, 53 และ 55 ข้อสอบที่ลำเอียงตรงกัน 2 วิธี มีจำนวน 24 ข้อ คือ 1, 3,
 4, 7, 12, 13, 14, 15, 19, 21, 25, 27, 30, 35, 36, 37, 38, 41,
 43, 45, 47, 49, 52 และ 56 ข้อสอบที่ลำเอียงเฉพาะวิธีใดวิธีหนึ่ง มีจำนวน
 20 ข้อ คือ 8, 9, 16, 20, 24, 26, 28, 31, 32, 33, 34, 39, 40,
 42, 46, 50, 54, 57, 58 และ 59 และข้อสอบที่ไม่ลำเอียงทั้ง 3 วิธี มีจำนวน
 8 ข้อ คือ 5, 6, 11, 17, 18, 22, 44 และ 51

จากตาราง 34 ในแบบทดสอบวิชาภาษาอังกฤษ กขค ฉบับที่ 1 พบว่า วิธีให้ทฤษฎีการตอบข้อความตัดสินจำนวนข้อสอบที่ลำเอียงมากกว่าวิธีแมนเทล - แชนส์เชล และวิธี SIBTEST โดยพบจำนวนข้อสอบที่ลำเอียงเป็น 22, 21 และ 6 ข้อ ตามลำดับ จำนวนข้อสอบที่ลำเอียงตรงกันทั้ง 3 วิธี มีจำนวน 4 ข้อ คือ 3, 13, 23 และ 34 ข้อสอบที่ลำเอียงตรงกัน 2 วิธี มีจำนวน 11 ข้อ คือ 2, 6, 9, 11, 12, 19, 24, 26, 30, 33 และ 36 ข้อสอบที่ลำเอียงเฉพาะวิธีใดวิธีหนึ่ง มีจำนวน 15 ข้อ คือ 1, 4, 8, 10, 14, 16, 17, 18, 20, 21, 25, 31, 32, 35 และ 37 และข้อสอบที่ไม่ลำเอียงทั้ง 3 วิธี มีจำนวน 6 ข้อ คือ 5, 15, 22, 27, 28 และ 29

ตาราง 35 แสดงข้อสอบที่ลำเอียงและไม่ลำเอียงที่ได้จากการวิเคราะห์ทาง 3 วิธี ของแบบทดสอบวิชา ภาษาอังกฤษ กขค ฉบับที่ 2

วิธี	IRT												SIB												MH											
ข้อสอบที่ ลำเอียง	1	2	3	4	5	6	7	8	9	-	-	-	3	-	5	-	-	-	-	1	-	3	-	5	-	-	8	-	10							
	11	12	13	14	15	16	17	18	19	-	-	12	-	-	-	-	-	-	-	11	12	13	14	15	16	17	-	-	20							
	21	22	-	24	25	26	27	28	29	30	-	22	-	-	-	27	-	-	-	-	22	23	-	-	-	27	28	-	30							
	31	32	33	-	35	36	37				-	-	-	-	-	-	-	-	-	-	32	33	34	35	-	-	-	-								
รวม	33												5												22											
ข้อสอบที่ ไม่ลำเอียง	-	-	-	-	-	-	-	-	-	10	1	2	-	4	-	6	7	8	9	10	-	2	-	4	-	6	7	-	9	-						
	-	-	-	-	-	-	-	-	-	20	11	-	13	14	15	16	17	18	19	20	-	-	-	-	-	-	-	18	19	-						
	-	-	23	-	-	-	-	-	-	-	21	-	23	24	25	26	-	28	29	30	21	-	-	24	25	26	-	-	29	-						
	-	-	-	34	-	-	-	-	-		31	32	33	34	35	36	37				31	-	-	-	-	36	37									
รวม	4												32												15											

4. ผลการทดสอบความแตกต่างของจำนวนข้อสอบที่ล่าเอียงจากการวิเคราะห์ด้วย 3 วิธีของแบบทดสอบแต่ละวิชา

จำนวนข้อสอบที่ล่าเอียงในแต่ละวิธีที่วิเคราะห์ได้ของแบบทดสอบแต่ละวิชาในตารางที่ 24-35 นั้น สามารถทดสอบความแตกต่างของจำนวนข้อสอบที่ล่าเอียงในแต่ละวิธี ได้โดยใช้สถิตินอนพาราเมตริก วิธี Cochran Q Test

ตาราง 36 แสดงค่า Q จากการทดสอบความแตกต่างของจำนวนข้อสอบที่ล่าเอียงทั้ง 3 วิธี ของแบบทดสอบแต่ละวิชา

แบบทดสอบวิชา	Cochran Q Test	
	ฉบับที่ 1	ฉบับที่ 2
ภาษาไทย ก	27.1578**	6.7407*
สังคมศึกษา ก	6.3913*	10.4571**
ภาษาอังกฤษ กข	24.6666**	10.8888**
ภาษาไทย กข	11.3529**	10.5714**
สังคมศึกษา กข	10.7916**	9.3777**
ภาษาอังกฤษ กขค	18.5384**	37.3125**

จากตาราง 36 แสดงว่าวิธีวิเคราะห์ความล่าเอียงทั้ง 3 วิธี ตัดสินจำนวนข้อสอบที่ล่าเอียงในแบบทดสอบแต่ละวิชาได้แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($\chi^2_{2, .01} = 9.21$) ในทุกฉบับยกเว้นแบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 1 และวิชาภาษาไทย ก ฉบับที่ 2 ที่ตัดสินจำนวนข้อสอบที่ล่าเอียงแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 ($\chi^2_{2, .05} = 5.99$) ซึ่งโดยรวมแล้ววิธีใช้ทฤษฎีการตอบข้อคำถามตัดสินจำนวนข้อสอบที่ล่าเอียงได้จำนวนมากที่สุด ซึ่งสามารถทดสอบความแตกต่างรายคู่ ด้วยวิธีไคสแควร์ ดังแสดงในตาราง 37

ตาราง 37 แสดงค่าไคสแควร์ในการทดสอบความแตกต่างของจำนวนข้อสอบที่
ลำเอียงของการวิเคราะห์แต่ละวิธี

แบบทดสอบวิชา	ฉบับที่	ค่าไคสแควร์		
		IRT - SIB	SIB - MH	MH - IRT
ภาษาไทย ก	1	32.26**	5.35*	13.28**
	2	5.86*	1.82	1.18
สังคมศึกษา ก	1	3.13	1.78	1.54
	2	6.13*	0.03	6.97**
ภาษาอังกฤษ กข	1	10.34**	18.51**	1.44
	2	3.30	14.20**	1.40
ภาษาไทย กข	1	12.30**	5.76*	1.38
	2	10.66**	2.14	3.48
สังคมศึกษา กข	1	28.66**	7.84**	0.14
	2	7.71**	6.72**	0.03
ภาษาอังกฤษ กขค	1	14.70**	6.56**	0.04
	2	42.42**	16.84**	8.56**

จากตาราง 37 พบว่าวิธีใช้ทฤษฎีการตอบข้อคำถามตัดสินจำนวนข้อสอบที่
ลำเอียงมากกว่าวิธี SIBTEST และวิธีแมนเทิล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่
ระดับ .01 ในวิชาภาษาไทย ก ฉบับที่ 1 สังคมศึกษา ก ฉบับที่ 2 ภาษาอังกฤษ
กข ฉบับที่ 1 ภาษาไทย กข ฉบับที่ 1 และ 2 สังคมศึกษา กข ฉบับที่ 1 และ 2
และวิชาภาษาอังกฤษ กขค ฉบับที่ 1 และ 2 ส่วนวิชาสังคมศึกษา ก ฉบับที่ 2
ตัดสินจำนวนข้อสอบที่ลำเอียงมากกว่าวิธี SIBTEST อย่างมีนัยสำคัญที่ระดับ .05
นอกนั้นไม่แตกต่างกัน ส่วนวิธี SIBTEST ตัดสินจำนวนข้อสอบที่ลำเอียงมากกว่า
วิธีแมนเทิล-แฮนส์เซลอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในวิชาภาษาอังกฤษ กข

ตาราง 38 แสดงสาเหตุความล้มเหลวของข้อสอบวิชาภาษาไทย ก จำนวน 9 ข้อ

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล้มเหลวของข้อสอบ	ผู้สอบ				นิสิตนักศึกษา				รวม			
			อันดับที่				อันดับที่				อันดับที่			
			1	2	3	%	1	2	3	%	1	2	3	%
1	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความแล้วย้อนไม่เกี่ยวข้อง กับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลขเป็นคำ/ข้อความที่พบเห็นหรือให้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	-	4	12		4	8	14		2	6	13	
(9)			22	26	16	32	18	28	27	22	12			
			-	22	18	6	4	2	3	13	10			
			-	8	-	2	10	6	1	9	3			
			32	18	20	18	36	20	25	26	20			
			-	-	-	-	-	4	-	2				
			46	20	30	38	24	24	42	22	27			
			-	2	4	-	-	2	-	1	3			

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าช้าของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม			
			อันดับที่				อันดับที่				อันดับที่			
			1	2	3	%	1	2	3	%	1	2	3	%
2	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเยือกไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	32	4	10		4	10	8	18	7	9		
(10)			20	14	8	52	24	16	36	19	12			
			6	12	6	6	6	6	6	9	6			
			6	36	6	16	18	8	11	27	7			
			12	22	36	4	22	26	8	22	31			
			-	-	-	2	2	10	1	1	5			
			24	10	30	12	16	28	18	13	29			
			-	2	4	4	-	4	2	1	4			

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอบ				นิสิตนักศึกษา				รวม			
			อันดับที่				อันดับที่				อันดับที่			
			1	2	3	%	1	2	3	%	1	2	3	%
3	หญิง	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อความที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อความที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความแล้วย้อนไม่เกี่ยวข้อง 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อความที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อความที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	2	2	2		-	8	6	1	5	4		
(15)			12	22	18	24	22	22	-	18	22	9		
			46	28	14	14	26	16	30	27	15			
			-	4	6	10	8	16	5	6	11			
			12	26	18	26	10	22	19	18	14			
			-	-	-	-	-	8	-	-	4			
			26	18	42	24	26	22	25	22	32			
			2	-	-	2	-	4	2	-	2			

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าเียงของข้อสอบ	ผู้สอน			นิสิตนักศึกษา			รวม			
			อันดับที่			อันดับที่			อันดับที่			
			1	2	3	1	2	3	1	2	3	
			%	%	%	%	%	%	%	%	%	
4	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความแล้วย้อนไม่เกี่ยวข้อง กับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	2	2	6	6	10	6	18	6	4	12
(29)			16	44	28	14	20	16	15	32	22	
			-	2	6	8	6	4	4	4	5	
			6	6	6	18	18	6	12	12	6	
			22	32	18	16	18	22	19	25	20	
			-	2	2	2	10	4	1	6	3	
			52	10	34	32	18	24	42	14	29	
			2	2	-	-	2	6	1	2	3	

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม			
			อันดับที่				อันดับที่				อันดับที่			
			1	2	3	%	1	2	3	%	1	2	3	%
5	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเขียนย่อไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	2	2	4	8	4	4	16	5	3	10		
(31)			22	38	20	14	28	28	18	33	24			
			28	2	2	6	8	17	4	5				
			4	6	12	8	12	6	9	12				
			14	32	10	24	20	16	19	26	13			
			-	2	6	-	10	-	6	3				
			28	16	46	34	20	31	18	33				
			2	2	-	6	-	4	1	-				

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน			นิสิตนักศึกษา			รวม		
			อันดับที่			อันดับที่			อันดับที่		
			1	2	3	1	2	3	1	2	3
			%	%	%	%	%	%	%	%	%
6 (34)	หญิง	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเย็นไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	2	6	10	8	2	6	5	4	6
			52	10	22	26	22	14	39	16	18
			6	6	2	12	16	6	9	11	4
			6	-	6	4	12	18	5	6	12
			4	32	28	36	6	10	20	19	19
			-	-	2	2	6	4	1	3	3
			28	46	24	12	34	32	20	40	28
			2	-	6	-	2	-	1	1	3

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม			
			อันดับที่				อันดับที่				อันดับที่			
			1	2	3	%	1	2	3	%	1	2	3	%
7	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเย็นไม่เกี่ยวข้อง กับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 9. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	6	6	6	28	12	8	17	9	7			
(38)			14	18	30	6	20	12	10	19	21			
			44	10	8	20	12	12	32	11	10			
			8	6	6	4	18	8	6	12	7			
			6	14	16	14	14	22	10	14	19			
			-	2	4	2	8	16	1	5	10			
			22	36	30	18	18	10	20	27	20			
			-	8	18	8	4	12	4	6	15			

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าช้าของข้อสอบ	ผู้สอน			นิสิตนักศึกษา			รวม		
			อันดับที่			อันดับที่			อันดับที่		
			1	2	3	1	2	3	1	2	3
			%	%	%	%	%	%	%	%	%
8	หญิง	1. ข้อสอบยากเกินไป ทำให้กลัวเสียงที่จะเจอ 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเย็นไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	2	10	6	14	6	10	8	8	8
(39)			38	26	12	26	8	20	32	17	16
			16	24	10	8	12	16	12	18	13
			10	6	4	22	12	6	16	9	5
			6	6	44	12	14	16	9	10	30
			-	2	4	-	8	8	-	5	6
			26	24	14	12	40	24	19	32	19
			2	2	4	6	-	-	4	1	2

ตาราง 38 (ต่อ)

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าช้าของข้อสอบ	ผู้สอน			นิสิตนักศึกษา			รวม		
			อันดับที่			อันดับที่			อันดับที่		
			1	2	3	%	1	2	3	%	1
9	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเยือกไม่เกี่ยวข้องกับข้อสอบ ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	6	2	10	16	14	14	11	8	12
(49)			14	16	20	30	14	10	22	15	15
			38	6	4	6	6	14	22	6	9
			20	10	8	8	6	16	14	8	12
			4	16	22	12	32	4	8	24	13
			-	2	10	-	8	10	-	5	10
			18	46	18	16	16	20	17	31	19
			-	2	8	12	4	12	6	3	10

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม			
			อันดับที่				อันดับที่				อันดับที่			
			1 %	2 %	3 %		1 %	2 %	3 %		1 %	2 %	3 %	
2	หญิง	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 5. ข้อคำถามให้อ่านข้อความเย็นเยื่อไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 6. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 7. เวลาในการทำแบบทดสอบน้อยเกินไป 8. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 9. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	-	-	2		2	-	14	1	-	8		
(11)			66	8	24	58		24	14	62	16	19		
			16	12	4	6		14	8	11	13	6		
			4	6	14	4		12	10	4	9	12		
			8	36	30	18		42	6	13	39	18		
			-	2	-	-		-	8	-	1	4		
			6	24	32	12		8	40	9	16	36		
			-	-	-	-		-	-	-	-	-		

ตาราง 39 (ต่อ)

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าช้าของข้อสอบ	ผู้สอน			นิสิตนักศึกษา			รวม		
			อันดับที่			อันดับที่			อันดับที่		
			1	2	3	1	2	3	1	2	3
			%	%	%	%	%	%	%	%	%
4 (20)	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา	20	4	10	22	8	-	21	6	5
		2. เป็นข้อความที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน	52	22	20	20	30	24	36	26	22
		3. เป็นข้อความที่ถามถึงความจำ	2	16	6	8	10	12	5	13	9
		4. ข้อคำถามให้อ่านข้อความแล้วย้อนไม่เกี่ยวข้อง									
		5. ข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ	2	-	2	8	2	6	5	1	4
		6. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม	12	30	14	10	24	24	11	27	19
		7. เวลาในการทำแบบทดสอบน้อยเกินไป	-	-	10	-	-	6	-	-	8
		8. เป็นข้อความที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า	12	28	38	26	18	22	19	23	30
		เป็นข้อความที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	-	-	-	6	4	6	3	2	3

ตาราง 39 (ต่อ)

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าช้าของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม			
			อันดับที่				อันดับที่				อันดับที่			
			1	2	3	%	1	2	3	%	1	2	3	%
5 (28)	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา	14	10	10		4	18	26	14	14	14	18	
		2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน	30	32	26	14	32	10	22	32	18			
		3. เป็นข้อคำถามที่ถามความจำ	28	8	4	50	10	6	39	9	5			
		4. ข้อคำถามให้อ่านข้อความเย็นเยื่อไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ	-	-	2	-	6	8	-	3	5			
		5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม	22	18	22	12	8	18	17	13	20			
		6. เวลาในการทำแบบทดสอบน้อยเกินไป	-	4	4	-	6	8	-	5	6			
		7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า	6	28	30	20	20	20	13	24	25			
		8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	-	-	2	-	-	4	-	-	3			

ตาราง 39 (ต่อ)

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าช้าของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม				
			อันดับที่				อันดับที่				อันดับที่				
			1	2	3	%	1	2	3	%	1	2	3	%	
6	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความแล้วย้อนไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	10	2	14	30	8	16	20	5	15				
(38)			34	28	32	6	12	14	20	20	23				
			12	18	10	18	28	26	15	23	18				
			-	4	4	8	18	2	4	11	3				
			34	20	20	14	14	6	24	17	13				
			-	-	-	2	6	10	1	3	5				
			10	28	16	22	14	26	16	21	21				
			-	-	2	-	-	-	-	-	-				

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม			
			อันดับที่				อันดับที่				อันดับที่			
			1	2	3	%	1	2	3	%	1	2	3	%
			%	%	%	%	%	%	%	%	%	%	%	%
7	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความแล้วย้อนไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	4	2	6	12	18	18	20	16	8	10	12	
(39)			24	38	24	8	18	20	16	23	22	7		
			62	6	8	60	6	16	6	61	11	7		
			-	-	6	4	14	6	2	7	6			
			2	18	22	4	12	12	3	15	17			
			-	-	2	2	2	4	1	1	3			
			8	34	32	10	18	30	9	26	31			
			-	2	-	-	-	2	-	2	-	2		

ตาราง 39 (ต่อ)

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าช้าของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม			
			อันดับที่				อันดับที่				อันดับที่			
			1	2	3	%	1	2	3	%	1	2	3	%
8	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความแล้วย้อนไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	-	8	4		8	2	14		4	5	9	
(40)			76	10	10	54	24	12	65	17	11			
			2	2	12	6	18	14	4	10	13			
			2	8	2	8	4	2	5	6	2			
			14	42	30	10	34	16	12	38	23			
			-	-	-	2	-	6	1	-	3			
			6	30	42	12	16	32	9	23	37			
			-	-	-	-	-	-	-	-	2			

ตาราง 39 (ต่อ)

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม			
			อันดับที่				อันดับที่				อันดับที่			
			1	2	3	%	1	2	3	%	1	2	3	%
9 (41)	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา	2	-	-	-	4	6	8	3	-	-	-	4
		2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน	44	42	14	58	16	6	51	29	10	10	10	10
		3. เป็นข้อคำถามที่ถามความจำ	30	12	12	4	18	22	17	15	17	17	17	17
		4. ข้อคำถามให้อ่านข้อความแล้วย้อนไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ	2	-	12	2	2	2	2	1	7	7	7	7
		5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม	10	38	10	8	8	26	18	9	32	14	14	14
		6. เวลาในการทำแบบทดสอบน้อยเกินไป	-	-	2	-	-	6	8	-	3	5	5	5
		7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า	12	6	40	18	26	26	15	16	33	33	33	33
		8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	-	2	-	-	6	-	10	3	1	5	5	5

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าเียงของข้อสอบ	ผู้สอน		นิสิตนักศึกษา				รวม		
			อันดับที่		อันดับที่				อันดับที่		
			1	2	3	1	2	3	1	2	3
			%	%	%	%	%	%	%	%	%
10	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเยื่อไม่เกี่ยวข้องกับข้อสอบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	8	4	-	10	10	12	9	7	6
(42)			50	26	6	20	22	14	35	24	10
			30	6	16	50	22	2	40	14	9
			6	-	4	2	12	16	4	6	10
			-	22	42	-	10	16	-	16	29
			-	2	6	4	4	12	2	3	9
			6	40	24	12	20	24	9	30	24
			-	-	2	2	-	2	1	-	2

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม			
			อันดับที่				อันดับที่				อันดับที่			
			1	2	3	%	1	2	3	%	1	2	3	%
11 (53)	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา	10	6	2	16	16	12	13	11	7			
		2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน	20	24	20	16	4	12	18	14	16			
		3. เป็นข้อคำถามที่ถามความจำ	52	8	6	40	34	16	46	21	11			
		4. ข้อคำถามให้อ่านข้อความแล้วย้อนไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ	2	6	2	2	8	-	2	7	1			
		5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม	4	26	34	6	10	20	5	18	27			
		6. เวลาในการทำแบบทดสอบน้อยเกินไป	-	2	6	-	10	12	-	6	9			
		7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า	12	26	28	12	18	24	12	22	26			
		8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	-	2	2	6	-	2	3	1	2			

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน			นิสิตนักศึกษา			รวม		
			อันดับที่			อันดับที่			อันดับที่		
			1	2	3	1	2	3	1	2	3
			%	%	%	%	%	%	%	%	%
12	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเยื่อไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	4	4	-	10	16	6	7	10	3
(55)			56	16	12	20	10	12	38	13	12
			2	2	18	50	20	12	26	11	15
			18	10	-	2	12	8	10	11	4
			6	50	12	-	18	10	3	34	11
			4	8	2	4	16	14	4	12	8
			8	4	36	12	4	30	10	4	33
			2	6	20	2	4	8	2	5	14

ตาราง 39 (ต่อ)

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าช้าของข้อสอบ	ผู้สอน		นิสิตนักศึกษา				รวม																				
			อันดับที่		อันดับที่				อันดับที่																				
			1	2	3	%	1	2	3	%	1	2	3	%															
13	ชาย	1. ข้อสอบยากเกินไป ทำให้กลัวสิ่งที่จะเจอ 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเยือกไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	4	32	10	12	8	14	8	20	12	4	32	10	12														
(63)			30	12	10	22	8	10	26	10	10	6	4	4	10	18	8	11	6										
			28	10	16	28	16	14	28	13	15	14	18	20	18	-	22	2	4	1	3								
			12	10	10	12	10	14	12	10	12	14	12	10	12	14	12	10	12	12	6	4	24	10	4	10	8	4	17

ตาราง 40 แสดงสาเหตุความล้มเหลวของข้อสอบวิชา ภาษาอังกฤษ กข จำนวน 6 ข้อ

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล้มเหลวของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม				
			อันดับที่				อันดับที่				อันดับที่				
			1	2	3	%	1	2	3	%	1	2	3	%	
1	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเยื่อไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	12	4	6		6	10	20		9	7	13		
(4)			24	34	18		38	14	-		31	24	9		
			14	2	8		8	14	10		11	8	9		
			2	8	4		8	6	8		5	7	6		
			14	28	24		12	30	16		13	29	20		
			-	4	-		8	6	8		4	5	4		
			30	20	36		20	20	32		25	20	34		
			4	-	4		-	-	2		2	-	3		

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม			
			อันดับที่				อันดับที่				อันดับที่			
			1	2	3	%	1	2	3	%	1	2	3	%
2	หญิง	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเช่นย่อไม่เกี่ยวข้อง 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	-	-	2	4	2	6	2	1	4			
(6)			16	38	30	20	14	20	18	26	25			
			8	10	14	16	6	18	12	8	16			
			8	-	8	28	16	-	18	8	4			
			14	26	30	20	20	24	17	23	27			
			-	4	-	4	-	2	2	2	1			
			52	22	10	8	40	18	30	31	14			
			2	-	6	-	2	2	1	1	4			

ตาราง 40 (ต่อ)

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม				
			อันดับที่				อันดับที่				อันดับที่				
			1	2	3	%	1	2	3	%	1	2	3	%	
3	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเยื่อไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	34	8	26	32	26	18	31	17	22				
(15)			12	28	4	26	26	6	19	27	5				
			4	12	2	6	4	6	5	8	4				
			4	4	12	14	2	-	9	3	6				
			12	6	18	4	14	28	8	10	23				
			4	16	20	6	4	2	5	10	11				
			28	22	10	12	24	22	20	23	16				
			2	4	8	-	-	12	1	2	10				

ตาราง 40 (ต่อ)

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน			นิสิตนักศึกษา			รวม		
			อันดับที่			อันดับที่			อันดับที่		
			1	2	3	1	2	3	1	2	3
			%	%	%	%	%	%	%	%	%
4 (20)	หญิง	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา	8	4	6	6	6	6	7	5	6
		2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน	22	32	14	38	10	16	30	21	15
		3. เป็นข้อคำถามที่ถามความจำ	12	6	14	12	10	4	12	8	14
		4. ข้อคำถามให้อ่านข้อความเย็นเยือกไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ	10	12	-	26	4	16	18	8	8
		5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม	22	28	24	6	26	16	14	27	20
		6. เวลาในการทำแบบทดสอบน้อยเกินไป	-	-	6	8	2	14	4	1	10
		7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า	26	18	34	4	40	26	15	29	30
		8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	-	-	2	-	2	2	-	1	2

ตาราง 40 (ต่อ)

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน			นิสิตนักศึกษา			รวม		
			อันดับที่			อันดับที่			อันดับที่		
			1	2	3	1	2	3	1	2	3
			%	%	%	%	%	%	%	%	%
5	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อความที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อความที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความแล้วย้อนไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อความที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อความที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	20	8	2	28	14	20	24	11	11
(35)			4	20	24	8	24	4	6	22	14
			26	20	12	22	12	16	24	18	14
			4	10	-	4	6	6	4	8	3
			16	14	24	2	2	30	9	8	27
			6	6	6	6	8	10	6	7	8
			22	16	20	22	18	8	22	17	14
			2	6	12	8	16	6	5	11	9

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม			
			อันดับที่				อันดับที่				อันดับที่			
			1	2	3	%	1	2	3	%	1	2	3	%
6	หญิง	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเยือกไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	10	2	10		6	6	18		8	4	14	
(38)			6	26	14	14	14	14	12	10	20	13		
			6	16	4	10	10	8	8	13	6			
			24	10	14	10	10	6	17	10	10			
			10	8	22	4	20	18	7	14	20			
			4	12	2	12	8	-	8	10	1			
			34	24	20	22	20	26	28	22	23			
			6	2	14	2	12	12	4	7	13			

ตาราง 41 แสดงสาเหตุความล้มเหลวของข้อสอบวิชา ภาษาไทย กข จำนวน 3 ข้อ

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล้มเหลวของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม			
			อันดับที่				อันดับที่				อันดับที่			
			1	2	3	%	1	2	3	%	1	2	3	%
1 (22)	ชาย	1. ข้อสอบยากเกินไป ทำให้กลัวเสี่ยงที่จะเดา	-	8	4	14	10	10	7	9	7	7		
		2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน	26	20	24	38	24	10	32	22	17			
		3. เป็นข้อคำถามที่ถามความจำ	-	-	2	6	-	2	3	-	2			
		4. ข้อคำถามให้อ่านข้อความแล้วย้อนไม่เกี่ยวข้อง												
		5. ตัวเลขเป็นค่า/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม	38	14	6	14	40	14	26	27	10			
		6. เวลาในการทำแบบทดสอบน้อยเกินไป	18	32	8	10	6	12	14	19	10			
		7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า	-	10	8	6	4	10	3	7	9			
		8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	16	14	42	8	14	34	12	14	38			
			2	2	6	4	2	8	3	2	7			

ตาราง 41 (ต่อ)

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าเยี่ยงของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม			
			อันดับที่				อันดับที่				อันดับที่			
			1	2	3	%	1	2	3	%	1	2	3	%
3 (38)	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเยื่อไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาดูขบถน้อย	4	4	12	24	20	20	20	14	12	16		
			12	18	12	4	6	8	8	12	10			
			22	22	24	28	12	4	25	17	14			
			4	10	10	12	8	4	8	9	7			
			4	24	18	10	16	22	7	20	20			
			-	-	6	-	18	14	-	9	10			
			52	22	10	14	14	22	33	18	16			
			2	-	8	8	6	6	5	3	7			

ตาราง 42 แสดงสาเหตุความล่าช้าของข้อสอบวิชาสังคมศึกษา กข จำนวน 8 ข้อ

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าช้าของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม			
			อันดับที่				อันดับที่				อันดับที่			
			1	2	3	%	1	2	3	%	1	2	3	%
1	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา	4	-	2		6	-	6		5	-	4	
(13,		2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน	18	40	22	38	16	4	28	28	13			
61)		3. เป็นข้อคำถามที่ถามความจำ	16	12	14	32	28	14	24	20	14			
		4. ข้อคำถามให้อ่านข้อความเย็นก่อนไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ	-	2	4		2	24	10	1	13	7		
		5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม	12	28	26		2	24	24	7	26	25		
		6. เวลาในการทำแบบทดสอบน้อยเกินไป	-	4	4		-	6	10	-	5	7		
		7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า	50	12	22	18	14	32	34	13	27			
		8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	-	2	6		2	6	-	1	4	3		

ตาราง 42 (ต่อ)

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าช้าของข้อสอบ	ผู้สอน			นิสิตนักศึกษา			รวม		
			อันดับที่			อันดับที่			อันดับที่		
			1	2	3	1	2	3	1	2	3
			%	%	%	%	%	%	%	%	%
2	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเยื่อไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	42	14	10	22	14	10	32	14	10
(14)			30	32	4	12	24	12	21	27	8
			6	8	10	28	14	10	17	11	10
			8	12	4	20	8	10	14	10	7
			12	8	22	-	18	16	6	13	19
			-	4	-	4	4	6	2	4	3
			2	24	50	14	18	32	8	21	41
			-	-	-	-	-	-	-	-	2
								4			

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าช้าของข้อสอบ	ผู้สอน			นิสิตนักศึกษา			รวม		
			อันดับที่			อันดับที่			อันดับที่		
			1	2	3	1	2	3	1	2	3
			%	%	%	%	%	%	%	%	%
3	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความแล้วย้อนไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้หมดเวลาตอบน้อย	4	2	4	6	8	20	5	5	12
(13, 25)			40	30	18	36	22	20	38	26	19
			40	16	6	24	22	10	32	19	8
			-	-	-	2	6	4	1	3	2
			6	30	26	12	26	18	9	28	22
			-	-	2	-	-	2	-	-	2
			10	22	42	18	16	26	14	19	34
			-	-	2	2	-	-	1	-	1

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าช้าของข้อสอบ	ผู้สอน			นิสิตนักศึกษา			รวม		
			อันดับที่			อันดับที่			อันดับที่		
			1	2	3	1	2	3	1	2	3
			%	%	%	%	%	%	%	%	%
4	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเยื่อไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	22	14	2	18	8	30	20	11	16
(29)			20	28	28	16	14	28	18	21	28
			12	8	12	26	30	6	17	19	9
			-	2	-	-	18	6	-	10	3
			10	34	16	12	10	10	11	22	13
			2	2	10	2	2	4	2	2	7
			34	12	28	24	16	14	29	14	21
			-	-	4	2	2	2	-	1	3

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าช้าของข้อสอบ	ผู้สอน			นิสิตนักศึกษา			รวม		
			อันดับที่			อันดับที่			อันดับที่		
			1	2	3	1	2	3	1	2	3
			%	%	%	%	%	%	%	%	%
5	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเยื่อไม่เกี่ยวข้อง กับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	26	16	8	14	14	32	20	15	20
(33)			16	12	16	6	14	8	11	13	12
			28	28	18	36	28	8	32	28	13
			16	10	8	26	14	10	21	12	9
			2	10	12	6	4	12	4	7	12
			-	2	8	6	4	8	3	3	8
			10	22	26	2	18	16	6	20	21
			2	-	4	4	4	6	3	2	5

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน			นิสิตนักศึกษา			รวม		
			อันดับที่			อันดับที่			อันดับที่		
			1	2	3	1	2	3	1	2	3
			%	%	%	%	%	%	%	%	%
6	ชาย	1. ข้อสอบยากเกินไป ทำให้กลัวสิ่งที่จะเจอ 2. เป็นข้อความที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อความที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเย็นไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อความที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อความที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	6	10	12	14	22	20	10	16	16
(37)			12	26	26	8	10	10	10	18	18
			42	6	14	50	20	16	46	13	15
			-	4	-	2	4	6	1	4	3
			24	8	28	12	6	4	18	7	16
			-	4	6	6	-	14	3	2	10
			16	42	12	6	32	26	11	37	19
			-	-	2	2	4	4	1	2	3

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน			นิสิตนักศึกษา			รวม		
			อันดับที่			อันดับที่			อันดับที่		
			1	2	3	1	2	3	1	2	3
			%	%	%	%	%	%	%	%	%
7	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อความที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อความที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความแล้วย้อนไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อความที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อความที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	4	4	28	6	12	34	5	8	31
(69)			6	12	8	4	8	10	5	10	9
			76	8	14	72	12	8	74	10	11
			-	2	-	2	10	14	1	6	7
			2	38	20	6	12	16	4	25	18
			-	-	4	4	10	12	2	5	8
			8	28	22	2	28	14	5	28	18
			4	8	4	4	8	10	4	8	7

ตาราง 42 (ต่อ)

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน			นิสิตนักศึกษา			รวม		
			อันดับที่			อันดับที่			อันดับที่		
			1	2	3	%	1	2	3	%	%
8 (76)	ชาย	ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน เป็นข้อคำถามที่ถามความจำ ข้อคำถามให้อ่านข้อความเย็นเยือกไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม เวลาในการทำแบบทดสอบน้อยเกินไป เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	12	14	4	12	16	20	12	17	10
			12	10	18	4	6	8	8	8	12
			58	16	10	60	10	20	59	13	15
			-	-	4	6	2	3	4	3	
			-	16	24	4	4	18	2	10	20
			-	4	4	4	8	18	2	6	11
			12	30	24	2	34	12	7	32	18
			6	10	12	8	10	8	7	10	10

ตาราง 43 แสดงสาเหตุความล้มเหลวของข้อสอบวิชา ภาษาอังกฤษ กขค จำนวน 6 ข้อ

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล้มเหลวของข้อสอบ	ผู้สอน				นิสิตนักศึกษา				รวม								
			อันดับที่				อันดับที่				อันดับที่								
			1	2	3	%	1	2	3	%	1	2	3	%					
1	(4)	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเยื่อไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	-	-	-	-	-	2	6	-	1	3	-	1	3				
(4)			22	44	18	20	36	8	21	40	13	8	24	18	16	24	13	20	20
			-	-	-	-	4	6	2	3	3	3	3	3	3	3	3	3	3
			28	18	22	22	16	16	25	17	19	17	19	19	19	19	19	19	
			-	4	-	4	-	-	6	2	3	2	3	3	3	3	3	3	
			40	10	40	32	24	34	36	17	37	17	37	37	37	37	37	37	
			2	-	2	-	-	-	-	1	-	1	-	1	-	1	-	1	

ตาราง 43 (ต่อ)

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน		นิสิตนักศึกษา				รวม		
			อันดับที่		อันดับที่				อันดับที่		
			1	2	3	1	2	3	1	2	3
3 (17)	หญิง	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเขียนย่อไม่เกี่ยวข้อง กับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	-	14	4	2	10	8	1	12	6
			6	24	44	28	24	6	17	24	34
			24	8	4	14	12	10	20	10	7
			8	8	4	20	10	16	14	9	7
			8	16	20	2	14	34	10	15	27
			4	2	10	8	4	2	6	3	6
			46	26	14	24	26	20	35	26	17
			4	2	-	2	-	4	3	1	2

ตาราง 43 (ต่อ)

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความล่าเียงของข้อสอบ	ผู้สอน		นิสิตนักศึกษา				รวม		
			อันดับที่		อันดับที่				อันดับที่		
			1	2	3	1	2	3	1	2	3
			%	%	%	%	%	%	%	%	%
4	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเช่นย่อไม่เกี่ยวข้อง 5. ข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 6. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 7. เวลาในการทำแบบทดสอบน้อยเกินไป 8. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 9. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	8	6	2	14	8	12	11	7	7
(18)			18	26	28	14	16	28	16	20	28
			32	2	-	6	12	4	20	7	2
			4	10	4	24	16	18	14	13	11
			4	18	30	10	10	4	7	14	17
			2	4	4	4	14	6	3	9	5
			28	30	28	28	18	22	28	24	25
			4	4	4	-	6	6	2	5	5

ข้อที่	กลุ่มที่ได้เปรียบ	สาเหตุความลำเอียงของข้อสอบ	ผู้สอน			นิสิตนักศึกษา			รวม		
			อันดับที่			อันดับที่			อันดับที่		
			1	2	3	1	2	3	1	2	3
			%	%	%	%	%	%	%	%	%
5	ชาย	1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา 2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มสนใจและพบเห็นในชีวิตประจำวัน 3. เป็นข้อคำถามที่ถามความจำ 4. ข้อคำถามให้อ่านข้อความเย็นเยือกไม่เกี่ยวข้อง กับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ 5. ตัวเลือกเป็นคำ/ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม 6. เวลาในการทำแบบทดสอบน้อยเกินไป 7. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า 8. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย	-	-	12	4	6	10	2	3	11
(33)			4	10	20	10	22	10	7	16	15
			16	10	10	-	8	6	8	9	8
			40	12	2	28	20	8	34	16	5
			4	16	16	10	18	6	7	17	11
			4	6	14	8	4	8	6	5	9
			32	42	16	20	14	30	26	31	15
			-	4	10	2	8	12	1	6	11

ตาราง 44 สาเหตุอันดับ 1 ของความลำเอียงของข้อสอบระหว่างชายและหญิง จำแนกตาม รายวิชาและผู้ระบุสาเหตุ

แบบทดสอบ วิชา	กลุ่ม ได้เปรียบ	ผู้ระบุสาเหตุ		
		อาจารย์ผู้สอน	นิสิตนักศึกษา	รวม
ภาษาไทย ก	ชาย (6 ข้อ)	เป็นข้อคำถามที่ผู้สอบ ได้รับการฝึกฝนเฉพาะ เรื่องนั้นๆ จะมีโอกาส ตอบถูกมากกว่า	เป็นข้อคำถามที่ผู้สอบ ได้รับการฝึกฝนเฉพาะ เรื่องนั้นๆ จะมีโอกาส ตอบถูกมากกว่า	เป็นข้อคำถามที่ผู้สอบ ได้รับการฝึกฝนเฉพาะ เรื่องนั้นๆ จะมีโอกาส ตอบถูกมากกว่า
	หญิง (3 ข้อ)	เป็นข้อคำถามที่ผู้หญิง สนใจและพบเห็นใน ชีวิตประจำวัน	เป็นข้อคำถามที่ผู้หญิง สนใจและพบเห็นใน ชีวิตประจำวัน	เป็นข้อคำถามที่ผู้หญิง สนใจและพบเห็นใน ชีวิตประจำวัน
สังคมศึกษา ก	ชาย (12 ข้อ)	เป็นข้อคำถามที่ผู้ชาย สนใจและพบเห็นใน ชีวิตประจำวัน	เป็นข้อคำถามที่ถาม ความจำ	เป็นข้อคำถามที่ผู้ชาย สนใจและพบเห็น ในชีวิต ประจำวัน
	หญิง (1 ข้อ)	เป็นข้อคำถามที่ผู้หญิง สนใจและพบเห็นใน ชีวิตประจำวัน	เป็นข้อคำถามที่ผู้หญิง สนใจและพบเห็นใน ชีวิตประจำวัน	เป็นข้อคำถามที่ผู้หญิง สนใจและพบเห็น ในชีวิต ประจำวัน
ภาษาอังกฤษ กข	ชาย (3 ข้อ)	เป็นข้อคำถามที่ผู้สอบ ได้รับการฝึกฝนเฉพาะ เรื่องนั้นๆ จะมีโอกาส ตอบถูกมากกว่า	เป็นข้อคำถามที่ผู้ชาย สนใจและพบเห็น ใน ชีวิตประจำวัน	เป็นข้อคำถามที่ผู้สอบ ได้รับการฝึกฝนเฉพาะ เรื่องนั้นๆ จะมีโอกาส ตอบถูกมากกว่า

ตาราง 44 (ต่อ)

แบบทดสอบ วิชา	กลุ่ม ได้เปรียบ	ผู้ระบุสาเหตุ		
		อาจารย์ผู้สอน	นิสิตนักศึกษา	รวม
	หญิง (3 ข้อ)	เป็นข้อคำถามที่ผู้สอบ ได้รับการฝึกฝนเฉพาะ เรื่องนั้นๆ จะมีโอกาส ตอบถูกมากกว่า	ข้อคำถามให้อ่าน ข้อความเย็นเยื่อ ไม่ เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่าน ได้เปรียบ	เป็นข้อคำถามที่ผู้สอบ ได้รับการฝึกฝนเฉพาะ เรื่องนั้นๆ จะมีโอกาส ตอบถูกมากกว่า
ภาษาไทย กข	ชาย (3 ข้อ)	เป็นข้อคำถามที่ผู้สอบ ได้รับการฝึกฝนเฉพาะ เรื่องนั้นๆ จะมีโอกาส ตอบถูกมากกว่า	เป็นข้อคำถามที่ผู้ชาย สนใจและพบเห็นใน ชีวิตประจำวัน	เป็นข้อคำถามที่ผู้สอบ ได้รับการฝึกฝนเฉพาะ เรื่องนั้นๆ จะมีโอกาส ตอบถูกมากกว่า
สังคมศึกษา กข	ชาย (6 ข้อ)	เป็นข้อคำถามที่ถาม ความจำ	เป็นข้อคำถามที่ถาม ความจำ	เป็นข้อคำถามที่ถาม ความจำ
ภาษาอังกฤษ กขค	หญิง (6 ข้อ)	เป็นข้อคำถามที่ผู้สอบ ได้รับการฝึกฝนเฉพาะ เรื่องนั้นๆ จะมีโอกาส ตอบถูกมากกว่า	เป็นข้อคำถามที่ผู้สอบ ได้รับการฝึกฝนเฉพาะ เรื่องนั้นๆ จะมีโอกาส ตอบถูกมากกว่า	เป็นข้อคำถามที่ผู้สอบ ได้รับการฝึกฝนเฉพาะ เรื่องนั้นๆ จะมีโอกาส ตอบถูกมากกว่า

จากตาราง 44 พบว่าในวิชาภาษาไทย ก ภาษาอังกฤษ กขค สังคมศึกษา กข และ สังคมศึกษา ก เฉพาะข้อสอบที่กลุ่มหญิงได้เปรียบสาเหตุของความลำเอียงในอันดับ 1 จะสอดคล้องกันทั้งอาจารย์ผู้สอนและนิสิต/นักศึกษา ซึ่งจะเป็นข้อคำถามที่พบนั่นสนใจพบเห็นในชีวิตประจำวัน ข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า และ

เป็นข้อคำถามที่ถามความจำ ส่วนวิชาสังคมศึกษา ก เฉพาะข้อที่ช่ายได้เปรียบอาจารย์ผู้สอนระบุว่าสาเหตุของความลำเอียงอันดับ 1 เกิดจากข้อคำถาม เป็นเรื่องที่ผู้ช่ายสนใจและพบเห็นในชีวิตประจำวัน แต่นิสิต/นักศึกษาระบุว่าเป็นข้อคำถามที่ถามความจำ แต่ถ้ารวมกันก็จะเป็นเรื่องที่ผู้ช่ายสนใจและพบเห็นในชีวิตประจำวันวิชาภาษาไทย กข และภาษาอังกฤษ กข ไม่ว่าจะเป็นเพศใดได้เปรียบ อาจารย์ผู้สอนระบุว่าสาเหตุมาจากข้อคำถามเป็นข้อที่ผู้สอนที่ได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า และรวมทั้งอาจารย์และนิสิต/นักศึกษา แต่เฉพาะนิสิต/นักศึกษา ระบุว่าในข้อสอบที่ช่ายได้เปรียบเกิดจากข้อคำถามให้อ่านข้อความเย็นเยื่อไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้ช่ายอ่านได้เปรียบ

6. ความสัมพันธ์ของผลการสอบที่คิดคะแนนมาตรฐานที่-ปกติ และคิดคะแนนน้ำหนักความสามารถที่ใช้ข้อสอบทั้งหมดกับใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง

จากการศึกษาผลการทดสอบที่พิจารณาการคิดคะแนนมาตรฐานที่-ปกติ และคะแนนน้ำหนักความสามารถ ทั้งที่ใช้จำนวนข้อสอบทั้งหมดกับใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง โดยศึกษาถึงความสัมพันธ์ของลำดับที่ของวิธีการคิดคะแนนทั้ง 2 วิธี กับจำนวนข้อสอบที่นำมาใช้คิดคะแนน ซึ่งค่าของความสัมพันธ์ลำดับที่ (Rank Correlation) ของแต่ละวิธีแสดงดังตาราง 45-46

ตาราง 45 แสดงค่าความสัมพันธ์ลำดับที่ของผลการสอบที่ให้คะแนน 2 วิธี และใช้ จำนวนข้อสอบทั้งหมดกับใช้เฉพาะข้อสอบที่ปราศจากความลำเอียงในผู้สอบสายวิทยาศาสตร์

	T-ALL	T-UB	θ-ALL	θ-UB
T-ALL	1.0000	.9999**	.9994**	.9992**
T-UB	-	1.0000	.9994**	.9992**
θ-ALL	-	-	1.0000	.9993**
θ-UB	-	-	-	1.0000

** มีนัยสำคัญที่ระดับ .01

จากตาราง 45 แสดงว่าลำดับที่ของผลการสอบในสายวิทยาศาสตร์ที่คิดคะแนนมาตรฐานที่-ปกติ คะแนนน้ำหนักความสามารถ และใช้จำนวนข้อสอบทั้งหมดกับใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง ต่างมีความสัมพันธ์ในลำดับที่อย่างมีนัยสำคัญที่ระดับ .01

ตาราง 46 แสดงค่าความสัมพันธ์ลำดับที่ของผลการสอบที่คิดคะแนน 2 วิธี และใช้จำนวนข้อสอบทั้งหมดกับใช้เฉพาะข้อสอบที่ปราศจากความลำเอียงในผู้สอบสายศิลปศาสตร์

	T-ALL	T-UB	θ-ALL	θ-UB
T-ALL	1.0000	.9999*	.9996	.9993
T-UB	-	1.0000	.9996	.9993
θ-ALL	-	-	1.0000	.9993
θ-UB	-	-	-	1.0000

** มีนัยสำคัญที่ระดับ .01

จากตาราง 46 แสดงว่าลำดับที่ของผลการสอบในสายศิลปศาสตร์ที่คิดคะแนนมาตรฐานที่-ปกติ คะแนนน้ำหนักความสามารถ และใช้จำนวนข้อสอบทั้งหมดกับใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง ต่างมีความสัมพันธ์ในลำดับที่อย่างมีนัยสำคัญที่ระดับ .01

7. ประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการเรียนของการคิดคะแนนมาตรฐานที่-ปกติ และคะแนนน้ำหนักความสามารถกับการใช้จำนวนข้อสอบทั้งหมดกับเฉพาะข้อสอบที่ปราศจากความลำเอียง ผลการวิเคราะห์ดังแสดงในตาราง 47-50

ตาราง 47 ค่าสัมประสิทธิ์ตัวทำนาย ค่าความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ตัวทำนาย ค่าความคลาดเคลื่อนมาตรฐานของการทำนาย ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ ค่าสัดส่วนความแปรปรวนที่ทำนายได้ และค่าทดสอบสัมประสิทธิ์ตัวทำนายในการทำนายผลสัมฤทธิ์ทางการเรียน ด้วยคะแนนสอบคัดเลือก 3 วิชา ที่คิดคะแนนมาตรฐานที่-ปกติ และใช้จำนวนข้อสอบทั้งหมด

สาย	เกณฑ์	ตัวทำนาย	b	β	SE_b	SE_{est}	R	R^2	F
วิทย	เกรดภาคเรียนที่ 1	ภาษาอังกฤษ กข	.004	.085	.002	.587	.135	.018	4.726**
		ภาษาไทย ก	.005	.104	.001				
		สังคมศึกษา ก	.001	.103	.001				
ศิลป์	เกรดภาคเรียนที่ 1	ภาษาอังกฤษ กขค	-1.951	-.142	.811	120.239	.159	.025	2.402
		ภาษาไทย กข	.621	.061	.597				
		สังคมศึกษา กข	.447	.037	.709				

** $p < .01$

$n = 763$

จากตาราง 47 เมื่อทำนายผลสัมฤทธิ์ทางการเรียนทั้งสายวิทยาศาสตร์ และสายศิลปศาสตร์ ด้วยคะแนนสอบคัดเลือก 3 วิชาที่คิดคะแนนมาตรฐานที่-ปกติ และใช้จำนวนข้อสอบทั้งหมดพบว่า

1. การทำนายผลสัมฤทธิ์ทางการเรียนในสายวิทยาศาสตร์ ตัวทำนายที่สามารถทำนายได้คือ วิชาภาษาอังกฤษ กข (X_3) และวิชาภาษาไทย ก (X_1) ได้ค่าสหสัมพันธ์พหุคูณ (R) เท่ากับ .135 ค่าสัดส่วนความแปรปรวนที่ทำนายได้ (R^2) เท่ากับ .018 และความคลาดเคลื่อนมาตรฐานของการทำนาย (SE_{est}) เท่ากับ .587 ซึ่งมีนัยสำคัญทางสถิติที่ระดับ .01 ($.01 F_{3,759} = 3.80$) สร้างสมการทำนายในรูปคะแนนมาตรฐานได้ด้วย

$$X'_1 = .085X_3 + .104X_1$$

2. การทำนายผลสัมฤทธิ์ทางการเรียนในสายศิลปศาสตรบัณฑิตทำนายที่สามารถทำนายได้ คือวิชาภาษาอังกฤษ กขค (X_6) ได้ค่าสหสัมพันธ์พหุคูณ (R) เท่ากับ .159 ค่าสัดส่วนความแปรปรวนที่ทำนายได้ (R^2) = .025 และความคลาดเคลื่อนมาตรฐานของการทำนาย (SE_{est}) เท่ากับ 120.239 ซึ่งไม่มีนัยสำคัญทางสถิติ สร้างสมการทำนายในรูปคะแนนมาตรฐานได้ด้วย

$$x'_2 = -.142x_6$$

ตาราง 48 ค่าสัมประสิทธิ์ตัวทำนาย ค่าความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ตัวทำนาย ค่าความคลาดเคลื่อนมาตรฐานของการทำนาย ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ ค่าสัดส่วนความแปรปรวนที่ทำนายได้ และค่าทดสอบสัมประสิทธิ์ตัวทำนายในการทำนายผลสัมฤทธิ์ทางการเรียน ด้วยคะแนนสอบคัดเลือก 3 วิชาที่คิดคะแนนมาตรฐานที่ปกติ และใช้ข้อสอบเฉพาะที่ปราศจากความลำเอียง

สาย	เกณฑ์	ตัวทำนาย	b	β	SE_b	SE_{est}	R	R^2	F
วิทย์	เกรดภาคเรียนที่ 1	ภาษาอังกฤษ กข	.004	.074	.001	.587	.129	.016	4.318
		ภาษาไทย ก	.005	.106	.001				
		สังคมศึกษา ก	.001	.018	.001				
ศิลป์	เกรดภาคเรียนที่ 1	ภาษาอังกฤษ กขค	-1.692	-.129	.775	120.395	.151	.022	2.156
		ภาษาไทย กข	.679	.066	.604				
		สังคมศึกษา กข	.508	.040	.739				

** $p < .01$

$n = 763$

จากตาราง 48 เมื่อทำนายผลสัมฤทธิ์ทางการเรียนทั้งสายวิทยาศาสตร์ และสายศิลปศาสตร์ ด้วยคะแนนสอบคัดเลือก 3 วิชาที่คิดคะแนนมาตรฐานที่ปกติ และใช้ข้อสอบเฉพาะที่ปราศจากความลำเอียง พบว่า

1. การทำนายผลสัมฤทธิ์ทางการเรียนในสายวิทยาศาสตร์ ตัวทำนายที่สามารถทำนายได้คือ วิชาภาษาอังกฤษ กข (X_3) และวิชาภาษาไทย ก (X_1) ได้ค่าสหสัมพันธ์พหุคูณ (R) เท่ากับ .129 ค่าสัดส่วนความแปรปรวนที่ทำนายได้ (R^2) เท่ากับ .016 และความคลาดเคลื่อนมาตรฐานของการทำนาย (SE_{est}) เท่ากับ .587 ซึ่งมีนัยสำคัญทางสถิติที่ระดับ .01 (.01 $F_{3,759} = 3.80$) สร้างสมการทำนายในรูปคะแนนมาตรฐานได้ด้วย

$$Z'_1 = .106Z_1 + .074Z_3$$

2. การทำนายผลสัมฤทธิ์ทางการเรียนในสายศิลป์ศาสตร์ ตัวทำนายที่สามารถทำนายได้ คือวิชาภาษาอังกฤษ กขค (X_6) ได้ค่าสหสัมพันธ์หาคู่ (R) เท่ากับ .151 ค่าสัดส่วนความแปรปรวนที่ทำนายได้ (R^2) = .022 และความคลาดเคลื่อนมาตรฐานของการทำนาย (SE_{est}) เท่ากับ 120.395 ซึ่งไม่มีนัยสำคัญทางสถิติ สร้างสมการทำนายในรูปคะแนนมาตรฐานได้ด้วย

$$X'_2 = -.129X_6$$

ตาราง 49 ค่าสัมประสิทธิ์ตัวทำนาย ค่าความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ตัวทำนาย ค่าความคลาดเคลื่อนมาตรฐานของการทำนาย ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ ค่าสัดส่วนความแปรปรวนที่ทำนายได้ และค่าทดสอบสัมประสิทธิ์ตัวทำนายในการทำนายผลสัมฤทธิ์ทางการเรียน ด้วยคะแนนสอบคัดเลือก 3 วิชาที่คิดคะแนนน้ำหนักความสามารถ และใช้จำนวนข้อสอบทั้งหมด

สาย	เกณฑ์	ตัวทำนาย	b	β	SE _b	SE _{est}	R	R ²	F
วิทย์	เกรดภาคเรียนที่ 1	ภาษาอังกฤษ กข	.034	.066	.025	.586	.141	.020	5.187**
		ภาษาไทย ก	.030	.051	.025				
		สังคมศึกษา ก	.026	.051	.024				
ศิลป์	เกรดภาคเรียนที่ 1	ภาษาอังกฤษ กขค	-13.48	-.126	7.909	121.112	.105	.011	1.040
		ภาษาไทย กข	6.493	.058	8.496				
		สังคมศึกษา กข	6.133	.058	8.028				

** $p < .01$

n = 763

จากตาราง 49 เมื่อทำนายผลสัมฤทธิ์ทางการเรียนทั้งสายวิทยาศาสตร์ และสายศิลปศาสตร์ ด้วยคะแนนสอบคัดเลือก 3 วิชาที่คิดคะแนนน้ำหนักความสามารถและใช้จำนวนข้อสอบทั้งหมด พบว่า

1. การทำนายผลสัมฤทธิ์ทางการเรียนในสายวิทยาศาสตร์ ไม่มีตัวแปรใดที่สามารถทำนายได้

2. การทำนายผลสัมฤทธิ์ทางการเรียนในสายศิลปศาสตร์ ไม่มีตัวแปรใดที่สามารถทำนายได้

ตาราง 50 ค่าสัมประสิทธิ์ตัวทำนาย ค่าความคลาดเคลื่อนมาตรฐานของสัมประสิทธิ์ตัวทำนาย ค่าความคลาดเคลื่อนมาตรฐานของการทำนาย ค่าสัมประสิทธิ์สหสัมพันธ์พหุคูณ ค่าสัดส่วนความแปรปรวนที่ทำนายได้ และค่าทดสอบสัมประสิทธิ์ตัวทำนายในการทำนายผลสัมฤทธิ์ทางการเรียน ด้วยคะแนนสอบคัดเลือก 3 วิชาที่คิดคะแนนน้ำหนักความสามารถ และใช้ข้อสอบเฉพาะที่ปราศจากความลำเอียง

สาย	เกณฑ์	ตัวทำนาย	b	β	SE_b	SE_{est}	R	R^2	F
วิทย์	เกรดภาคเรียนที่ 1	ภาษาอังกฤษ กข	-.011	-.021	.027	.583	.174	.030	7.979**
		ภาษาไทย ก	.062	.111	.025				
		สังคมศึกษา ก	.051	.104	.025				
ศิลป์	เกรดภาคเรียนที่ 1	ภาษาอังกฤษ กขค	-1.824	-.017	7.947	121.608	.055	.003	.281
		ภาษาไทย กข	5.371	.047	8.940				
		สังคมศึกษา กข	-5.861	-.059	7.675				

** $p < .01$ $n = 763$

จากตาราง 50 เมื่อทำนายผลสัมฤทธิ์ทางการเรียนทั้งสายวิทยาศาสตร์ และสายศิลป์ศาสตร์ ด้วยคะแนนสอบคัดเลือก 3 วิชาที่คิดคะแนนน้ำหนักความสามารถและใช้ข้อสอบเฉพาะที่ปราศจากความลำเอียง พบว่า

1. การทำนายผลสัมฤทธิ์ทางการเรียนในสายวิทยาศาสตร์ ตัวทำนายที่สามารถทำนายได้คือ วิชาภาษาอังกฤษ กข (X_1) และวิชาสังคมศึกษา ก (X_2) ได้ค่าสหสัมพันธ์พหุคูณ (R) เท่ากับ .174 ค่าสัดส่วนความแปรปรวนที่ทำนายได้ (R^2) เท่ากับ .030 และความคลาดเคลื่อนมาตรฐานของการทำนาย (SE_{est}) เท่ากับ .583 ซึ่งมีนัยสำคัญทางสถิติที่ระดับ .01 ($.01 F_{3,759} = 3.80$) สร้างสมการทำนายในรูปคะแนนมาตรฐานได้ด้วย

$$X'_1 = .111X_1 + .104X_2$$

2. การทำนายผลสัมฤทธิ์ทางการเรียนในสายศิลป์ศาสตร์ ไม่มีตัวแปรใดที่สามารถทำนายได้

สรุปผลการวิเคราะห์สหสัมพันธ์พหุคูณของการทำนายผลสัมฤทธิ์ทางการเรียนในแต่ละสาย การคิดคะแนน และการใช้จำนวนข้อสอบที่ต่างกัน แสดงในตาราง 51

ตาราง 51 ตัวแปรทำนายที่ทำนายผลสัมฤทธิ์ทางการเรียนในแต่ละสายที่คิดคะแนน และการใช้จำนวนข้อสอบที่ต่างกัน

สาย	วิธีคิดคะแนน	การใช้จำนวนข้อสอบ	ตัวทำนาย	β	R	SE _{cst}
วิทย	T	ALL	ภาษาไทย	.104	.135	.587
			ภาษาอังกฤษ กข	.085		
	T	UB	ภาษาไทย ก	.106	.129	.587
			ภาษาอังกฤษ กข	.074		
	๑	ALL	ไม่มี	-	-	-
	๑	UB	ภาษาไทย ก	.111	.174	.583
			สังคมศึกษา ก	.104		
ศิลป์	T	ALL	ภาษาอังกฤษ กขค	-.142	.159	120.239
	T	UB	ภาษาอังกฤษ กขค	-.129	.151	120.395
	๑	ALL	ไม่มี	-	-	-
	๑	UB	ไม่มี	-	-	-

จากตาราง 51 พบว่าวิธีคิดคะแนนตัวแปรทำนายที่ทำนายผลสัมฤทธิ์ทางการเรียนในแต่ละสายที่คิดคะแนน และใช้จำนวนข้อสอบที่ต่างกันได้ดีที่สุดคือ

1. สายวิทยาศาสตร์
 - 1.1 การคิดคะแนนมาตรฐานที-ปกติ และการใช้ข้อสอบทั้งหมดสามารถทำนายผลสัมฤทธิ์ทางการเรียนได้ ซึ่งตัวแปรทำนายดังกล่าวคือ วิชาภาษาไทย (X_1) และวิชาภาษาอังกฤษ กข (X_3) มีค่าสหสัมพันธ์พหุคูณ (R) เท่ากับ .135 สร้างสมการทำนายในรูปคะแนนมาตรฐานได้ดังนี้

$$Z' = .104Z + .085Z$$

1.2 การคิดคะแนนมาตรฐานที่-ปกติ และการใช้เฉพาะข้อสอบปราศจากความลำเอียง สามารถทำนายผลสัมฤทธิ์ทางการเรียนได้ ซึ่งตัวแปรทำนายดังกล่าวคือวิชาภาษาไทย ก (X_1) และภาษาอังกฤษ กข (X_3) มีค่าสหสัมพันธ์พหุคูณ (R) เท่ากับ .129 สร้างสมการทำนายในรูปคะแนนมาตรฐานได้ดังนี้

$$X'_2 = .106X_1 + .074X_2$$

1.3 การคิดคะแนนน้ำหนักความสามารถและใช้ข้อสอบทั้งหมดไม่มีวิชาใดที่สามารถทำนายผลสัมฤทธิ์ทางการเรียนได้

1.4 การคิดคะแนนน้ำหนักความสามารถและใช้เฉพาะข้อสอบปราศจากความลำเอียง สามารถทำนายผลสัมฤทธิ์ทางการเรียนได้ ซึ่งตัวแปรทำนายดังกล่าวคือวิชาภาษาไทย ก (X_1) และสังคมศึกษา ก (X_2) มีค่าสหสัมพันธ์พหุคูณ (R) เท่ากับ .174 สร้างสมการทำนายในรูปคะแนนมาตรฐานได้ดังนี้

$$X'_4 = .111X_1 + .104X_2$$

2. สายศิลปศาสตร์

2.1 การคิดคะแนนมาตรฐานที่-ปกติ และการใช้ข้อสอบทั้งหมดสามารถทำนายผลสัมฤทธิ์ทางการเรียนได้ ซึ่งตัวแปรทำนายดังกล่าวคือ วิชาภาษาอังกฤษ กขค (X_6) มีค่าสหสัมพันธ์พหุคูณ (R) เท่ากับ .159 สร้างสมการทำนายในรูปคะแนนมาตรฐานได้ดังนี้

$$X'_1 = -.142X_6$$

2.2 การคิดคะแนนมาตรฐานที่-ปกติ และการใช้เฉพาะข้อสอบปราศจากความลำเอียง สามารถทำนายผลสัมฤทธิ์ทางการเรียนได้ ซึ่งตัวแปรทำนายดังกล่าวคือวิชาภาษาอังกฤษ กขค (X_6) มีค่าสหสัมพันธ์พหุคูณ (R) เท่ากับ .151 สร้างสมการทำนายในรูปของคะแนนมาตรฐานได้ดังนี้

$$X'_2 = -.129X_6$$

2.3 การคิดคะแนนน้ำหนักความสามารถและใช้ข้อสอบทั้งหมดไม่มีวิชาใดที่สามารถทำนายผลสัมฤทธิ์ทางการเรียนได้

2.4 การคิดคะแนนน้ำหนักความสามารถและใช้เฉพาะข้อสอบปราศจากความลำเอียง ไม่มีวิชาใดที่สามารถทำนายผลสัมฤทธิ์ทางการเรียนได้

บทที่ 5

สรุปผล อภิปรายผล และข้อเสนอแนะ

จุดมุ่งหมายของการวิจัย

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่วิเคราะห์ความลำเอียงและการคิดคะแนนด้วยวิธีต่างกันโดยศึกษา

1. จำนวนข้อสอบที่ลำเอียงต่อเพศชายและเพศหญิงในวิชาต่าง ๆ เมื่อใช้วิธีการวิเคราะห์ความลำเอียงด้วยวิธีการต่างกัน
2. ความสอดคล้องของการตัดสินผลการสอบที่ใช้วิธีการคิดคะแนนต่างกัน
3. สาเหตุความลำเอียงของข้อสอบที่ลำเอียงต่อเพศชายและเพศหญิง

สมมติฐานในการวิจัย

1. เมื่อใช้วิธีการวิเคราะห์ความลำเอียงด้วยวิธีการใช้ทฤษฎีการตอบข้อคำถามวิธีแมนเทิล-แฮนส์เซล และวิธี SIBTEST ระหว่างเพศชายและหญิงต่างตัดสินความลำเอียงของข้อสอบแตกต่างกัน
2. การตัดสินผลการสอบที่คิดคะแนนมาตรฐานที่-ปกติ และคะแนนน้ำหนักความสามารถจากข้อสอบทั้งหมดกับคะแนนมาตรฐานที่-ปกติ และคะแนนน้ำหนักความสามารถจากข้อสอบที่ปราศจากความลำเอียงระหว่างเพศชายและเพศไม่สอดคล้องกัน

วิธีดำเนินการวิจัย

1. ติดต่อกับฝ่ายวิชาการของมหาวิทยาลัยสงขลานครินทร์ เพื่อขอใช้ข้อมูลการรับสมัครผลการตอบข้อสอบ ผลการสอบคัดเลือก และระดับคะแนนผลการเรียนในรายวิชาต่าง ๆ ในภาคเรียนที่ 1 ของปีการศึกษา 2538
2. นำผลการตอบข้อสอบคัดเลือกทั้งหมดวิเคราะห์ด้วยโปรแกรม SPSS-PC⁺ และวิเคราะห์องค์ประกอบเพื่อตรวจสอบความเป็นมิติเดียวของแบบทดสอบแต่ละฉบับ

3. วิเคราะห์หาคุณภาพของข้อสอบเป็นรายข้อในทุกวิชาโดย

3.1 วิเคราะห์หาคุณภาพรายข้อตามทฤษฎีการตอบข้อคำถาม ด้วยโปรแกรมสำเร็จรูป BILOG

3.2 วิเคราะห์ความลำเอียงของข้อสอบในแต่ละข้อ แต่ละวิชา ตามทฤษฎีการตอบข้อคำถาม วิธีแมนเทิล-แฮนส์เซล และวิธี SIBTEST ด้วยโปรแกรมสำเร็จรูป IRTDIF และ SIBTEST

3.3 ทดสอบความแตกต่างของจำนวนข้อสอบที่ลำเอียงในแต่ละฉบับ แต่ละวิชาทั้ง 3 วิธีของแบบทดสอบด้วยสถิติอนุพาราเมตริกแบบ Cochran Q Test และทดสอบรายคู่ ด้วยวิธีไคสแควร์ (χ^2)

4. นำข้อสอบที่ตรวจพบว่ามีคุณภาพต่ำลงทั้ง 3 วิธีไปให้ผู้ระบุสาเหตุ ระบุสาเหตุของความลำเอียงต่อเพศชายและหญิง แล้วสรุปสาเหตุของความลำเอียงข้อสอบแต่ละข้อ

5. นำคะแนนจากผลการสอบของข้อสอบทั้งหมดกับคะแนนจากผลการสอบที่คิดเฉพาะข้อสอบที่ปราศจากความลำเอียง มาหาความสัมพันธ์ในรูปของคะแนนมาตรฐานที่-ปกติ และคะแนนน้ำหนักความสามารถในแต่ละวิชา และหาความสัมพันธ์ตามลำดับที่ (Rank Correlation) ในคะแนนรวม

6. คำนวณค่าการถดถอยพหุคูณ (Multiple Regression Analysis) ของคะแนนสอบคัดเลือกกับระดับคะแนนที่เรียนในชั้นปีที่ 1 ของนักเรียนที่ได้รับการคัดเลือก โดยการคิดคะแนนมาตรฐานที่-ปกติ และคะแนนน้ำหนักความสามารถ

การวิเคราะห์ข้อมูล

ผู้วิจัยวิเคราะห์ข้อมูลในแต่ละวิชาโดยมีขั้นตอนเหมือนกันดังนี้

1. ตรวจสอบความเป็นมิติเดียวกันของแบบทดสอบแต่ละวิชา
2. วิเคราะห์คุณภาพรายข้อของแบบทดสอบแต่ละวิชา
3. วิเคราะห์ความลำเอียงของข้อสอบแต่ละวิชาทั้ง 3 วิธี
4. ทดสอบความแตกต่างของจำนวนข้อสอบที่ลำเอียงจากการวิเคราะห์ด้วย 3 วิธี
5. สรุปสาเหตุความลำเอียงของข้อสอบแต่ละวิชา
6. ความสัมพันธ์ของผลการสอบที่คิดคะแนนมาตรฐานที่-ปกติและคิดคะแนนน้ำหนักความสามารถที่ใช้ข้อสอบทั้งหมดกับใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง

7. ประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการเรียนของการคิดคะแนน และการใช้จำนวนข้อสอบ

สรุปผลการวิจัย

1. ผลการตรวจสอบความเป็นมิติเดียวของแบบทดสอบแต่ละฉบับพบว่าแบบทดสอบทั้ง 6 วิชา 12 ฉบับ มีความเป็นมิติเดียวกัน โดยมีค่าไอเกนในองค์ประกอบที่ 1 สูงกว่าค่าไอเกนขององค์ประกอบที่ 2, 3 และ 4 ประมาณ 3 เท่า

2. คุณภาพรายข้อของแบบทดสอบแต่ละฉบับ พบว่าเมื่อพิจารณารวมทุกกลุ่ม ทุกวิชา มีค่าความยากง่ายเฉลี่ย ระหว่าง 1.216-2.587 ค่าอำนาจจำแนกเฉลี่ย ระหว่าง .436-.893 และค่าสัมประสิทธิ์การเดาเฉลี่ยระหว่าง .174-.461 ในกลุ่มชาย มีค่าความยากง่ายเฉลี่ยระหว่าง 1.277-2.922 ค่าอำนาจจำแนกเฉลี่ย ระหว่าง .456-.923 และค่าสัมประสิทธิ์การเดาเฉลี่ยระหว่าง .185-.246 ในกลุ่มหญิงมีค่าความยากง่ายเฉลี่ยระหว่าง 1.304-2.589 ค่าอำนาจจำแนกเฉลี่ย ระหว่าง .483-.856 และค่าสัมประสิทธิ์การเดาเฉลี่ยระหว่าง .180-.228

3. ดัชนีความลำเอียงของแบบทดสอบแต่ละฉบับพบว่าทั้ง 6 วิชา 12 ฉบับ มีค่าดัชนีความลำเอียงของวิธีการวิเคราะห์ความลำเอียงด้วยวิธีใช้ทฤษฎีการตอบข้อคำถามอยู่ระหว่าง .000-3.390 วิธีแมนเทิล-แฮนส์เซล อยู่ระหว่าง .520-1.710 และวิธี SIBTEST อยู่ระหว่าง .000-.160

4. ผลการตรวจสอบความแตกต่างของจำนวนข้อสอบที่ลำเอียงจากการตรวจสอบด้วยวิธีการต่างกันพบว่าทั้ง 3 วิธีตัดสินจำนวนข้อสอบที่ลำเอียงในแบบทดสอบแต่ละวิชาแตกต่างกันโดยที่แบบทดสอบวิชาสังคมศึกษา ก ฉบับที่ 1 และภาษาไทย กข ฉบับที่ 2 แตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 นอกนั้นแตกต่างกันในระดับนัยสำคัญที่ระดับ .01 โดยที่วิธีใช้ทฤษฎีการตอบข้อคำถามตัดสินจำนวนที่ลำเอียงมากที่สุด

5. สาเหตุของความลำเอียงพบว่า ข้อสอบที่ชายได้เปรียบจะเกิดจากสาเหตุที่ว่า เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า เป็นเรื่องราวที่ผู้ชายสนใจและพบเห็นในชีวิตประจำวัน และเป็นข้อสอบที่ถามความจำ ส่วนข้อสอบที่หญิงได้เปรียบจะเกิดจากสาเหตุที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า และเป็นเรื่องราวที่ผู้หญิงสนใจและพบเห็นในชีวิตประจำวัน

6. ความสัมพันธ์ลำดับที่ของการสอบที่คิดคะแนนมาตรฐานที่-ปกติ และคิดคะแนนน้ำหนักความสามารถ โดยใช้ข้อสอบจำนวนทั้งหมดกับใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง พบว่า ความสัมพันธ์ของลำดับที่ของผลการสอบไม่ว่าจะคิดคะแนนมาตรฐานที่-ปกติ หรือคิดคะแนนน้ำหนักความสามารถ และใช้ข้อสอบจำนวนทั้งหมด หรือใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง ต่างมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

7. ประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการเรียนของการคิดคะแนนและการใช้จำนวนข้อสอบพบว่า ในสายวิทยาศาสตร์ การคิดคะแนนน้ำหนักความสามารถ ทั้งใช้ข้อสอบจำนวนทั้งหมด และใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง มีประสิทธิภาพในการทำนายสูงกว่าการคิดคะแนนมาตรฐานที่-ปกติที่ใช้ข้อสอบจำนวนทั้งหมด และใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง ส่วนสายศิลปศาสตร์ พบว่าการคิดคะแนนมาตรฐานที่-ปกติ ทั้งที่ใช้ข้อสอบจำนวนทั้งหมด และใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง มีประสิทธิภาพในการทำนายสูงกว่าการคิดคะแนนน้ำหนักความสามารถทั้งที่ใช้ข้อสอบจำนวนทั้งหมด และใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง

อภิปรายผล

1. จำนวนข้อสอบที่ลำเอียงจากการตรวจสอบด้วยวิธีการใช้ทฤษฎีการตอบข้อคำถามวิธีแมนเทิล-แฮนส์เซล และวิธี SIBTEST มีความแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 และ .01 โดยที่วิธีการใช้ทฤษฎีการตอบข้อคำถามค้นพบข้อสอบที่มีความลำเอียงมากที่สุด ซึ่งสอดคล้องกับการศึกษาของ ทศนีย์ หิรมนตรี (2530 : 76-78) งานวิจัยของ สุรศักดิ์ อมรรัตนศักดิ์ (2531 : 200-207) และงานวิจัยของ กาญจนา วัฒนสุนทร (2537 : 267) ที่พบว่าวิธีการใช้ทฤษฎีการตอบข้อคำถามค้นพบข้อสอบที่ลำเอียงมากกว่าวิธีแมนเทิล-แฮนส์เซล และวิธี SIBTEST ทั้งนี้เนื่องจากวิธีการใช้ทฤษฎีการตอบข้อคำถามมีความไวในการตรวจสอบความลำเอียง เพราะเป็นวิธีการที่ใช้ความแตกต่างของพื้นที่ระหว่างโค้งลักษณะข้อสอบของกลุ่มเพศ (Kim and Cohen. 1991) และขึ้นอยู่กับเกณฑ์ที่กำหนดขนาดของดัชนีความลำเอียงที่ถือว่าเป็นข้อสอบที่ลำเอียงด้วย ส่วนสาเหตุที่วิธีแมนเทิล-แฮนส์เซล และวิธี SIBTEST ตรวจสอบจำนวนข้อสอบที่ลำเอียงได้น้อยกว่าและไม่แตกต่างกันนั้นเป็นเพราะว่าเกณฑ์ในการตัดสินข้อสอบที่ลำเอียง ใช้ผลการทดสอบ X ซึ่งจะคงที่

มากกว่าการกำหนดเกณฑ์ที่ใช้ความแตกต่างของพื้นที่ตามทฤษฎีการตอบข้อคำถาม (Raju and Others. 1993)

2. ความสอดคล้องของการตัดสินใจผลการสอบที่ใช้วิธีการคิดคะแนนต่างกัน นั้น เมื่อพิจารณาจากความสัมพันธ์ลำดับที่ของการสอบที่คิดคะแนนมาตรฐานที่-ปกติ และคะแนนน้ำหนักความสามารถ โดยใช้ข้อสอบจำนวนทั้งหมดกับใช้เฉพาะข้อสอบ ปราศจากความลำเอียง มีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 นั้น เนื่องจากการคิดคะแนนแต่ละวิธี และการใช้จำนวนข้อสอบ ให้ค่าความมากน้อยใน ลำดับที่ในลักษณะไปในทิศทางเดียวกันหรือซ้ำซ้อนกัน (Multicollinearity) (Kerlinger and Pedhazur. 1973 : 369) ซึ่งสอดคล้องกับการศึกษาวิจัยของ จำเนียร สุขหลาย (2530 : 162-165) ที่พบว่าความสัมพันธ์ของลำดับที่ของการ คิดคะแนนดิบ คะแนนมาตรฐานที่-ปกติ คะแนนถ่วงน้ำหนัก และคะแนนความสามารถ แท้ มีความสัมพันธ์กันสูงในทางบวกอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และ .01

3. ประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการเรียนของการคิดคะแนน และการใช้จำนวนข้อสอบพบว่าในสายวิทยาศาสตร์ การคิดคะแนนน้ำหนักความสามารถทั้งใช้ข้อสอบจำนวนทั้งหมด และใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง มีประสิทธิภาพในการทำนายสูงกว่าการคิดคะแนนมาตรฐานที่-ปกติที่ใช้ข้อสอบจำนวน ทั้งหมด และใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง ส่วนสายศิลปศาสตร์พบว่า การคิดคะแนนมาตรฐานที่-ปกติ ทั้งที่ใช้ข้อสอบจำนวนทั้งหมด และใช้เฉพาะข้อสอบที่ ปราศจากความลำเอียง มีประสิทธิภาพในการทำนายสูงกว่าการคิดคะแนนน้ำหนัก ความสามารถทั้งที่ใช้ข้อสอบจำนวนทั้งหมด และใช้เฉพาะข้อสอบที่ปราศจากความ ลำเอียง ทั้งนี้เนื่องจากการทำนายผลสัมฤทธิ์ทางการเรียนที่คิดคะแนนน้ำหนักความ สามารถนั้นในการกะประมาณคะแนนน้ำหนักความสามารถตามทฤษฎีการตอบข้อคำถาม จะต้องใช้จำนวนผู้สอบมาก (มากกว่า 1,000 คน) จึงจะกะประมาณค่าความ สามารถได้คงที่กว่าจำนวนผู้สอบน้อย (Hutten. 1982 : 4799-A) ซึ่งในสาย วิทยาศาสตร์ มีจำนวน 765 คน ส่วนสายศิลปศาสตร์ มีจำนวน 286 คน จึงเป็น ไปได้ทั้งในสายวิทยาศาสตร์ การคิดคะแนนน้ำหนักความสามารถจะมีประสิทธิภาพใน การทำนายสูงกว่าการคิดคะแนนมาตรฐานที่-ปกติที่ไม่ขึ้นอยู่กับจำนวนผู้สอบ ส่วนสาย ศิลปศาสตร์มีจำนวนผู้สอบน้อย (286 คน) จึงทำให้การคิดคะแนนมาตรฐานที่-ปกติ มีประสิทธิภาพในการทำนายสูงกว่า

4. สาเหตุความลำเอียงของข้อสอบที่ทั้งกลุ่มชายและหญิงได้เปรียบ ก็มาจากสาเหตุที่เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ มีโอกาสตอบถูกมากกว่า และเป็นเรื่องราวที่กลุ่มนั้นสนใจและพบเห็นในชีวิตประจำวันนั้น ซึ่งก็สอดคล้องกับธรรมชาติของการเรียนรู้ เพราะว่าผู้ฝึกหรือพบเห็นบ่อย ๆ ย่อมจะได้เปรียบที่จะเกิดการเรียนรู้ โดยเฉพาะที่เกี่ยวข้องกับความจำด้วย

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. ในการสอบคัดเลือกควรพิจารณาถึงความลำเอียงของข้อสอบ และควรวิเคราะห์ความลำเอียงด้วยวิธีที่เหมาะสมกับจำนวนผู้สอบ กล่าวคือถ้ากลุ่มผู้สอบ มีจำนวนน้อยกว่า 1,000 คน ควรใช้วิธีแมนเทิล-แฮนส์เชล หรือวิธี SIBTEST แต่ถ้ากลุ่มผู้สอบมีจำนวนมากกว่า 1,000 คนแล้ว ควรเลือกใช้วิธีใช้ทฤษฎีการตอบข้อคำถาม ทั้งนี้เพื่อความเหมาะสมระหว่างวิธีการกับจำนวนกลุ่มผู้สอบและให้เกิดความยุติธรรมกับกลุ่มผู้สอบทุกกลุ่มที่สนใจ
2. การคิดคะแนนควรเลือกใช้คะแนนมาตรฐานที่-ปกติสำหรับกรณีที่มีผู้สอบน้อยกว่า 3,000 คน และควรเลือกใช้การคิดคะแนนน้ำหนักความสามารถ (0) กรณีที่มีผู้สอบมากกว่า 3,000 คน เพราะมีความเที่ยงตรงเชิงพยากรณ์สูงกว่า

ข้อเสนอแนะในการวิจัยในครั้งต่อไป

1. ควรศึกษาความเที่ยงตรงของแบบทดสอบคัดเลือกระดับอื่น ๆ ที่ศึกษาทุกวิชาและศึกษาความเที่ยงตรงเชิงพยากรณ์กับเกรดที่สำเร็จการศึกษา
 2. ควรศึกษาตัวแปรกลุ่มอื่น ๆ เช่น ที่ตั้งของโรงเรียน ระดับการศึกษา
 3. ควรศึกษาเปรียบเทียบผลการตรวจสอบความลำเอียงด้วยวิธีการอื่น ๆ
- เพิ่มเติม

บรรณานุกรม

- ✓ พิชรี ปิยภัณท์. การวิเคราะห์ลำเอียงของข้อสอบจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนชั้นประถมศึกษาปีที่ 6. ปรินทิพนิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2531. อัดสำเนา.
- ✓ วิรัช วรพรรัตน์. การตรวจสอบคุณภาพเครื่องมือในการวัดผลและการวิจัย. กรุงเทพฯ : สำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2531. อัดสำเนา.
- ✓ สุรศักดิ์ อมรรัตนศักดิ์. การศึกษาเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่ต่างกัน 4 วิธี. วิทยานิพนธ์ ค.ด. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2531. อัดสำเนา.
- ✓ สุพัฒน์ สุกมลสันต์. การวิเคราะห์ความลำเอียงของข้อทดสอบภาษาอังกฤษเข้ามหาวิทยาลัยปี 2531-2533. กรุงเทพฯ : สถาบันภาษา จุฬาลงกรณ์มหาวิทยาลัย, 2534.
- อนันต์ จันทร์กวี. "เล็ก ๆ น้อย ๆ เกี่ยวกับการสอบคัดเลือกเข้าศึกษาต่อในมหาวิทยาลัย," วารสารการวัดผลการศึกษา. 8(22) : 12-18 ; พฤษภาคม - สิงหาคม 2529.

Ackerman, T.A. "A Diactic Explanation of Item Bias, Item Impact, and Item Validity from a Multidimensional Perspective," Journal of Educational Measurement. 29(1) : 67-91 ; Spring, 1992.

Ackerman, T.A. and Evans, J.A. An Investigation of Relationship between Reliability, Power, and the Type I Error Rate of the Mantel-Haenszel and Simultaneous Item Bias Detection Procedures. Paper presented at the Annual Meeting of the National Council on Measurement in Education. San Francisco, CA, April 21-23 ; 1992.

Baeza, Jesus, Jr. Test Item Bias and American Indian Students. Unpublished Doctoral Dissertation, The University of Iowa, 1989.

Baghi, Heibatillah 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.

Baghi, H. and Ferrara, S.F. Detecting Differential Item Functioning Using IRT and Mantel-Haenszel Techniques : Implementing Procedures and Comparing Results. 14-17 ; February, 1990.

Berk, R.A. (Ed.) Handbook of Methods for Detecting Item Bias.
Baltimore, Md : John 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.

Carton, S.T. and Marco G.L. "Method Used by Test Publishers," In Handbook of Methods for Detecting Test Bias. Baltimore and London : The Johns Hopkins University Press, 1982.

(7)

Chang, L. "The Effects of Instructional Differences on Item Bias Indices and a Comparison of Item Bias Detection Procedures," Dissertation Abstracts International. 55(04) : 330-A ; October, 1993.

Chen, T.H. "A Logistic Regression, Paired Comparisons Method for Testing Rasch Model Parameter Invariance : A Simulation Study," Dissertation Abstracts International. 52(10) : 3558-A ; April, 1992.

Clauser, B.E. "Factors Influencing the Performance of the Mantel-Haenszel Procedure in Identifying Differential Item Functioning," Dissertation Abstracts International. 54(02) : 493-A ; August, 1993.

(3)

Clauser, Brian E. and Others. Examination of Various Influence on the Mantel-Haenszel Statistic. Paper presented at the Annual Meeting of the American Educational Research Association. Chicago, Illinois, April 3-7 ; 1991.

- _____. "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.
- Cohen, Allan and Kim, Seock-Ho. "A Comparison of Lord's X² 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.
- Diamond, J.J. "A Preliminary Study of the Reliability and Validity of a Scoring Procedure Based Upon Confidence and Partial Information," Journal of Educational Measurement. 12(2) : 129-135 ; Summer, 1975.
- 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. and Cleary, T. Anne. "Gender-Based Differential Item Performance in Mathematics Achievement Items," Journal of Educational Measurement. 24(2) : 157-166 ; Summer, 1987.
- Freedle, Roy and Kostin, Irene. "Item Difficulty of Four Verbal Item Types and an Index of Differential Item Functioning for Black and White Examinees," Journal of Educational Measurement. 27(4) : 329-343 ; Winter, 1990.
- ④ Furgon. "The Use of Flexible Logistic Regression in Assessing Differential Item Functioning," Dissertation Abstracts International. 54(09) : 3411-A ; March, 1994.
- Green, Bert F., Grone, Carolyn R. and Folk, Valerie Greaud. "A Method of Studying Differential Distractor Functioning," Journal of Educational Measurement. 26(2) : 147-160 ; Summer, 1989.

- Hakstian, A.R. and Kansup, W. "A Comparison of Several Methods of Assessing Partial Knowledge in Multiple-Choice Test : Testing Procedure," Journal of Educational Measurement. 12(3) : 201-211 ; Fall, 1975.
- 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 Rogers, H. Jane. "Detecting Potential Biased Test Items. Comparison of IRT Area and Meantel-Haenszel Methods," Applied Measurement in Education. 2(4) : 313-334 ; 1989.
- Harris, Abigail M. and Carlton, Sydell T. Patterns of Gender. "Differences on Mathematics Items on the Scholastic Aptitude Test," Applied Measurement in Education. 6(2) : 137-151 ; 1993.
- Hulin, Charles L., Drasgow, Fritz and Parsons, Charles K. Item Response Theory Application to Psychological Measurement. Illinois : Dow Jones Irwin, 1983.
- Humphreys, Lloyd G. and Taber, Thomas. "Ability Factors as a Function of Advantaged and Disadvantaged Groups," Journal of Educational Measurement. 10(2) : 107-115 ; Summer, 1973.
- Hutten, Leah R. "The Fit of Empirical Data to Two Latent Trait Models," Dissertation Abstracts International. 42(11) : 4799-A ; May, 1982.
- Ibrahim, A.K. "Distribution and Power of Selected Item Bias Indices : A Monte Carlo Study," Dissertation Abstracts International. 55(03) : 540-A ; September, 1994.
- Intasuwan, P. A Comparison of Three Approaches for Determining Item Bias in Gross-National Testing. Dector's Thesis The University of Pittsburgh, 1979.

- Kim, Seock-Ho and Cohen, S. "A Comparison of Two Area Measures for Detecting Differential Item Functioning," Applied Psychological Measurement. 15(3) : 269-278 ; September, 1991.
- Kok, Frank G., Mellenbergh, Gideon J. and Van Der. Flier, Henk. "Detecting Experimentally Induced Item Bias Using the Iterative Logit Method," Journal of Educational Measurement. 22(4) : 295-303 ; Winter, 1985.
- Lautenschlager, Gary J. and Park, Dong-Gun. "IRT Item Bias Detection Procedure : Issue 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., 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. and Novick, M.R. Application of Item Response Theory to Practical Testing Problems. Hillsdale, N.J. : Lawrence Erlbaum Associates, 1980.
- _____. Statistical Theory of Mental Test Scores. Reading. Massachusette : Addison-Wesley, 1968.

Majumdar, D.G. "Comparison of Johnson-Neyman, Mantel-Haenszel and Partial Correlation Procedures for Detecting Differential Item Performance for Moderate Sample Size," Dissertation Abstracts International. 53(01) : 106-A ; July, 1992.

④

Maller, S.J. "Validity and Item Bias of the WISC-III with Deaf Children," Dissertation Abstracts International. 55(05) : 1249-A ; November, 1994.

Maracuillo, Leonard A. and Marycllem Mc Sweeney Nonparametric and Distribution-Free Methods for the Social Sciences. Wadsworth Publishing Company, Inc., Belmont California : 1977, 555p.

Marascuilo, Leonard A. and Slaughter, Robert E. "Statistical Procedures for Identifying Possible Sources of Item Biased on X2 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.

McGrogan, Harold James, Jr. The Determination of Cultural Item Bias in the California Achievement Tests. Unpublished Doctoral Dissertation, The University of Arizona, 1989.

Miller, M.D. and Oshima, T.C. "Effect of Sample Size, Number of Biased Item and Magnitude of Bias on a Two-Stage Item Bias Estimation Method. Applied Psychological of Measurement. 16(4) : 381-388 ; December, 1992.

⑨

Moon, T.R. "Comparing two Extensions on the Mantel-Haenszel Statistic for Detecting Differential Item Functioning on a Direct Writing Assessment," Dissertation Abstracts International. 55(05) : 1249-A ; November, 1994.

②

- Narayanan, P. and Swaminathan, H. Performance of the Mantel-Haenszel and Simultaneous Item Bias Procedures for Detecting Differential Item Functioning, Laboratory of Psychometric and Evaluative Research Report No.252. Paper Present at the Annual Meeting of the National Council on Measurement in Education. Atlanta, Ga, April 13-15 ; 1993.
- O'Neal, M.R. A Comparison of Methods for Detecting Item Bias. Paper Presented at the Annual Meeting of the Mid-South Educational Research Association. 20th, Lexington, KY, November 12-15 ; 1991.
- Oshima, T.C. and Miller, M.D. "Multidimensionality and Item Bias in Item Response Theory," Applied Psychological Measurement. 16(3) : 237-248 ; September, 1992.
- ✓Oshima, Takako, The Effect of Multidimensionality on Item Bias Detection Based on Item Response Theory. Unpublished Doctoral Dissertation, University of Florida, 1989.
- Oshima, Takako C. and Miller, M. David. "Multidimensionality and IRT-Based Item Invariance Indexes : The Effect of Between Group Variation in Trait Correlation," Journal of Educational Measurement. 27(3) : 273-283 ; Fall, 1990.
- Osterlind, Steven J. Test Item Bias. California : Sage Publications, Inc., 1983.
- Park, Dong-Gun and Lautenschlager, Gary J. "Improving IRT Item Bias Detection With Iterative Linking and Ability Scale Purification," Applied Psychological Measurement. 14(2) : 163-173 ; June, 1990.
- Perlman, Carole L. and Others. Investigating the Stability of Four Methods for Estimating Item Bias. Paper Presented at the Annual Meeting of the National Council for Measurement in Education. New Orleans, Los Angeles, April 5-9 ; 1988.

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.

Raju, Nambury S. Gode, Drasgow, Fritz 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.

Reynolds, Cecil R. "An Examination for Bias in a Preschool Test Battery Across Race and Sex," Journal of Educational Measurement. 17(2): 137-146 ; Summer, 1980.

Reynolds, Trudy Lynn. An Application of Several Differential Item Functioning Procedures to Assess the Quality of a Translated Test. Unpublished Doctoral Dissertation, Southern Illinois University at Carbondale, 1989.

Rogers, Helen Jane. A Logistic Regression Procedure for Detecting Item Bias. Unpublished Doctoral Dissertation, University of Massachusetts, 1989.

Roussos, Louis A. and Stout, William F. Simulation Studies of Effects of Small Sample Size and Studies Item Parameters on SIBTEST and Mantel-Haenszel Type I Error Performance. Paper Presented at the 1993 AERA Annual Meeting, Atlanta, Georgia, April 16 ; 1993.

Rubin, Donald B. "Discussion," In Test Validity. p.241-253 ; Hillsdale, New Jersey : Lawrence Erlbaum Associates, 1988.

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) : 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.

Scheuneman, Janice Dowd. "A Posteriori Analysis of Biased Item" In Handbook of Methods for Detecting Item bias. p. 181-198. John Hopkins University Press, 1982.

Scheuneman, Janice Dowd. and Gerritz, Kalle. "Using Differential Item Functioning Procedures to Explore Sources of Item Difficulty and Group Performance Characteristics," Journal of Educational Measurement. 27(2) : 109-131 ; Summer, 1990.

Schmitt, Alicia P. and Dorans, Neil J. "Differential Item Functioning for Minority Examinees on the SAT," Journal of Educational Measurement. 27(1) : 67-81 ; Spring, 1990.

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.

①

Shealy, Robin and Stout, William. A Model-Based Standardization Approach that Separates True Bias/DIF from Group Ability Differences and Detects Test Bias/DTF as Well as Item Bias/DIF. Research Report for the Office of Naval Research Cognitive and Neural Sciences Grant, 1992.

Shepard, Lorrie, Camilli, Gregory and Williams, David 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.

- Skaggs, Gary and Lissitz, Robert W. "The Consistency of Detecting Item Bias Across Different Test Administration : Implications of Another Failure," Journal of Educational Measurement. 29(3) : 227-242 ; Fall, 1992.
- Smith, J.B. A Study of Item Bias in the Maine Educational Assessment Test. Doctor's Dissertation. Boston : Boston College, 1993.
- Subkoviak, Michael J., Mack, Joanne S., Ironson, Gial H. and Craig, Robert D. "Empirical Comparison of Deleted Item Bias Detection Prodecures with Bias Manipulation," Journal of Educational Measurement. 21(1) : 49-59 ; 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 April 8-11 ; 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 Responce Theory in the Study of Group Differences in Trace Lines," In Test Validity. p.147-170. Hillsdale, New Jersey : Lawrence Erlbaum Associates, 1988.
- Uttaro, T.E. "Factors Influencing the Mantel-Haenszel Procedure in the Detection of Differential Item Functioning," Dissertation Abstracts International. 53(04) : 1105-A ; October, 1992.
- Van Der Flier, Henk, Mellanbergh, Gideon J., Ader, Herman J. and Wijn, Marina. "An Iterative Item Bias Detection Method," Journal of Educational Measurement. 21(2) : 131-145 ; Summer, 1984.

Veale, James R. and Foreman, Dale I. "Assessing Cultural Bias Using Foil Response Data : Cultural Variation," Journal of Educational Measurement. 20(3) : 249-258 ; Fall, 1983.

Wainer, Howard, Sireci, Stephen G. and Thissen, David. "Differential Testlet Functioning : Definitions and Detection," Journal of Educational Measurement. 28(3) : 197-219 ; Fall, 1991.

Wang, M.W. and Stanley, J.C. "Differential Weighting : A Review of Methods and Empirical Studies," Review of Educational Research. 40(5) : 663-705 ; December, 1970.

②

Welch, C. and Hoover, H.D. "Procedures for Extending Item Bias Detection Techniques to Polytomously Scored Items," Applied Measurement in Education. 6(1) : 1-19 ; 1993.

⑤

Westers, P. The Solution-Error Response-Error Model : A Method for the Examination of Test Item Bias. Doctor's Dissertation. Twente University, 1993.

Zeidner, Moshe. "Sociocultural Differences in Examinees' Attitudes Toward Scholastic Ability Exams," Journal of Educational Measurement. 25(1) : 67-77 ; Spring, 1988.

บรรณานุกรม

- ✓ กาญจนา วัฒนสุนทร. การพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ. วิทยานิพนธ์ ค.ค. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2538. อัดสำเนา.
- คณะกรรมการการศึกษาแห่งชาติ, สำนักงาน. วัตถุประสงค์ นโยบาย และมาตรการของแผนพัฒนาการศึกษาแห่งชาติ ฉบับที่ 7 (พ.ศ.2535-2539). กรุงเทพฯ : สำนักงานคณะกรรมการการศึกษาแห่งชาติ สำนักงานนายกรัฐมนตรี, 2534.
- จำเนียร สุขหลาย. ผลของวิธีการคิดคะแนนต่อความตรงเชิงทำนายของแบบทดสอบคัดเลือกเข้าศึกษาต่อในสหวิทยาลัยศรีอยุธยา. วิทยานิพนธ์ ค.ค. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2530. อัดสำเนา.
- ชอลดา พันธุ์เสนา และคนอื่น ๆ. "สรุปรายงานการวิจัยประสิทธิภาพด้านความตรงเชิงพยากรณ์ของข้อสอบคัดเลือกเข้าศึกษาต่อในมหาวิทยาลัยสงขลานครินทร์โดยวิธีรับตรง ปีการศึกษา 2534," วารสารสงขลานครินทร์. 1(1) : 1-10 ; กันยายน - ธันวาคม 2537.
- ชูศักดิ์ ชัมภลิต. "IRT กับผลการสอบคัดเลือก," การวัดผลการศึกษา. 4(2) : 33-44 ; กันยายน - ธันวาคม 2535.
- ✓ ชัชชัย เผ่าพงศ์. การวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์และภาษา ระดับชั้นมัธยมศึกษาตอนต้น. วิทยานิพนธ์ ศศ.ม. ขอนแก่น : มหาวิทยาลัยขอนแก่น, 2527. อัดสำเนา.
- ✓ นิรมล ชัยสวัสดิ์. การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของแบบทดสอบความเข้าใจในการอ่านภาษาไทยตามทฤษฎีคลาสสิกอลที่ใช้วิธีวิเคราะห์ต่างกัน. วิทยานิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2537. อัดสำเนา.
- ✓ ทศนีย์ หิรมนตรี. การวิเคราะห์ความลำเอียงของแบบทดสอบวิชาคณิตศาสตร์ โครงการตรวจสอบคุณภาพการศึกษาชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526. วิทยานิพนธ์ ค.ม. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2530. อัดสำเนา.
- "เผยแพร่ให้นักเรียนได้มีอัตราแข่งขัน 8 ต่อ 1 ในการสอบคัดเลือกเข้ามหาวิทยาลัย," สารมหาวิทยาลัยสงขลานครินทร์. 4(3) : 24-25 ; กุมภาพันธ์ - มีนาคม 2538.

ภาคผนวก

ตัวอย่าง

แบบสำรวจสาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบ
ที่มีต่อกลุ่มผู้สอบเพศชายและ/หรือเพศหญิง

เพื่อการตรวจสอบความเที่ยงตรงของแบบทดสอบคัดเลือก ดังนั้นขอให้ท่านพิจารณาข้อสอบแต่ละข้อที่ได้รับการวิเคราะห์ทางสถิติแล้วพบว่า มีผู้สอบบางกลุ่มมีการได้เปรียบ เสียเปรียบ นั้น สิ่งที่เกิดความได้เปรียบ เสียเปรียบ น่าจะเกิดมาจากสาเหตุมากที่สุด ซึ่งโดยทั่วไปแล้วสาเหตุที่ก่อให้เกิดการได้เปรียบ เสียเปรียบในการตอบข้อสอบของผู้สอบแต่ละกลุ่มนั้นเนื่องมาจากสาเหตุต่าง ๆ เช่น

1. ข้อสอบยากเกินไป ทำให้กล้าเสี่ยงที่จะเดา
2. เป็นข้อคำถามที่ผู้สอบเฉพาะกลุ่มพบเห็นในชีวิตประจำวัน
3. เป็นข้อคำถามที่ถามความจำ
4. ข้อคำถามเน้นการคำนวณทั้ง ๆ ที่ เรื่องนั้น ๆ ไม่จำเป็นต้องใช้
5. ข้อคำถามให้อ่านข้อความเย็นเย็นไม่เกี่ยวข้องกับข้อคำถาม ทำให้ผู้สอบอ่านได้เปรียบ
6. ตัวเลือกเป็นคำ / ข้อความที่พบเห็นหรือใช้เฉพาะกลุ่ม
7. เวลาในการทำแบบทดสอบน้อยเกินไป
8. เป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะเรื่องนั้น ๆ จะมีโอกาสตอบถูกมากกว่า
9. เป็นข้อคำถามที่เรียงอยู่ข้อหลัง ๆ ทำให้มีเวลาตอบน้อย
10. อื่น ๆ (ระบุ).....

ในการพิจารณาขอให้ท่านอ่านข้อสอบแล้วตัดสินใจว่า ข้อสอบข้อนั้นเกิดการได้เปรียบมาจากสาเหตุใด “มากที่สุด” ใน 3 อันดับ” โดยนำหมายเลขหน้าข้อของสาเหตุที่สรุปไว้ข้างต้นไปเติมลงในช่องลำดับของสาเหตุที่เกิดความลำเอียง ดังตัวอย่าง

ตัวอย่าง 1

แบบสำรวจสาเหตุที่ก่อให้เกิดความลำเอียงของข้อสอบที่มีต่อกลุ่มผู้สอบเพศชายและ/หรือผู้สอบเพศหญิง
แบบทดสอบวิชา ภาษาไทย จำนวน 50 ข้อ ให้เวลาทำ 55 นาที

ข้อสอบ ข้อที่	ข้อสอบ	กลุ่มที่ได้ เปรียบ	อันดับของ สาเหตุที่ก่อ ให้เกิดความ ลำเอียง			ข้อเสนอแนะ เพิ่มเติม
			1	2	3	
45	ข้อใดไม่แสดงทัศนะ 1. ความรักทำให้ตาบอด 2. สวรรค์อยู่ในอกนรกอยู่ในใจ 3. ไม่เห็นโลงศพไม่หลั่งน้ำตา 4. แรงกิริยาเท่ากับแรงปฏิกิริยา	หญิง	6	2	8	

จากตัวอย่าง 1 แสดงว่า ท่านพิจารณาแล้ว สรุปว่า ข้อสอบข้อนี้ เพศหญิงได้เปรียบมากกว่าเพศชาย เพราะมาจากสาเหตุที่ว่า เป็นข้อคำถามที่มีตัวเลือกที่เพศหญิงพบเห็นเสมอเป็นอันดับ 1 เป็นข้อความที่เพศหญิงพบมากในชีวิตประจำวัน เป็นอันดับ 2 และเป็นข้อคำถามที่เพศหญิงคุ้นเคย ง่าย จึงมีโอกาสตอบถูกมากกว่าเป็นอันดับ 3

ตัวอย่าง 2

แบบสำรวจสาเหตุที่ก่อให้เกิดความล่าช้าของข้อสอบที่มีต่อกลุ่มผู้สอบเพศชายและ/หรือผู้สอบเพศหญิง
แบบทดสอบวิชา สังคมศึกษา จำนวน 80 ข้อ ให้เวลาทำ 55 นาที

ข้อสอบ ข้อที่	ข้อสอบ	กลุ่มที่ได้ เปรียบ	อันดับของ สาเหตุที่ก่อให้เกิดความ ล่าช้า			ข้อเสนอแนะ เพิ่มเติม
			1	2	3	
78	D = 10 - P เป็นสมการเส้นตรง S = 2P - 5 เป็นสมการเส้นโค้งหลาย เมื่อ P คือ ราคาสินค้า ถ้าตีมันเท่ากับข้อผิดพลาดมีค่าเท่าใด? ①. 5 2. 10 3. 15 4. 20	ชาย	9	1	4	

จากตัวอย่าง 2 แสดงว่าท่านพิจารณาแล้วสรุปว่า ข้อสอบข้อนี้เพศชายได้เปรียบมากกว่าเพศหญิงเพราะมาจากสาเหตุที่ว่าเป็นข้อคำถามที่อยู่
ข้อหลัง ๆ มีเวลา ตอบน้อย และเพศชายตัดสินใจได้รวดเร็ว เป็นอันดับ 1 ข้อสอบยากเกินไป ทำให้เพศชายกล้าเสี่ยงที่จะเดาเป็นอันดับ 2 และเป็น
ข้อคำถามที่เน้นการคำนวณทั้ง ๆ ที่ไม่จำเป็นต้องใช้เป็นอันดับ 3
ผลการพิจารณาอย่างรอบคอบของท่านมีความสำคัญและเป็นประโยชน์ต่อการศึกษายิ่งยิ่ง ขอขอบคุนในความอนุเคราะห์

นางสาวเรวดี อินทสระ

นิสิตปริญญาเอก สาขาการทดสอบและวัดผลการศึกษา

มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

ข้อสอบ ข้อที่	ข้อสอบ	กลุ่มที่ได้ เปรียบ	อันดับของ สาเหตุที่ก่อ ให้เกิดความ ลำเอียง			ข้อเสนอแนะ เพิ่มเติม
			1	2	3	
9	ข้อความที่ขีดเส้นใต้ข้อความหน้าที่เป็นประธานของประโยค 1. อาหารเข้ายังไม่รับประทานกันเลขใช้ใหม่ 2. ช่วงเวลาสงกรานต์มีกิจกรรมพัฒนา 3. แพ้ชนะในเกมเป็นเรื่องธรรมดาของนักกีฬา 4. ที่บ้านปลูกต้นไม้หลายชนิดทั้งไม้ดอกและไม้ผล	ชาย				
10.	ข้อใดเป็นประโยคที่สมบูรณ์ 1. บ้านเรือนชนบทที่ขงน้ำไว้นุ่มเพื่ออาบกับภายใน ครอบครัวยังมีอยู่ 2. การอธิบายที่มุ่งสื่อสารไปยังประชาชนที่ยังขาดความรู้ความเข้าใจเรื่องโรคเอดส์ 3. แนวทางให้สามารถวิจารณ์วรรณคดีได้อย่างมีหลักเกณฑ์เมื่อเรียนสูงขึ้น 4. กระบวนการเรียนรู้ที่ดีที่สุดที่สามารถสร้างความรู้ความเข้าใจที่แตก่างกันได้	ชาย				
15.	ข้อใดมีคำสะกดผิดปนอยู่ด้วย 1. ดานขโมยเป็นชื่อโรคพยาธิลำไส้อย่างหนึ่งมักเกิดแก่เด็ก 2. โรคที่เกิดจากตัวเชื้อโรคเข้าสู่แผลที่ทำให้มีอาการกระดูกคือบาดทะยัก 3. ขยักขยอนเป็นอาการที่พะอืดพะอมจนจะอาเจียน 4. ลักษณะที่เกิดเป็นเม็ดขึ้นในปากในคอ ล้นเป็นฝ้า เรียกว่า ขาง	หญิง				

ข้อสอบ ข้อที่	ข้อสอบ	กลุ่มที่ ได้ เปรียบ	อันดับของ สาเหตุที่ก่อ ให้เกิดความ ลำเอียง			ข้อเสนอแนะ เพิ่มเติม
			1	2	3	
9.	รูปแบบการค้าของไทยภายหลังการทำสนธิสัญญาเบาริ่งเป็นรูปแบบใด 1. ทุนนิยม 2. เสรีนิยม 3. สังคมนิยม 4. กึ่งศักดินากึ่งทุนนิยม	ชาย				
11.	จำนวนผู้ป่วยโรคเอดส์จำแนกตามกลุ่มอาชีพระหว่าง พ.ศ. 2530-2536 นั้น พบ ว่าอาชีพใดมีอัตราผู้ป่วยเป็นโรคเอดส์มากที่สุด 1. เกษตรกร 2. ข้าราชการ 3. ผู้ใช้แรงงาน 4. หญิงบริการทางเพศ	หญิง				

ข้อสอบข้อที่	ข้อสอบ	กลุ่มที่ได้เปรียบ	อันดับของสาเหตุที่ก่อให้เกิดความลำเอียง			ข้อเสนอแนะเพิ่มเติม
			1	2	3	
4.	Bob is moving out of his parents' house into an apartment. Liz, his girlfriend, is helping him. Bob: Oh! Liz : Bob!(4)..... Bob: My arms are hurting. Liz : I'm not surprised. Why don't you stop for a moment?(5)..... Bob:(6)..... I think I will. 1. What is that? 2. What hurts you? 3. What's happening? ④ 4. What's the matter?	หญิง				

ประวัติย่อของผู้วิจัย

ชื่อ นางสาวเรวดี ชื่อสกุล อินทสระ
เกิดวันที่ 26 เดือน ตุลาคม พุทธศักราช 2503
สถานที่เกิด อำเภอ ตะโหมด จังหวัด พัทลุง
สถานที่อยู่ปัจจุบัน 108/84 หมู่ 2 ซอยประภาส ถนนกาญจนวนิช
ตำบลเขารูปช้าง อำเภอเมือง จังหวัดสงขลา 90000
ตำแหน่งหน้าที่การงานปัจจุบัน อาจารย์
สถานที่ทำงานปัจจุบัน ภาควิชาพื้นฐานของการศึกษา
คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
ภาคใต้ อำเภอเมือง จังหวัดสงขลา 90000
ประวัติการศึกษา
พ.ศ. 2521 มัธยมศึกษาปีที่ 5 จากโรงเรียนสตรีพัทลุง
พ.ศ. 2523 ป.กศ.สูง (วิชาเอกคณิตศาสตร์ วิชาโทวิทยาศาสตร์
ทั่วไป) จากวิทยาลัยครูสงขลา
พ.ศ. 2525 กศ.บ. (วิชาเอกคณิตศาสตร์ วิชาโทการวัดผลการ
ศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ สงขลา
พ.ศ. 2530 กศ.ม. (การวัดผลการศึกษา) จาก มหาวิทยาลัย
ศรีนครินทรวิโรฒ ประสานมิตร
พ.ศ. 2539 กศ.ด. (การทดสอบและวัดผลการศึกษา) จาก
มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร

ผลการตรวจสอบความลำเอียงของข้อสอบต่อการศึกษาความเที่ยงตรงเชิงพยากรณ์
ของแบบทดสอบคัดเลือกที่คิดคะแนนต่างกัน

บทคัดย่อ
ของ
เรวดี อินทสระ

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษาดุซงึ่บัณฑิต สาขาการทดสอบและวัดผลการศึกษา
มิถุนายน 2539

การศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่วิเคราะห์ความลำเอียงต่อเพศด้วยวิธีใช้ทฤษฎีการตอบข้อคำถาม (IRT) วิธีแมนเทิล-แฮนส์เซล (MH) และวิธี SIBTEST การตัดสินผลการสอบที่คิดคะแนนมาตรฐานที่-ปกติและคะแนนน้ำหนักความสามารถและสาเหตุความลำเอียงของข้อสอบ โดยศึกษาความลำเอียงของข้อสอบคัดเลือกเข้าศึกษาในชั้นปีที่ 1 ประเภทรับตรง ปีการศึกษา 2538 ของมหาวิทยาลัยสงขลานครินทร์ในวิชาภาษาไทย ก วิชาสังคมศึกษา ก วิชาภาษาอังกฤษ กข วิชาละ 8,127 คน (ชาย 2,722 คน หญิง 5,405 คน) วิชาภาษาไทย กข วิชาสังคมศึกษา กข และวิชาภาษาอังกฤษ กขค วิชาละ 5,415 คน (ชาย 1,454 คน หญิง 3,961 คน) ความเที่ยงตรงเชิงพยากรณ์ศึกษาจากคะแนนสอบคัดเลือกกับเกรดภาคเรียนที่ 1 ปีการศึกษา 2538 ของนักศึกษาที่ได้รับการคัดเลือกจากประเภทรับตรง สายวิทยาศาสตร์ 763 คน และสายศิลปศาสตร์ 281 คนและสาเหตุความลำเอียงของข้อสอบจากการระบุสาเหตุของนักวัดผลการศึกษาหรืออาจารย์ผู้สอน จำนวน 50 คน และนิสิตนักศึกษาที่เรียนในสาขาวิชานั้น ๆ วิชาละ 50 คน

ผลการศึกษาพบว่า วิธีการตรวจสอบความลำเอียงทั้ง 3 วิธี ตัดสินจำนวนข้อสอบที่ลำเอียงแตกต่างกันในวิชาภาษาไทย ก ฉบับที่ 2 และวิชาสังคมศึกษา ก ฉบับที่ 1 ที่ระดับนัยสำคัญทางสถิติที่ .05 นอกนั้นต่างกันที่ระดับนัยสำคัญทางสถิติที่ .01 โดยที่วิธีใช้ทฤษฎีการตอบข้อคำถามตัดสินจำนวนข้อสอบที่ลำเอียงมากที่สุด ความสัมพันธ์ของลำดับที่ของการสอบไม่ว่าจะคิดคะแนนมาตรฐานที่-ปกติหรือคิดคะแนนน้ำหนักความสามารถและใช้ข้อสอบจำนวนทั้งหมดหรือใช้เฉพาะข้อสอบที่ปราศจากความลำเอียงต่างมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการเรียนในสายวิทยาศาสตร์ การคิดคะแนนน้ำหนักความสามารถทั้งใช้ข้อสอบจำนวนทั้งหมดและใช้เฉพาะข้อสอบที่ปราศจากความลำเอียงมีประสิทธิภาพในการทำนายสูงกว่าการคิดคะแนนมาตรฐานที่-ปกติ ส่วนสายศิลปศาสตร์การคิดคะแนนมาตรฐานที่-ปกติทั้งที่ใช้ข้อสอบจำนวนทั้งหมดและใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง มีประสิทธิภาพในการทำนายสูงกว่าการคิดคะแนนน้ำหนักความสามารถและสาเหตุของความลำเอียงของข้อสอบต่อเพศทั้งชายและหญิงเกิดจากข้อสอบเป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะจะมีโอกาสตอบถูกมากกว่า เป็นเรื่องราวที่กลุ่มนั้น ๆ สนใจและเป็นข้อสอบที่ถามความจริง

THE EFFECT OF ITEM BIAS DETECTION ON PREDICTIVE VALIDITY OF
ENTRANCE EXAMINATION TESTS USING DIFFERENT SCORING PROCEDURES

AN ABSTRACT
BY
RAYVADEE INTASARA

Presented in partial fulfillment of the requirements for the
Doctor of Education degree in Testing and Measurement
at Srinakharinwirot University
June 1996

Predictive validity of Prince of Songkla University Entrance Examination tests for the year 1995 were investigated by using three types of item biased detection : Item Response Theory (IRT), Mantel-Hanszel (MH), and SIBTEST. Two scoring procedures with total number of items, and, only the items that were not biased were applied to determine the examination scores. They were standard normalized T-score and ability score (θ). The data consisted of 8,127 students (2722 male, and 5405 female) who took Thai A, Social Science A, and English AB tests. And 5,415 students (1454 male and 3961 female) who took Thai AB, Social Science AB, and English ABC tests. Grade-point-average of Freshy year first semester 1995 of 763 Science students and 281 Arts students were used as criterion score for the predictive validity of the Entrance examination tests. The sources of item biased determined by 50 evaluators and/or teachers and 50 students, majoring in Thai, Social Science, and English respectively.

The result showed that the different of number of item biased from each pair of tests were statistically difference at .01, except between the test Thai A form II and Social Science A form I was statistically difference at .05. Item Response Theory procedure detected the highest number of item biasness. All rank of scores, with or without biased items, between any scoring procedures correlated with .01 level of significant.

Entrance examination for Science students in ability scores for either all items or unbiased items had higher predictive efficiency than standard normalized T-scores, while for the Arts students, the result was on the other way around. Finally, the sources of sex biased were defined by past experiences, concentration, and items that called for memory.