

การเปรียบเทียบผลของความลำเอียงของแบบทดสอบวัดความเข้าใจทางภาษา
โดยใช้ขนาดกลุ่มผู้สอบและวิธีวิเคราะห์ต่างกัน

ปริญญานิพนธ์
ของ
พิพุฒิพงษ์ แนวทอง

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา
สิงหาคม 2544
ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

๕๕๓.๖๕๖(๒)

๖๓๔๓๓๐

๓ : การเปรียบเทียบผลของความลำเอียงของแบบทดสอบวัดความเข้าใจทางภาษา
โดยใช้ขนาดกลุ่มผู้สอบและวิธีวิเคราะห์ต่างกัน

บทคัดย่อ

ของ

พิรุณพงษ์ แนวทอง

๕๕๓.๖๕๖(๒)
๖๓๔๓๓๐

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา

สิงหาคม ๒๕๔๔

พิรุณพงษ์ แนวทอง. (2544). การเปรียบเทียบผลของความลำเอียงของแบบทดสอบวัดความเข้าใจทางภาษาโดยใช้ขนาดกลุ่มผู้สอบและวิธีวิเคราะห์ต่างกัน. ปรินซ์นิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ.
คณะกรรมการควบคุม : รองศาสตราจารย์อังคณา สายยศ, อาจารย์ชวลิต รวยอาจิน

การศึกษาค้นคว้าครั้งนี้มีจุดมุ่งหมายเพื่อเปรียบเทียบผลของความลำเอียงของแบบทดสอบวัดความเข้าใจทางภาษา โดยใช้ขนาดผู้สอบต่างกัน คือ กลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:250, 500:500 และ 250:500 ด้วยวิธีวิเคราะห์ต่างกัน คือ วิธีHW3 วิธีไบซีเรียลและวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด โดยเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงและเปรียบเทียบค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว ในการศึกษาครั้งนี้มีกลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2543 ของโรงเรียนสังกัดกรมสามัญศึกษา กลุ่มที่ 4 ส่วนกลาง จำนวน 2,800 คน ซึ่งเลือกมาโดยการสุ่มแบบสองขั้นตอน ผลการวิจัยได้ผลดังนี้

1.จำนวนข้อที่ลำเอียงระหว่างที่วิเคราะห์ด้วยวิธีHW3กับวิธีไบซีเรียล และวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด เมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น250:250และ500:500 มีจำนวนข้อที่ลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01และ.05 ตามลำดับ ส่วนวิธีHW3กับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด เมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:500 มีจำนวนข้อที่ลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

2.จำนวนข้อที่ลำเอียง เมื่อวิเคราะห์ด้วยวิธีไบซีเรียลและวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด ระหว่างที่วิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น500:500กับ250:500 มีจำนวนข้อที่ลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 ส่วนวิธีHW3 มีจำนวนข้อที่ลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทุกกลุ่มผู้สอบที่เป็นกลุ่มอ้างอิง:กลุ่มเปรียบเทียบ

3.ค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อที่ลำเอียงออกแล้ว ระหว่างที่วิเคราะห์ด้วยวิธีHW3กับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด เมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น250:250และ500:500 มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01และ.05 ตามลำดับ ส่วนค่าความเชื่อมั่นระหว่างวิธีHW3กับวิธีไบซีเรียล และ วิธีไบซีเรียลกับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด เมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น250:500 มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01

4.ค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อที่ลำเอียงออกแล้ว เมื่อวิเคราะห์วิธีไบซีเรียล ระหว่างที่วิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น500:500กับ250:500 มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01 และวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด ระหว่างที่วิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น250:250กับ500:500 มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 ส่วนวิธีHW3 มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทุกกลุ่มผู้สอบที่เป็นกลุ่มอ้างอิง:กลุ่มเปรียบเทียบ

A COMPARISON OF THE ITEM BIAS IN VERBAL COMPREHENSIVE TEST APPROACHES
BY DIFFERENT SAMPLE SIZES AND METHODS

AN ABSTRACT
BY
PIPUTPONG NAWTONG

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

Piputpong Nawtong. (2001). *A Comparison of the item bias in verbal comprehensive test approaches by different sample sizes and methods*. Master Thesis, M.Ed. (Educational Measurement). Bangkok : Graduate School, Srinakharinwirot University. Advisor Committee : Assoc.Prof.Aungkana Saiyos, Mr.Chawalit Rouyajin.

The purpose of this research was to compare the items bias in verbal comprehensive test using three different sample sizes from the reference group: the focal group were 250:250, 250:500 and 500:500. The three different methods were HW3, Biserial and Item-total Correlation. The comparison of the item bias was to determine numbers of the items bias and Split-Half reliabilities of the test after the item bias was deleted. The samples for the studies were 2,800 Mathayom Suksa 2 students of secondary school under Department Of General Education who were studying their 2nd semester in the academic year 2000, selected by two-stage random sampling.

The results of the study were as following:

1. The number of items bias analysed between HW3 and Biserial method, Biserial and Item-total Correlation method and analysed by the sample size from the reference group: the focal group of 250:250 and 500:500 were significantly different at .01 and .05 levels respectively. The sample size from the reference group: the focal group were 250:500 had the number of items bias between HW3 and Item-total Correlation method were not significantly different.

2. The number of items bias analysed by Biserial and Item-total Correlation method between sample sizes from the reference group: the focal group of 500:500 and 250:500 were significantly different at .05 level, but HW3 method were not significantly different in both sample sizes.

3. The Split-Half reliability coefficient of the test deleted items bias analysed between HW3 and Item-total Correlation method and analysed by the sample sizes from the reference group: the focal group of 250:250 and 500:500 were significantly different at .05 and .01 levels respectively. The sample size from the reference group: the focal group were 250:500 had the Split-Half reliability coefficient of the test deleted items bias between HW3 and Biserial method, Biserial and Item-total Correlation method were significantly different at .01 level.

4. The Split-Half reliability coefficient of the test deleted items bias analysed by Biserial method between sample sizes from the reference group: the focal group of 500:500 and 250:500 were significantly different at .01 level, and between the sample sizes from the reference group: the focal group of 250:250 and 500:500 had the Split-Half reliability coefficient of the test deleted items bias by Item-total Correlation method were significantly different at .05 level, but HW3 method were not significantly different in both sample sizes.

ปริญญานิพนธ์ฉบับนี้ได้รับทุนสนับสนุนการทำปริญญานิพนธ์
จากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ
จึงขอขอบพระคุณยิ่ง

ปริญญานิพนธ์

เรื่อง

การเปรียบเทียบผลของความลำเอียงของแบบทดสอบวัดความเข้าใจทางภาษา
โดยใช้ขนาดกลุ่มผู้สอบและวิธีวิเคราะห์ต่างกัน

ของ

นายพิรุณพงษ์ แนวทอง

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา
ของมหาวิทยาลัยศรีนรินทรวิโรฒ

.....คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ ดร.เสริมศักดิ์ วิศวลาภรณ์)
วันที่...17...เดือน...สิงหาคม.... 2544

คณะกรรมการสอบปริญญานิพนธ์

.....ประธาน
(รองศาสตราจารย์อังคณา สายยศ)

.....กรรมการ
(อาจารย์ชวลิต รวยอาจิณ)

.....กรรมการที่แต่งตั้งเพิ่มเติม
(รองศาสตราจารย์วิญญา วิศวลาภรณ์)

.....กรรมการที่แต่งตั้งเพิ่มเติม
(รองศาสตราจารย์ชูศรี วงศ์รัตนะ)

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้ สำเร็จลุล่วงได้ด้วยความช่วยเหลืออย่างดียิ่งจากรองศาสตราจารย์ อังคณา สายยศ และ อาจารย์ชวลิต รวยอาจิณ ประธานและกรรมการที่ปรึกษาปริญญานิพนธ์ และ รองศาสตราจารย์วิญญา วิศาลาภรณ์,รองศาสตราจารย์ชูศรี วงศ์รัตน์ กรรมการสอบปริญญานิพนธ์ ที่ให้คำปรึกษาแนะนำทางและข้อคิดเห็นต่าง ๆ ตลอดเวลาที่ศึกษาและทำปริญญานิพนธ์มาโดยตลอดจนสำเร็จ ผู้วิจัยกราบขอขอบคุณเป็นอย่างสูง

ในการตรวจคุณภาพของเครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้ได้รับความอนุเคราะห์จากรองศาสตราจารย์วิญญา วิศาลาภรณ์ รองศาสตราจารย์นิภา ศรีไพโรจน์ อาจารย์ประชุมพร จันทศิริ อาจารย์วาณี โชติพัฒนกุล และอาจารย์ศิวรัตน์ กาญจนพฤกษ์ ผู้วิจัยขอกราบขอขอบคุณ

ขอกราบขอบคุณ กรุงเทพมหานคร ที่ช่วยอนุเคราะห์จัดสรรเงินงบประมาณแผ่นดินให้เป็นทุนในการศึกษาต่อในครั้งนี้

ขอกราบขอบคุณ บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ที่ช่วยอนุเคราะห์จัดสรรเงินสนับสนุนการทำปริญญานิพนธ์ในครั้งนี้

ในการเก็บรวบรวมข้อมูลในการวิจัยได้รับความกรุณาจากผู้บริหารโรงเรียนมัธยมศึกษา ส่วนกลาง กลุ่มที่ 4 ที่อำนวยความสะดวกในการเก็บรวบรวมข้อมูล และได้รับความกรุณาจากผู้บริหารโรงเรียน ผู้ช่วยฝ่ายวิชาการ คณะอาจารย์ทุกท่านที่เกี่ยวข้อง ได้รับความช่วยเหลือ อาจารย์วรรณทิพา สิงหา และอาจารย์นวลองค์ บำรุง ในการเก็บรวบรวมข้อมูล

ขอขอบคุณอาจารย์ กาญจนา จิตต์สุพร และ Mrs. Abigail Salmon นักศึกษาฝึกงานแห่ง Cambridge University ,London ,England ที่ช่วยตรวจสอบบทคัดย่อภาษาอังกฤษ

ขอขอบใจนักเรียนในโรงเรียนกลุ่มตัวอย่าง ที่ให้ความร่วมมือในการเก็บรวบรวมข้อมูลเป็นอย่างดี

ท้ายสุดนี้ ผู้วิจัยของน้อมรำลึกถึงพระคุณของบิดา มารดา ครู อาจารย์ ที่ให้การอบรมสั่งสอน ให้ความเมตตาช่วยเหลือ และเป็นกำลังใจ สนับสนุนการศึกษาของผู้วิจัยตั้งแต่อดีตจนถึงปัจจุบัน

พิรุฬพงษ์ แนวทอง

สารบัญ

บทที่	หน้า
1 บทนำ	1
ภูมิหลัง	1
ความมุ่งหมายของการศึกษาค้นคว้า.....	6
ความสำคัญของการศึกษาค้นคว้า.....	6
ขอบเขตของการศึกษาค้นคว้า.....	7
ประชากรและกลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้า	7
ตัวแปรที่ศึกษา.....	7
นิยามศัพท์เฉพาะ.....	8
สมมติฐานในการวิจัย	11
2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	12
ความหมายของความลำเอียงของข้อสอบ.....	12
ทฤษฎีเกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบ.....	14
หลักการวิเคราะห์ความลำเอียง.....	31
งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบ.....	33
งานวิจัยภายในประเทศ.....	33
งานวิจัยต่างประเทศ.....	40
3 วิธีดำเนินการศึกษาค้นคว้า.....	48
การกำหนดประชากรและและการเลือกกลุ่มตัวอย่าง.....	48
การสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	50
การจัดกระทำและการวิเคราะห์ข้อมูล.....	55
วิธีการเก็บรวบรวมข้อมูล.....	54
4 ผลการวิเคราะห์ข้อมูล.....	63
สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล.....	63
การเสนอผลการวิเคราะห์ข้อมูล.....	64
ผลการวิเคราะห์.....	64

สารบัญ(ต่อ)

บทที่	หน้า
5	
สรุปผล อภิปรายผล และข้อเสนอแนะ.....	73
สังเขปความมุ่งหมาย สมมติฐานและวิธีดำเนินการศึกษาค้นคว้า.....	73
สรุปผลการศึกษาค้นคว้า.....	75
อภิปรายผล.....	77
ข้อเสนอแนะ.....	79
 บรรณานุกรม.....	 81
 ภาคผนวก.....	 96
 ประวัติย่อผู้วิจัย.....	 113

บัญชีตาราง

ตาราง		หน้า
1	แสดงข้อมูลผลการสอบของกลุ่มตัวอย่างกลุ่ม R และ F ที่มีคะแนนรวมอยู่ในช่วงคะแนน j ข้อที่ i.....	19
2	ข้อมูลในระดับความสามารถที่ k จากตัวแปร.....	21
3	จำนวนนักเรียนในกลุ่มตัวอย่างที่ใช้ในการวิจัย จำแนกตามขนาดโรงเรียน.....	49
4	ค่าความลำเอียงและจำนวนข้อสอบที่ลำเอียงจากแบบทดสอบวัดความเข้าใจทางภาษา จำแนกตามวิธีวิเคราะห์ความลำเอียงและลักษณะกลุ่มผู้สอบทั้ง 3 ลักษณะ.....	65
5	ผลการเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียง ระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบเดียวกัน เมื่อวิเคราะห์ความลำเอียงวิธีต่างกัน.....	66
6	ผลการเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบต่างกัน เมื่อวิเคราะห์ความลำเอียงวิธีเดียวกัน	67
7	จำนวนข้อสอบหลังจากคัดเลือกข้อสอบที่มีมีความลำเอียงออกแล้ว ค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับ ค่าฟิชเชอร์-ซีและการทดสอบด้วยสถิติไคสแควร์ระหว่างวิธีวิเคราะห์.....	69
8	ผลการเปรียบเทียบค่าความเชื่อมั่นในรูปค่าฟิชเชอร์-ซีของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบเดียวกัน เมื่อวิเคราะห์ความลำเอียงวิธีต่างกัน.....	70
9	ผลการเปรียบเทียบค่าความเชื่อมั่นในรูปค่าฟิชเชอร์-ซีของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบต่างกัน เมื่อวิเคราะห์ความลำเอียงวิธีเดียวกัน.....	71
10	ค่าความยาก(p) และค่าอำนาจจำแนก (r) รายข้อของแบบทดสอบวัดความเข้าใจทางภาษา จากการทดสอบครั้งที่ 1 ฉบับที่ 1.....	92
11	ค่าความยาก(p) และค่าอำนาจจำแนก (r) รายข้อของแบบทดสอบวัดความเข้าใจทางภาษา จากการทดสอบครั้งที่ 1 ฉบับที่ 2.....	93

บัญชีตาราง(ต่อ)

ตาราง		หน้า
12	ค่าความล่าเียงจากการวิเคราะห์ความล่าเียงของข้อสอบด้วยวิธีค่าความ สัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด จำแนกตามลักษณะ กลุ่มผู้สอบต่างกัน.....	95
13	ค่าความล่าเียงจากการวิเคราะห์ความล่าเียงของข้อสอบด้วยวิธีไบชีเรียล จำแนกตามลักษณะกลุ่มผู้สอบต่างกัน.....	97
14	ค่าความล่าเียงจากการวิเคราะห์ความล่าเียงของข้อสอบด้วยวิธี HW3 จำแนกตามลักษณะกลุ่มผู้สอบต่างกัน.....	99

บัญชีภาพประกอบ

ภาพประกอบ	หน้า
1 กราฟไพบาริเอทของค่าเฉลี่ยของคณพืวดำกลุ่ม 1 และคณพืวดำกลุ่ม 2 จากแบบทดสอบวัดความถนัดทางการเรียนด้านภาษา.....	16
2 กราฟไพบาริเอทของค่าเฉลี่ยของคณพืวดำกลุ่ม 5 และคณพืวดำกลุ่ม 1 จากแบบทดสอบวัดความถนัดทางการเรียนด้านภาษา.....	17
3 ความหมายของค่าพารามิเตอร์ข้อสอบตามโมเดลโลจิสติกที่มีพารามิเตอร์ 1 ตัว.....	28
4 ความหมายของค่าพารามิเตอร์ข้อสอบตามโมเดลโลจิสติกที่มีพารามิเตอร์ 2 ตัว.....	29
5 ความหมายของค่าพารามิเตอร์ข้อสอบตามโมเดลโลจิสติกที่มีพารามิเตอร์ 3 ตัว.....	30
6 ลำดับขั้นตอนในการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	50

ภูมิหลัง

การวัดผลและการประเมินผลการศึกษา นับเป็นเครื่องมือสำคัญที่ช่วยในการปรับปรุงคุณภาพของการศึกษา เพราะผลจากการวัดผลและประเมินผลการศึกษา จะเป็นพื้นฐานสำหรับการตัดสินใจของครูและ นักการศึกษา เพื่อใช้ในการปรับปรุงวิธีการสอน การใช้สื่อการสอน แบบเรียน หลักสูตร การแนะแนว การจัดระบบบริหารของโรงเรียน ตลอดจนการปรับปรุงวิธีการเรียน ของผู้เรียนให้มีประสิทธิภาพยิ่งขึ้นจะส่งผลให้การดำเนินการจัดการศึกษาเป็นไปตามแนวทางและบรรลุผลตามจุดมุ่งหมายที่หลักสูตรต้องการ (อนันต์ ศรีโสภณ. 2524 : 1 ; ไพฑูรย์ เวทการ. 2536 : 7)

ในการวัดผลการศึกษาและจิตวิทยานั้นสิ่งที่วัด และ ตรวจสอบส่วนใหญ่เป็นคุณลักษณะภายใน (Latent Traits) ที่มีความซับซ้อนและแปรเปลี่ยนไปตามสถานการณ์ ทำให้ไม่สามารถวัดคุณลักษณะดังกล่าวได้โดยตรง ดังนั้นจึงต้องอาศัยวิธีการวัดทางอ้อมโดยกำหนดสิ่งเร้า (Stimuli) เช่น แบบทดสอบ ซึ่งจะวัดสิ่งที่มีความสัมพันธ์เชิงปริมาณกับสิ่งที่ต้องการจะวัด แล้วนำผลที่ได้จากการวัดไปอ้างอิง (Infer) หรืออธิบายกลับไปสู่สิ่งที่ต้องการวัดจริงหรือทำนายความสามารถที่แท้จริง (True ability or True score) อันเป็นคุณลักษณะภายในของบุคคลที่ถูกวัด (Lord . 1980 : 3) เช่น คุณลักษณะด้านพุทธิพิสัย จิตพิสัยและทักษะพิสัย เป็นต้น และเครื่องมือที่ใช้ในการวัดผลการเรียนรู้ของนักเรียนเพื่อดึงเอาสิ่งที่แฝงเร้นอยู่ภายในเป็นสิ่งที่เราต้องการวัดออกมาได้ เช่น การสังเกต การสัมภาษณ์ การใช้แบบทดสอบ การจัดอันดับคุณภาพ เป็นต้น (ล้วน สายยศและอังคณา สายยศ . 2539 : 21 – 29) แต่เครื่องมือที่นิยมใช้กันมากในวงการศึกษาคือ แบบทดสอบ (อเนก เพียรอนุกุลบุตร. 2524 : 148) สาเหตุที่แบบทดสอบเป็นที่นิยมใช้ในการวัดผลเกือบทุกระดับเนื่องจากมีความสะดวกในการนำมาใช้ ครูและนักการศึกษาส่วนใหญ่มีความเข้าใจและคุ้นเคยกับเครื่องมือประเภทนี้รวมถึงขั้นตอนการสร้างก็ไม่ยุ่งยากเกินไป นอกจากนี้แบบทดสอบยังใช้วัดความสามารถได้ตั้งแต่พฤติกรรมขั้นความรู้ความจำจนถึงพฤติกรรมขั้นการสังเคราะห์และการประเมินค่า และยังมีความเป็นปรนัย (Objectivity) สูง หมายถึง การเข้าใจคำตรงกัน การตรวจให้คะแนนตรงกัน และการแปลผลคะแนนที่ได้ตรงกัน ส่วนในด้านคุณภาพของแบบทดสอบจะเน้นหนักในด้านความเชื่อมั่น (Reliability) ความเที่ยงตรง (Validity) ความยากง่าย และค่าอำนาจจำแนก แบบทดสอบเลือกตอบมีประโยชน์และคุณค่าต่อการวัดผลก็จริง แต่ยังมีข้อสงสัยว่านักเรียนที่เลือกข้อนั้นมีความรู้ในเรื่องนั้นจริงหรือว่าตอบแบบเดาหรือนักเรียนตอบข้อสอบนั้นผิดเพราะไม่มีความรู้ในเรื่องนั้นจากปัญหาดังกล่าว ทำให้มีการพัฒนาแบบทดสอบมาเป็นแบบทดสอบเลือกตอบชนิดคำตอบซ้อน (Double Multiple Choice) และแบบทดสอบเลือกตอบชนิดถูกผิดหลายตัวเลือกหรือแบบทดสอบเลือกตอบถูกผิด (Multiple True False : MTF) และแบบทดสอบเลือกแบบหลายตัวเลือกนี้สามารถแก้การเดาข้อสอบได้ ทั้งนี้เพราะว่าข้อสอบเลือกตอบแบบถูกผิดนักเรียนจะต้องคิดทุกตัวเลือก

เพื่อพิจารณาว่าตัวเลือกใดถูกและตัวเลือกใดผิดไม่เหมือนกับข้อสอบแบบธรรมดาเพราะถ้าเลือกได้ตัวเลือกถูกแล้วก็ไม่ต้องไปพิจารณาตัวเลือกอื่นทำให้เกิดการสับสนตัวเลือกที่เห็นว่าผิดออกที่ละตัวแล้วเดาตัวเลือกสุดท้ายได้ถูกต้อง ซึ่งเป็นจุดอ่อนของข้อสอบเลือกตอบธรรมดา ในข้อสอบเลือกตอบแบบถูกผิดนี้ได้เริ่มใช้เป็นเครื่องมือวัดผลในกลุ่มนักเรียนแพทย์ ซึ่งปรากฏว่าใช้วัดความรู้นักเรียนได้ผลดี จึงนิยมใช้ข้อสอบเลือกตอบถูกผิดแทนข้อสอบเลือกตอบธรรมดา(อังคณา สายยศ. 2538 : 42)

นอกจากคุณลักษณะที่กล่าวมาแล้วยังมีคุณลักษณะที่สำคัญประการหนึ่ง คือ คุณลักษณะของข้อคำถามที่ให้ความเสมอภาค ความยุติธรรม (Fairness) ของผู้สอบไม่เกิดการได้เปรียบเสียเปรียบกับกลุ่มผู้สอบกลุ่มใดกลุ่มหนึ่งโดยเฉพาะ ไม่ว่าจะเป็นความแตกต่างกันในด้าน เพศ ภาษา เชื้อชาติ วัฒนธรรม ศาสนา ฐานะทางเศรษฐกิจของครอบครัว ภูมิภาคหรือสภาพทางภูมิศาสตร์ ซึ่งคุณลักษณะดังกล่าวนี้เป็นคุณลักษณะของข้อคำถามที่ปราศจากความลำเอียง (Bias) ข้อคำถามที่ดีจึงควรเป็นข้อคำถามที่มีทั้งคุณภาพและความเสมอภาค ปราศจากความลำเอียง ในเรื่องความยุติธรรมในการสอบได้มีการศึกษาและได้รับการยอมรับอย่างจริงจังในช่วงปี 1960 โดยมีการเสนอวิธีการต่าง ๆ เพื่อตรวจสอบความลำเอียง (Test Bias) หรือความลำเอียงในการคัดเลือกผู้สอบ (Selection Bias) ในขณะเดียวกันกับผู้สร้างแบบทดสอบก็สนใจวิธีการจำแนกข้อสอบที่ไม่เหมาะสมในแบบทดสอบ เพื่อนำไปปรับปรุงหรือตัดข้อสอบออกจากแบบทดสอบ (สมศักดิ์ จันผ่อง. 2541 : 81) ซึ่งจากงานวิจัยของกรีน (Intasuwan . 1979 : 1 ; citing Green. 1972) พบว่าข้อคำถามบางข้อคำถามของแบบทดสอบวัดผลสัมฤทธิ์ของแคลิฟอร์เนีย (California Achievement Test) ซึ่งเป็นข้อคำถามที่มีคุณภาพ เมื่อไปทดสอบกับกลุ่มผู้สอบที่มีความแตกต่างกันทางเชื้อชาติ ภาษา สภาพเศรษฐกิจ และสภาพภูมิอากาศ จะให้ผลการสอบที่แตกต่างกัน ซึ่งอาจจะมีสาเหตุมาจากวัฒนธรรมและสิ่งแวดล้อมที่ต่างกันมีผลต่อความเข้าใจภาษาที่ใช้ในข้อคำถามต่างกันและเรียกข้อคำถามประเภทนี้ว่า ข้อคำถามลำเอียงหรือข้อสอบลำเอียง (Item Bias or Differential Item Function : DIF) และจากผลงานวิจัยของดูลิทเทิลและเคลียร์ (Doolittle and Cleary .1987 :157 - 166) พบว่าเมื่อนำแบบทดสอบคำนวณทางคณิตศาสตร์ เรขาคณิต และคณิตศาสตร์การใช้เหตุผลไปทดสอบกับกลุ่มตัวอย่างเพศชายและเพศหญิงและข้อความบางข้อความนั้นลำเอียงต่อเพศ กล่าวคือมีความลำเอียงต่อนักเรียนเพศชาย นอกจากนั้นงานวิจัยของ สบง ลักษณะ (Burrill .1982 : 168 ; Citing Lasana.1979) ได้ศึกษาถึงความลำเอียงทางเพศและผิวของแบบทดสอบทางด้านภาษาและคณิตศาสตร์ก็พบว่าความลำเอียงของข้อสอบจากแบบทดสอบทางด้านภาษาและคณิตศาสตร์จำนวนหนึ่งระหว่างกลุ่มนักเรียนชายและนักเรียนหญิงและนักเรียนผิวดำและนักเรียนผิวขาวทั้งนี้อาจเนื่องมาจากข้อสอบที่ใช้ เช่น ข้อสอบวิชาคณิตศาสตร์จำเป็นต้องอาศัยความสามารถด้านการอ่านใจหยาบปัญหา ถ้ากลุ่มนักเรียนเพศชายหรือนักเรียนหญิง และนักเรียนผิวดำหรือนักเรียนผิวขาว มีความสามารถทางการอ่านไม่เท่ากัน ข้อสอบข้อนี้ย่อมมีความลำเอียงทางการทดสอบ

หรือข้อสอบที่ใช้วัดความสามารถทางภาษา ถ้ากลุ่มนักเรียนเพศชายหรือนักเรียนหญิง และนักเรียนผิวดำหรือนักเรียนผิวขาว มีการได้เปรียบเสียเปรียบเกี่ยวกับภาษาที่ใช้แตกต่างกันย่อมทำให้มีความลำเอียงทางการทดสอบได้เช่นกัน ซึ่งสอดคล้องกับงานวิจัยของชัชชัย เผ่าพงษ์ (2527 : 73 – 76) นิรมล ชัยชวลิต (2537 : 57 – 59) พรรณี จินตมาศ (2540 : 90– 91) เชปาร์ด (Shepard . 1982 : 11) และ ปอปแฮม (Popham . 1981 : 181) ที่ระบุว่าภาษา เพศ เป็นแหล่งที่ก่อให้เกิดความลำเอียง ส่วน คมศักดิ์ ชื่นชม (2539 : บทคัดย่อ) จิตสุตา ธารพร (2539 :131 – 132) ระบุว่าศาสนาที่นักเรียนนับถือเป็นแหล่งที่ก่อให้เกิดความลำเอียงและ พัชรี ปิยภักดิ์ (2531 : 65 – 67) ระบุว่านักเรียนที่มีสภาพทางภูมิศาสตร์หรือมีพื้นที่ต่างกัน เป็นแหล่งก่อให้เกิดความลำเอียงของแบบทดสอบ

จากที่กล่าวมาข้างต้นจะเห็นว่าความแตกต่างของเชื้อชาติ ศาสนา วัฒนธรรม สภาพทางภูมิศาสตร์ สภาพทางสังคม ล้วนแล้วแต่เป็นแหล่งที่ทำให้ข้อคำถามบางข้อที่มีคุณภาพเกิดความลำเอียงกับกลุ่มผู้สอบบางกลุ่ม นอกจากนี้ภาษาและเพศก็เป็นแหล่งที่ก่อให้เกิดความลำเอียงของข้อคำถามด้วยเช่นกัน ฉะนั้นในการสร้างและพัฒนาแบบทดสอบไม่ว่าจะเป็นแบบทดสอบวัดผลสัมฤทธิ์ แบบทดสอบวัดความถนัด แบบทดสอบวัดเจตคติ และแบบทดสอบวัดผลภาคปฏิบัติ ควรได้รับการวิเคราะห์ความลำเอียงก่อนที่จะไปใช้ในการสอบจริง (Berk . 1982 : 3) โดยเฉพาะแบบทดสอบที่มีภาษาเข้ามาเกี่ยวข้อง เช่น แบบทดสอบวัดความเข้าใจทางภาษาที่ผู้สอบจะต้องใช้ความเข้าใจทางภาษาเพื่อแปลความหมาย ตีความหมาย ขยายความและจับใจความสำคัญหรือสรุปสาระสำคัญ ตลอดจนใช้เหตุผลพิจารณาตัดสินใจเพื่อสรุปในสถานการณ์ที่กำหนดให้ได้ ด้วยเหตุผลดังกล่าวจึงควรศึกษาทั้งคุณภาพและความลำเอียงของข้อสอบระหว่างกลุ่มผู้สอบที่มีเพศต่างกัน เพื่อให้แน่ใจว่าผลที่ได้จากการวัดมีความถูกต้อง แม่นยำ เป็นการวัดความสามารถที่แท้จริงของผู้สอบโดยไม่เกิดการได้เปรียบเสียเปรียบต่อกลุ่มผู้สอบเพศใดเพศหนึ่งโดยเฉพาะ

สำหรับวิธีการวิเคราะห์ความลำเอียงของข้อสอบนั้น มีวิธีวิเคราะห์หลายวิธี ดังนี้วิธีที่ใช้หลักของทฤษฎีการวัดผลแบบดั้งเดิม (Classical Test Theory) เช่น วิธีแปลงค่าความยาก (Transformed Item Difficulty) การวิเคราะห์ความแปรปรวน (Analysis of variance) การวิเคราะห์ด้วยไคสแควร์ (Chi – Square) วิธีวิเคราะห์องค์ประกอบ (Factor Analysis) วิธีวิเคราะห์การถดถอย (Regression Analysis) วิธีลอกลีเนียร์ (Log – Linear) วิธีกระบวนการมาตรฐาน (Standardization Procedure) วิธี Golden Rule(Golden Rule Procedure) วิธีการถดถอยลอจิสติก (Logistic Regression Method) วิธีการวิเคราะห์ตัวลวง (Distractor Analysis) วิธีค่าอำนาจจำแนก (Item Discrimination Procedure) เป็นต้น และ วิธีวิเคราะห์ที่ใช้หลักของทฤษฎีการตอบสนองข้อคำถาม (Item Response Theory) เช่น วิธีโค้งลักษณะข้อสอบ-3 พารามิเตอร์(Item Characteristic Curve-3 Parameter) วิธีโค้งลักษณะข้อสอบ-2 พารามิเตอร์ (Item Characteristic Curve-2 Parameter) วิธีโค้งลักษณะข้อสอบ-1 พารามิเตอร์ (Item Characteristic Curve-1 Parameter)

ซึ่งผลการวิเคราะห์ความลำเอียงของข้อสอบโดยหลักของทฤษฎีการตอบข้อคำถาม พบว่า เป็นวิธีที่ให้ผลการวิเคราะห์ถูกต้องที่สุด แต่การใช้ทฤษฎี การตอบข้อคำถามในการวิเคราะห์ความลำเอียงของข้อสอบเป็นวิธีการที่ใช้การคำนวณยุ่งยากซับซ้อน ใช้กลุ่มตัวอย่างขนาดใหญ่ และสิ้นเปลืองค่าใช้จ่ายมาก (Narayanan and Swaminathan. 1994 : 315) ทำให้เกิดความยุ่งยากในทางปฏิบัติ จากเหตุผลที่กล่าวมาข้างต้น จึงมีการศึกษาวิธีการวิเคราะห์ความลำเอียงของข้อสอบ โดยหลีกเลี่ยงการใช้ทฤษฎีการตอบข้อคำถาม สำหรับวิธีที่มีผู้เสนอ คือ วิธีที่ใช้สถิติแบบนอน-พารามेटริก (Non-Parametric) ในการวิเคราะห์ความลำเอียงของข้อสอบ เช่น วิธีแมนเทล-แฮนส์เซล (Mantel – Haenszel:MH) วิธีชิปเทสท์ (SIBTEST:SIB) และวิธี HW 3 (HW 3 Method) ซึ่งทั้ง 3 วิธีนี้เป็นวิธีการคำนวณที่ไม่ยุ่งยากซับซ้อนและสิ้นเปลืองค่าใช้จ่ายไม่มากนักเหมาะที่จะนำมาใช้ในทางปฏิบัติจึงเป็นที่สนใจและนิยมใช้ในปัจจุบัน

มีนักวิจัยหลายท่านได้ทำการศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบแต่ละวิธี ซึ่งวิธีวิเคราะห์ความลำเอียงของข้อสอบแต่ละวิธีที่กล่าวมามีจุดเด่น, จุดด้อย และได้ผลแตกต่างกัน เช่น เชปาร์ดและคณะ (Berk .1982 : 169 ; Citing Shepard,Camilli and Averrill . 1980) สับโคเวียตและคณะ (Subkoviak. 1984 : 49 – 59) พบว่า วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ เป็นวิธีที่เหมาะสมในทางทฤษฎี ถูกต้องและตรงที่สุด แต่ คมศักดิ์ ชื่นชม (2539) พบว่า วิธีค่าอำนาจจำแนก,วิธีวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และวิธีวิเคราะห์องค์ประกอบ ให้ผลการวิเคราะห์ความลำเอียงไม่แตกต่างกันในทางสถิติ สำหรับนารายานานและสวานินาธาน (Narayanan and Swaminathan. 1994 : 315) พบว่า วิธี SIB จะให้ผลการวิเคราะห์ที่ถูกต้องกว่าวิธี MH เมื่อกลุ่มตัวอย่างมีการแจกแจงความสามารถไม่เท่ากัน พรรรณี จินตมาศ(2540 : บทคัดย่อ) ได้ทำการศึกษาเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบโดยใช้ขนาดกลุ่มผู้สอบและวิธีวิเคราะห์ต่างกัน คือ วิธีแปลงค่าความยาก วิธีแมนเทล-แฮนส์เซล และวิธีชิปเทสท์ พบว่า วิธีแมนเทล-แฮนส์เซล กับ วิธีชิปเทสท์ ให้ผลในการค้นพบความลำเอียงของแบบทดสอบได้ใกล้เคียงกัน ส่วนเวลช์และมิลเลอร์ (Welch and Miller. 1995:170 – 174) แวงและเลนี่ (Wang and Lane.1996 :153 - 189) พบว่าวิธี HW 3 พบข้อสอบที่มีความลำเอียงมากที่สุด

นอกจากนี้ยังมีงานวิจัยที่ศึกษาผลกระทบของขนาดกลุ่มตัวอย่างในการวิเคราะห์ความลำเอียงของข้อสอบ จะพบว่าวิธีการวิเคราะห์ข้อสอบที่มีความลำเอียงได้เพิ่มขึ้นเมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้น (Wise. 1987;Hill.1990 : 5 ; Rogers. 1993 : 105 ; Narayanan and Swaminathan. 1994 : 315) นอกจากนี้ ฮิลล์ (Hill.1990 : 5) ได้เสนอว่าในแต่ละกลุ่มจะมีประสิทธิภาพในการวิเคราะห์ความลำเอียงได้ดี เมื่อกลุ่มตัวอย่างแต่ละกลุ่มมีขนาด 300 คน และสวานินาธานและนารายานาน (Narayanan and Swaminathan. 1994 : 315) ยังพบว่าในทางปฏิบัติอัตราการค้นความลำเอียงของข้อสอบ มีผลต่อกลุ่มเปรียบเทียบที่มีขนาดเล็กมากกว่ากลุ่มอ้างอิงที่มีขนาดเล็ก เช่น กลุ่มเปรียบเทียบ 300 คน ($N_R=300$) กับ กลุ่มอ้างอิง 100, 200 ,300 คน ($N_F = 100 , 200 , 300$)

มีอัตราการค้นหาความลำเอียงอยู่ระหว่าง 62 % ถึง 84 % และกลุ่มเปรียบเทียบ 100 คน ($N_R = 100$) กับกลุ่มอ้างอิง 300 , 500 , 1,000 คน ($N_F = 300 , 500 , 1,000$) มีอัตราการค้นหาความลำเอียงอยู่ระหว่าง 62 % ถึง 66 % ส่วน พรณี จินตมาศ (2540) พบว่า เมื่อวิเคราะห์โดยใช้กลุ่มผู้สอบขนาด 500 คน จำนวนข้อสอบที่ลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทุกวิเคราะห์ ส่วนการศึกษาผลกระทบของขนาดกลุ่มตัวอย่างในการวิเคราะห์ความลำเอียงของข้อสอบแบบวิธี HW3 นั้น เวลช์และฮูเวอร์ (Wing and Lane. 1996 : 183–184 ; citing Wech and Hoover. 1993) ได้ทำการเปรียบเทียบวิธีการวิเคราะห์ความลำเอียงสองวิธีคือ วิธีHW3กับ วิธีไคสแควร์แบบแมนเทิล ภายใต้เงื่อนไขของข้อมูลหลาย ๆ เงื่อนไข ซึ่งผลการศึกษาพบว่าดัชนีของ HW3 ใช้ได้ดีกว่าวิธีของแมนเทิลภายใต้อัตราความคลาดเคลื่อนประเภทที่ 1 และดัชนีHW3 มีแนวโน้มที่จะค้นหาความลำเอียงได้ดีกว่าวิธีแมนเทิล เมื่อขนาดกลุ่มผู้สอบเท่ากับหรือมากกว่า 500 คน ขึ้นไป ลักษณะพิเศษอีกอย่างหนึ่งคือในทางปฏิบัติสถิติไคสแควร์แบบแมนเทิลมีผลน้อยกว่า วิธี HW3 เมื่อขนาดกลุ่มผู้สอบประมาณ 250 คน ถึง 500 ข้อ โดยการเปรียบเทียบระดับความสามารถของขนาดกลุ่มผู้สอบที่ได้จากการสังเกต วิธีHW 3 มีอำนาจในการค้นหามากกว่าวิธีดังกล่าวกับกลุ่มผู้สอบ 250 : 250 คน

จากที่กล่าวมาข้างต้นและจากการศึกษารายงานการวิจัยที่ผ่านมา จะเห็นว่ามีการศึกษาความลำเอียงของข้อสอบมีหลายวิธี เช่น วิธีแปลงค่าความยาก (Transformed Item Difficulty) การวิเคราะห์ความแปรปรวน(Analysis of variance) การวิเคราะห์ด้วยไคสแควร์ (Chi – Square) วิธีวิเคราะห์การถดถอย (Regression Analysis) วิธีกระบวนการมาตรฐาน Standardization Procedure) วิธีการถดถอยลอจิสติก (Logistic Regression Method) วิธีค่าอำนาจจำแนก (Item Discrimination Procedure) วิธีโค้งลักษณะข้อสอบ-3 พารามิเตอร์ (Item Characteristic Curve-3 Parameter) วิธีโค้งลักษณะข้อสอบ-2 พารามิเตอร์ (Item Characteristic Curve-2 Parameter) วิธีโค้งลักษณะข้อสอบ-1 พารามิเตอร์ (Item Characteristic Curve-1 Parameter) วิธีแมนเทิล – แฮนส์เซล (Mantel –Haenszel : MH) วิธีซิปเทสท์ (SIBTEST : SIB) และ วิธีHW3 (HW 3 Method) แต่ยังไม่มีการวิจัยที่ศึกษาเปรียบเทียบความลำเอียงของข้อสอบเมื่อขนาดกลุ่มผู้สอบในกลุ่มเปรียบเทียบและกลุ่มอ้างอิงต่างกัน โดยวิธีการค่าอำนาจจำแนก (Item Discrimination Procedure) ด้วยวิธี HW3 (HW3 Method) วิธีไบซีเรียล และวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับผลรวมทั้งหมด (Item-total Correlation) ซึ่งเป็นวิธีที่ใช้ได้ดีในทางปฏิบัติ ง่ายต่อการคำนวณ และการแปลผล

ด้วยเหตุผลดังกล่าวข้างต้น ผู้วิจัยจึงสนใจที่จะศึกษาผลการวิเคราะห์ความลำเอียงของแบบทดสอบต่อกลุ่มผู้สอบที่มีเพศแตกต่างกัน โดยใช้วิธีการค่าอำนาจจำแนก (Item Discrimination Procedure) ทั้ง 3 วิธีคือวิธีHW 3 (HW 3 Method) วิธีไบซีเรียลและวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนทั้งหมด (Item-total Correlation) และขนาดกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่ม

เปรียบเทียบในการวิเคราะห์ต่างกัน เนื่องจากวิธีการวิเคราะห์ทั้ง 3 วิธีนี้ เป็นวิธีที่เหมาะสมกับการวิเคราะห์ความลำเอียงกับข้อมูลที่มีการตรวจให้คะแนนแบบไม่เป็น 0 – 1 ดังนั้นผู้วิจัยจึงใช้เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้จึงเป็นแบบทดสอบแบบตัวเลือกถูกผิดที่มีการตรวจให้คะแนนทุกตัวเลือกโดยศึกษาจำนวนข้อสอบที่มีความลำเอียงและค่าเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว อันจะเป็นแนวทางในการนำเอาวิธีการวิเคราะห์ความลำเอียงมาใช้ในทางปฏิบัติ ตลอดจนปรับปรุง แก้ไขข้อสอบที่มีความลำเอียงให้สามารถนำไปใช้ได้อย่างยุติธรรมต่อไป

จุดมุ่งหมายของการศึกษาค้นคว้า

ในการศึกษาค้นคว้าครั้งนี้ ผู้วิจัยได้ตั้งจุดมุ่งหมาย ได้ดังนี้

1. เพื่อเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียงด้วยวิธีต่างกัน เมื่อมีขนาดกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบเดียวกัน
2. เพื่อเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบต่างกัน เมื่อวิเคราะห์ความลำเอียงด้วยวิธีเดียวกัน
3. เพื่อเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วระหว่างวิธีวิเคราะห์ความลำเอียงด้วยวิธีต่างกัน เมื่อมีขนาดกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบเดียวกัน
4. เพื่อเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบต่างกัน เมื่อวิเคราะห์ความลำเอียงด้วยวิธีเดียวกัน

ความสำคัญของการศึกษาค้นคว้า

ผลจากการศึกษาค้นคว้าครั้งนี้ จะทำให้ทราบว่าวิธีวิเคราะห์ความลำเอียงของข้อสอบที่ใช้ลักษณะตัวแปรในการศึกษาความลำเอียงที่เป็นเพศ และ ขนาดกลุ่มผู้สอบที่เป็นกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีจำนวนต่างกันในการวิเคราะห์ความลำเอียง วิธีใดจะพบจำนวนข้อสอบที่มีความลำเอียงมากที่สุดและทราบถึงค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว สำหรับใช้เป็นแนวทางในการตัดสินใจเลือกใช้วิธีวิเคราะห์ความลำเอียงที่เหมาะสมมาใช้ในทางปฏิบัติพัฒนาแบบทดสอบให้มีคุณภาพและมีความยุติธรรมต่อไป

ขอบเขตของการศึกษาค้นคว้า

ประชากรและกลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้า

ประชากรที่ใช้ในการศึกษาค้นคว้า

ประชากรที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 1 ปีการศึกษา 2543 นักเรียนในโรงเรียนสังกัดกรมสามัญศึกษาจังหวัดกรุงเทพมหานคร กลุ่มที่ 4 จำนวน 14 โรงเรียน มีจำนวน 174 ห้องเรียน มีจำนวนนักเรียนทั้งหมด 7,095 คน แยกเป็นเพศชาย 3,435 คน เพศหญิง 3,660 คน

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้า

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนที่มีเพศต่างกัน ชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 1 ปีการศึกษา 2543 ซึ่งเลือกมาโดยการสุ่มแบบสองขั้นตอน (Two – stage Random Sampling) โดยมีขนาดโรงเรียนเป็นชั้น (Strata) โรงเรียนเป็นหน่วยการสุ่ม (Sampling Unit) ได้จำนวนโรงเรียน 7 มีจำนวนห้อง 56 ห้องเรียน มีจำนวนนักเรียน 2,800 คน กลุ่มตัวอย่างแบ่งเป็น 2 กลุ่ม คือ

1. กลุ่มตัวอย่างที่ใช้ในการศึกษาผลการวิเคราะห์ความลำเอียง ด้านจำนวนข้อสอบที่มีความลำเอียง จำนวน 1,000 คน เป็นนักเรียนที่มีเพศแตกต่างกัน คือ นักเรียนเพศชายจำนวน 500 คน และเป็นนักเรียนเพศหญิงจำนวน 500 คน

2. กลุ่มตัวอย่างที่ใช้ในการศึกษาผลวิเคราะห์ความลำเอียงด้านความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว จำนวน 1,800 คน

ตัวแปรที่ศึกษา

1. ตัวแปรอิสระ ได้แก่

1.1 วิธีวิเคราะห์ความลำเอียงด้วยวิธีค่าอำนาจจำแนกที่แตกต่างกัน 3 วิธี คือ

1.1.1 วิธี HW3

1.1.2 วิธีไบซีเรียล

1.1.3 วิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด (Item – total Correlation)

1.2 ขนาดกลุ่มผู้สอบที่ใช้ในการวิเคราะห์มี 2 กลุ่ม คือกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ซึ่งสามารถแบ่งตามขนาดกลุ่มผู้สอบที่วิเคราะห์มีลักษณะแตกต่างกัน 3 รูปแบบ ดังนี้

1.2.1 จำนวนกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:250

1.2.2 จำนวนกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 500:500

1.2.3 จำนวนกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:500

2.ตัวแปรตาม ได้แก่

2.1 จำนวนข้อสอบที่มีความลำเอียง

2.2 ค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียง
ออกแล้ว

3.ตัวแปรที่ใช้ในการวิเคราะห์ความลำเอียงเป็นเพศของนักเรียน ได้แก่ เพศชาย
และ เพศหญิง

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้ คือ แบบทดสอบวัดความเข้าใจทางภาษา
จำนวน 40 ข้อ ชนิด 5 ตัวเลือก ซึ่งเป็นแบบทดสอบแบบตัวเลือกตอบถูกผิด ที่ผู้วิจัยสร้างขึ้นเอง

นิยามศัพท์เฉพาะ

1. ความเข้าใจทางภาษา หมายถึง ความสามารถในการแปลความ ตีความ ขยายความ
และจับใจความสำคัญหรือสรุปสาระสำคัญและทำการตัดสินใจเพื่อสรุปสถานการณ์ของเนื้อเรื่อง
หรือบทความประเภทร้อยแก้วและร้อยกรองที่อ่านได้

2. แบบทดสอบวัดความเข้าใจทางภาษา หมายถึง ชุดของข้อคำถามที่วัดความเข้าใจทาง
ภาษา มีจำนวน 40 ข้อ และเป็นแบบทดสอบที่ผู้วิจัยสร้างขึ้นเอง ซึ่งแบบทดสอบวัดความเข้าใจทาง
ภาษาในครั้งนี้เป็นแบบตัวเลือกถูก-ผิด (Multiple True - False) ซึ่งมีลักษณะดังนี้

2.1 แบบทดสอบแบบตัวเลือกถูก-ผิด (Multiple True-False) หมายถึง แบบ
ทดสอบที่ไม่มีตัวเลือกถูกเลยหรือมีตัวเลือกถูกตั้งแต่ 1 ตัวเลือกขึ้นไป ในการตอบให้ผู้สอบพิจารณา
ทุกตัวเลือก โดยทำเครื่องหมายถูก (✓) ให้ตรงกับตัวเลือกที่นักเรียนคิดว่าถูก และทำเครื่องหมาย
ผิด (X) ให้ตรงกับตัวเลือกที่คิดว่าผิด ส่วนการตรวจให้คะแนนจะใช้วิธีการตรวจให้คะแนนแบบให้
คะแนนทุกตัวเลือก กล่าวคือ ในข้อหนึ่งมีคะแนนเต็ม 5 คะแนน หากคำตอบของผู้สอบในตัวเลือก
ใดตรงกับเฉลยไม่ว่าตัวเลือกนั้นจะถูกหรือผิดก็ตามให้ตัวเลือกละ 1 คะแนน ถ้าหากไม่ตรงกับเฉลย
เลยให้ 0 คะแนน

3. ข้อสอบที่มีความลำเอียง หมายถึง ข้อสอบที่วัดความสามารถหรือ คุณลักษณะเดียวกัน
แต่เมื่อนำไปสอบกับผู้สอบที่มีเพศต่างกัน 2 กลุ่ม แต่มีความสามารถเท่ากัน แล้วได้ผลการสอบต่าง
กัน

4. ดัชนีความลำเอียงของข้อสอบ หมายถึงค่าสถิติที่ใช้แสดงความแตกต่างของผลการตอบข้อสอบระหว่างกลุ่มผู้สอบที่มีเพศต่างกัน

5. วิธีการวิเคราะห์ความลำเอียงของข้อสอบ หมายถึง กระบวนการคำนวณค่าดัชนีความลำเอียงของข้อสอบ ซึ่งในที่นี้ใช้วิธีวิเคราะห์ความลำเอียงด้วยวิธีค่าอำนาจจำแนก 3 วิธี ดังนี้

5.1 วิธี HW3 หมายถึง การวิเคราะห์หาความลำเอียงของข้อสอบโดยการพิจารณาจากค่าเฉลี่ยระหว่างกลุ่มผู้สอบ 2 กลุ่ม โดยใช้สถิติ t -test independent เนื่องจากค่า t -test ที่คำนวณได้นี้เหมาะสำหรับการหาความลำเอียงกับกลุ่มผู้สอบขนาดเล็ก (Welch and Hoover.1993) ในการศึกษาครั้งนี้ผู้วิจัยได้ทำการศึกษาความลำเอียงกับกลุ่มตัวอย่างที่มีขนาดแตกต่างกันและมีขนาดใหญ่ จึงทำการปรับแก้ค่า t -test ที่มีผลมาจากขนาดกลุ่มผู้สอบ(d_k) และ คำนวณค่าความคลาดเคลื่อนมาตรฐานของช่วงพิสัยค่าน้ำหนักของแต่ละกลุ่ม (S_d) และแล้วทดสอบด้วย Z ถ้าค่า Z แตกต่างจาก 0 อย่างมีนัยสำคัญ แสดงว่าข้อสอบข้อนั้นมีความลำเอียง (Hedges. et al .1989)

5.2 วิธีไบซีเรียล หมายถึง การวิเคราะห์หาดัชนีความลำเอียงของข้อสอบ โดยพิจารณาความแตกต่างของค่าอำนาจจำแนกของข้อสอบ ที่ได้มาจากการคำนวณด้วยวิธีไบซีเรียลระหว่างกลุ่มผู้สอบต่างลักษณะกัน หากพบที่มีความแตกต่างอย่างมีนัยสำคัญทางสถิติ แสดงว่าข้อสอบข้อนั้นมีความลำเอียง

5.3 วิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด (Item-total Correlation) หมายถึง การวิเคราะห์หาดัชนีความลำเอียงของข้อสอบ โดยพิจารณาความแตกต่างของค่าอำนาจจำแนกของข้อสอบที่คำนวณจากค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด (Item-total Correlation) หากพบที่มีความแตกต่างอย่างมีนัยสำคัญทางสถิติ แสดงว่าข้อสอบข้อนั้นมีความลำเอียง

6. ขนาดของกลุ่มผู้สอบ หมายถึง จำนวนกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่ตอบข้อสอบจากแบบทดสอบที่ผู้วิจัยสร้างขึ้น โดยแบ่งเป็น 4 กลุ่ม ดังนี้

6.1 กลุ่มอ้างอิง จำนวน 250 คน

6.2 กลุ่มอ้างอิง จำนวน 500 คน

6.3 กลุ่มเปรียบเทียบ จำนวน 250 คน

6.4 กลุ่มเปรียบเทียบ จำนวน 500 คน

ในการวิเคราะห์ความลำเอียงของแบบทดสอบระหว่างผู้สอบในกลุ่มอ้างอิง:กลุ่มเปรียบเทียบ ในการวิเคราะห์ครั้งนี้ มี 3 รูปแบบ ดังนี้

รูปแบบที่ 1 จำนวนกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเท่ากัน เป็น 250:250

รูปแบบที่ 2 จำนวนกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเท่ากัน เป็น 500:500

รูปแบบที่ 3 จำนวนกลุ่มอ้างอิง:กลุ่มเปรียบเทียบต่างกัน เป็น 250:500

7. กลุ่มอ้างอิง (Reference Group : R) หมายถึง กลุ่มผู้สอบที่คาดว่าจะจะเป็นกลุ่มที่ได้ประโยชน์จากข้อสอบวัดความเข้าใจทางภาษา คือเป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องมากกว่าผู้สอบอีกกลุ่มหนึ่งจากแบบทดสอบวัดความเข้าใจทางภาษาฉบับเดียวกัน ในการวิจัยครั้งนี้ กลุ่มอ้างอิงคือนักเรียนเพศหญิง

8. กลุ่มเปรียบเทียบ (Focal Group : F) หมายถึง กลุ่มผู้สอบที่คาดว่าจะจะเป็นกลุ่มเสียประโยชน์จากข้อสอบวัดความเข้าใจทางภาษา คือ เป็นกลุ่มที่มีความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องต่ำกว่าผู้สอบกลุ่มอ้างอิงจากแบบทดสอบวัดความเข้าใจทางภาษาฉบับเดียวกัน ในการวิจัยครั้งนี้ กลุ่มเปรียบเทียบคือนักเรียนเพศชาย

9. ผลการวิเคราะห์ความลำเอียง หมายถึง ผลที่ได้จากการคำนวณด้วยวิธีวิเคราะห์ความลำเอียงโดยวิธีค่าอำนาจจำแนกที่แตกต่างกัน 3 วิธี ซึ่งได้แก่ วิธี HW3 วิธีไบซีเรียล และ วิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด (Item-total Correlation) ซึ่งในการศึกษาค้นคว้าครั้งนี้ หมายถึง

9.1 จำนวนข้อสอบที่มีความลำเอียง หมายถึง จำนวนข้อสอบจากแบบทดสอบวัดความเข้าใจทางภาษาที่ผู้วิจัยสร้างขึ้นเอง ให้ผลการสอบแตกต่างกันระหว่างกลุ่มผู้สอบที่มีเพศต่างกันระหว่างนักเรียนเพศชายและหญิงแต่มีความสามารถเท่ากัน ที่วิเคราะห์ความลำเอียงด้วยวิธีวิเคราะห์ต่างกัน 2 วิธี จำแนกในแต่ละวิธีวิเคราะห์ความลำเอียงและขนาดกลุ่มผู้สอบ

9.2 ค่าความเชื่อมั่นของแบบทดสอบ หมายถึง คุณสมบัติของแบบทดสอบที่วัดความรู้ความสามารถของนักเรียนได้คงที่แน่นอน ซึ่งในการวิจัยครั้งนี้ศึกษาความเชื่อมั่นแบบครึ่งฉบับ (Split - Half) โดยคำนวณจากสูตรสหสัมพันธ์แบบเพียร์สัน แล้วทำการขยายความเชื่อมั่นของแบบทดสอบทั้งฉบับ โดยใช้สูตรขยายของสเปียร์แมนบราวน์ (Spearman-Brown Formula) ซึ่งการคำนวณค่าความเชื่อมั่นนี้ คำนวณหลังจากการคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว

10. ผู้เชี่ยวชาญ หมายถึง ครูผู้สอนวิชาภาษาไทยในระดับชั้นมัธยมศึกษาปีที่ 2 ที่มีประสบการณ์ อย่างน้อย 5 ปี จำนวน 3 คน และนักวัดผลการศึกษานักเรียน 2 คน

สมมติฐานในการวิจัย

1. จำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียงด้วยวิธีต่างกัน เมื่อมีขนาดกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบเดียวกัน จะมีจำนวนแตกต่างกัน
2. จำนวนข้อสอบที่มีความลำเอียงระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงด้วยวิธีเดียวกัน จะมีจำนวนแตกต่างกัน
3. ค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว ระหว่างวิธีวิเคราะห์ความลำเอียงด้วยวิธีต่างกัน เมื่อมีขนาดกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบเดียวกัน จะมีค่าแตกต่างกัน
4. ค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว ระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบต่างกัน เมื่อวิเคราะห์ความลำเอียงด้วยวิธีเดียวกัน จะมีค่าแตกต่างกัน

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องกับการศึกษาค้นคว้าครั้งนี้ ซึ่งผู้วิจัยได้ศึกษาเนื้อหาต่าง ๆ ดังหัวข้อที่จะเสนอต่อไปนี้

1. ความหมายของความลำเอียงของข้อสอบ
2. ทฤษฎีที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบ
3. หลักการวิเคราะห์ความลำเอียง
4. งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบ

ความหมายของความลำเอียงของข้อสอบ

ได้มีผู้ให้ความหมายของคำว่า ความลำเอียงของข้อสอบ ไว้หลายท่านด้วยกัน เช่น เอลเลนและเยน (Allen and Yen.1979:129) ได้ให้ความหมายว่าข้อสอบลำเอียงเป็นข้อสอบที่วัดความสามารถหรือคุณลักษณะทางจิตวิทยาของผู้สอบที่มีระดับความสามารถเท่ากัน แล้วได้ผลต่างกัน โดยพิจารณาจากโครงสร้างข้อสอบของผู้สอบในแต่ละกลุ่มจะต่างกัน เช่นเดียวกับ เคลียร์และฮิลตัน (Cleary and Hilton.1968:61) แมคแคนลีย์และเมนโดซา (Mc Canley and Mendoza.1985 : 89) ไวนเนอร์และบรัน (Wainer and Braum.1988:129) แฮมเบิลตันและคณะ (Hambleton et al. 1991:110) อังคณา สายยศ (2539:6) ได้ให้ความหมายของความลำเอียงของข้อสอบได้สอดคล้องกันว่าเป็นข้อสอบที่พิจารณาจากผลการสอบที่แตกต่างกันจากผู้สอบสองกลุ่ม เช่น กลุ่มที่มีความแตกต่างกันทางเพศ ภาษา สังคม สภาพทางภูมิศาสตร์ ศาสนา เป็นต้น โดยที่กลุ่มผู้สอบนั้นมีความสามารถเท่ากัน สำหรับ ฮอลแลนด์และไวนเนอร์ (Holland and Wainer.1993:5) บุญชม ศรีสะอาด (2535:195) ให้ความหมายว่าข้อสอบลำเอียงเป็นข้อสอบที่ให้ผลการสอบแตกต่างกันจากผู้สอบที่มีความสามารถไม่ต่างกัน และเสนอว่าผลการสอบที่แตกต่างกันนั้นมาจากองค์ประกอบอย่างบาง เช่น ภาษา เพศ วัฒนธรรม ศาสนา เชื้อชาติ สภาพทางภูมิศาสตร์ สภาพทางเศรษฐกิจและสังคม ประสบการณ์ส่วนตัว เป็นต้น

ส่วนลอร์ด (Lord.1980:212) ฮูลิน,ดราสโกและพาร์สัน (Hulin,Drasgow and Parsons. 1983:152) ซับโคเวียคและคณะ (Subkoviak et al.1984:49) ไพน์ (Cohen , Kim and Wollack. 1996:15 ; citing Pine.1977) ได้ให้ความหมายสอดคล้องกันในรูปของความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้สอบที่มีความสามารถเท่ากัน แต่อยู่ต่างกลุ่มกัน แต่ผลลัพธ์จะไม่เท่ากัน

นอกจากนี้ยังมีผู้ให้ความหมายของความลำเอียงที่แตกต่างกันออกไป เช่น เมเรนส์และเลห์มานน์ (Mehrens and Lehmann.1978:435 - 544) ได้ให้ความหมายของข้อสอบที่มีความลำเอียงด้านวัฒนธรรม (Cultural Bias) ว่าหมายถึง ข้อสอบที่ทำให้กลุ่มผู้สอบที่มีความแตกต่างทางวัฒนธรรมที่มีคะแนนเฉลี่ยแตกต่างกันหรือเป็นข้อสอบที่วัดคุณลักษณะทางจิตวิทยาหรือผลสัมฤทธิ์ในเนื้อหาเดียวกันไม่ตรงกัน และได้ให้ความหมายของข้อสอบที่มีความลำเอียงระหว่างเพศ (Sex Bias) ว่าหมายถึง ข้อสอบที่มีเนื้อหาลำเอียงต่อเพศใดเพศหนึ่งโดยเฉพาะและมีผลต่อการทำข้อสอบแต่ละข้อ ซึ่งสอดคล้องกับ เอลส์และคณะ (Jensen.1980 :369 ; citing Eells and others. 1951) ที่ได้ให้ความหมายของข้อสอบที่มีความลำเอียงทางวัฒนธรรม ว่าหมายถึง ความแตกต่างกันในขอบเขตของการได้รู้จักและคุ้นเคยกับเนื้อหาวิชาเฉพาะหรือกระบวนการเฉพาะ ซึ่งเด็กที่ถูกทดสอบนั้นหากมีกลุ่มใดได้รู้จักและคุ้นเคยมาก่อนย่อมมีโอกาสตอบถูกมากกว่าผู้สอบกลุ่มอื่นที่มีลักษณะของกลุ่มแตกต่างกัน ส่วนกรีนและดราเปอร์ (Shepard,Camilli and Averilli.1981:324 ; citing Green and Draper. 1972) ได้ให้ความหมายของข้อสอบที่มีความลำเอียงว่า หมายถึงข้อสอบที่มีครั้งหนึ่งของการจำแนกที่ดีที่สุดของข้อสอบต่าง ๆ สำหรับกลุ่มใดกลุ่มหนึ่งและมีครั้งมีการจำแนกที่ต่ำมากที่สุดสำหรับกลุ่มอื่น ๆ สำหรับออสเตอร์ลายด์ (Osterlind.1983:10) ไวริสมาและจัวร์ส (Wiersma and Jurs.1990 : 334) คามิลลีและเชปาร์ต (Camilli and Shepard.1994:8) ได้ให้ความหมายของ ความลำเอียงของข้อสอบ ไว้สอดคล้องกันว่า เป็นความคลาดเคลื่อนอย่างมีระบบทางการวัดของกลุ่มผู้สอบ 2 กลุ่มมากกว่าจะเป็นผลของแต่ละบุคคล ปอปแฮม (Popham.1990:178) สมัคคีต จันผ่อง (2541:83) ได้ให้ความหมายของความลำเอียงของข้อสอบไว้สอดคล้องกันว่าเป็นรูปผลการสอบที่ไม่ยุติธรรมเนื่องมาจากความชอบหรือความโอเนียงของบุคคล ส่วนมิลล์แซปและอีเวอร์สัน (Millsap and Everson.1993:297) ได้ให้ความหมายของความลำเอียงของข้อสอบไว้ว่า ความลำเอียงของข้อสอบเกิดจากระบบการวัดที่ไม่แน่นอน โคลและมอส (Robert L.Linn.1989:205 ; citing Cole and Moss.) ได้ให้ความหมายของความลำเอียงว่า เป็นความแตกต่างของความเที่ยงตรงของความสัมพันธ์ของคะแนนสอบกับความสามารถอื่น ๆ หรือ ความสัมพันธ์ของกลุ่มย่อยกับชุดแบบทดสอบ

จากนิยามดังกล่าวข้างต้นพอสรุปได้ว่า ความลำเอียงของข้อสอบ หมายถึง ข้อสอบที่วัดความสามารถหรือคุณลักษณะทางจิตวิทยาเดียวกัน เช่น ใช้แบบทดสอบด้านพุทธิพิสัย จิตพิสัย และ ทักษะพิสัย แต่เมื่อนำไปทดสอบกับกลุ่มผู้สอบ 2 กลุ่ม ที่มีความแตกต่างกันในด้าน เพศ ภาษา วัฒนธรรม ศาสนา เชื้อชาติ สภาพทางภูมิศาสตร์ สภาพทางเศรษฐกิจและสังคม ประสบการณ์ส่วนตัว แต่ที่มีความสามารถไม่แตกต่างกัน แล้วได้ผลการวัดที่แตกต่างกัน

ทฤษฎีที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบ

ได้มีผู้เสนอแนะวิธีการวิเคราะห์ความลำเอียงของข้อสอบไว้หลายท่าน ซึ่งเมื่อนำมาแบ่งเป็นประเภทต่าง ๆ โดยยึดทฤษฎีทางการวัดเป็นเกณฑ์ในการแบ่ง จะได้ 2 ประเภทใหญ่ ๆ คือ

1. วิธีวิเคราะห์ความลำเอียงโดยยึดทฤษฎีการวัดผลแบบดั้งเดิม (Classical Test Theory)
2. วิธีวิเคราะห์ความลำเอียงโดยยึดทฤษฎีการตอบข้อสอบ (Item Response Theory)

วิธีวิเคราะห์ความลำเอียงโดยยึดทฤษฎีการวัดผลแบบดั้งเดิม (Classical Test Theory)

เป็นทฤษฎีวัดผลที่ใช้กันทั่วไป ทฤษฎีนี้อธิบายอิทธิพลของความคลาดเคลื่อนของการวัดที่มีต่อคะแนน โดยอาศัยข้อตกลงเบื้องต้นและข้อสรุปบางประการ ที่ได้จากการใช้หลักการทางสถิติและคณิตศาสตร์ในการพิสูจน์และหาข้อสรุป โดยมีเงื่อนไขว่า ถ้าข้อตกลงเบื้องต้นสมเหตุสมผลแล้ว ข้อสรุปที่ได้จะสมเหตุสมผลด้วย แต่ถ้าข้อตกลงไม่สมเหตุสมผลจะทำให้ได้ข้อสรุปที่ผิดพลาดเช่นกัน ซึ่งมีข้อตกลงดังนี้ (Allen and Yen.1979:56-60 ; Ford.1980:4-5)

$$1. X = T + E$$

คะแนนที่ได้จากการสอบ (คะแนนที่สังเกตได้ ; X) เป็นผลรวมของคะแนนจริง(T) กับคะแนนความคลาดเคลื่อน (E) โดยที่คะแนนจริงของผู้สอบจะเป็นค่าคงที่ ในขณะที่คะแนนจากการสอบและคะแนนความคลาดเคลื่อนจะเป็นค่าที่แปรเปลี่ยนไปในการสอบแต่ละครั้ง

$$2. E(X) = T$$

ค่าความคาดหวังหรือค่าเฉลี่ยของประชากร (μ_x) ของคะแนนจากการสอบของผู้สอบคนหนึ่งเท่ากับคะแนนจริงของผู้สอบนั้น โดยทำการสอบซ้ำหลาย ๆ ครั้ง และการทดสอบซ้ำแต่ละครั้งต้องเป็นอิสระจากกัน

$$3. \rho_{ET} = 0$$

คะแนนความคลาดเคลื่อนและคะแนนจริงที่ได้จากประชากรผู้สอบจากแบบทดสอบฉบับหนึ่งไม่มีความสัมพันธ์กัน

$$4. \rho_{E1E2} = 0$$

คะแนนความคลาดเคลื่อนจากแบบทดสอบ 2 ฉบับ จะไม่มีความสัมพันธ์กัน

$$5. \rho_{E1T2} = 0$$

คะแนนความคลาดเคลื่อนในการสอบแบบทดสอบฉบับที่หนึ่ง ไม่มีความสัมพันธ์กับคะแนนจริงในแบบทดสอบฉบับอื่น

6. ข้อตกลงนี้เป็นข้อตกลงที่กล่าวถึงระดับความคู่ขนานของแบบทดสอบ 2 ฉบับ โดยกล่าวว่า ถ้าแบบทดสอบทั้ง 2 ฉบับมีความคู่ขนาน (Classical Parallel Test) แบบทดสอบทั้ง

2 ฉบับนั้น จะต้องมีความเป็นเอกพันธ์ในเนื้อหา (Homogeneous in Content) คะแนนจริงจากแบบทดสอบทั้ง 2 ฉบับ จะมีค่าเท่ากัน ($T_1 = T_2$) ความแปรปรวนของคะแนนความคลาดเคลื่อนจากแบบทดสอบทั้ง 2 ฉบับ จะมีค่าเท่ากัน ($\sigma^2_{E1} = \sigma^2_{E2}$) คะแนนสอบเฉลี่ยและความแปรปรวนของคะแนนสอบจากแบบทดสอบทั้ง 2 ฉบับ จะมีค่าเท่ากัน ($\mu_1 = \mu_2$ และ $\sigma^2_1 = \sigma^2_2$) ความแปรปรวนร่วมของแบบทดสอบทั้งสองฉบับจะเท่ากัน และความแปรปรวนร่วมกับเกณฑ์ภายนอกของแบบทดสอบจะเท่ากันด้วย

7. ข้อตกลงนี้กล่าวถึง ระดับความคู่ขนานอีกระดับหนึ่งของแบบทดสอบฉบับ คือ แบบทดสอบ 2 ฉบับจะมีค่าความเท่าเทียม (Tau - Equivalent) เมื่อแบบทดสอบทั้ง 2 ฉบับ นั้น มีความเป็นเอกพันธ์ในเนื้อหา และ คะแนนจริงจากแบบทดสอบฉบับที่หนึ่งมีค่าเท่ากับผลรวมของคะแนนจริงจากแบบทดสอบฉบับที่สองกับค่าคงที่ ($T_1 = T_2 + C_{12}$)

ข้อตกลงเบื้องต้นทั้ง 7 ข้อนี้ นำไปสู่ข้อสรุปอีกมากมายที่ใช้ในการอธิบายความสัมพันธ์ระหว่างคะแนนความคลาดเคลื่อน คะแนนจริง และคะแนนจากการสอบ และการแปลผลจากแบบทดสอบที่สร้างขึ้น เพื่อนำผลที่ได้มาอธิบายความสามารถของผู้สอบ โดยพยายามค้นหาคะแนนจริงของผู้สอบ และการหาความคลาดเคลื่อนมาตรฐานในการวัด นอกจากนี้ยังนำข้อตกลงและข้อสรุปมาประยุกต์หาคุณภาพของแบบทดสอบที่สร้างขึ้น เช่น ค่าความเชื่อมั่นและค่าความเที่ยงตรงของแบบทดสอบ ซึ่งคุณภาพเหล่านี้จะมีความเกี่ยวพันอยู่กับค่าสหสัมพันธ์ของความแปรปรวน และความแปรปรวนร่วมของคะแนนจริง คะแนนที่ได้จากการสอบและคะแนนความคลาดเคลื่อน

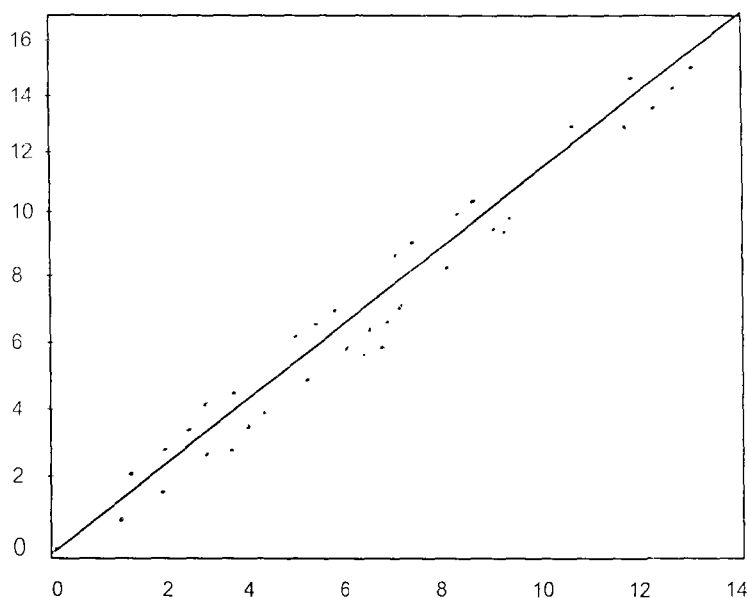
ส่วนวิธีการวิเคราะห์ความลำเอียงของข้อสอบโดยใช้ทฤษฎีแบบดั้งเดิม มีหลายวิธีที่มีผู้นิยมทำการศึกษาไว้ เช่น การแปลงค่าความยาก (Transformed Item Difficulty) วิธีไคสแควร์ (Chi-square Methods) วิธีกระบวนการมาตรฐาน (Standardization Procedure) วิธีการวิเคราะห์ความแปรปรวน (ANOVA) และวิธีกระบวนการค่าอำนาจจำแนก (Item Discrimination Procedures) ซึ่งวิธีที่กล่าวมาทั้งหมดเป็นการวิเคราะห์ความลำเอียงโดยอาศัยข้อตกลงและข้อสรุปของทฤษฎีการวัดผลแบบดั้งเดิม

วิธีการแปลงค่าความยาก (Transformed Item Difficulty)

ผู้เสนอการวิเคราะห์ความลำเอียงวิธีนี้ คือ แองกอฟฟ์ (Angoff. 1972) โดยนิยามข้อสอบที่มีความลำเอียงว่าเป็นข้อสอบที่เบี่ยงเบนไปจากเส้นแกนหลัก วิธีนี้ใช้ค่าความยาก (p) ของข้อสอบแต่ละข้อ จากกลุ่มผู้สอบแต่ละกลุ่ม (ดังนั้นค่าความยากของข้อสอบจะมี 2 ค่า) มาแปลงให้เป็นคะแนนมาตรฐานซี (z-values) แล้วแปลงคะแนนซีเป็นค่าความยากมาตรฐานหรือค่าเดลต้า (Δ) โดยใช้ สมการ $\Delta = 13 + 4z$ นำค่าเดลต้า (Delta) แต่ละค่ามาลงจุดคู่อันดับบนกราฟไปวาริเอท

จุดต่าง ๆ บนกราฟไบนารีเอทซึ่งแทนค่าเดลต้าจะอยู่ในรูปวงรีแล้วทำการคำนวณหาระยะทางที่จุดต่าง ๆ เหล่านี้ห่างจากเส้นแกนหลัก (Major Axis) ข้อสอบข้อใดที่ห่างจากเส้นแกนหลักมากกว่าหรือเท่ากับ $+ 3 S_d$ หรือข้อที่มีระยะห่างตั้งฉากมากกว่าหรือเท่ากับ 0.75 หรือ น้อยกว่าหรือเท่ากับ $- 0.75$ ถือว่าข้อสอบข้อนั้นมีความลำเอียง ดังภาพประกอบต่อไปนี้

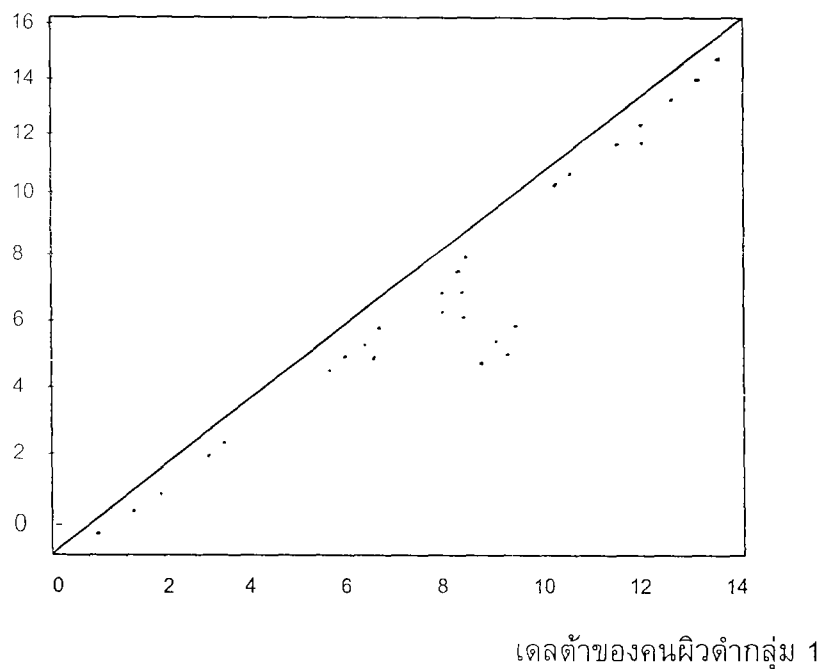
เดลต้าของคนผิวดำกลุ่ม 2



เดลต้าของคนผิวดำกลุ่ม 1

ภาพประกอบ 1 แสดงกราฟไบนารีเอทของค่าเดลต้าของคนผิวดำกลุ่ม 1 และคนผิวดำของกลุ่ม 2 จากแบบทดสอบวัดความถนัดทางการเรียนด้านภาษา (Preliminary Scholastic Aptitude) (Angoff and Ford, 1973 : 98)

เดลด้าของคนผิวขาวกลุ่ม 5



ภาพประกอบ 2 แสดงกราฟไบวาริเอทของค่าเดลด้าของคนผิวขาวกลุ่ม 5 และคนผิวดำของกลุ่ม 1 จากแบบทดสอบวัดความถนัดทางการเรียนด้านภาษา (Preliminary Scholastic Aptitude) (Angoff and Ford. 1973 : 100)

เมื่อลงจุดคู่อันดับค่าเดลด้าของข้อสอบแต่ละข้อของผู้เข้าสอบแต่ละกลุ่มข้อสอบที่ไม่มี ความลำเอียงจะได้กราฟไบวาริเอท ดังภาพประกอบ 1 ซึ่งภาพประกอบ 1 เป็นกราฟไบวาริเอท ของค่าเดลด้าจากกลุ่มตัวอย่างคนผิวดำ 2 กลุ่ม โดยใช้แบบทดสอบวัดความถนัดทางการเรียน (Berk.1982:99 ; citing Angoff and Ford.1973) จะเห็นว่าจุดต่าง ๆ ของค่าเดลด้าของกลุ่มผู้ เข้าสอบ 2 กลุ่ม จะกระจายอยู่ในแนวเส้นแกนหลัก คือ ไม่มีการเบี่ยงเบนไปจากเส้นแกนหลัก ส่วนภาพประกอบ 2 เป็นกราฟไบวาริเอทของค่าเดลด้า ที่ได้จากการทดสอบกลุ่มผู้สอบคนผิวดำ และคนผิวขาวโดยใช้แบบทดสอบชุดเดียวกันจะเห็นว่าจุดต่าง ๆ ของค่าเดลด้าของผู้เข้าสอบ 2 กลุ่มนี้กระจายห่างจากเส้นแกนหลัก คือมีการเปลี่ยนแปลงไปจากเส้นแกนหลัก ข้อสอบเหล่านี้ยาก ต่อกลุ่มผู้สอบผิวดำ ดังนั้นแบบทดสอบชุดนี้มีปฏิสัมพันธ์ระหว่างข้อสอบกับเชื้อชาติ (Item Race Interaction) และเรียกได้ว่าเป็นข้อสอบที่มีความลำเอียงต่อเชื้อชาติ

วิธีไคสแควร์ (Chi – Square : χ^2)

วิธีไคสแควร์มีหลายวิธี แต่ที่นิยมมีอยู่ 2 วิธี คือ

1. วิธีไคสแควร์ของชีนีแมน (Scheneman. 1979) เป็นการใช้ค่าของไคสแควร์ที่ใช้กันโดยทั่วไปเป็นดัชนีความลำเอียงของข้อสอบ โดยการแบ่งคะแนนรวมของผู้สอบในแต่ละกลุ่มออกเป็นช่วง (ประมาณ 3 ถึง 5 ช่วง) ก็สามารถคำนวณค่าดัชนีความลำเอียงได้ ในวิธีนี้ชีนีแมนได้กำหนดให้ C2 เป็นค่าดัชนีความลำเอียงที่ขึ้นอยู่กับเฉพาะความถี่ของการตอบข้อสอบได้ถูกต้องเท่านั้น โดยการคำนวณค่า C2 จากสูตร

$$C2 = \sum_{j=1}^J \frac{(E_{j1} - O_{j1})^2}{E_{j1}} + \sum_{j=1}^J \frac{(E_{j2} - O_{j2})^2}{E_{j2}}$$

เมื่อ C2 แทน ค่าไคสแควร์แบบชีนีแมน

O_{jk} แทน ความถี่ที่สังเกตได้ของการตอบข้อสอบได้ถูกต้องในช่วงคะแนนที่ j ของกลุ่มที่ k

$E_{jk} = P_j N_{jk}$ แทน ความถี่ที่คาดหวังไว้ในการตอบข้อสอบได้ถูกต้องในช่วงคะแนนที่ j ของกลุ่มที่ k

P_j แทน สัดส่วนของผู้สอบที่ตอบข้อสอบได้ถูกต้องในช่วงคะแนนที่ j

N_{jk} แทน จำนวนของผู้สอบในช่วงคะแนนที่ j ของกลุ่มที่ k

2. วิธีไคสแควร์เต็มรูป วิธีนี้ ชีนีแมน (Scheneman.1981) ได้สร้างขึ้นมาเพื่อแก้ไขข้อบกพร่องของค่าดัชนี C2 ที่ได้จากวิธีแรกในด้านการแจกแจงที่ไม่เป็นการแจกแจงอย่างสม่ำเสมอ และการทดสอบนัยสำคัญที่ไม่เหมาะสม (Camilli and Shepard. 1994) ชีนีแมนได้แก้ปัญหานี้โดยการนำเอาสัดส่วนของการตอบผิดในแต่ละช่วงรวมเข้าไปด้วย ทำให้สูตรเป็นดังนี้

$$\chi^2 full = \sum_{j=1}^J \frac{(E_{j1} - O_{j1})^2}{E_{j1} - (1 - P_{j0})} + \sum_{j=1}^J \frac{(E_{j2} - O_{j2})^2}{E_{j2} - (1 - P_{j0})}$$

โดยทั่วไปวิธีไคสแควร์มีข้อดีหลายข้อ เช่น ใช้กับกลุ่มตัวอย่างที่มีขนาดเล็กมีความสับสนเล็กน้อย คำนวณได้ง่าย และที่สำคัญคือ ใช้การจับคู่ระหว่างกลุ่มต่าง ๆ กับระดับความสามารถ แต่ผลจากการวิจัย พบว่า ค่าดัชนี C2 และไคสแควร์เต็มรูป มีแนวโน้มให้ผลที่เหมือนกันมาก (Scheneman. 1986) แต่อาจจะแตกต่างกันที่จุดทศนิยม แต่เมื่อกลุ่มตัวอย่างมีขนาดใหญ่วิธีไคสแควร์เต็มรูปมีแนวโน้มที่ให้ค่าบางอย่างที่ดีกว่า และกลุ่มตัวอย่างต้องมีจำนวนไม่น้อย

กว่า 300 คน ส่วนดัชนี C2 ใช้กับกลุ่มตัวอย่างตั้งแต่ 100 คน ขึ้นไปโดยประมาณ วิธีไคสแควร์เต็มรูปใช้ได้ดีในการวิเคราะห์ความลำเอียงของข้อสอบข้อที่ยาก ๆ ส่วนดัชนี C2 เหมาะสำหรับการวิเคราะห์ความลำเอียงของข้อสอบที่ง่าย จะใช้ได้ดี แต่ทั้งสองวิธีได้ถูกวิจารณ์ว่าแต่ละกลุ่มที่นำมาเปรียบเทียบอาจจะมีความสามารถไม่เท่ากัน ภายในความกว้างของแต่ละช่วงคะแนน

วิธีของแมนเทล-แฮนส์เซล (Mantel – Haenszel : MH)

เทคนิคแมนเทล-แฮนส์เซล (Millsap and Everson.1993:302-304) เป็นวิธีการที่แมนเทลและแฮนส์เซล (Mantel and Haenszel) ได้เสนอขึ้นใช้ตั้งแต่ปี 1959 แต่ฮอลแลนด์และไทเยอร์ (Holland and Thayer) เพิ่งนำเสนอมาใช้เพื่อตรวจสอบความลำเอียงของข้อสอบในปี 1988 และในปีนั้นก็ได้มีการตรวจสอบความลำเอียงของข้อสอบด้วยเทคนิค MH โดย ดูแรน , แฮมเบิลตัน, โรเจอร์, ฮอลแลนด์ และไรท์ เทคนิค MH เป็นเทคนิคที่มีความคล้ายคลึงกับวิธีไคสแควร์ที่เสนอโดยชูนเนแมน (Scheuneman.1979) มาร์สคูโลและสลาughter (Marascuilo and Slaughter. 1981) และ เมลเลนเบิร์ก (Mellenberg.1982) โดยการเปรียบเทียบผลการตอบข้อสอบของกลุ่มผู้สอบ 2 กลุ่ม หรือฮอลแลนด์ เรียกว่า กลุ่มอ้างอิงและกลุ่มเปรียบเทียบ (Reference and Focal Group) โดยที่กลุ่มแรกจะใช้อ้างอิงถึงกลุ่มที่คาดว่าจะได้ประโยชน์จากข้อสอบและกลุ่มหลังเป็นกลุ่มเสียประโยชน์จากข้อสอบในกรณีที่ข้อสอบมีความลำเอียงโดยมีการตรวจสอบทุก ๆ ระดับคะแนนรวมจากการสอบ ข้อสอบใดที่ผู้สอบทั้ง 2 กลุ่ม ทำได้เท่า ๆ กัน จะเป็นข้อสอบที่ถือว่าไม่มีความลำเอียงต่อกลุ่มใดกลุ่มหนึ่ง วิธีการตรวจสอบ มีดังนี้

ถ้าให้ N_{Rj} และ N_{Fj} แทนจำนวนผู้สอบในกลุ่มอ้างอิง และกลุ่มเปรียบเทียบที่มีคะแนนรวมในช่วงคะแนน j และให้ N_j ซึ่ง $= N_{Rj} + N_{Fj}$ เป็นจำนวนผู้สอบรวมทั้ง 2 กลุ่ม ที่ได้คะแนนรวมจากการสอบข้อสอบข้อที่ i ในช่วงคะแนน j แล้วนำมาเขียนในรูปตาราง 2 ทาง แสดงผลการตอบถูก (1) และผิด (0) ดังตาราง 1 (Narayanan and Swaminathan.1994:315-317)

ตาราง 1 แสดงข้อมูลผลการสอบของกลุ่มตัวอย่างกลุ่ม R และ F ที่มีคะแนนรวมอยู่ในช่วงคะแนน j ข้อที่ i

กลุ่ม	คะแนนที่ได้จากข้อสอบที่ต้องการตรวจสอบ		
	1	0	รวม
กลุ่มอ้างอิง (R)	A_j	B_j	N_{Rj}
กลุ่มเปรียบเทียบ (F)	C_j	D_j	N_{Fj}
รวม	N_{1j}	N_{0j}	N_j

เมื่อ A_j, B_j, C_j และ D_j เป็นความถี่ที่สังเกตได้ของการตอบถูก (1) และผิด (0) ในช่วงคะแนน j

จากตาราง จะได้ว่า ในช่วงคะแนน j A_j และ B_j คือ จำนวนผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบข้อที่ i ถูกต้อง และไม่ถูกต้องตามลำดับ ขณะที่ C_j และ D_j คือ จำนวนผู้สอบกลุ่มเปรียบเทียบที่ตอบข้อสอบข้อที่ i ถูกต้อง และไม่ถูกต้องตามลำดับ เช่นกัน และคำนวณค่าความน่าจะเป็นในรูปของค่าสัดส่วนการตอบข้อสอบถูก-ผิด ระหว่างกลุ่ม โดยกำหนดให้เป็นค่า α ดังนี้

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

ส่วนสถิติในการทดสอบสมมติฐาน คือ

$$H_o : \left[\pi_{Rj} / (1 - \pi_{Rj}) \right] = \left[\pi_{Fj} / (1 - \pi_{Fj}) \right]$$

$$H_1 : \left[\pi_{Rj} / (1 - \pi_{Rj}) \right] = \alpha \left[\pi_{Fj} / (1 - \pi_{Fj}) \right]$$

เมื่อ $j = 1, 2, 3 \dots K$, $\alpha \neq 1$

และ ค่า π_{Rj} เป็นค่าความจะเป็นที่กลุ่มอ้างอิงตอบข้อสอบข้อนั้นถูกในระดับช่วงคะแนน j

ค่า π_{Fj} เป็นค่าความจะเป็นที่กลุ่มเปรียบเทียบตอบข้อสอบข้อนั้นถูกในระดับช่วง

คะแนน j

ซึ่งสมมติฐานทั้ง 2 นี้ ใช้ในการตรวจสอบความลำเอียงของข้อสอบแบบยูนิฟอร์ม (Uniform DIF) แมนเทลและแฮนส์เซล (Mantel and Haenszel, 1959) ได้เสนอค่าสถิติไคสแควร์เพื่อทดสอบค่าที่ได้ว่ามีความแตกต่างจาก 1.00 อย่างมีนัยสำคัญหรือไม่ที่ระดับชั้นแห่งความเป็นอิสระ = 1 ดังนี้

$$\chi^2_{MH} = \frac{\left[\left| \sum A_j - \sum E(A_j) \right| - 0.5 \right]^2}{\sum \text{var}(A_j)}$$

นอกจากนั้น Education Testing Service's ได้เสนอให้แปลงค่า α_{MH} ให้เป็นคะแนนมาตรฐานในรูปเดลต้าที่มีค่าเฉลี่ยเป็น 0 ดูแรนและฮอลแลนด์ (Doran and Hooland, 1993) เรียกค่าที่แปลงนี้ว่า MH D.DIF ที่มีสมการดังนี้ (Millsap and Everson, 1993 : 303)

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

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

วิธีแมนเทิล-แฮนส์เซลที่ประยุกต์ใช้กับข้อมูลแบบเรียงลำดับ (Zwick , Donoghue and Grima. 1993 : 238 – 239)

แมนเทิล (1963) ได้เสนอชั้นแห่งความเป็นอิสระในการทดสอบไว้ที่ 1 ในเงื่อนไขของลักษณะของข้อมูลแบบเรียงลำดับ การประยุกต์วิธีการวิเคราะห์ความลำเอียงในด้านเนื้อหาเกี่ยวกับตัวดัชนีของลักษณะการตอบสนองและเปรียบเทียบข้อสอบจากจำนวนกลุ่มเปรียบเทียบและกลุ่มอ้างอิง โดยการจับคู่แล้วทำการเปรียบเทียบความสามารถ ขาง,มาสซีโอ และรอสสอส (Chang,Mazzeo and Roussos.1993) เวลช์และฮูเวอร์ (Welch and Hoover.1993) ได้ทำการศึกษาค่าสถิติที่ใช้ในการวิเคราะห์ความลำเอียงในข้อมูลที่จำลองขึ้นมา และ ฮอลแลนด์ และไทเยอร์ (Holland.1991) ได้ทำการศึกษาการจัดข้อมูลลงในตารางจรณแบบ $2 \times T \times K$ เมื่อ T เป็นจำนวนลักษณะการตอบสนอง และ K เป็นจำนวนของระดับการจับคู่ของตัวแปรในแต่ละระดับของ k ข้อสอบสามารถตอบสนองเท่ากับตารางแบบ $2 \times T$ ซึ่งจะแสดงให้เห็นในตาราง 2

ตาราง 2 ข้อมูลในระดับความสามารถที่ k จากตัวแปรต่างๆ

กลุ่ม	คะแนนการสอบ					รวม
	y1	y2	y3	...	yt	
อ้างอิง	N_{R1k}	N_{R2k}	N_{R3k}	...	N_{Rtk}	N_{R+k}
เปรียบเทียบ	N_{F1k}	N_{F2k}	N_{F3k}	...	N_{Ftk}	N_{F+k}
รวม	N_{+1k}	N_{+2k}	N_{+3k}		N_{+tk}	N_{++k}

ค่าของ $y_1, y_2, \dots, y_t$ เป็นสัญลักษณ์ของคะแนน T นั้น สามารถดูได้จากข้อสอบ ภายในตารางจะมีค่า N_{Rtk} และ N_{Ftk} ซึ่งเป็นจำนวนผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตามลำดับ ระดับความสามารถที่ k ของตัวแปรที่จับคู่และคะแนนการสอบที่ได้คือ y_t เครื่องหมาย + คือ ค่าดัชนีผลรวมการปฏิบัติ จากตาราง N_{F+k} คือ จำนวนรวมของกลุ่มเปรียบเทียบในระดับความสามารถที่ k

ค่าสถิติที่เสนอโดยแมนเทิล คือ

$$\chi^2_{MH} = \frac{\left[\sum_k Fk - \sum_k E(Fk) \right]^2}{\sum_k Var(Fk)}$$

เมื่อ F_k เป็นผลรวมของกลุ่มเปรียบเทียบในระดับความสามารถที่ k ของการจับคู่กับตัวแปร ซึ่งสามารถหาได้จาก

$$F_k = \sum_i y_i n_{Fik}$$

ความคาดหวังของ F_k ภายใต้สมมติฐานศูนย์ (H_0) คือ

$$E_H(F_k) = \frac{n_{F+k}}{n_{++k}} \sum_i y_i n_{+ik}$$

และความแปรปรวนของ F_k ภายใต้สมมติฐาน H_0 คือ

$$Var_H(F_k) = \frac{n_{R+k} n_{F+k}}{n_{++k}^2 (n_{++k} - 1)} \left\{ \left(n_{++k} \sum_i y_i^2 n_{+ik} \right) - \left(\sum_i y_i n_{+ik} \right)^2 \right\}$$

ภายใต้สมมติฐาน H_0 สถิติไคสแควร์กับจำนวนแห่งชั้นความเป็นอิสระเท่ากับ 1 ในส่วนของแบบทดสอบที่มีการตรวจให้คะแนนแบบ 0-1 สถิติที่ใช้เหมือนกันทุกอย่าง คือใช้วิธีแมนเทิล-แอนส์เซล (1959) เป็นสถิติที่ไม่มีความสัมพันธ์ที่ไม่ต่อเนื่อง การประยุกต์การหาความลำเอียง การเสนอแนะการปฏิเสธ H_0 ดูจากจำนวนกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ กลุ่มใดมีความเป็นอันหนึ่งอันเดียวกันในแนวโน้มความแตกต่างในค่าเฉลี่ยของการสอบของข้อสอบที่จะศึกษา

วิธีการวิเคราะห์ความแปรปรวน (Analysis Of Variance)

การวิเคราะห์ความลำเอียงโดยวิธีนี้ เป็นการทดสอบนัยสำคัญทางสถิติของผลปฏิสัมพันธ์ (Interaction Effect) ทั้งหมดระหว่างข้อสอบกับกลุ่มผู้สอบ โดยมีข้อตกลงเบื้องต้นว่า แบบทดสอบจะต้องมีความเอกพันธ์ในเนื้อหาความแปรปรวนของข้อสอบต้องเท่ากันและ ความ

แปรปรวนร่วมของข้อสอบจะต้องเท่ากันด้วย สำหรับการทดสอบปฏิสัมพันธ์ระหว่างข้อสอบกับกลุ่มผู้สอบมีรูปแบบการวิเคราะห์ ดังนี้ (Laksana.1979:12)

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

เมื่อ μ แทน ค่าเฉลี่ย

α_i แทน อิทธิพลจากข้อสอบที่ i ($i = 1, 2, 3, \dots, I$)

β_j แทน อิทธิพลจากกลุ่มผู้สอบ กลุ่มที่ j ($j = 1, 2, 3, \dots, J$)

$(\alpha\beta)_{ij}$ แทน ปฏิสัมพันธ์ระหว่างข้อสอบกับกลุ่มผู้สอบ

ε_{ij} แทน ความคลาดเคลื่อนแบบสุ่มภายในเซลล์ i, j

ขั้นตอนในการวิเคราะห์ความลำเอียง

1. จัดข้อสอบและกลุ่มผู้สอบ ลงตามแนวสดมภ์ (Column) และแนวนอน (Row)
2. คำนวณค่าสัดส่วนในการตอบถูกในแต่ละข้อระหว่างกลุ่มผู้สอบ
3. ทดสอบปฏิสัมพันธ์ทั้งหมดระหว่างข้อสอบกับกลุ่มผู้สอบหากพบว่ามีปฏิสัมพันธ์อย่างมีนัยสำคัญทางสถิติ แสดงว่าข้อสอบข้อนั้นมีความลำเอียง
4. หากผลการทดสอบในข้อ 3 มีปฏิสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติแล้วทำการทดสอบรายข้อว่า มีข้อใดบ้างที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ และ ถือว่าข้อสอบข้อนั้นเป็นข้อสอบที่มีความลำเอียง

จากขั้นตอนการวิเคราะห์โดยวิธีนี้ จะเห็นว่า ข้อสอบที่ถูกคัดเลือกกว่าเป็นข้อสอบลำเอียงนั้น จะเป็นข้อสอบที่ให้ผลการวัดแตกต่างกันใน 2 กลุ่ม คือ อาจจะเป็นข้อสอบที่ง่ายสำหรับกลุ่มหนึ่ง แต่ยากสำหรับกลุ่มหนึ่ง หรือข้อสอบข้อนั้นอาจไม่ได้วัดความสามารถเดียวกันในผู้สอบแต่ละกลุ่ม

วิธีกระบวนการค่าอำนาจจำแนกของข้อสอบ (Item Discrimination Procedures)

เป็นการนำค่าอำนาจจำแนกของข้อสอบข้อเดียวกันที่หามาจากกลุ่มต่าง ๆ ด้วยวิธีสหสัมพันธ์พอยต์ไบซีเรียลหรือไบซีเรียลมาทดสอบความแตกต่างของดัชนีอำนาจจำแนกระหว่างกลุ่มทางจริยธรรมหรือเพศ หากพบว่ามีค่าแตกต่างอย่างมีนัยสำคัญทางสถิติ แสดงว่าข้อสอบข้อนั้นมีความลำเอียง ฮูเวอร์และโคลเลน (Hoover and Kolen.1984) ได้ใช้สัมพัทธ์ไบซีเรียลหาค่าอำนาจจำแนกของข้อสอบในแบบทดสอบ โดยแบ่งกลุ่มสูงและกลุ่มต่ำ แล้วนำค่าอำนาจจำแนกของแต่ละกลุ่มย่อยมาทดสอบหาความแตกต่างสัมบูรณ์ วิธีการนี้มีข้อได้เปรียบที่เหมือนกับวิธีแปลงค่าความยากคือ วิธีการที่ง่าย ไม่แพง ง่ายต่อการอธิบาย และ กลุ่มตัวอย่างไม่จำเป็นต้องมีขนาดใหญ่

วิธี HW 3 (HW 3 method)

วิธี HW 3 เป็นวิธีที่เสนอโดย Welch and Hoover เมื่อปี 1993 (Welch and Miller.1995:166–167 ; Potenza and Dorans.1995:27 ; citing Welch and Hoover.1993) เป็นวิธีการทดสอบความแตกต่างกันระหว่างค่าเฉลี่ยของกลุ่มผู้สอบ 2 กลุ่มที่เป็นอิสระจากกัน (k-indepent test) โดย $k = 1, 2, 3, \dots, k$ อ้างอิงถึงแต่ละระดับของการวัดที่มีนัยสำคัญทางสถิติในแต่ละระดับของความสามารถ การคำนวณ t - test แบบแยกกลุ่ม เป็นคำนวณเพื่อทดสอบสมมติฐาน ค่าเฉลี่ยที่ได้จาก 2 กลุ่มย่อย ในข้อที่ i จะคำนวณ มีสูตร ดังนี้

$$t_k = \frac{\bar{X}_{Fk} - \bar{X}_{Rk}}{\sqrt{\frac{S_{Fk}^2 n_{Fk} + S_{Rk}^2 n_{Rk}}{m_k} \left\{ \frac{1}{n_{Fk}} + \frac{1}{n_{Rk}} \right\}}}$$

เมื่อ	n_{Fk} แทน	จำนวนของกลุ่มเปรียบเทียบในระดับการตอบที่ k
	n_{Bk} แทน	จำนวนของกลุ่มอ้างอิงในระดับการตอบที่ k
	$\bar{X}_{Fk}$ แทน	ค่าเฉลี่ยของกลุ่มเปรียบเทียบในระดับการตอบที่ k
	$\bar{X}_{Bk}$ แทน	ค่าเฉลี่ยของกลุ่มอ้างอิงในระดับการตอบที่ k
	S_{Fk}^2 แทน	ค่าความแปรปรวนของกลุ่มเปรียบเทียบในระดับการตอบที่ k
	S_{Bk}^2 แทน	ค่าความแปรปรวนของกลุ่มอ้างอิงในระดับการตอบที่ k
	m_k แทน	ชั้นแห่งความเป็นอิสระ เท่ากับ $n_{Fk} + n_{Bk} - 2$

เนื่องจากสูตร HW3 ที่เวลช์และฮูเวอร์ (Welch and Hoover.1993) เสนอไว้นั้น เป็นการทดสอบหาความลำเอียงกับกลุ่มผู้สอบที่มีขนาดเล็ก ฮีดกีส ,ไซแมนสกายและวูดเวิร์ธ (Hedges,Shymansky and Woodworth.1989) ได้เสนอสถิติที่ใช้ปรับแก้ค่า t -test เกี่ยวกับจะใช้กระบวนการให้ค่านำหนักเพื่อความแม่นยำของขนาดกลุ่มเปรียบเทียบและกลุ่มอ้างอิงที่มีขนาดแตกต่างกัน และค่าของน้ำหนักของความแปรปรวนของกลุ่มตัวอย่าง

ค่าความคลาดเคลื่อนมาตรฐานของแต่ละกลุ่มจากน้ำหนักของการเปรียบเทียบ การเพิ่มขึ้นของขนาดกลุ่มตัวอย่างในการทดสอบ และเส้นสารสนเทศของการคำนวณความคลาดเคลื่อนมาตรฐานของการเปรียบเทียบขนาดกลุ่ม (Hedges et al.,1989)

ค่าสถิติ t ที่ปรับแก้ความลำเอียงที่มีจากขนาดของกลุ่มตัวอย่าง คือ

$$g_k = \sqrt{\frac{n_{Fk} + n_{Rk}}{n_{Fk} n_{Rk}}} \times t_k$$

อย่างไรก็ตาม ค่า g_k เป็นประมาณค่าของประชากรที่ปราศจากความลำเอียง เมื่อขนาดของกลุ่มตัวอย่างไม่ใหญ่ (Hedges et al., 1989) และความสัมพันธ์ขององค์ประกอบ , ค่า J_k เป็นการใช้ในการประมาณค่าของขนาดกลุ่มที่ปราศจากความลำเอียง เมื่อ $d_k = J_k \times g_k$ โดย

$$J_k = 1 - \frac{3}{4m_k - 1}$$

ดังนั้น ขนาดของกลุ่มที่ไม่ลำเอียงในการปรับแก้ค่า t - test คำนวณได้จาก

$$d_k = J_k \times \sqrt{\frac{n_{Fk} + n_{Rk}}{n_{Fk} n_{Rk}}} \times t_k$$

โดยมีข้อตกลงว่า $d_1, \dots, d_k$ เป็นค่าที่ปรับแก้ค่า t - test จากขนาดกลุ่มผู้สอบ กับค่าความคลาดเคลื่อนมาตรฐานของ $S_1, \dots, S_k$ เมื่อช่วงน้ำหนักของระดับการตอบที่ k ที่เป็นอิสระจากขนาดของกลุ่มซึ่งมาจาก

$$D = \frac{\left(\frac{d_1}{S_1^2} + \frac{d_2}{S_2^2} + \dots + \frac{d_k}{S_k^2} \right)}{\left(\frac{1}{S_1^2} + \frac{1}{S_2^2} + \dots + \frac{1}{S_k^2} \right)}$$

เมื่อ

$$S_k^2 = J_k^2 \frac{m_k}{m_k - 2} \left\{ \frac{n_{Fk} + n_{Rk}}{n_{Fk} n_{Rk}} \right\} + d_k^2 \left\{ \frac{J_k^2 m_k}{m_k - 2} - 1 \right\}$$

ค่าความคลาดเคลื่อนมาตรฐานของช่วงพิสัยน้ำหนัก คือ

$$S_d = \frac{1}{\sqrt{\left(\frac{1}{S_1^2} + \frac{1}{S_2^2} + \dots + \frac{1}{S_K^2} \right)}}$$

วิธีการทดสอบสมมติฐานว่ามีหรือไม่มีความลำเอียงระหว่างกลุ่มอ้างอิงกับกลุ่มเปรียบเทียบในข้อสอบแต่ละข้อ

$$Z_{HW3} = \frac{D}{S_d}$$

ซึ่งค่า Z เป็นค่าที่มีการแจกแจงอยู่ในช่วง $N(0,1)$ ถ้า Z มีความแตกต่างจากสมมติฐาน H_0 อย่างมีนัยสำคัญทางสถิติ แสดงว่าข้อสอบข้อนั้นมีความลำเอียง

วิธีการวิเคราะห์ความลำเอียงของข้อสอบ โดยใช้ทฤษฎีการวัดผลแบบดั้งเดิม เป็นวิธีที่ง่ายต่อการคำนวณและการแปลผล ไม่ซับซ้อน ใช้ได้ดีในทางปฏิบัติ ไม่ต้องการข้อมูลจำนวนมาก การวิเคราะห์ข้อสอบที่มีความลำเอียง อาศัยค่าสถิติที่ใช้ทดสอบนัยสำคัญทางสถิติ เป็นค่าดัชนีความลำเอียงของข้อสอบ แต่วิธีเหล่านี้มีข้อจำกัดในทางทฤษฎี เนื่องจากผลการวิเคราะห์จะมีค่าไม่คงที่ ขึ้นอยู่กับความสามารถของผู้สอบแต่ละกลุ่ม

วิธีวิเคราะห์ความลำเอียงโดยยึดทฤษฎีการตอบข้อคำถาม (Item Response Theory)

ทฤษฎีการตอบข้อคำถามเป็นทฤษฎีที่อธิบายถึงถึงความสัมพันธ์ระหว่างลักษณะหรือความสามารถที่มีอยู่ในตัวบุคคล กับพฤติกรรมการตอบสนองต่อข้อสอบของบุคคลนั้น โดยทฤษฎีนี้มีความเชื่อว่า พฤติกรรมการตอบสนองต่อข้อสอบของบุคคล จะถูกกำหนดโดยลักษณะหรือความสามารถที่มีอยู่ในตัวบุคคล (Lord and Novick.1968:358) ซึ่งไม่สามารถจะสังเกตได้ ทฤษฎีนี้พยายามอธิบายความสัมพันธ์ระหว่างลักษณะภายในตัวบุคคลกับพฤติกรรมที่บุคคลตอบสนองต่อข้อสอบ โดยใช้แบบจำลองทางคณิตศาสตร์ ดังนี้

ให้ P แทน ผลการสอบ (Performance)
 θ แทน ความสามารถ (Ability or Trait)
 f แทน ฟังก์ชัน (Function)

แล้วจะได้ว่า

$$P = f(\theta)$$

ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อคำถาม

1. ข้อตกลงเกี่ยวกับความเป็นมิติเดียว (Dimensionality)

ทฤษฎีมีข้อตกลงว่า ความสามารถของมนุษย์มีอยู่ทั้งหมด k อย่าง แต่จะมีความสามารถเดียว (Only one trait) อธิบายการตอบข้อสอบของแบบทดสอบได้ หรือคะแนนจากผู้สอบสามารถอธิบายได้เพียงความสามารถเดียว ดังนั้นแบบทดสอบที่ใช้จึงควรเป็นแบบทดสอบกับกลุ่มตัวอย่าง r กลุ่ม ที่มีความสามารถเท่ากับ θ นำคะแนนการสอบทั้ง r กลุ่มมาแจกแจงความถี่แล้ว คะแนนมีลักษณะการแจกแจงเป็นอย่างเดียว การสร้างแบบทดสอบให้เป็นความเป็นมิติเดียวทำได้โดยการสร้างคำถามที่วัดความสามารถเดียว แล้วทำการวิเคราะห์หองค์ประกอบเพื่อรวมคำถามที่มีลักษณะองค์ประกอบเดียวกันไว้ด้วยกัน

2. ข้อตกลงเกี่ยวกับความเป็นอิสระของข้อสอบ (Local Independence)

ข้อตกลงนี้ กล่าวว่า โอกาสที่ผู้สอบตอบคำถามข้อหนึ่งถูกนั้นไม่เกี่ยวกับการตอบคำถามข้ออื่น ๆ ในแบบทดสอบนั้น

3. ข้อตกลงเกี่ยวกับลักษณะโค้งจากผลการสอบของข้อสอบ (Item Characteristic Curve : ICC)

ข้อตกลงนี้กล่าวว่าโอกาสที่ผู้สอบจะตอบข้อสอบข้อใดข้อหนึ่งได้ถูกต้องหรือไม่ ขึ้นอยู่กับโค้งลักษณะเฉพาะของข้อสอบแต่ละแบบที่ใช้ ไม่ใช่ขึ้นอยู่กับลักษณะการแจกแจงความสามารถของประชากรนั้น นั่นคือ รูปร่างของโค้งลักษณะเฉพาะของข้อสอบไม่ได้ขึ้นอยู่กับลักษณะการกระจายของความสามารถของกลุ่มผู้สอบหรือแปรเปลี่ยนตามกลุ่มผู้สอบ

จากข้อตกลงเบื้องต้น 3 ประการนี้ สามารถนำมาใช้เป็นแนวทางของการวิเคราะห์ความลำเอียงของข้อสอบได้หลายวิธี แต่นิยมใช้กัน คือ วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ วิธีโค้งลักษณะข้อสอบ 2 พารามิเตอร์ และวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์

วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ หรือ การวิเคราะห์ความลำเอียงตามแบบจำลองราสช์ (Rasch Model)

วิธีดังกล่าวนี้ กำหนดให้ข้อสอบทุกข้อมีค่าอำนาจจำแนก (a) เป็น 1.00 ทุกข้อ และค่าการเดา (c) มีค่าเป็น 0 ทุกข้อ แล้วทำการคำนวณค่าความยาก (b) ของข้อสอบจากโค้งลักษณะข้อสอบแต่ละกลุ่ม โดยมีรูปแบบทางคณิตศาสตร์ ดังนี้ (Hambleton, Swaminathan and Roger, 1991:12)

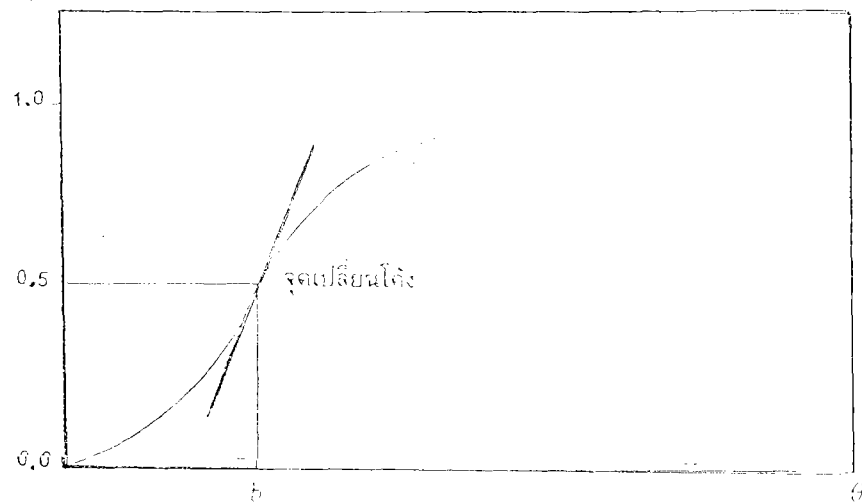
$$P_i(\theta) = \frac{e^{(\theta - b_i)}}{1 + e^{(\theta - b_i)}}$$

เมื่อ $P_i(\theta)$ แทน โอกาสของนักเรียนคนหนึ่งที่มีระดับความสามารถ (θ) ตอบข้อสอบ
ข้อ i ถูก

θ แทน ความสามารถของผู้สอบ

b_i แทน ค่าความยากของข้อสอบที่ i .

$P_i(\theta)$



ภาพประกอบ 3 แสดงความหมายของค่าพารามิเตอร์ข้อสอบตามโมเดลโลจิสติกที่มีพารามิเตอร์
1 ตัว

จากนั้น นำค่าความยากของแต่ละกลุ่มมาทดสอบนัยสำคัญ โดยใช้ค่าสถิติ t

จากสูตร

$$t = \frac{d_1 - d_2}{\sqrt{SE_1 + SE_2}}$$

เมื่อ d_1 , d_2 แทน ค่าความยากของข้อสอบที่ได้จากการวิเคราะห์ในกลุ่มที่ 1 และ
กลุ่มที่ 2

SE_1 , SE_2 แทน ค่าความคลาดเคลื่อนมาตรฐานของข้อสอบที่ได้จากการวิเคราะห์

ค่า t ที่ได้ เป็นดัชนีความลำเอียงของข้อสอบ ซึ่ง ดร.บ. (Berk. 1982 : 138 ;
citing Draba.1977) เสนอให้ใช้ค่า $t \geq 2.4$ เป็นค่าในการกำหนดว่าข้อสอบมีความลำเอียง

วิธีโค้งลักษณะ 2 พารามิเตอร์

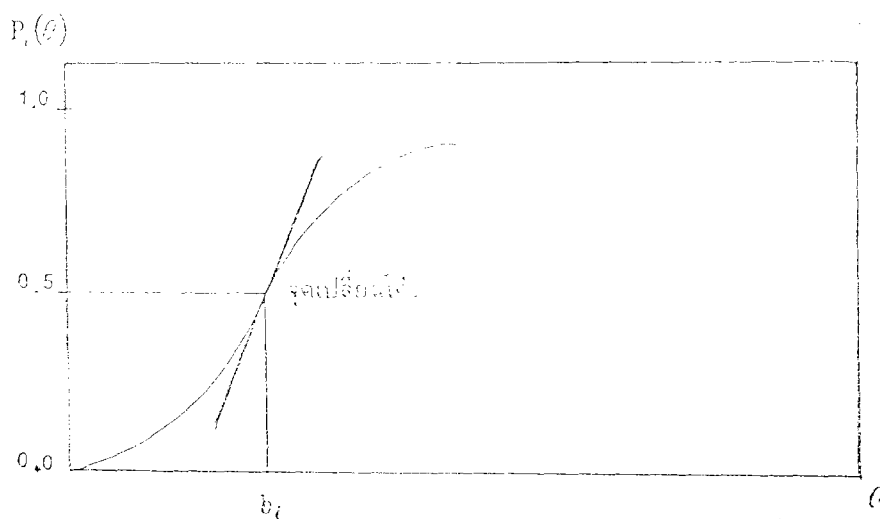
วิธีนี้จะกำหนดให้ค่าการเดา (c) ของข้อสอบทุกข้อเป็น 0 ส่วนค่าอำนาจจำแนก และค่าความยากจะแปรผันไปตามกลุ่มย่อยแต่ละกลุ่มซึ่งมีรูปแบบทางคณิตศาสตร์ ดังนี้ (Hambleton, Ronald and Swaminathan, 1985:36)

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

เมื่อ b_i แทน ค่าความยากที่ระดับความสามารถที่แท้จริงที่จุดโค้งชันที่สุด

a_i แทน ค่าอำนาจจำแนก ซึ่งมีค่าเป็นสัดส่วนโดยตรงกับค่าความชันของโค้ง ณ จุดเปลี่ยนโค้ง

D แทน ค่า Scaling Factor ซึ่งมีค่า 1.7140



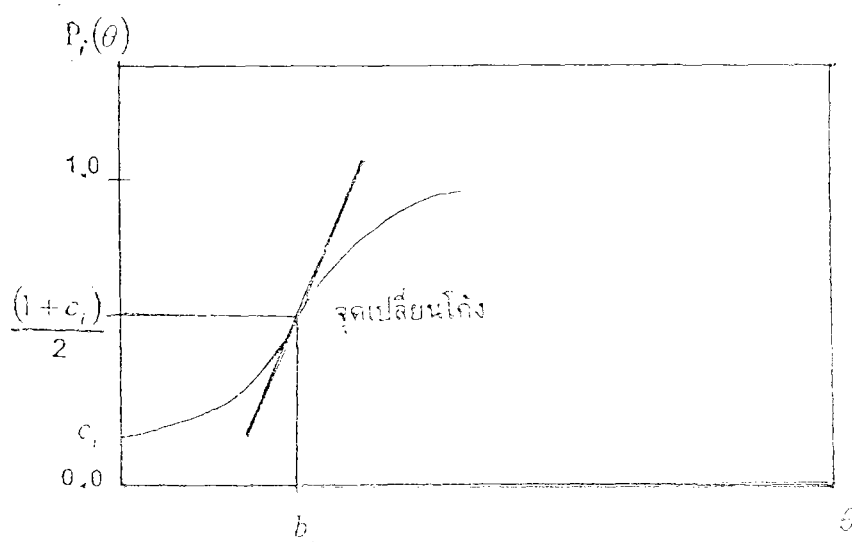
ภาพประกอบ 4 แสดงความหมายของค่าพารามิเตอร์ข้อสอบตามโมเดลโลจิสติกที่มีพารามิเตอร์ 2 ตัว

จากนั้นนำลักษณะโค้งลักษณะของข้อสอบแต่ละกลุ่มที่ศึกษาไปทดสอบด้วยสถิติ การแจกแจง F ถ้า F มีค่าอย่างมีนัยสำคัญทางสถิติ แสดงว่าข้อสอบข้อนั้นมีความลำเอียง

วิธีโค้งลักษณะ 3 พารามิเตอร์

วิธีนี้กำหนดให้ ค่าอำนาจจำแนก ค่าความยาก และค่าการเดาแปรผันไปตามกลุ่มย่อยแต่ละกลุ่ม มีรูปแบบทางคณิตศาสตร์ดังนี้ (Hambleton, Ronald and Swaminathan. 1985:37)

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



ภาพประกอบ 5 แสดงความหมายของค่าพารามิเตอร์ข้อสอบตามโมเดลโลจิสติกที่มีพารามิเตอร์ 3 ตัว

วิธีการวิเคราะห์ความลำเอียงของข้อสอบโดยใช้โค้งลักษณะข้อสอบนั้น จะถือว่าข้อสอบมีความลำเอียง เมื่อผู้สอบที่มีระดับความสามารถเท่ากัน แต่อยู่ต่างกลุ่มกัน มีโอกาสจะตอบถูกไม่เท่ากัน ซึ่งความลำเอียงของข้อสอบจะแสดงให้เห็นด้วยโค้งลักษณะข้อสอบที่แตกต่างกันของข้อสอบเดียวกันจากผู้สอบต่างกลุ่มกันมีลักษณะโค้งลักษณะเหมือนกัน ย่อมแสดงว่า ข้อสอบนั้นไม่ลำเอียง ดังนั้น ค่าดัชนีความลำเอียงของข้อสอบ จึงคำนวณโดยการหาความแตกต่างของพื้นที่คำนวณจากโค้งลักษณะข้อสอบจากผู้สอบแต่ละกลุ่ม สำหรับการประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบ จะใช้โปรแกรมสำเร็จในการประมาณค่า

การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้ทฤษฎีการตอบข้อคำถาม จะให้ค่าพารามิเตอร์ของข้อสอบที่เป็นจริงและเหมาะสมกับข้อมูลเนื่องจากการไม่แปรเปลี่ยนของค่าพารามิเตอร์ ซึ่งเป็นลักษณะเด่นของทฤษฎีนี้ จึงเป็นวิธีวิเคราะห์ที่ใช้ได้ดีตามทฤษฎี แต่ในทางปฏิบัตินั้นจะต้องใช้จำนวนกลุ่มตัวอย่างขนาดใหญ่มากและการวิเคราะห์เป็นวิธีการที่ยุ่งยากซับซ้อน จำเป็นต้องใช้เครื่องคอมพิวเตอร์ในการคำนวณ ซึ่งทำให้ค่าใช้จ่ายในการวิเคราะห์สูง (Narayanan and Swaminathan, 1994 : 315) จึงมีการศึกษาความลำเอียงของข้อสอบโดยหลีกเลี่ยงการใช้ทฤษฎีการตอบข้อคำถาม ซึ่งผู้เสนอ คือ วิธีการใช้สถิตินอนพารามิเตอร์ (Non - Parametric) ในการวิเคราะห์ความลำเอียง จากเหตุผลดังกล่าวมานี้ ผู้วิจัยจึงทำการวิเคราะห์ความลำเอียง โดยยึดหลักทฤษฎีการวิเคราะห์ความลำเอียงแบบดั้งเดิมซึ่งใช้วิธี HW 3 (HW 3 method) และวิธีค่าอำนาจจำแนก (Item Discrimination Procedure) ซึ่งเป็นวิธีการคำนวณที่ไม่ยุ่งยากซับซ้อนและสิ้นเปลืองค่าใช้จ่ายไม่มากนักในการคำนวณ และเหมาะที่จะนำมาใช้ในทางปฏิบัติ

หลักการวิเคราะห์ความลำเอียงของข้อสอบ

หลักการวิเคราะห์ความลำเอียงของข้อสอบ แบ่งเป็น 2 หลัก คือ การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้เกณฑ์ภายนอกและเกณฑ์ภายใน (Runder, Getson and Knight, 1980 : 1-10) ดังรายละเอียดดังนี้

1. การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้เกณฑ์ภายนอก

การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้เกณฑ์ภายนอกนี้ สามารถวิเคราะห์ได้ทั้งรายข้อและรายฉบับ วิธีการวิเคราะห์จะเขียนกราฟแสดงความสัมพันธ์ระหว่างคะแนนจากตัวแปรเกณฑ์ภายนอกกับตัวแปรทำนาย วิธีนี้มีจุดมุ่งหมายที่จะวิเคราะห์ความถดถอยของตัวแปรทั้งสองแล้วทำการเปรียบเทียบค่าความชัน ค่าการตัดแกนของเส้นกราฟ ในการวิเคราะห์ความลำเอียงของข้อสอบทั้งฉบับ จะใช้คะแนนรวมเป็นตัวแปรทำนาย แต่ถ้าเป็นการวิเคราะห์รายข้อจะใช้ค่าความยาก (p) ของข้อสอบแต่ละข้อ เป็นตัวแปรทำนาย ส่วนตัวแปรเกณฑ์ภายนอกจะใช้คะแนนรวมหรือเกรดเกี่ยวกับผลสัมฤทธิ์ของงานบางอย่างที่ให้ทำ (Cronbach, 1970 ; Anastssi, 1976)

การวิเคราะห์ความลำเอียงของข้อสอบด้วยหลักการนี้ จึงเป็นการเปรียบเทียบเส้นกราฟระหว่างกลุ่มผู้สอบ โดยถ้าเส้นกราฟดังกล่าวมีค่าความชันและค่าการตัดแกนแตกต่างกันในแต่ละกลุ่มแล้ว ข้อสอบข้อนั้นหรือแบบทดสอบฉบับนั้นก็就会有ความลำเอียงต่อกลุ่มผู้สอบที่มีค่าการตัดแกนและค่าความชันมากกว่า (Haebara, 1979)

การวิเคราะห์ความลำเอียงของข้อสอบตามหลักการที่กล่าวมานี้ มีจุดอ่อนตรงที่ในทางปฏิบัติแล้วเป็นการยากมากที่จะหาตัวแปรเกณฑ์ภายนอกที่มีความเที่ยงตรงเชิงพยากรณ์ และมีความยุติธรรม ซึ่งหากตัวแปรเกณฑ์ภายนอกขาดคุณสมบัติดังกล่าว จะทำให้ผลการวิเคราะห์ความลำเอียงของข้อสอบขาดความสมบูรณ์และถูกต้องเท่าที่ควร

2. การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้เกณฑ์ภายใน

การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้เกณฑ์ภายในนี้ ยึดหลักการที่ว่า ข้อสอบที่ผู้สอบที่มีระดับความสามารถเท่ากัน แต่อยู่ต่างกลุ่มกัน มีคะแนนจริงจากผลการสอบแตกต่างกันเป็นข้อสอบที่มีความลำเอียง การวิเคราะห์ความลำเอียงโดยหลักการนี้อาจแบ่งได้เป็น 3 หลักใหญ่ ๆ (Laksana and coffman.1980) ดังนี้

2.1 ใช้หลักการวัดเบี่ยงเบนสัมพัทธ์ (Relative Deviation) ของข้อสอบแต่ละข้อจากแนวโน้มเข้าสู่ส่วนกลาง วิธีนี้มีข้อด้อยเบื้องต้นว่า ข้อสอบส่วนใหญ่ในแบบทดสอบมีความเป็นเอกพันธ์ในการวัดความสามารถใดความสามารถหนึ่งในผู้สอบต่างกลุ่มกัน ข้อสอบใดที่เบี่ยงเบนไปจากส่วนกลางมากกว่าที่คาดหวังไว้ ก็อาจตั้งข้อสงสัยได้ว่าเป็นข้อสอบที่มีความลำเอียงวิธีหนึ่งที่ใช้กันบ่อย ๆ ก็คือ การวิเคราะห์ความแปรปรวน(ANOVA) ข้อสอบที่ลำเอียงคือข้อสอบที่มีปฏิสัมพันธ์ระหว่างข้อสอบกับกลุ่มผู้สอบมีนัยสำคัญทางสถิติ ส่วนอีกวิธีหนึ่งได้แก่ การเขียนกราฟไบวาริเอท วิธีการนี้ค่าพารามิเตอร์เกี่ยวกับค่าความยาก เช่น ค่า p ค่าอาร์คไซน์ ค่าเดลต้า ค่า b จากกลุ่มหนึ่งจะลงจุดคู่อันดับ คู่กับค่าพารามิเตอร์จากอีกกลุ่มหนึ่ง ข้อสอบข้อใดที่เบี่ยงเบนไปจากเส้นแกนหลักมาก ถือว่าเป็นข้อสอบที่มีความลำเอียง

2.2 ใช้หลักการประเมินความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) ของแบบทดสอบวิธีการหนึ่งที่ใช้กันมาก คือการวิเคราะห์องค์ประกอบ ดัชนีความลำเอียงของข้อสอบตามวิธีการนี้คือ ความแตกต่าง (Discrepancy) ของน้ำหนักองค์ประกอบ (Factor Loading) จากผู้สอบต่างกลุ่มกันหรือความแตกต่างอย่างมีนัยสำคัญ ระหว่างค่าเฉลี่ยของคะแนนตัวประกอบ (Factor scores) จากแต่ละกลุ่มที่นำมาเปรียบเทียบกัน ดัชนีความลำเอียงทั้งสองที่กล่าวมาเป็นตัวชี้ว่า ข้อสอบไม่ได้วัดสิ่งเดียวกันในผู้สอบต่างกลุ่มกัน

2.3 ใช้หลักการประมาณค่าโอกาสในการตอบข้อสอบแต่ละข้อถูก วิธีการที่ใช้กันทั่ว ๆ ไป คือวิธีไคสแควร์ (Chi-Square) และวิธีโค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC) วิธีการทั้งสองนี้มีความคล้ายคลึงกันในการใช้โอกาสในการตอบถูกที่แตกต่างกันจากผู้สอบ 2 กลุ่มหรือมากกว่า เป็นดัชนีความลำเอียงของข้อสอบ ส่วนข้อแตกต่างระหว่างวิธีการทั้งสอง ก็คือวิธีไคสแควร์จะประมาณความสามารถของผู้สอบโดยใช้คะแนนดิบ ส่วนวิธีโค้งลักษณะข้อสอบจะประมาณความสามารถของผู้สอบจากคุณลักษณะแฝง

งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบ

งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบนั้น แบ่งเป็นงานวิจัยในประเทศและงานวิจัยต่างประเทศ ดังนี้

งานวิจัยในประเทศ

จากการศึกษารายงานการวิจัยภายในประเทศที่ผ่านมา มีผู้ทำการวิจัยเกี่ยวกับการวิเคราะห์ความลำเอียงของข้อสอบไว้หลายท่าน โดยศึกษาในรูปการเปรียบเทียบผลการวิเคราะห์ความลำเอียงด้วยวิธีต่าง ๆ กับกลุ่มย่อยที่มีลักษณะแตกต่างกันในด้านภาษา เพศ ศาสนาและสภาพภูมิศาสตร์ รวมทั้งการเปรียบเทียบคุณภาพของแบบทดสอบหลังคัดข้อสอบที่มีความลำเอียงออกแล้ว ซึ่งผู้วิจัยจะได้นำเสนอ ดังนี้

ชัชชัย เผ่าพงศ์ (2527:73-76) ได้ทำการวิเคราะห์ความลำเอียงของข้อสอบมาตรฐานวัดความถนัดทางการเรียนด้านคณิตศาสตร์และภาษาในระดับมัธยมศึกษาตอนต้นระหว่างกลุ่มผู้สอบเพศชายและเพศหญิงโดยการเปรียบเทียบโค้งลักษณะข้อสอบด้วยวิธีแบบจำลองโลจิสติกแบบ 3 พารามิเตอร์ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปีการศึกษา 2524 เพศชายและหญิง ซึ่งประกอบด้วยทุกภูมิภาคทั่วประเทศ 5 ภาค คือ ภาคเหนือ ภาคกลาง ภาคใต้ ภาคตะวันออกและภาคตะวันออกเฉียงเหนือกลุ่มซึ่งทำการวิเคราะห์ด้วยแบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์ เป็นนักเรียนชาย 1,610 คน และนักเรียนหญิง 1,337 คน ส่วนอีกกลุ่มหนึ่งเป็นนักเรียนชาย 1,316 คน และนักเรียนหญิง 985 คน ทำการวิเคราะห์ด้วยแบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์ จำนวน 30 ข้อ มีข้อสอบที่ลำเอียงต่อกลุ่มใดกลุ่มหนึ่งโดยเฉพาะ 9 ข้อ คือ ลำเอียงต่อกลุ่มนักเรียนชาย 7 คน และลำเอียงต่อกลุ่มนักเรียนหญิง 2 ข้อ ข้อสอบที่มีความลำเอียงต่อกลุ่มนักเรียนชายและหญิงในระดับปานกลางขึ้นไปมีจำนวน 5 ข้อ ซึ่งวัดในเรื่องร้อยละ การหาปริมาตรและการหาความยาวของเส้นในด้านของรูปสามเหลี่ยมจำนวนเรื่องละ 1 ข้อ และเรื่องเกี่ยวกับโจทย์ปัญหา จำนวน 2 ข้อ จากข้อสอบจำนวน 5 ข้อ ดังกล่าวนี้นี้เป็นข้อสอบที่มีความลำเอียงต่อกลุ่มนักเรียนหญิงในช่วงความสามารถแรกและลำเอียงต่อกลุ่มนักเรียนชายในช่วงความสามารถต่อมาจำนวน 1 ข้อ ส่วนแบบทดสอบวัดความถนัดทางการเรียนด้านภาษาจำนวน 40 ข้อ มีข้อสอบที่ลำเอียงต่อกลุ่มนักเรียนชาย โดยเฉพาะ 3 ข้อ และลำเอียงต่อกลุ่มนักเรียนหญิงโดยเฉพาะมี 8 ข้อ ข้อสอบที่มีความลำเอียงในระดับปานกลางขึ้นไปมีจำนวน 9 ข้อ ซึ่งวัดความเข้าใจเกี่ยวกับการอ่านข้อความจำนวน 7 ข้อ จากข้อสอบจำนวน 9 ข้อนี้เป็นข้อที่มีความลำเอียงต่อกลุ่มนักเรียนชายโดยเฉพาะจำนวน 1 ข้อและเป็นข้อสอบที่มีความลำเอียงต่อกลุ่มนักเรียนหญิงโดยเฉพาะจำนวน 6 ข้อและมีข้อสอบจำนวน 2 ข้อ มีความลำเอียงทางการทดสอบมากกว่า 1 ช่วงความสามารถ

ทัศนีย์ พิรมนตรี(2530:77) ได้ทำการวิเคราะห์ความลำเอียงของแบบทดสอบวิชาคณิตศาสตร์ โครงการตรวจสอบคุณภาพทางการศึกษาของสำนักงานทดสอบทางการศึกษา กรมวิชาการ กระทรวงศึกษาธิการ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 จำนวน 7,036 คน โดยเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียง ระหว่างกลุ่มนักเรียนในกรุงเทพมหานครกับกลุ่มนักเรียนในภูมิภาคทั้ง 5 ภาค คือ ภาคเหนือ ภาคกลาง ภาคใต้ ภาคตะวันออก และภาคตะวันออกเฉียงเหนือด้วยวิธีวิเคราะห์ 3 วิธี คือ วิธีกำหนดจุดเดสต้า วิธีทดสอบความแตกต่างระหว่างกลุ่มด้วยสถิติไคสแควร์ในโมเดลล็อกลิเนียร์ 2 โมเดล คือ โมเดลที่ไม่มีพารามิเตอร์ผลร่วมระหว่างคะแนนกับกลุ่มและโมเดลที่ไม่มีพารามิเตอร์ของผลหลักที่เกิดจากกลุ่มและวิธีโค้งลักษณะข้อสอบที่ใช้พารามิเตอร์ 3 ตัว พบข้อสอบที่มีความลำเอียงจากวิธีโค้งลักษณะข้อสอบที่ใช้พารามิเตอร์ 3 ตัว มากที่สุด และในแต่ละวิธีวิเคราะห์พบว่ามีข้อสอบที่ลำเอียงซ้ำกันระหว่างกลุ่มนักเรียนในกรุงเทพมหานครกับกลุ่มนักเรียนในทุกภาค

สุรศักดิ์ อมรรตณศักดิ์ (2531:200–207) ได้ทำการศึกษาเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่แตกต่างกัน 4 วิธี คือ วิธีวิเคราะห์ความแปรปรวน วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 1 ตัว และวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว โดยหาสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียง 4 วิธี และเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อนและหลังการศึกษาความลำเอียงของข้อสอบ ตามวิธีการคิดคะแนนรวมที่แตกต่างกัน 6 วิธี ในด้านจำนวนผู้ได้รับการคัดเลือก สัดส่วนชาย:หญิง ที่ได้รับการคัดเลือกและความเชื่อมั่นของแบบทดสอบโดยกลุ่มตัวอย่าง คือ ผู้เข้าสอบแข่งขันเพื่อบรรจุเข้ารับราชการเป็นราชการครู ในปี พ.ศ. 2529 จำนวน 2,340 คน เป็นชาย 1,170 คน หญิง 1,170 คน เครื่องมือที่ใช้ในการรวบรวมข้อมูล ได้แก่ แบบทดสอบที่ใช้ในการสอบแข่งขันเพื่อบรรจุเข้ารับราชการเป็นราชการครูใน ปี พ.ศ. 2529 ซึ่งมี 4 ฉบับ คือ ฉบับที่ 1 แบบทดสอบวิชาเอก (ความรู้เกี่ยวกับการประถมศึกษา) ฉบับที่ 2 แบบทดสอบวิชาการศึกษาและกฎหมาย ฉบับที่ 3 แบบทดสอบภาษาไทย ฉบับที่ 4 แบบทดสอบความสามารถทั่วไป ซึ่งผลการศึกษา วิธีวิเคราะห์ความลำเอียงแต่ละวิธีพบจำนวนข้อที่ลำเอียงแตกต่างกัน โดยวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว พบจำนวนข้อสอบที่ลำเอียงมากที่สุด รองลงมาได้แก่ วิธีวิเคราะห์ความแปรปรวนและวิธีแปลงค่าความยากของข้อสอบ พบจำนวนข้อสอบที่มีความลำเอียงน้อยที่สุด ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียง 4 วิธี พบว่า มีค่าอยู่ระหว่าง.7535ถึง.9921 ในด้านจำนวนผู้ได้รับการคัดเลือกเมื่อเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อนและหลังการศึกษาความลำเอียงของข้อสอบตามวิธีคิดคะแนนดิบกับวิธีรวมคะแนนแบบอื่น ๆ อีก 5 วิธี มีจำนวนผู้ได้รับการคัดเลือกแตกต่างกันประมาณ 4 % ถึง 24 % ส่วนวิธีแปลงคะแนนดิบเป็นคะแนนมาตรฐานที่ปกติกับวิธีรวมคะแนนแบบอื่น ๆ อีก 4 วิธี มีจำนวนผู้ได้รับการคัดเลือกแตกต่างกันประมาณ 6% ถึง 23%ส่วนค่าความเชื่อมั่นพบว่าลดลงจากเดิมเล็กน้อย

พัชรีย์ ปิยภักดิ์ (2531:64–69) ได้ทำการวิเคราะห์ความลำเอียงของข้อสอบวิชาคณิตศาสตร์ เรื่องสมการ ชั้นประถมศึกษาปีที่ 6 จำนวน 45 ข้อ ที่ผู้วิจัยสร้างขึ้นเองระหว่างกลุ่มผู้สอบที่มีเพศ และสภาพภูมิศาสตร์ต่างกัน ด้วยวิธีวิเคราะห์ 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีไคสแควร์ และวิธีไคงลักษณะของข้อสอบ 1 พารามิเตอร์ โดยทำการเปรียบเทียบจำนวนข้อสอบที่ ลำเอียง การวิเคราะห์หาสัมประสิทธิ์สหสัมพันธ์ของวิธีวิเคราะห์ทั้ง 3 วิธี และศึกษาเปรียบเทียบ ค่าความเชื่อมั่นของแบบทดสอบก่อนและหลังคัดข้อสอบที่มีความลำเอียงออกแล้ว กลุ่มตัวอย่าง ที่ใช้เป็นนักเรียนชายและหญิง ชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2530 ของโรงเรียนสังกัด สำนักงานการประถมศึกษาแห่งชาติในกรุงเทพมหานครและจังหวัดสมุทรสาคร จำนวน 1,422 คน ซึ่งเลือกโดยวิธีสุ่มแบบแบ่งชั้น ผลการวิเคราะห์พบว่า วิธีไคงลักษณะข้อสอบ 1 พารามิเตอร์ พบ จำนวนข้อสอบที่ลำเอียงต่อเพศมากที่สุด จำนวน 7 ข้อ รองลงมา คือ วิธีไคสแควร์และวิธีแปลงค่า ความยากของข้อสอบ โดยค้นพบได้จำนวน 3 ข้อ และ 2 ข้อ ตามลำดับ สำหรับความแตกต่างกัน ในด้านสภาพภูมิศาสตร์ ได้ผลการค้นพบเช่นเดียวกัน กล่าวคือ วิธีไคงลักษณะข้อสอบ 1 พารา มิเตอร์ พบจำนวนข้อสอบที่ลำเอียงต่อเพศมากที่สุด จำนวน 18 ข้อ รองลงมา คือ วิธีไคสแควร์ และวิธีแปลงค่าความยากของข้อสอบ โดยค้นพบได้จำนวน 3 ข้อ และ 10 ข้อ ตามลำดับ ค่าสัม ประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงต่อเพศ ระหว่างวิธีวิเคราะห์ทั้ง 3 วิธี มีค่าระหว่าง .1713ถึง.5613 และค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงต่อนักเรียนกรุงเทพมหานคร และ นักเรียนสมุทรสาคร ระหว่างวิธีวิเคราะห์ทั้ง 3 วิธี มีค่าระหว่าง.1868 ถึง.6009 สำหรับค่า ความเชื่อมั่นของแบบทดสอบก่อนและก่อนการคัดข้อสอบที่ลำเอียงออกแล้ว พบว่าแตกต่างกัน อย่างไม่มีนัยสำคัญทางสถิติ

นิรมล ชัยขวลิต (2537:57–58) ได้ทำการเปรียบเทียบผลการวิเคราะห์ความลำเอียงของแบบ ทดสอบความเข้าใจในการอ่านภาษาไทย จำนวน 50 ข้อ ที่ผู้วิจัยสร้างขึ้นเองระหว่างกลุ่มผู้สอบที่มี เพศต่างกัน ด้วยวิธีวิเคราะห์ความลำเอียง 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีไคสแควร์ และวิธีวิเคราะห์ความแปรปรวน โดยศึกษาเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงและศึกษา เปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังคัดข้อสอบที่มีความลำเอียงออกแล้วที่คำนวณ ด้วยสูตร $KR - 20$ และวิธีแบ่งครึ่งฉบับ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชายและหญิง ชั้นมัธยม ศึกษาปีที่ 1 ปีการศึกษา 2536 ของโรงเรียนในสังกัดกรมสามัญศึกษาส่วนกลาง กลุ่มที่ 5 กรุงเทพมหานคร จำนวน 1,006 คน ซึ่งได้มาโดยการสุ่มอย่างง่าย ผลการวิเคราะห์พบว่า การ วิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากของข้อสอบกับวิธีไคสแควร์และวิธีแปลงค่าความ ยากของข้อสอบกับวิธีวิเคราะห์ความแปรปรวนมีจำนวนข้อที่มีความลำเอียงแตกต่างกันอย่างมีนัย สำคัญทางสถิติที่ระดับ.05 ส่วนวิธีไคสแควร์กับวิธีวิเคราะห์ความแปรปรวนพบจำนวนข้อที่มีความ ลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ.05 สำหรับค่าความเชื่อมั่นที่คำนวณด้วย สูตร $KR - 20$ และแบบแบ่งครึ่งฉบับที่วิเคราะห์ทั้ง 3 วิธีแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

วิชาภาษาไทย กข และวิชาภาษาอังกฤษ กขค วิชาละ 5,4175 คน (ชาย1,454คน หญิง3,961 คน) ความเที่ยงตรงเชิงพยากรณ์ศึกษาจากคะแนนสอบคัดเลือกกับระดับผลการเรียนภาคเรียนที่ 1 ปีการศึกษา 2538 ของนักศึกษาที่ได้รับการคัดเลือกจากประเภทรับตรง สายวิทยาศาสตร์ 763 คน และสายศิลปศาสตร์ 281 คนและสาเหตุความลำเอียงของข้อสอบจากการระบุสาเหตุของนักวัดผล การศึกษาหรืออาจารย์ผู้สอนจำนวน 50คนและนิสิตนักศึกษาที่เรียนในสาขาวิชานั้นๆวิชาละ50 คน ผลการศึกษาพบว่า วิธีตรวจสอบความลำเอียงทั้ง 3 วิธี ตัดสินจำนวนข้อสอบที่ลำเอียงแตกต่างกัน ในวิชาภาษาไทย ก ฉบับที่ 2 และวิชาสังคมศึกษา ก ฉบับที่ 1 ที่ระดับนัยสำคัญทางสถิติที่ .05 นอกจากนั้นต่างกันที่ระดับนัยสำคัญทางสถิติที่ .01 โดยใช้ทฤษฎีการตอบข้อคำถามตัดสินจำนวน ข้อสอบที่ลำเอียงได้มากที่สุด ความสัมพันธ์ของลำดับที่ของการสอบไม่ว่าจะคิดคะแนนมาตรฐานที่-ปกติ หรือ คิดคะแนนน้ำหนักความสามารถและใช้ข้อสอบจำนวนทั้งหมด หรือ ใช้เฉพาะข้อสอบที่ ปราศจากความลำเอียงต่างมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการเรียนในสายวิทยาศาสตร์ การคิดคะแนนน้ำหนักความสามารถทั้งใช้ข้อสอบจำนวนทั้งหมดและใช้เฉพาะข้อสอบที่ปราศจากความลำเอียงมีประสิทธิภาพในการทำนายสูงกว่าการคิดคะแนนมาตรฐานที่-ปกติ ส่วนสายศิลปศาสตร์ การคิดคะแนนมาตรฐานที่-ปกติ ใช้ข้อสอบทั้งหมด และใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง มีประสิทธิภาพในการทำนายสูงกว่าการคิดคะแนนน้ำหนักความสามารถ และสาเหตุของความลำเอียงของข้อสอบ ต่อเพศทั้งเพศชายและหญิง เกิดจากข้อสอบเป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะจะมี โอกาสตอบถูกมากกว่า เป็นเรื่องราวที่กลุ่มนั้น ๆ สนใจ และเป็นข้อสอบที่ถามความจำ

อังคณา สายยศ (2539:26) ได้ศึกษาผลการวิเคราะห์ความลำเอียงของแบบทดสอบความ เข้าใจในการอ่านตามลักษณะการใช้ภาษาพูดในครอบครัวเป็นภาษาลาว เขมร และภาษาส่วย โดยศึกษาเปรียบเทียบจำนวนข้อที่ลำเอียง ความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อที่มีความ ลำเอียงออกแล้ว และคะแนนที่ได้จากการสอบหลังคัดข้อที่ลำเอียงออกแล้ว กลุ่มตัวอย่างเป็นนัก เรียนชั้นประถมศึกษาปีที่ 6 ในจังหวัดสุรินทร์ ได้มาโดยการสุ่มแบบแบ่งชั้น อำเภอเป็นชั้นและ โรงเรียนเป็นหน่วยการสุ่ม ได้จำนวนโรงเรียนทั้งสิ้น 29 โรงเรียน นักเรียน 860 คน ทำการ วิเคราะห์ความลำเอียงด้วยวิธีแมนเทล-แฮนส์เซล ผลการวิจัยพบว่า จำนวนข้อสอบที่ลำเอียงของ แบบทดสอบความเข้าใจในการอ่านระหว่างกลุ่มผู้สอบที่ใช้ภาษาพูดเป็นภาษาลาว ภาษาเขมร และ ภาษาส่วย แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ สำหรับค่าความเชื่อมั่นของแบบทดสอบหลัง คัดข้อที่มีความลำเอียงออกแล้วพบว่าค่าความเชื่อมั่นระหว่างกลุ่มผู้สอบที่ใช้ภาษาพูดเป็นภาษา ลาว ภาษาเขมร และ ภาษาส่วย แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ส่วนคะแนนที่ได้รับ จากการสอบด้วยแบบทดสอบที่คัดข้อที่มีความลำเอียงออกแล้วมีคะแนนระหว่างกลุ่มผู้สอบที่พูด ภาษาลาวกับภาษาส่วย ภาษาเขมรกับภาษาส่วย มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ส่วนภาษาลาวกับภาษาเขมร มีคะแนนแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

จิตสุตา ธารพร (2539:129 – 134) ได้ทำการศึกษาเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบโจทยปัญหาทางคณิตศาสตร์และแบบทดสอบวัดความเข้าใจทางภาษาโดยใช้การวิเคราะห์ 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และวิธีแมนเทิล-แฮนส์เซล โดยทำการเปรียบเทียบกันในด้านจำนวนข้อที่มีความลำเอียง และค่าความเชื่อมั่นหลังคัดข้อที่มีความลำเอียงออกแล้ว ในการศึกษาครั้งนี้ใช้นักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2538 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดกระบี่ จำนวน 2,689 คน ซึ่งเลือกมาโดยการสุ่มอย่างง่าย ผลการวิจัยพบว่าวิธีวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบโจทยปัญหาทางคณิตศาสตร์และแบบทดสอบวัดความเข้าใจทางภาษาแต่ละวิธีพบจำนวนข้อสอบที่มีความลำเอียงแตกต่างกัน ซึ่งวิธีแมนเทิล-แฮนส์เซล พบจำนวนข้อสอบที่มีความลำเอียงมากที่สุดและวิธีแปลงค่าความยากของข้อสอบพบจำนวนข้อที่มีความลำเอียงน้อยที่สุด โดยที่การวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากของข้อสอบกับวิธีโค้งลักษณะของข้อสอบ 3 พารามิเตอร์ และวิธีแปลงค่าความยากของข้อสอบกับวิธีแมนเทิล-แฮนส์เซล มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.001 ส่วนวิธีโค้งลักษณะของข้อสอบ 3 พารามิเตอร์กับวิธีแมนเทิล-แฮนส์เซล มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติ

สำหรับค่าความเชื่อมั่นของข้อสอบจากแบบทดสอบโจทยปัญหาทางคณิตศาสตร์ที่คำนวณโดยใช้สูตรแบบ KR-20 หลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วจากวิธีแปลงค่าความยากของข้อสอบกับวิธีโค้งลักษณะของข้อสอบ 3 พารามิเตอร์ และวิธีแปลงค่าความยากของข้อสอบกับวิธีแมนเทิล-แฮนส์เซล มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ส่วนค่าความเชื่อมั่นของที่คำนวณโดยใช้สูตรแบบแบ่งครึ่งฉบับ พบว่า วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะของข้อสอบ 3 พารามิเตอร์ และ วิธีแมนเทิล-แฮนส์เซล มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ส่วนแบบทดสอบวัดความเข้าใจทางภาษาที่คำนวณโดยใช้สูตรแบบKR-20 และสูตรแบบแบ่งครึ่งฉบับ ฉบับ พบว่า วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะของข้อสอบ 3 พารามิเตอร์ และ วิธีแมนเทิล-แฮนส์เซล มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

คมศักดิ์ ชื่นชม (2539:บทคัดย่อ) ได้ทำการการศึกษาผลการวิเคราะห์ความลำเอียงที่ใช้วิธีต่างกันของแบบทดสอบวัดจริยธรรมด้านความซื่อสัตย์ ตามทฤษฎีคลาสสิกอลที่ใช้ต่างกัน 3 วิธี คือ วิธีค่าอำนาจจำแนกของข้อสอบ วิธีวิเคราะห์องค์ประกอบ และวิธีวิเคราะห์ถดถอย โดยเปรียบเทียบจำนวนข้อที่มีความลำเอียงและค่าความเชื่อมั่นแบบสัมประสิทธิ์แอลฟาหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2538 จังหวัดยะลา จำนวน 1,081 คน ซึ่งได้มาโดยการสุ่มแบบแบ่งชั้น ผลการวิจัยพบว่าผลการวิเคราะห์ความลำเอียงของข้อสอบ ด้วยวิธีค่าอำนาจจำแนกของข้อสอบ วิธีวิเคราะห์องค์ประกอบและวิธีวิเคราะห์การถดถอยมีจำนวนข้อที่มีความลำเอียงน้อยและไม่แตกต่างกันที่

ระดับนัยสำคัญทางสถิติ .05 ส่วนค่าความเชื่อมั่นของแบบทดสอบวัดจริยธรรมด้านความซื่อสัตย์ ที่ใช้สูตรคำนวณแบบสัมประสิทธิ์แอลฟา หลังคัดเลือกข้อที่มีความลำเอียงออก จากการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีค่าอำนาจจำแนกของข้อสอบ วิธีวิเคราะห์องค์ประกอบ และ วิธีวิเคราะห์การถดถอยมีค่าไม่แตกต่างกัน ที่ระดับนัยสำคัญทางสถิติ .05

พรรณี จินตมาศ (2540:87-92) ได้ทำการศึกษาเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบโดยใช้ขนาดกลุ่มผู้สอบต่างกัน คือ ขนาด 500 และ 1,000 คน และวิธีวิเคราะห์ 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีแมนเทิล-แฮนส์เทล และ วิธีชิปเทสท์ โดยทำการเปรียบเทียบในด้านจำนวนข้อที่มีความลำเอียง และค่าความเชื่อมั่นหลังคัดข้อที่มีความลำเอียงออกแล้ว ในการศึกษาครั้งนี้ใช้นักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2539 ของโรงเรียนสังกัดกรมสามัญศึกษา ส่วนกลาง จำนวน 2,200 คน ซึ่งเลือกมาโดยการสุ่มแบบแบ่งชั้น มีขนาดโรงเรียนเป็นชั้นและโรงเรียนเป็นหน่วยการสุ่ม ผลการวิจัยพบว่า เมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 500 คน วิธีชิปเทสท์พบข้อสอบที่มีความลำเอียงมากที่สุด และ วิธีแปลงค่าความยากของข้อสอบพบข้อสอบที่มีความลำเอียงน้อยที่สุด โดยจำนวนข้อสอบที่ลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทุกวิธีวิเคราะห์ และเมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 1,000 คน วิธีแมนเทิล-แฮนส์เทลพบข้อสอบที่มีความลำเอียงมากที่สุด วิธีแปลงค่าความยากของข้อสอบไม่พบข้อสอบที่มีความลำเอียง โดยจำนวนข้อสอบที่ลำเอียงจากการวิเคราะห์ด้วยวิธีแมนเทิล-แฮนส์เทลกับวิธีแปลงค่าความยากของข้อสอบ และ วิธีแปลงค่าความยากของข้อสอบกับวิธีชิปเทสท์แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ สำหรับจำนวนข้อสอบที่ลำเอียงจากการวิเคราะห์ด้วยวิธีแปลงค่าความยากของข้อสอบระหว่างกลุ่มผู้สอบขนาด 500 คน และกลุ่มผู้สอบขนาด 1,000 คน จะมีจำนวนข้อสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

ค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว เมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 500 คน จากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากของข้อสอบกับวิธีชิปเทสท์ มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ สำหรับค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่ลำเอียงออกแล้วจากการวิเคราะห์ด้วยวิธีชิปเทสท์ระหว่างกลุ่มผู้สอบ 500 คน และกลุ่มผู้สอบ 1,000 คน มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

งานวิจัยในต่างประเทศ

สำหรับงานวิจัยในต่างประเทศที่เกี่ยวกับการวิเคราะห์ความลำเอียงของข้อสอบนั้น มีผู้วิจัยทำวิจัยไว้หลายท่านด้วยกัน ซึ่งผู้วิจัยได้รวบรวมและสรุปไว้ดังนี้

ผจงจิต อินสุวรรณ (Intasuwan.1979:92) ศึกษาวิธีวิเคราะห์ความลำเอียงที่ต่างกัน 3 วิธี คือ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ วิธีราสซัส และวิธีไคสแควร์ ของแบบทดสอบ IEA (The International Association for Evaluation of Education Achivement) กลุ่มตัวอย่างที่ใช้ศึกษา เป็นนักเรียนอายุ 14 ปี ชาวอังกฤษ อเมริกัน และนิวซีแลนด์ จำนวน 8,601 คน ผลการวิจัยพบว่า สหสัมพันธ์ระหว่างวิธีทั้ง 3 วิธีมีค่าอยู่ระหว่าง.051–0.98 และวิธีไคสแควร์กับวิธีราสซัส มีสหสัมพันธ์กันสูงถึง 0.98

สงบ ลักษณะ (Buriril.1982:168 ; citing Laksana. 979) ได้วิเคราะห์ความลำเอียงต่อเพศและผิวของข้อสอบในแบบทดสอบ ITBS (IOWA Test of Basic Skills)ฉบับภาษาต้นคำศัพท์ และฉบับคณิตศาสตร์ ด้วยวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และวิธีวิเคราะห์ความแปรปรวน ผลการวิจัยพบว่า วิธีวิเคราะห์ความแปรปรวนเป็นวิธีที่ครอบคลุมกว่า แต่ต้องใช้กลุ่มตัวอย่างขนาดใหญ่และการคำนวณยุ่งยากซับซ้อน

รัตเนอร์และคณะ (Subkovaik.1984:51 ; citing Rudner et al.1980 ได้ทำการวิเคราะห์ความลำเอียงของข้อสอบ 7 วิธี คือวิธีแปลงค่าความยากของข้อสอบ 2 วิธี วิธีไคสแควร์ 2 วิธี และวิธีโค้งลักษณะข้อสอบ 3 วิธี กับข้อมูลที่จำลองขึ้น ผลการวิจัยพบว่าวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ เป็นวิธีที่วิเคราะห์ความลำเอียงของข้อสอบได้ดีที่สุด รองลงมาได้แก่หนึ่งในสองของวิธีวิเคราะห์ไคสแควร์ ส่วนวิธีแปลงค่าความยากของข้อสอบ ผู้วิจัยแนะนำให้ใช้ได้ทั้งในทางปฏิบัติ

ไอรอนสัน,ฮูแมน,วิลลิสและซิงเนอร์ (Iroson,Homan,Willis and Signer.1984:391-396) ได้ทำการศึกษาเกี่ยวกับความเที่ยงตรงของเทคนิคการวิเคราะห์ความลำเอียงกับแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์ โดยใช้วิเคราะห์การวิเคราะห์ความลำเอียงต่างกัน 3 วิธี คือวิธีแปลงค่าความยาก วิธีไคสแควร์ของคามิลลีและวิธีแบบ IRT ของลินน์และแอร์นีสท์ ในการศึกษาค้นคว้าครั้งนี้ใช้แบบทดสอบที่เกี่ยวกับความสามารถในการอ่านและคณิตศาสตร์(Comprehensive Test of Basic Skill:CTBS ; 1973) ซึ่งมี 3 รูปแบบ คือ รูปแบบที่ 1 มีจำนวน 10 ข้อ เป็นแบบทดสอบเกี่ยวกับอ่านสำหรับนักเรียนเกรด 2 และคณิตศาสตร์ระดับเกรด 2 รูปแบบที่ 2 มีจำนวน 10 ข้อ เป็นแบบทดสอบเกี่ยวกับอ่านสำหรับนักเรียนเกรด 2 และคณิตศาสตร์ระดับเกรด 4 รูปแบบที่ 3 มีจำนวน 10 ข้อ เป็นแบบทดสอบเกี่ยวกับอ่านสำหรับนักเรียนเกรด 4 และคณิตศาสตร์ระดับเกรด 4 กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นนักเรียนเกรด 2 และ เกรด 4 จำนวน 1,064 คน โดยแบ่งนักเรียนออกเป็น 2 กลุ่ม คือ กลุ่มขนาดใหญ่จำนวน 916 คน และกลุ่มขนาดเล็กจำนวน 148 คน ซึ่งผลการศึกษาพบว่าวิธีแบบ IRT ของลินน์และแอร์นีสท์ใช้ได้ดีในทางปฏิบัติกว่าวิธีอื่น ๆ

ฮูเวอร์ และ โคลเลน (Hoover and Kolan.1984:173-181) ได้ทำการศึกษาเกี่ยวกับความเชื่อมั่นของดัชนีความลำเอียงทางเพศ ซึ่งใช้วิธีการวิเคราะห์ความลำเอียง 6 วิธี คือวิธีแปลงค่าความยาก 2 วิธี, วิธีไบซีเรียล,วิธีพอยต์ไบซีเรียล,วิธีไคสแควร์และวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ เครื่องมือที่ใช้ในการศึกษาครั้งนี้เป็นแบบทดสอบ ITBS (Iowa Tests of Basic Skill) กลุ่มตัวอย่างที่ใช้เป็นนักเรียนเกรด 5 จำนวน 800 คน ในปี 1977 โดยแบ่งเป็นกลุ่มย่อย ทุละ 200 คน คือ กลุ่มนักเรียนชายผิวดำ,กลุ่มนักเรียนหญิงผิวขาว,กลุ่มนักเรียนหญิงผิวดำและกลุ่มนักเรียนชายผิวขาว ผลการศึกษาพบว่าดัชนีความลำเอียงของข้อสอบที่มีความยุติธรรมจะพบในการศึกษากับกลุ่มเปรียบเทียบ จำนวน 200 คน และกลุ่มอ้างอิงจำนวน 200 คน การใช้ดัชนีความลำเอียง ของข้อสอบในแบบทดสอบวัดผลสัมฤทธิ์ไม่สามารถที่จะคาดหวังในการตัดสินใจเกี่ยวกับข้อสอบที่มีความลำเอียงของข้อสอบได้

ชุนแมน (Scheuneman.1987:97-118) ศึกษาและสำรวจเกี่ยวกับสาเหตุของความลำเอียงของข้อคำถามในแบบทดสอบ GRE (Graduate Record Examination Genral Test) โดยการศึกษาครั้งนี้ มีสมมติฐานหลัก 6 ข้อ และสมมติฐานย่อย 16 ข้อ เกี่ยวกับความน่าจะเป็นของสาเหตุที่ก่อให้เกิดความลำเอียงของข้อคำถามในแบบทดสอบจากนักเรียนผิวขาวและนักเรียนผิวดำ โดยแบบทดสอบนี้พัฒนาเป็นรูปแบบคู่กัน ให้สอดคล้องกับสมมติฐานการทดสอบ ซึ่งแต่ละข้อจะใช้แบบสอบ 2 รูปแบบ คือแบบทดสอบเลือกตอบธรรมดาและแบบทดสอบตัวเลือกซ้อนมีความสอดคล้องในด้านข้อคำถามและการเฉลย แล้วทำการวิเคราะห์ด้วยวิธีล็อกไลเนียร์ ซึ่งผลการศึกษาว่าสมมติฐาน 10 ใน 16 ข้อ จะแสดงถึงปฏิสัมพันธ์ระหว่างกลุ่มย่อยและข้อคำถามจะแสดงให้เห็นถึงความแตกต่างของข้อคำถามที่โอนเอียงในการตอบของนักเรียนผิวขาวและผิวดำ ความซับซ้อนของปฏิสัมพันธ์ที่เกิดขึ้น อย่างไรก็ตาม เขาได้เสนอแนะว่า องค์ประกอบที่ไม่ได้ควบคุมจะมีผลต่อการสอบด้วยเช่นกัน

ซวิก,โดโนฮู และกริมา (Zwick,Donoghue and Grima.1993:233-257) ได้ทำการประเมินความลำเอียงของข้อสอบภาคปฏิบัติโดยมีวัตถุประสงค์เพื่ออธิบายการวิเคราะห์ความลำเอียงของแบบทดสอบที่มีการตรวจให้คะแนนแบบ 1-0 (dichotomous item) ด้วยวิธีวิเคราะห์แบบแมนเทิล-แฮนส์เซล ที่พัฒนาโดยฮอลแลนด์และไทเยอร์ (1988) วิธีสัมประสิทธิ์การสรุปอ้างอิงแบบแมนเทิล-แฮนส์เซล เพื่อที่จะพัฒนาวิธีวิเคราะห์ความลำเอียงจากข้อสอบภาคปฏิบัติ โดยข้อมูลที่ใช้ในการศึกษาเป็นข้อมูลแบบจำลองแบบ 54 เงื่อนไข (2 X 27) ในการข้ามกลุ่ม 4 ทาง ในการจับคู่ตัวแปรคำนวณ ซึ่งผลการศึกษาพบว่า วิธีการของแมนเทิล-แฮนส์เซล ดีกว่าวิธี GMH ในบางกรณี

เวลช์และฮูเวอร์ (Wing and Lane.1996 183–184 ; citing Wech and Hoover.1993)ได้ทำการเปรียบเทียบวิธีการวิเคราะห์ความลำเอียงสองวิธี คือ วิธีHW3 กับวิธีไคสแควร์แบบแมนเทิลภายใต้เงื่อนไขของข้อมูลหลาย ๆ เงื่อนไข ซึ่งผลการศึกษาพบว่า ดัชนีของHW3 ใช้ได้ดีกว่าวิธีของแมนเทิลภายใต้อัตราความคลาดเคลื่อนประเภทที่ 1 และดัชนีHW3 มีแนวโน้มที่จะค้นหาความลำเอียงได้ดีกว่าวิธีแมนเทิลเมื่อขนาดกลุ่มผู้สอบเท่ากับหรือมากกว่า 500 คนขึ้นไป ลักษณะพิเศษอีกอย่างหนึ่งคือ ในทางปฏิบัติสถิติไคสแควร์แบบแมนเทิลมีผลน้อยกว่าวิธี HW3 เมื่อขนาดกลุ่มผู้สอบประมาณ 250 ถึง 500 คนต่อกลุ่ม โดยการเปรียบเทียบระดับความสามารถของขนาดกลุ่ม ผู้สอบที่ได้จากการสังเกต วิธีHW3 มีอำนาจในการค้นหามากกว่าวิธีดังกล่าวกับกลุ่มผู้สอบ250: 250

มัลเลอร์ (Maller.1994) ได้พัฒนาแบบทดสอบวัดเชาว์ปัญญาสำหรับเด็กของเวลช์เลอร์ที่ปรับปรุงครั้งที่ 3 (WISH-III) ซึ่งเดิมสร้างไว้สำหรับเด็กที่สามารถพูดภาษาอังกฤษได้ ทำให้เกิดปัญหาว่าถ้าจะใช้กับเด็กหูตึงจะมีความเที่ยงตรงเชิงโครงสร้างเพียงพอหรือไม่ ข้อสอบจะทำหน้าที่เบี่ยงเบนไปหรือไม่และธรรมชาติของเนื้อหาของข้อสอบแต่ละข้อจะทำหน้าที่ลำเอียงหรือไม่ ซึ่งแบบทดสอบถูกแปลเป็นภาษาสัญลักษณ์ (ภาษามือ) และทดลองสอบกับเด็กนักเรียนหูตึง 110 คนจาก 3 แหล่ง อายุระหว่าง 8–11 ปี ที่หูตึงโดยกำเนิด ใช้ภาษามือและมี IQ ใกล้เคียงกัน ตรวจสอบความเที่ยงตรงเชิงโครงสร้างด้วยโปรแกรม LISREL วิเคราะห์ความแปรปรวนร่วม ตรวจสอบความลำเอียงโดยราชสีโมเดลในกลุ่มย่อยของแบบทดสอบภาพสำเร็จ การสื่อสาร ความคล้ายคลึงทางเรขาคณิต ศัพท์ และบทสนทนา ในทุกแบบทดสอบย่อยจะมี DIF และ DIF จะเพิ่มความแตกต่างในข้อสอบที่ยาก ยกเว้นคำศัพท์ ความลำเอียงของข้อสอบอาจจะมาจากสาเหตุการแปลความและหลักสูตร

นารายานานและสวามินาทาน (Narayanan and Swaminathan.1994:315-328) ได้ศึกษาเปรียบเทียบผลการวิเคราะห์ความลำเอียงแบบยูนิฟอร์ม ด้วยวิธี MH และวิธี SIB โดยใช้ข้อมูลที่จำลองภายใต้เงื่อนไข 1,296 เงื่อนไข ($9 \times 3 \times 2 \times 4 \times 6$) แล้วทำการวิเคราะห์ซ้ำ 100 ครั้ง ในแต่ละเงื่อนไข เงื่อนไขในการศึกษาประกอบด้วย

1. ขนาดกลุ่มตัวอย่างที่ต่างกัน 3 ขนาด คือกลุ่มเปรียบเทียบ ใช้ขนาด 100, 200 และ 300 คน กลุ่มอ้างอิง ใช้ขนาด 300 , 500 และ 1,000 คน วางไว้ด้วยกันได้กลุ่มตัวอย่างทั้งหมด 9 กลุ่ม
2. ความแตกต่างในการแจกแจงความสามารถ 3 ระดับ ระดับที่ 1 ให้ทั้ง 2 กลุ่มมีค่าเฉลี่ยการแจกแจงความสามารถอยู่ที่ 0.0 ระดับที่ 2 ให้ค่าเฉลี่ยของความสามารถในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 0.0 และ 0.5 ตามลำดับ และระดับที่ 3 ค่าเฉลี่ยความสามารถในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 0.0 และ -1.0 ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานของทั้ง 3 ระดับเป็น 1.0
3. ร้อยละของข้อสอบที่ลำเอียงมี 2 ขนาด คือ ร้อยละ 10 และร้อยละ 20 จากแบบทดสอบจำนวน 40 ข้อ
4. ขนาดของความลำเอียง 4 ระดับ คือ 0.4 , 0.6 , 0.8 และ 1.0

5. ชนิดของแบบทดสอบ 6 ชนิด คือ ชนิดของแบบทดสอบที่ประกอบด้วยค่าความยากและค่าอำนาจจำแนก 3 ระดับ คือ ระดับสูง ปานกลาง ต่ำ วางไว้ด้วยกัน แต่คัดแบบทดสอบที่ประกอบด้วยค่าความยากและค่าอำนาจจำแนกสูง ปานกลาง ต่ำ ทั้ง 2 ค่า ออก จำนวน 3 ฉบับ

ผลการวิจัยพบว่า วิธี MH และวิธี SIB สามารถตรวจค้นความลำเอียงได้ใกล้เคียงโดยจะตรวจค้นได้ดีในขนาดกลุ่มตัวอย่าง 600 คน (กลุ่มอ้างอิงและกลุ่มเปรียบเทียบ กลุ่มละ 300 คน) ขนาดของความลำเอียงทั้ง 4 ขนาดมีผลต่อการตรวจค้นความลำเอียงทั้ง 2 วิธีอย่างมีนัยสำคัญทางสถิติ ความแตกต่างในการแจกแจงความสามารถไม่มีผลต่อวิธีการวิเคราะห์ความลำเอียงแบบ SIB แต่มีผลต่อวิธี MH ที่เป็นเช่นนี้ เพราะวิธี SIB ได้มีการปรับแก้การถดถอย ขนาดแบบทดสอบและร้อยละของข้อสอบที่ลำเอียงไม่มีผลต่อทั้งสอง สำหรับการเกิดความคลาดเคลื่อนแบบที่ 1 วิธี MH จะมีอัตราการเกิดอยู่ในระดับปกติ แต่วิธี SIB จะเพิ่มขึ้นเล็กน้อยในกลุ่มที่มีความสามารถเท่ากัน แต่ในกลุ่มที่มีความสามารถไม่เท่ากันการเกิดความคลาดเคลื่อนแบบที่ 1 จะมีอัตราการเกิดสูงขึ้นทั้ง 2 วิธี ซึ่งการเพิ่มความยาวของแบบทดสอบจะทำให้เกิดความเชื่อมั่นสูงซึ่งมีผลให้ความคลาดเคลื่อนแบบที่ 1 น่าจะอัตราการเกิดลดลง

เวลช์และมิลเลอร์ (Welch and Miller.1995:163-178) ได้ทำการศึกษาความลำเอียงของแบบทดสอบเกี่ยวกับตัวอย่างปัญหาซึ่งแบบทดสอบเป็นแบบเขียนตอบ โดยการศึกษาครั้งนี้วัตถุประสงค์ของการศึกษามีอยู่ 2 ประการ เพื่อเปรียบเทียบผลลัพธ์ของการตรวจสอบความลำเอียงของแบบทดสอบที่มีการตรวจให้คะแนนแบบหลายค่า(Polytomous) ภายใต้ความแตกต่างของตัวแปรจากภายในและภายนอก ซึ่งการวิเคราะห์ความลำเอียงครั้งนี้ใช้ทั้งหมด 3 วิธีคือวิธีHW3 วิธีGMH และวิธีLogistic discriminant function analysis (LDFD) โดยกลุ่มตัวอย่างที่ใช้เป็นนักเรียนเกรด 8 ปี 1992 ตัวแปรที่ใช้ในการวิเคราะห์ความลำเอียง คือ เชื้อชาติ และเพศ ซึ่งผลการศึกษาพบว่า วิธีวิเคราะห์ความลำเอียงแบบ LDFD ค้นพบข้อสอบที่มีความลำเอียงมากที่สุดภายใต้เงื่อนไขของแบบทดสอบแบบยูนิฟอร์ม (Unifrom) เท่านั้น

มาโซร์,แคนจี และ คลาเซอร์ (Mazor,Kanrjee and Clauser.1995:131-144) ได้ใช้วิธี Logistic Regression และวิธีแมนเทล-แฮนส์เซล กับการประมาณค่าความสามารถหลายอย่าง (Multiple Ability Estimates) ในการค้นหาความลำเอียงของข้อสอบ ข้อมูลที่ใช้ในการศึกษาค้นคว้าเป็นแบบทดสอบวิชาเคมีและวิชาประวัติศาสตร์ ตัวแปรที่ใช้ในการวิเคราะห์ความลำเอียงมี 2 ตัวแปร คือ ตัวแปรแรกเป็นเพศของกลุ่มผู้สอบ ได้แก่ เพศชายและหญิง ตัวแปรที่สองเป็นกลุ่มผู้สอบที่มีความสามารถในการทำแบบฝึกหัดและรายงานภาษาอังกฤษได้ดี และ อีกกลุ่มหนึ่งคือกลุ่มผู้สอบที่มีความสามารถในการรายงานด้วยภาษามากกว่าการใช้ภาษา ซึ่งผลการศึกษาพบว่า วิธีทั้งสองวิธีการวิเคราะห์ความลำเอียงของข้อสอบกับตัวแปรที่ใช้ในการวิเคราะห์แตกต่างกัน ผลการวิเคราะห์ความลำเอียงให้ผลสอดคล้องกัน

ซวิกและไทเยอร์ (Zwick and Thayer.1996:187-201) ได้ทำการศึกษาเกี่ยวกับการประเมินความลำเอียงความลำเอียงของข้อสอบที่มีคะแนนเป็นแบบ Polytomous กับข้อมูลที่จำลองขึ้นมาที่มีความคลาดเคลื่อนมาตรฐานในกระบวนการทดสอบสมมติฐานโดยใช้วิธีแมนเทิลและวิธีการกระบวนการมาตรฐานของดูแรนและสมิทท์(1991) ผลการศึกษาพบว่าความคลาดเคลื่อนมาตรฐานจะใช้ในได้ดีในทางปฏิบัติในรูปแบบวิธีของแมนเทิลและความคลาดเคลื่อนอื่นๆในรูปแบบ multinomial แนวโน้มของค่า yield values จะมีในกลุ่มผู้สอบที่มีขนาดเล็ก

คลาเซอร์,นังกาสเตอร์และสวามินาธาน (Clauser,Nungester and Swaminathan.1996 : 453-464) ได้ทำการศึกษาเกี่ยวกับการค้นหาความลำเอียงภายใต้เงื่อนไขตัวแปรคะแนนการสอบและภูมิหลังของการศึกษาของกลุ่มตัวอย่าง โดยใช้วิธี Logistic Regression ในการวิเคราะห์ความลำเอียง โดยกลุ่มตัวอย่างที่ใช้ในการศึกษาค้างนี้เป็นนักศึกษาวิชาแพทยศาสตร์ปีที่ 1 ในโปรแกรมของการฝึกภาคปฏิบัติที่ฝึกปฏิบัติต่างกัน 4 กลุ่ม คือ กลุ่มเภสัชศาสตร์,กลุ่มศัลยแพทย์,กลุ่มกุมารเวชศาสตร์และกลุ่มสูตินารี แบ่งเป็นเพศชาย 1,000 คนและเพศหญิง 1,000 คน ซึ่งผลการศึกษาพบว่าจำนวนข้อสอบที่มีความลำเอียงมีจำนวนน้อยอย่างมีนัยสำคัญทางสถิติภายใต้เงื่อนไขคะแนนรวมและตัวแปรทางด้านพื้นฐานการศึกษากับคะแนนรวมหรืออย่างใดอย่างหนึ่ง

แวง และ เลนนี่ (Wang and Lane.1996:175-199) ได้ทำการศึกษาเกี่ยวกับการค้นหาความลำเอียงของแบบทดสอบภาคปฏิบัติวิชาคณิตศาสตร์โดยใช้วิธีการวิเคราะห์ความลำเอียงแตกต่างกัน 3 วิธี คือ วิธีLogistic Regression (LOG) วิธีLogistic discriminant function analysis (LDFD) และวิธี HW3 กับแบบทดสอบภาคปฏิบัติวิชาคณิตศาสตร์จำนวน 36 ข้อ ซึ่งมีอยู่ 4 ฟอรั่ม (ฟอรั่ม AถึงD) กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นนักเรียนเกรด 6-7 ในปี 1991ถึง 1992 จำนวน 1,782 คน พบว่าวิธีHW3 ค้นพบความลำเอียงของข้อสอบที่มีความลำเอียง 7 ข้อ ที่ระดับนัยสำคัญทางสถิติ.05 และ 2 ข้อ ที่ระดับนัยสำคัญทางสถิติ.005 วิธี Logistic discriminant function analysis (LDFD) ค้นพบข้อสอบที่มีความลำเอียงแบบอันยูนิฟอรั่ม (nonuniform) จากแบบทดสอบจำนวน 1 ข้อ และ 6 ข้อ จากการวิเคราะห์แบบยูนิฟอรั่ม(uniform) ที่ระดับนัยสำคัญทางสถิติ .05 และ 2 ข้อ ในการค้นหาแบบยูนิฟอรั่ม (uniform DIF) ที่ระดับนัยสำคัญทางสถิติ .005 และไม่พบข้อสอบที่มีความลำเอียงในการค้นหาแบบนอนยูนิฟอรั่ม (nonuniform DIF) ส่วนวิธี Logitic Regression (LOG) ในการค้นหาแบบ nonuniform DIF ไม่พบข้อสอบที่มีความลำเอียง และ ในการค้นหาแบบ uniform DIF พบข้อสอบที่มีความลำเอียง 9 ข้อ ที่ระดับนัยสำคัญทางสถิติ .05 และ 3 ข้อ ที่ระดับนัยสำคัญทางสถิติ .005 ในด้านเปอร์เซ็นต์ของการค้นหาความลำเอียงของข้อสอบทั้ง 3 วิธีจากแบบทดสอบ QCAI พบว่ามีความสัมพันธ์กันอยู่ในระดับสูง โดยมีช่วงพิสัยอยู่ระหว่าง 88 % ถึง 97 % ทั้งในการค้นหาแบบuniform และ nonuniform DIF ที่ระดับนัยสำคัญทางสถิติ .05

ซาง,มัสซิโอและรอสอส (Chang,Mazzeo and Roussos.1996:333-353) ได้ทำการศึกษาเกี่ยวกับการค้นหาความลำเอียงของแบบทดสอบที่ตรวจให้คะแนนแบบหลายค่า (Polytomously) โดยใช้วิธีSIBTEST โดยใช้ข้อมูลจำลอง มีวัตถุประสงค์เพื่อเปรียบเทียบวิธี Modified SIBTEST กับวิธี Mantel (MH) และวิธีStandradized mean difference (SMD) ในการศึกษาครั้งนี้ใน 2 ครั้งคือ

การศึกษาครั้งที่ 1 ทำการศึกษาเปรียบเทียบวิธีทั้ง 2 วิธีการ ภายใต้เงื่อนไขของการศึกษาแบบMH และ แบบSMD จะมีการค้นหาความลำเอียงที่ดีหรือไม่ ผลลัพธ์ของการศึกษาพบว่า วิธี Modified SIBTEST ใช้ในทางปฏิบัติที่เหมาะสมมากกว่าวิธีอื่น แต่วิธีMH และ วิธีSMD ใช้ได้ดีกับข้อมูลหายา บ ่

การศึกษาครั้งที่2 ใช้ข้อมูลภายใต้เงื่อนไขวิธีการทดสอบหาความลำเอียงของคะแนนที่สังเกตได้ ของแบบทดสอบที่มีการตรวจให้คะแนนแบบ 0-1 (Dichotomous) ผลลัพธ์ของการศึกษาครั้งที่ 2 พบว่า ภายใต้เงื่อนไขทั้งหมด วิธี Modified SIBTEST ใช้ได้ดีภายใต้ความคลาดเคลื่อนแบบ I มากกว่ากระบวนการอื่น ๆ หรือวิธีการอื่นๆ

ซวิก,ไทเยอร์ และมาสซิโอ (Zwick,Thyer and Mazzeo.1997:321-344) ได้ทำการศึกษาเกี่ยวกับการค้นหาความลำเอียงจากแบบทดสอบที่มีการตรวจให้คะแนนแบบหลายค่า โดยใช้วิธีวิเคราะห์ 5 วิธีด้วยกัน คือSMD 2 วิธี ได้แก่ SMD-H , SMD-M วิธีMH และวิธีSIBTEST 2 วิธี ได้แก่วิธีStandrad SIBTESTและวิธีModified SIBTEST โดยใช้เงื่อนไขแบบทดสอบที่มีการตรวจให้คะแนนแบบ0-1 จำนวน 50 ข้อ และแบบทดสอบที่มีการตรวจให้คะแนนแบบหลายค่า จำนวน 18 ข้อ ใช้กับกลุ่มตัวอย่าง 1,000 คน แบ่งเป็นกลุ่มอ้างอิง 500 คน และ กลุ่มเปรียบเทียบ 500 คน โดยมีการจับคู่ระหว่างข้อสอบแบบให้คะแนน2ค่ากับแบบให้คะแนนหลายค่า โดยใช้โมเดลลอจิตติสส์ 3 พารามิเตอร์ จากข้อมูลจำนวน 75 ข้อ ที่มีค่าอำนาจจำแนกในช่วง.74 ถึง 1.00 ค่าความยากอยู่ระหว่าง-1.95 ถึง 1.95 ค่าการเดาอยู่ที่ .15 และสถานการณ์ที่จำลองขึ้นมากับกลุ่มตัวอย่างทั้ง2กลุ่มที่มีการแจกแจงดัชนีการปฏิบัติที่ดีจากSMD เมื่อความแตกต่างของค่าเฉลี่ยของกลุ่มที่มีความคลาดเคลื่อนมาตรฐานเป็น 1 วิธีการ Modified SIBTEST เป็นการวัดผลกระทบของขนาดกลุ่มตัวอย่างค่อนข้างดี จาก 5 วิธีการ ในทางปฏิบัติจะไม่สามารถมองเห็นความแตกต่างของกลุ่มย่อยทั้ง 2 กลุ่ม จะมีการแจกแจงเบ้ไปด้านใดด้านหนึ่ง เมื่อกลุ่มมีการแจกแจงที่แตกต่างกันและข้อสอบที่ศึกษามีค่าอำนาจจำแนกสูง วิธีการ SIBTEST จะแสดงให้เห็นดีกว่าวิธีอื่นในรูปแบบความคลาดเคลื่อนประเภทที่ 1 กว่าวิธี SMD และ วิธีแมนเทลจากแบบทดสอบตอบสั้นๆ ค่าอำนาจจำแนกจาก 5 วิธี จะไม่เหมือนกันเป็นความแตกต่างในการวิเคราะห์หาความลำเอียงและองค์ประกอบอื่นๆ วิธีการที่ใช้กันบ่อยๆ ในการประยุกต์ใช้หาความลำเอียงจากแบบทดสอบแบบPolytomousเมื่อมีการสอบตัวแปรที่คือการจับคู่กันระหว่างการสอบกับแบบฝึกหัดอัตราความคลาดเคลื่อนประเภทที่1 จากวิธีแมนเทลและวิธี SMD จะเกี่ยวกับภายใต้เงื่อนไขของเกณฑ์

วิธีการปัจจุบันของ SIBTEST ไม่สามารถใช้กับการจับคู่ที่ง่าย ๆ ได้ นั่นคือ ใช้กับแบบทดสอบแบบ ถูก-ผิดไม่ได้ งานวิจัยในปัจจุบันจึงไม่นิยมกัน

อัลลาลอฟ,แฮมมิลตัน และ ซีรีซี (Allalovf,Hambleton and Sireci.1999:185-198) ได้ทำการศึกษาถึงสาเหตุของความลำเอียงในแบบทดสอบแปลความทางภาษารวมทั้งเสนอแนะรูปแบบของข้อสอบที่เป็นสาเหตุของความลำเอียงและสามารถที่จะเข้าใจถึงสาเหตุของความลำเอียงและเป็นแบบแผนในการเขียนข้อสอบที่ปราศจากความลำเอียง โดยวัตถุประสงค์ของการศึกษานี้จะศึกษาถึงสาเหตุของความลำเอียงของข้อสอบแปลความทางภาษา โดยแบ่งเป็น 2 ส่วน คือ ส่วนแรกศึกษาเกี่ยวกับรูปแบบของข้อสอบทางภาษากับการแสดงถึงความลำเอียงของข้อสอบ เมื่อมีการแปลความจากภาษาหนึ่งไปยังอีกภาษาหนึ่ง ส่วนที่สองเป็นเรื่องเกี่ยวกับการค้นหาสาเหตุของความลำเอียงจากวิธีการวิเคราะห์ความลำเอียงซึ่งมี 2 วิธี คือวิธีแรกเป็นการวิเคราะห์ความลำเอียงโดยดูจากค่าสถิติ (จากกลุ่มที่ทำข้อสอบและรูปแบบของข้อสอบ) และวิธีที่สองการวิเคราะห์โดยนักแปลภาษาในด้านรูปแบบและเนื้อหา จากข้อสอบที่มีความลำเอียง

ลักษณะที่สำคัญโดยเฉพาะ 3 อย่าง คือ (1) การศึกษาขนาดกลุ่มตัวอย่างที่ค่อนข้างขนาดใหญ่ที่มีความสามารถปกติ (2) ข้อสอบจะแสดงให้เห็นความลำเอียงมากในแบบทดสอบแปลความข้ามภาษา และ (3) กลุ่มผู้สอบขนาดเล็กจะมีความสามารถที่แตกต่างกัน เครื่องมือที่ใช้เป็นแบบทดสอบแปลความทางภาษาจากภาษาฮีบรู (Hebrew)ไปเป็นภาษารัสเซีย ซึ่งมีด้านย่อยๆ 4 ด้าน คือด้านเหตุผล จำนวน 26 ข้อ ด้านความเข้าใจประโยค จำนวน 33 ข้อ ด้านตรรกศาสตร์ จำนวน 36 ข้อ และด้านความเข้าใจในการอ่าน จำนวน 30 ข้อ รวมทั้งหมด 125 ข้อ ซึ่งวิธีการวิเคราะห์ใช้วิธีของแมนเทล-แฮนส์เซล โดยคำนวณจากโปรแกรม DICHODIF ซึ่งผลการศึกษาพบว่า ข้อสอบแปลความทางภาษา 34% มีความลำเอียงระหว่างภาษา โดยแบบทดสอบด้านเหตุผลจะแสดงให้เห็นถึงความลำเอียงถึง 65% ในแบบฝึกหัดการพูดภาษารัสเซีย ข้อสอบความเข้าใจประโยคจะมีปัญหา (มีความลำเอียงถึง 45%) เหตุผลหลักหรือสาเหตุของความลำเอียง ซึ่งได้จากระบุของนักแปลจำนวน 5 คน และนักวิจัยเกี่ยวกับการพูดภาษาฮีบรู จำนวน 3 คน รวม ทั้งหมด 8 คน พบว่าเกิดจาก การเปลี่ยนความยากของคำ , การเปลี่ยนรูปแบบของข้อสอบ , ความแตกต่างของวัฒนธรรม และ การเปลี่ยนในด้านเนื้อหา

จากการศึกษาผลการวิจัยข้างต้นจะเห็นได้ว่าการศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบมีหลายวิธี บางวิธีเป็นวิธีที่ยึดทฤษฎีการวัดผลแบบดั้งเดิม เช่น วิธีแปลงค่าความยาก วิธีค่าอำนาจจำแนก วิธีโคสแควร์ วิธีวิเคราะห์ความแปรปรวน เป็นต้นบางวิธียึดทฤษฎีการตอบข้อคำถาม เช่น วิธีโค้งลักษณะข้อสอบ เป็นต้น และบางวิธีก็ใช้สถิติแบบนอนพาราเมตริก เช่น วิธีแมนเทล-แฮนส์เซล,วิธีชิปเทสท์ ,วิธีGMH,วิธี HW3 , วิธี HW1,วิธี Modifed SIBTEST และวิธี Polytomous STND เป็นต้น จากวิธีต่าง ๆ ดังกล่าว บางวิธีเป็นวิธีที่ให้ผลการวิเคราะห์ที่ได้ถูกต้อง แม่นยำ แต่ไม่เหมาะที่จะนำมาใช้ในทางปฏิบัติเพราะต้องใช้กลุ่มตัวอย่างขนาดใหญ่

การคำนวณและการแปลผลซับซ้อน ยุ่งยาก ต้องเสียค่าใช้จ่ายสูง ในขณะที่บางวิธีเป็นวิธีที่ใช้กลุ่มตัวอย่างไม่มาก การคำนวณและการแปลผลง่าย ไม่ซับซ้อน แต่ให้ผลการวิเคราะห์ที่มีความถูกต้องน้อยกว่าวิธีอื่น ๆ และบางวิธีใช้กลุ่มตัวอย่างไม่มากการคำนวณและการแปลผลง่าย เหมาะสมที่จะนำมาใช้ในทางปฏิบัติได้ และให้ผลการวิเคราะห์ที่มีความถูกต้อง สำหรับตัวแปรที่ใช้ในการวิเคราะห์ความลำเอียงที่นิยมทำการศึกษามีดังนี้ ภาษา เพศ วัฒนธรรม ศาสนา สภาพทางภูมิศาสตร์ เป็นต้น นอกจากนี้ยังมีการศึกษาถึงสาเหตุที่เป็นแหล่งที่ก่อให้เกิดความลำเอียงโดยให้ผู้เชี่ยวชาญเฉพาะสาขาเป็นผู้ระบุความลำเอียงของแบบทดสอบ เป็นต้น ซึ่งผลการวิจัยที่ผ่านมา ยังไม่มีผู้ศึกษาผลการวิเคราะห์ความลำเอียงของแบบทดสอบ ด้วยวิธีค่าอำนาจจำแนก กล่าวคือวิธีHW 3 , วิธีไบซีเรียลและวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด ที่ระบุว่าเป็นวิธีที่ใช้ได้ดีในทางปฏิบัติแต่วิธีใดที่จะให้ผลการวิเคราะห์ได้ถูกต้องที่สุดและเหมาะสมที่สุด ยังไม่ปรากฏว่ามีผู้ศึกษา ผู้วิจัยจึงทำการสนใจศึกษาในครั้งนี้

บทที่ 3

วิธีดำเนินการศึกษาค้นคว้า

ในการศึกษาค้นคว้าครั้งนี้ ผู้วิจัยได้ดำเนินการตามขั้นตอนดังนี้

1. การกำหนดประชากรและการสุ่มกลุ่มตัวอย่าง
2. การสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า
3. การเก็บรวบรวมข้อมูล
4. การจัดกระทำและการวิเคราะห์ข้อมูล

การกำหนดประชากรและการเลือกสุ่มกลุ่มตัวอย่าง

ประชากร

ประชากรที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2543 ของโรงเรียนในสังกัดกรมสามัญศึกษา ส่วนกลาง กลุ่มที่ 4 จำนวน 14 โรงเรียน จำนวน 174 ห้องเรียน มีจำนวนนักเรียน 7,095 คน จำแนกตามขนาดของโรงเรียนมี 2 ขนาด คือ โรงเรียนขนาดใหญ่พิเศษ จำนวน 11 โรงเรียน มีห้องเรียนจำนวน 135 ห้องเรียน มีนักเรียนจำนวน 5,805 คน แยกเป็นเพศชาย 2,835 คน เพศหญิง 2,970 คน และ ขนาดใหญ่ 3 โรงเรียน มีห้องเรียนจำนวน 30 ห้องเรียน มีนักเรียนจำนวน 1,290 คน แยกเป็นเพศชาย 600 คน เพศหญิง 690 คน

การเลือกกลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2543 ของโรงเรียนในสังกัดกรมสามัญศึกษา ส่วนกลาง กลุ่มที่ 4 จำนวน 2,800 คน ซึ่งมีความเชื่อมั่น 99 เปอร์เซนต์ ($\alpha=.01$) (ส่วน สายยศและอังคณา สายยศ.2540 400 ; อ้างอิงมาจาก Yamane.1976) โดยสุ่มกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์ข้อสอบที่มีความลำเอียงจำนวน 1,000 คน เป็นนักเรียนที่มีเพศต่างกัน คือ นักเรียนเพศชายจำนวน 250 และ 500 คน ตามลำดับ และนักเรียนเพศหญิงจำนวน 250 และ 500 คน ตามลำดับ สำหรับกลุ่มตัวอย่างที่เหลือใช้เป็นกลุ่มตัวอย่างในการคำนวณค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว จำนวน 9 ฉบับๆละ 200 คน รวมเป็น 1,800 คน โดยการใช้การสุ่มกลุ่มตัวอย่างแบบสองขั้นตอน (Two-Stage Random Sampling) ซึ่งมีลำดับขั้นตอนในการสุ่ม ดังนี้

ขั้นที่ 1 สุ่มแบบแบ่งชั้น(Stratified Random Sampling) มีขนาดโรงเรียนเป็นชั้น(Strata) และมีโรงเรียนเป็นหน่วยการสุ่ม (Sampling Unit) จากแต่ละขนาดของโรงเรียนด้วยสัดส่วน 2 : 1 ได้โรงเรียนขนาดพิเศษ 5 โรงเรียน ขนาดใหญ่ 2 โรงเรียน

ขั้นที่ 2 สุ่มห้องเรียนจากโรงเรียนที่สุ่มไว้ในแต่ละขนาด โดยวิธีการสุ่มอย่างง่าย (Simple Random Sampling) และมีห้องเรียนเป็นหน่วยการสุ่ม (Sampling Unit) ในสัดส่วน 2:1 ได้โรงเรียนขนาดพิเศษ 42 ห้องเรียนมีนักเรียน 1,964 คน แยกเป็นเพศชายจำนวน 965 คน เพศหญิงจำนวน 999 คน โรงเรียนขนาดใหญ่ 14 ห้องเรียน มีนักเรียนทั้งหมด 1,036 คน แยกเป็นเพศชาย 502 คน เพศหญิง 534 คน รวมได้จำนวนห้องเรียนทั้งหมด 56 ห้องเรียน นักเรียนทั้งหมด 2,800 คน ดังแสดงในตาราง 3

ตาราง 3 จำนวนนักเรียนในกลุ่มตัวอย่างที่ใช้ในการวิจัย จำแนกตามขนาดโรงเรียน

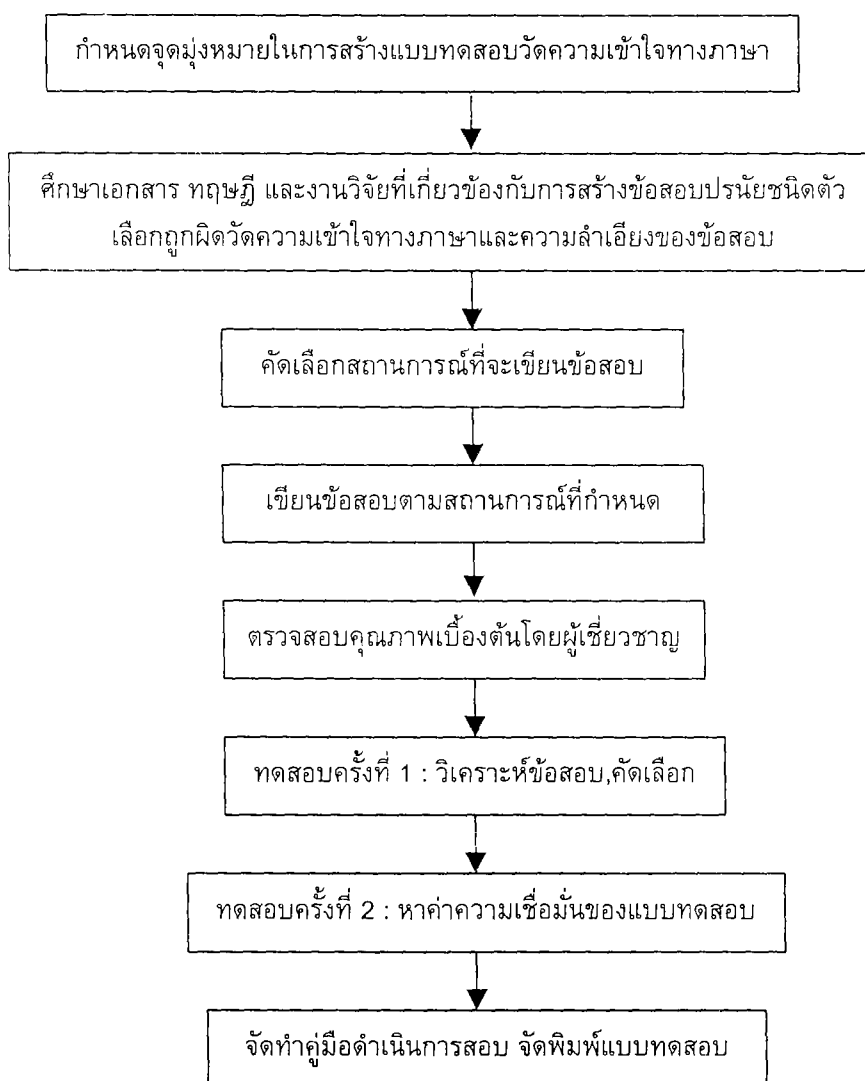
ขนาด โรงเรียน	โรงเรียน	จำนวนนักเรียนที่ใช้ในการวิจัย	
		เปรียบเทียบจำนวน ข้อสอบที่มีความลำเอียง (ห้อง)	เปรียบเทียบค่าความเชื่อมั่นหลัง จากคัดเลือกข้อสอบที่มีความ ลำเอียงออกแล้ว (ห้อง)
ใหญ่พิเศษ	พระโขนงพิทยาลัย	263 (6)	193 (4)
	วชิรธรรมสาริต	360 (8)	-
	สิริรัตนาร	243 (6)	-
	สุรศักดิ์มนตรี	-	527 (9)
	ศรีพฤฒา	-	421 (7)
ใหญ่	กุนนทีรุทธารามวิทยาคม	-	471 (8)
	ประชากรสมุทรอุปถัมภ์	144 (4)	188 (4)
รวม		1,000 (24)	1,800 (32)

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้ คือ แบบทดสอบวัดความเข้าใจทางภาษาสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2 เป็นแบบทดสอบเลือกตอบแบบตัวเลือกถูกผิด ชนิด 5 ตัวเลือก จำนวน 40 ข้อ ที่ผู้วิจัยสร้างเอง

การสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า

ในการสร้างเครื่องมือสำหรับการศึกษาค้นคว้า ผู้วิจัยได้ดำเนินการสร้างตามขั้นตอนดังภาพประกอบ 6



ภาพประกอบ 6 ลำดับขั้นตอนในการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า

จากแผนภูมิ ลำดับขั้นตอนในการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า มีรายละเอียดดังนี้

1. กำหนดจุดมุ่งหมายในการสร้างแบบทดสอบวัดความเข้าใจทางภาษา โดยสร้างขึ้นเพื่อใช้ความเข้าใจทางภาษาเพื่อใช้ในการเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบในด้านจำนวนข้อที่มีความลำเอียงที่ใช้วิธีวิเคราะห์และขนาดกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบต่างกัน โดยใช้เพศของผู้สอบเป็นตัวแปรที่ใช้ในการวิเคราะห์

2. ศึกษาเอกสาร เอกสาร ทฤษฎี และงานวิจัยที่เกี่ยวข้องกับวิธีการสร้างแบบทดสอบเลือกแบบตัวเลือกถูกผิดวัดความเข้าใจทางภาษาจากหนังสือวัดผลการเรียนรู้ของล้วน สายยศ และอังคณา สายยศ ,วารสารการวัดผลการศึกษาและวิธีวิเคราะห์ความลำเอียง งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของแบบทดสอบ จากวารสารทั้งในประเทศและต่างประเทศ

3. คัดเลือกสถานการณ์ที่จะนำมาเขียนข้อสอบวัดความเข้าใจทางภาษาสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2 จากวารสารและเอกสารต่าง ๆ โดยสถานการณ์ที่จะนำมาสร้างจะเป็นบทร้องกรอง เช่น บทสนทนา , บทความ , ข่าว , คำโฆษณา เป็นต้น และ บทร้อยแก้ว เช่น กาพย์ , กลอน , โคลง , ฉันท์ เป็นต้น มาจำนวน 80 สถานการณ์และนำสถานการณ์ที่เลือกไว้ไปให้ครูผู้สอนรายวิชาภาษาไทยในระดับชั้นมัธยมศึกษาปีที่ 2 จำนวน 3 คน เป็นผู้คัดเลือกสถานการณ์โดยพิจารณาความเหมาะสมของสถานการณ์ที่จะนำมาเขียนข้อสอบกับระดับความสามารถของนักเรียนชั้นมัธยมศึกษาปีที่ 2 เป็นรายสถานการณ์ และสถานการณ์ที่ถือว่าเป็นสถานการณ์ที่เหมาะสมกับระดับความสามารถของนักเรียนนั้นจะต้องได้รับการเห็นชอบจากครูผู้สอนจำนวน 2 ท่านขึ้นไป จากการคัดเลือกสถานการณ์ของครู พบว่า มีสถานการณ์ที่ครูผู้สอนเห็นว่ามีความเหมาะสมกับระดับความสามารถของนักเรียน มีจำนวนทั้งหมด 64 สถานการณ์ เนื่องจากได้จำนวนสถานการณ์มากกว่าที่กำหนดไว้ ผู้วิจัยจึงได้คัดเลือกให้เหลือ 20 สถานการณ์ โดยยึดสถานการณ์ที่ครูผู้สอนเห็นสอดคล้องกันทั้ง 3 คน เพื่อนำไปเขียนข้อสอบวัดความเข้าใจทางภาษาตามสถานการณ์ต่อไป

4. เขียนข้อสอบวัดความเข้าใจทางภาษาเป็นแบบตัวเลือกถูกผิดชนิด 5 ตัวเลือก ตามสถานการณ์ที่กำหนด สถานการณ์ละ 4 – 6 ข้อ รวมจำนวนทั้งหมด 87 ข้อ

5. นำแบบทดสอบที่สร้างขึ้นให้ผู้เชี่ยวชาญ ซึ่งมีความรู้ในเนื้อหาและการเขียนข้อสอบในวิชาภาษาไทยและมีประสบการณ์สอนในวิชาภาษาไทยระดับมัธยมศึกษาปีที่ 2 มาไม่น้อยกว่า 5 ปี จำนวน 3 ท่านและนักวัดผลการศึกษา จำนวน 2 ท่าน เป็นผู้ตรวจสอบความเที่ยงตรงเชิงน้อยเนื้อหา (Content Validity) ของแบบทดสอบเป็นรายข้อ โดยพิจารณาความสอดคล้องระหว่างข้อสอบแต่ละข้อกับเนื้อหาและสถานการณ์ที่เขียนขึ้น (IOC) และข้อสอบที่ถือว่าวัดได้ตรงตามสถานการณ์จะต้องมีค่า IOC ตั้งแต่ .50 ขึ้นไป จึงถือว่าข้อสอบข้อนั้นใช้ได้ จากการตรวจสอบความเที่ยงตรงเชิงเนื้อหาของผู้เชี่ยวชาญ พบว่า มีข้อสอบที่มีค่า IOC ตามเกณฑ์ จำนวน

83 ข้อ ส่วนอีก 4 ข้อ ที่เหลือนั้น ผู้วิจัยได้ทำการแก้ไขปรับปรุงด้านภาษาที่ใช้ให้เหมาะสมกับระดับความสามารถของนักเรียน หลังจากนั้นก็นำข้อสอบไปให้ประธานผู้ควบคุมปริญาานิพนธ์พิจารณาตัดสินอีกครั้งหนึ่ง และเนื่องจากแบบทดสอบมีจำนวนข้อมาก เมื่อนำไปทดลองเพื่อหาคุณภาพขั้นต้นกับกลุ่มตัวอย่าง อาจจะทำให้นักเรียนที่ทำข้อสอบเกิดความเบื่อหน่ายและไม่ตั้งใจในการทำข้อสอบ ดังนั้นผู้วิจัยจึงทำการแบ่งออกเป็น 2 ฉบับ คือ ฉบับที่ 1 จำนวน 42 ข้อ และฉบับที่ 2 จำนวน 45 ข้อ

6. นำแบบทดสอบจำนวน 2 ฉบับ ที่ผู้เชี่ยวชาญพิจารณาแล้วไปทดสอบครั้งที่ 1 กับกลุ่มตัวอย่างจำนวน 458 คน โดยสุ่มจากประชากรที่เหลือจากกลุ่มตัวอย่างที่ใช้ในการศึกษา โดยวิธีการสุ่มอย่างง่าย (Simple Random Sampling) ใช้ห้องเรียนเป็นหน่วยการสุ่ม (Sampling Unit) ได้โรงเรียนจันทร์หุ่นบำเพ็ญ จำนวน 6 ห้องเรียน มีนักเรียนจำนวน 270 คน และโรงเรียนมัธยมวัดธาตุทอง จำนวน 5 ห้องเรียน มีนักเรียนจำนวน 188 คน แล้วนำผลการทดสอบครั้งที่ 1 มาตรวจให้คะแนนโดยให้คะแนนทุกตัวเลือก แล้วนำผลการสอบไปวิเคราะห์ค่าความยากและค่าอำนาจจำแนกของข้อสอบ ซึ่งวิเคราะห์เป็นรายข้อใช้วิธีวิธีนี้และซาเบอร์ แล้วทำการพิจารณาคัดเลือกข้อสอบที่มีคุณภาพ คือข้อสอบที่มีค่าความยากอยู่ระหว่าง 0.20–0.80 และค่าอำนาจจำแนกตั้งแต่ 0.20 ขึ้นไปได้ โดยคัดเลือกตามสถานการณ์ซึ่งให้ได้อย่างน้อยสถานการณ์ละ 2 ข้อ ให้ได้ 40 ข้อ ซึ่งได้รายละเอียดดังนี้

ฉบับที่ 1 มี 9 สถานการณ์ จำนวน 42 ข้อ นำไปทดสอบกับกลุ่มตัวอย่างจำนวน 235 คน ซึ่งมีค่าความยากระหว่าง 0.16–0.57 และค่าอำนาจจำแนกระหว่าง 0.01–0.37 และคัดเลือกไว้ 8 สถานการณ์ มีจำนวน 19 ข้อ ซึ่งมีค่าความยาก ตั้งแต่ 0.23 – 0.52 และค่าอำนาจจำแนกตั้งแต่ 0.20–0.37

ฉบับที่ 2 มี 11 สถานการณ์ จำนวน 45 ข้อ นำไปทดสอบกับกลุ่มตัวอย่างจำนวน 223 คน ซึ่งมีค่าความยากระหว่าง 0.17–0.52 และค่าอำนาจจำแนกระหว่าง 0.00–0.53 และคัดเลือกไว้ 8 สถานการณ์ มีจำนวน 21 ข้อ ซึ่งมีค่าความยาก ตั้งแต่ 0.20–0.52 และค่าอำนาจจำแนกตั้งแต่ 0.21–0.53

และทำการกำหนดเวลาในการทำแบบทดสอบของนักเรียน โดยยึดเวลาที่นักเรียนส่วนใหญ่ทำเสร็จ 80 เปอร์เซ็นต์ขึ้นไปเป็นเกณฑ์ในการกำหนดเวลา ซึ่งได้ช่วงเวลาที่เหมาะสมเท่ากับ 60 นาที แล้วข้อสอบที่คัดเลือกไว้ ทั้ง 2 ฉบับ มาจัดเรียงข้อสอบใหม่โดยการสุ่ม ให้เป็นฉบับเดียว เพื่อนำไปทดสอบหาค่าความเชื่อมั่นต่อไป

7. นำแบบทดสอบที่คัดเลือกไว้ จำนวน 40 ข้อ ไปทดสอบครั้งที่ 2 เพื่อหาค่าความเชื่อมั่นของเครื่องมือ โดยใช้สูตรความเชื่อมั่นแบบฮอยท์ (Hoyt's ANOVA procedure) กับกลุ่มตัวอย่างจำนวน 509 คนที่เลือกมาโดยสุ่มจากประชากรที่เหลือจากกลุ่มตัวอย่างที่ใช้ในการศึกษา โดยวิธีการสุ่มอย่างง่าย (Simple Random Sampling) ใช้ห้องเรียนเป็นหน่วยการสุ่ม (Sampling Unit)

ได้โรงเรียนปทุมคงคา จำนวน 6 ห้องเรียน มีนักเรียนจำนวน 275 คน และโรงเรียนพิบูลประชาสรรค์ จำนวน 5 ห้องเรียน มีนักเรียนจำนวน 234 คน และทำการคำนวณหาค่าความเชื่อมั่นของแบบทดสอบทั้งฉบับมีค่าเท่ากับ 0.8860

8. นำข้อสอบจำนวน 40 ข้อ มาจัดพิมพ์เป็นรูปเล่มและจัดทำคู่มือดำเนินการสอบ เพื่อใช้ในการทำวิจัยต่อไป

ลักษณะของเครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้ คือ แบบทดสอบวัดความเข้าใจทางภาษาเป็นแบบทดสอบเลือกตอบถูกผิดชนิด 5 ตัวเลือก จำนวน 40 ข้อ ที่ผู้วิจัยสร้างขึ้น ซึ่งมีลักษณะดังนี้

ตัวอย่าง

คำชี้แจง แบบทดสอบนี้เป็นแบบทดสอบเลือกตอบแบบตัวเลือกถูกผิดชนิด 5 ตัวเลือก วัดความเข้าใจทางภาษา โดยในแต่ละข้อจะกำหนดสถานการณ์มาให้ให้นักเรียนอ่านก่อน นักเรียนจะต้องพิจารณาเลศนัยใจความหรือข้อความให้ดี ซึ่งตัวเลือกในแต่ละข้อนั้นมี **ตัวเลือกถูกและตัวเลือกผิดหลายตัว** ดังนั้นนักเรียนจะต้องพิจารณาหาตัวเลือกถูกและตัวเลือกผิดให้ดีแล้วไปตอบในกระดาษคำตอบ โดยทำเครื่องหมายถูก (✓) ในช่องที่ตรงกับตัวเลือกที่เห็นว่าถูก และทำเครื่องหมายกากบาท (✕) ในช่องที่ตรงกับตัวเลือกที่เห็นว่าผิด ดังตัวอย่างข้อ (๐) ดังนี้

ให้นักเรียนอ่านข้อความต่อไปนี้ แล้วตอบคำถามข้อ ๐

“ไม่ขอรับทรัพย์สินจากใครหมด ทั้งยศศักดิ์สติลาภไพศาล
ไม่หวังให้ใครหนุนบุญบันดาล ชอบผลงานทุกอย่างสร้างด้วยตัว”

(๐). คำพูดดังกล่าวตรงกับสำนวนใด

- ก. ชักแม่น้ำทั้งห้า
- ข. กระชอกโอ๊กฮาก
- ค. กำปั้นทุบดิน
- ง. ขวานผ่าซาก
- จ. กลับกลอก

ในตัวอย่างข้อ ๐ จะเห็นว่าคำพูดดังกล่าวตรงกับ

สำนวน ในตัวเลือกข้อ ค และ ง ดังนั้นข้อ ค และ ง จึงเป็นคำตอบที่ถูกต้องส่วนตัวเลือกข้อ ก ,ข และ จ เป็นคำตอบที่ผิด จึงไปตอบในกระดาษคำตอบ ดังนี้

	ก	ข	ค	ง	จ
(๐)	✕	✕	✓	✓	✕

การตรวจให้คะแนน

การตรวจให้คะแนนเป็นแบบคะแนนทุกตัวเลือก โดยพิจารณาจากคำตอบของผู้สอบในตัวเลือกตัวใดตรงกับเฉลยให้ตัวเลือกตัวละ 1 คะแนน ไม่ว่าจะทำเครื่องหมายถูกหรือเครื่องหมายผิดก็ตาม และคำตอบในตัวเลือกใดไม่ตรงกับเฉลยเลยหรือไม่ตอบ ให้ 0 คะแนนในตัวเลือกนั้น

วิธีการเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลเพื่อทำการวิจัยในครั้งนี้ ผู้วิจัยได้ดำเนินการตามลำดับขั้นตอน ดังนี้

1. นำหนังสือจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ติดต่อขออนุญาตจากผู้บริหารโรงเรียนที่ใช้เป็นกลุ่มตัวอย่างและประสานงานกับอาจารย์ผู้เกี่ยวข้องในการดำเนินการสอบเพื่อแจ้งวัตถุประสงค์พร้อมทั้งกำหนดวัน – เวลากล่าวคือ อยู่ในช่วงเดือน ธันวาคม 2543 ถึง มกราคม 2544 กับนักเรียนในระดับชั้นมัธยมศึกษาปีที่ 2

2. จัดเตรียมอุปกรณ์ ได้แก่ แบบทดสอบ กระดาษคำตอบ ให้มากกว่าจำนวนผู้เข้าสอบ

3. ไปถึงสถานที่สอบในวันเวลาที่ทำการสอบ ก่อนกำหนดเวลาสอบ

4. ชี้แจงวัตถุประสงค์ของการสอบและพูดโน้มน้าวจิตใจผู้สอบให้มีความกระตือรือร้นที่จะทำข้อสอบด้วยความตั้งใจอย่างเต็มความสามารถและอธิบายคำชี้แจงในการทำข้อสอบให้ผู้สอบทุกคนเข้าใจและทำความเข้าใจกับผู้ช่วยดำเนินการสอบให้เข้าใจตรงกัน

5. นำแบบทดสอบไปทดสอบกับกลุ่มตัวอย่างครั้งที่ 1 จำนวน 458 คน เพื่อนำมาวิเคราะห์ค่าความยากและค่าอำนาจจำแนกด้วยวิธีวิธีนี้และซาเบอร์และทำการคัดเลือกข้อสอบที่อยู่ในเกณฑ์ที่ตั้งไว้

6. นำแบบทดสอบไปทดสอบกับกลุ่มตัวอย่างครั้งที่ 2 กับกลุ่มตัวอย่างจำนวน 500 คน แล้วทำการวิเคราะห์หาค่าความเชื่อมั่นของแบบทดสอบด้วยวิธีของฮอยท์ (Hoyt's ANOVA procedure)

7. นำแบบทดสอบที่มีคุณภาพจำนวน 40 ข้อ ไปทดสอบกับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 2 ที่เป็นกลุ่มตัวอย่าง จำนวน 2,800 คน ซึ่งแบ่งเป็น 2 กลุ่ม ดังนี้

7.1. กลุ่มตัวอย่างที่ใช้ในการศึกษาผลการวิเคราะห์ความลำเอียง ด้านจำนวนข้อสอบที่มีความลำเอียง จำนวน 1,000 คน เป็นนักเรียนที่มีเพศแตกต่างกัน คือ นักเรียนเพศชาย จำนวน 500 คน และเป็นนักเรียนเพศหญิงจำนวน 500 คน

7.2. กลุ่มตัวอย่างที่ใช้ในการศึกษาผลวิเคราะห์ความลำเอียงด้านความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว จำนวน 1,800 คน

เพื่อไปทำการวิเคราะห์ข้อมูลที่จะนำไปสู่การตอบจุดมุ่งหมายและสมมติฐานของการศึกษาค้นคว้าที่ตั้งไว้

การจัดกระทำและการวิเคราะห์ข้อมูล

1. หลังจากนำแบบทดสอบที่มีคุณภาพจำนวน 40 ข้อ ไปทดสอบกับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 2 ที่เป็นกลุ่มตัวอย่าง จำนวน 1,000 คน แบ่งเป็นนักเรียนเพศชาย 500 คน เพศหญิง 500 คนและทำการตรวจให้คะแนนของแบบทดสอบวัดความเข้าใจทางภาษา โดยตรวจให้คะแนนใช้วิธีการตรวจแบบให้คะแนนทุกตัวเลือก

2. ทำการสุ่มกลุ่มผู้สอบในกลุ่มอ้างอิง (เพศชาย) และกลุ่มเปรียบเทียบ (เพศหญิง) แล้วทำการแบ่งผู้สอบออกเป็น 4 กลุ่ม คือ กลุ่มผู้สอบที่เป็นเพศชาย 2 กลุ่ม คือ จำนวน 250 , 500 คน ตามลำดับ และกลุ่มผู้สอบเพศหญิง 2 กลุ่ม คือ จำนวน 250 , 500 คน ตามลำดับ

3. ทำการวิเคราะห์ความลำเอียงตามวิธีการวิเคราะห์ความลำเอียงทั้ง 3 วิธี เพื่อศึกษาจำนวนข้อสอบที่มีความลำเอียง จำแนกในแต่ละวิธีวิเคราะห์และลักษณะกลุ่มผู้สอบ เพื่อตอบคำถามในจุดมุ่งหมายข้อ 1 และ 2 แล้วทำการทดสอบสมมติฐานข้อที่ 1 และ 2

4. นำผลการวิเคราะห์ความลำเอียงตามข้อ 3 มาคัดเลือกข้อที่ลำเอียงออกแล้ว ทำการหาค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว ในแต่ละวิธีวิเคราะห์ความลำเอียง ตามขนาดกลุ่มผู้สอบ กลุ่มละ 200 คน รวมจำนวนทั้งหมด 1,800 คน โดยใช้วิธีการหาค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับ (Split-Half) เพื่อตอบคำถามในจุดมุ่งหมายข้อ 3 และ 4 แล้วทำการทดสอบสมมติฐานข้อที่ 3 และ 4

5. สถิติที่ใช้ในการวิเคราะห์ข้อมูล

5.1 ค่าสถิติที่ใช้ในการหาคุณภาพของเครื่องมือ

5.1.1 สถิติพื้นฐาน คือ ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน

5.1.2 ค่าความยากของแบบทดสอบ โดยใช้เทคนิค 25% ของวิทนีและซาเบอร์

(Whitney and Sabers.1970 ; อ้างอิงมาจาก ล้วน สายยศและอังคณา สายยศ.2539:199–200)

$$P_E = \frac{S_U + S_L - (2NX_{\min})}{2N(X_{\max} - X_{\min})}$$

เมื่อ	P_E	แทน ดัชนีค่าความยากง่าย
	S_U	แทน ผลรวมของคะแนนกลุ่มเก่ง
	S_L	แทน ผลรวมของคะแนนกลุ่มอ่อน
	N	แทน จำนวนผู้เข้าสอบของกลุ่มเก่งหรือกลุ่มอ่อน
	$X_{\max}$	แทน คะแนนที่นักเรียนทำได้สูงสุด
	$X_{\min}$	แทน คะแนนที่นักเรียนทำได้ต่ำสุด

5.1.3 ค่าอำนาจจำแนกของแบบทดสอบ ของวิทนีและซาเบอร์ (Whitney and Sabers.1970 ; อ้างอิงมาจาก ล้วน สายยศและอังคณา สายยศ. 2539:201)

$$D = \frac{S_U - S_L}{2N(X_{\max} - X_{\min})}$$

เมื่อ	D	แทน ดัชนีค่าความยากง่าย
	S_U	แทน ผลรวมของคะแนนกลุ่มเก่ง
	S_L	แทน ผลรวมของคะแนนกลุ่มอ่อน
	N	แทน จำนวนผู้เข้าสอบของกลุ่มเก่งหรือกลุ่มอ่อน
	$X_{\max}$	แทน คะแนนที่นักเรียนทำได้สูงสุด
	$X_{\min}$	แทน คะแนนที่นักเรียนทำได้ต่ำสุด

5.1.4 ค่าความเชื่อมั่นของแบบทดสอบด้วยวิธีของฮอยท์ (ล้วน สายยศและอังคณา สายยศ. 2539:220)

$$r_u = 1 - \frac{MS_E}{MS_p}$$

เมื่อ	r_u	แทน ค่าความเชื่อมั่น
	MS_E	แทน คะแนนความแปรปรวนของความคลาดเคลื่อน
	MS_p	แทน คะแนนความแปรปรวนระหว่างคน

5.2 สถิติที่ใช้ในการวิจัย

5.2.1 สถิติที่ใช้ในการวิเคราะห์ความลำเอียง

1. วิธีค่าอำนาจจำแนกแบบวิธี HW 3 มีขั้นตอนการคำนวณ ดังนี้ (Welch and Miller.1995:166–167 , Wang and Lane.1996:181–182 ; Citing Welch and Hoover.1993)

1. 1 คำนวณค่า t – test ของข้อสอบในแต่ละระดับการตอบ ซึ่งผู้วิจัยแบ่งออกเป็น 3 ระดับ คือ ระดับที่ 1 ช่วงคะแนน 0 –1 คะแนน ระดับที่ 2 ช่วงคะแนน 2 – 3 คะแนน และระดับที่ 3 ช่วงคะแนน 4 –5 คะแนน โดยคำนวณตามสูตร ดังนี้

$$t_k = \frac{\bar{X}_{Fk} - \bar{X}_{Rk}}{\sqrt{\frac{S_{Fk}^2 n_{Fk} + S_{Rk}^2 n_{Rk}}{m_k} \left\{ \frac{1}{n_{Fk}} + \frac{1}{n_{Rk}} \right\}}}$$

เมื่อ	t_k	แทน ค่าที่บ่งบอกถึงค่าดัชนีความลำเอียงในระดับที่ k
	n_{Fk}	แทน จำนวนของกลุ่มเปรียบเทียบในระดับที่ k
	n_{Rk}	แทน จำนวนของกลุ่มอ้างอิงในระดับที่ k
	$\overline{X}_{Fk}$	แทน ค่าเฉลี่ยของกลุ่มเปรียบเทียบในระดับที่ k
	$\overline{X}_{Rk}$	แทน ค่าเฉลี่ยของกลุ่มอ้างอิงในระดับที่ k
	S^2_{Fk}	แทน ค่าความแปรปรวนของกลุ่มเปรียบเทียบในระดับที่ k
	S^2_{Rk}	แทน ค่าความแปรปรวนของกลุ่มอ้างอิงในระดับที่ k
	df	แทน ชั้นแห่งความเป็นอิสระซึ่งหาได้จาก $n_{Fk} + n_{Rk} - 2$

1.2 เนื่องจากค่า t -test ที่คำนวณได้เป็นค่าความลำเอียงสำหรับกลุ่มตัวอย่างขนาดเล็ก (Potenza and Dorans.1995:27) ในการศึกษาครั้งนี้ใช้กลุ่มตัวอย่างที่มีขนาดใหญ่ และมีความแตกต่างกัน จึงได้ใช้สถิติที่ปรับแก้ค่า t -test ที่มีผลมาจากขนาดกลุ่มตัวอย่าง แล้วทดสอบค่าซี (Z-test) ดังนี้ (Hedges. Et al.1989)

1.2.1 คำนวณหาค่า d_k จากสูตร

$$d_k = J_k \times \sqrt{\frac{n_{Fk} + n_{Rk}}{n_{Fk} n_{Rk}}} \times t_k$$

เมื่อ d_k แทน ค่าที่ปรับแก้ค่า t - test ที่มีผลมาจากขนาดของกลุ่มผู้สอบในแต่ละระดับการตอบ
 J_k แทน ค่าที่ปรับแก้ความลำเอียงที่มีผลมาจากขนาดของกลุ่มผู้สอบในแต่ละข้อ ซึ่งหาได้จาก

$$J_k = 1 - \frac{3}{4m_k - 1}$$

t_k แทน ค่าที่บ่งบอกถึงค่าดัชนีความลำเอียงในระดับการตอบที่ k

1.2.2 คำนวณหาค่า S_k^2 จากสูตร

$$S_k^2 = J_k^2 \frac{m_k}{m_k - 2} \left\{ \frac{n_{Fk} + n_{Rk}}{n_{Fk} n_{Rk}} \right\} + d_k^2 \left\{ \frac{J_k^2 m_k}{m_k - 2} - 1 \right\}$$

เมื่อ S_k^2 แทน ค่าความแปรปรวนของข้อสอบในแต่ละระดับการตอบ

1.2.3 คำนวณหาค่า S_d จากสูตร

$$S_d = \frac{1}{\sqrt{\left(\frac{1}{S^2_1} + \frac{1}{S^2_2} + \dots + \frac{1}{S^2_K}\right)}}$$

เมื่อ S_d แทน ค่าความคาดเคลื่อนมาตรฐานของ d_K

1.2.4 ทดสอบค่า z-test ของข้อสอบในแต่ละข้อ

$$Z_{HW3} = \frac{D}{S_d}$$

เมื่อ Z_{HW3} แทน ค่าดัชนีความลำเอียงหลังจากปรับค่า t-test แล้ว
 D แทน คำนำน้หนักคะแนนเฉลี่ยของ d_k ซึ่งคำนวณได้จาก

$$D = \frac{\left(\frac{d_1}{S^2_1} + \frac{d_2}{S^2_2} + \dots + \frac{d_k}{S^2_k}\right)}{\left(\frac{1}{S^2_1} + \frac{1}{S^2_2} + \dots + \frac{1}{S^2_k}\right)}$$

S_d แทน ค่าความคาดเคลื่อนมาตรฐานของ d_K

1.3. เกณฑ์ที่ใช้ในการตัดสินความลำเอียงของข้อสอบ คือ Z_{HW3} ที่คำนวณแตกต่างจาก H_0 อย่างมีนัยสำคัญ แสดงว่าข้อนั้นมี ความลำเอียง

2. วิธีหาค่าอำนาจจำแนกแบบวิธีไบซีเรียล มีขั้นตอนในการคำนวณดังนี้
 (Green and Draper.1972)

2.1. คำนวณค่าอำนาจจำแนกแบบไบซีเรียลของข้อสอบแต่ละข้อ ซึ่งมีสูตรดังนี้ (Garrett and Woodworth.1966:378)

$$r_{bis} = \frac{\mu_p - \mu_q}{\sigma} \cdot \frac{pq}{y}$$

เมื่อ	r_{bis}	แทน ค่าอำนาจจำแนกของแบบทดสอบที่คำนวณด้วยวิธีไบซีเรียล
	μ_p	แทน คะแนนเฉลี่ยของผู้สอบในกลุ่มสูง
	μ_q	แทน คะแนนเฉลี่ยของผู้สอบในกลุ่มต่ำ
	σ	แทน คะแนนเบี่ยงเบนมาตรฐานของแบบทดสอบฉบับนั้น
	p	แทน สัดส่วนของกลุ่มสูง
	q	แทน สัดส่วนของกลุ่มต่ำ (1- p)
	y	แทน ค่า Ordinate เปิดได้จากตารางโค้งปกติ

2.2. ดัชนีความลำเอียงของข้อสอบ โดยการทดสอบความแตกต่างของค่าอำนาจจำแนกระหว่างกลุ่มต่าง ๆ ในข้อสอบแต่ละข้อ โดยใช้สถิติซี (Berk.1982:209-210)

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

เมื่อ	$Z_{r_{bis}}$	แทน ค่าดัชนีความลำเอียงของข้อสอบ
	Z_{r1}	แทน ค่าอำนาจจำแนกของข้อสอบที่แปลงเป็นคะแนนมาตรฐาน โดยใช้ตาราง Fisher's ของกลุ่มสูง
	Z_{r2}	แทน ค่าอำนาจจำแนกของข้อสอบที่แปลงเป็นคะแนนมาตรฐาน โดยใช้ตาราง Fisher's ของกลุ่มต่ำ
	N_1, N_2	แทน จำนวนนักเรียนในกลุ่มสูงและต่ำ ตามลำดับ

3. วิธีคำนวณอำนาจจำแนกแบบวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด (Item-total Correlation) มีขั้นตอนดังนี้

3.1. คำนวณค่าอำนาจจำแนกด้วยวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด (Garrett and Woodworth.1966:382)

$$r_{xy} = \frac{N \sum XY - \sum X \sum Y}{\sqrt{[N \sum X^2 - (\sum X)][N \sum Y^2 - (\sum Y)]}}$$

เมื่อ	r_{xy}	แทน ค่าอำนาจจำแนกที่คำนวณจากสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน
-------	----------	--

N	แทน จำนวนกลุ่มตัวอย่าง
$\sum X$	แทน ผลรวมของคะแนนรายข้อ
$\sum Y$	แทน ผลรวมของคะแนนรวม ที่ไม่รวมข้อที่คำนวณ
$\sum XY$	แทน ผลรวมของผลคูณของคะแนน X และคะแนน Y
$\sum X^2$	แทน ผลรวมของกำลังสองของคะแนน X
$\sum Y^2$	แทน ผลรวมของกำลังสองของคะแนน Y

3.2 ดัชนีความลำเอียงของข้อสอบ โดยการทดสอบความแตกต่างของค่า
อำนาจจำแนกระหว่างกลุ่มต่าง ๆ ในข้อสอบแต่ละข้อ โดยใช้สถิติซี (Berk.1982:209-210)

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

เมื่อ	$Z_{r_{xy}}$	แทน ค่าดัชนีความลำเอียงของข้อสอบ
	Z_{r1}	แทน ค่าอำนาจจำแนกของข้อสอบที่แปลงเป็นคะแนน มาตรฐาน โดยใช้ตาราง Fisher's Z ของกลุ่มสูง
	Z_{r2}	แทน ค่าอำนาจจำแนกของข้อสอบที่แปลงเป็นคะแนน มาตรฐาน โดยใช้ตาราง Fisher's Z ของกลุ่มต่ำ
	N	แทน จำนวนนักเรียนในกลุ่มสูงและต่ำ ตามลำดับ

5.2.2 สถิติที่ใช้ในการทดสอบความแตกต่างของจำนวนข้อสอบที่มีความลำเอียงจากการ
วิเคราะห์ความลำเอียงด้วยวิธีต่างกัน (ล้วน สายยศและอังคณา สายยศ.2540:253-254)

$$Z = \frac{P_1 - P_2}{S_{p1-p2}}$$

เมื่อ	Z	แทน ค่าการแจกแจงของ Z
	P_1	แทน สัดส่วนของข้อสอบของ
	$P_1 = \frac{f_1}{n_1}$	
	P_2	แทน สัดส่วนของข้อสอบ

$$P_2 = \frac{f_2}{n_2}$$

S_{p1-p2} แทน ค่าความคลาดเคลื่อนมาตรฐานของความแตกต่าง
ระหว่างสัดส่วนที่วิเคราะห์ความลำเอียงของ 2 กลุ่ม
ซึ่งมีสูตรดังนี้

$$S_{p1-p2} = \sqrt{pq \frac{1}{n_1} + \frac{1}{n_2}}$$

เมื่อ P แทน สัดส่วนของข้อสอบที่มีความลำเอียงที่วิเคราะห์ด้วยวิธี
วิเคราะห์ทั้ง 2 วิธี ซึ่งคำนวณ

$$p = \frac{f_1 + f_2}{n_1 + n_2}$$

และ

$$q = 1 - p$$

5.2.3 ค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วด้วย
วิธีแบ่งครึ่งฉบับ (Split-Half) (ล้วน สายยศและอังคณา สายยศ.2536:166)

$$r_{tt} = \frac{2r_{\frac{1}{2}\cdot\frac{1}{2}}}{1 + r_{\frac{1}{2}\cdot\frac{1}{2}}}$$

เมื่อ r_{tt} แทน ค่าความเชื่อมั่นของแบบทดสอบ
 $r_{\frac{1}{2}\cdot\frac{1}{2}}$ แทน ค่าความเชื่อมั่นของแบบทดสอบครึ่งฉบับ คำนวณโดยใช้
สูตร Pearson Product-Moment Correlation Coefficient

5.2.4 สถิติที่ใช้ทดสอบความแตกต่างของค่าความเชื่อมั่นหลังคัดเลือกข้อสอบที่มีความ
ลำเอียงออกแล้ว จากวิธีการวิเคราะห์ความลำเอียงต่างกัน ใช้ทดสอบไคสแควร์ (χ^2) ดังนี้
(Shedcor and Cochran.1967:187)

$$\chi^2 = \sum (n_j - 3) Z^2_i \frac{\{\sum (n_i - 3) Z^2_i\}^2}{\sum (n_i - 3)}$$

เมื่อ	χ^2	แทน ค่าไคสแควร์
	Z^2_i	แทน คะแนนมาตรฐานซึ่งแปลงจากสัมประสิทธิ์ความเชื่อมั่น ตาราง Fisher' Z
	n_i	แทน จำนวนผู้สอบในแต่ละกลุ่ม
	k	แทน จำนวนแบบทดสอบ
	df	แทน ชั้นแห่งความเป็นอิสระ มีค่าเท่ากับ $k - 1$

5.2.5 สถิติที่ใช้ทดสอบความแตกต่างของค่าความเชื่อมั่นเป็นรายคู่ (Ferguson.1981 :
196)

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

เมื่อ	Z	แทน คะแนนมาตรฐานโค้งปกติ
	Z_{r1}, Z_{r2}	แทน คะแนนมาตรฐานที่แปลงเป็นคะแนนมาตรฐาน โดยใช้ ตาราง Fisher's Z
	N_1, N_2	แทน จำนวนนักเรียนในกลุ่มสูงและต่ำ ตามลำดับ

บทที่ 4

ผลการวิเคราะห์ข้อมูล

ในการเสนอผลการวิเคราะห์ข้อมูลจากการศึกษาครั้งนี้ เพื่อความสะดวกและความเข้าใจในการนำเสนอและการแปลความหมายของผลการวิเคราะห์ข้อมูล ผู้วิจัยจึงได้กำหนดสัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล ดังต่อไปนี้

สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล

N_R	แทน จำนวนผู้สอบในกลุ่มอ้างอิง (เพศหญิง)
N_F	แทน จำนวนผู้สอบในกลุ่มเปรียบเทียบ (เพศชาย)
K	แทน จำนวนข้อสอบทั้งหมด
k_{D1}	แทน จำนวนข้อสอบที่ลำเอียงจากการวิเคราะห์ด้วยวิธีHW3
k_{D2}	แทน จำนวนข้อสอบที่ลำเอียงจากการวิเคราะห์ด้วยวิธีไบซีเรียล
k_{D3}	แทน จำนวนข้อสอบที่ลำเอียงจากการวิเคราะห์ด้วยวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด
P	แทน ค่าความยากของข้อสอบที่ได้จากการวิเคราะห์โดยใช้ทฤษฎีดั้งเดิม
r	แทน ค่าอำนาจจำแนกของข้อสอบที่ได้จากการวิเคราะห์โดยใช้ทฤษฎีดั้งเดิม
$HW3$	แทน การวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีHW3
r_{bis}	แทน การวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีไบซีเรียล
r_{xy}	แทน การวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด
Z_{HW3}	แทน ค่าที่แสดงถึงความลำเอียงของข้อสอบจากการวิเคราะห์ความลำเอียงด้วยวิธีHW3 ที่ใช้ค่าซี
$Z_{r_{bis}}$	แทน ค่าที่แสดงถึงความลำเอียงของข้อสอบจากการวิเคราะห์ความลำเอียงด้วยวิธีไบซีเรียล ที่ใช้ค่าซี
$Z_{r_{xy}}$	แทน ค่าที่แสดงถึงความลำเอียงของข้อสอบจากการวิเคราะห์ความลำเอียงด้วยวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด ที่ใช้ค่าซี
r_{tt}	แทน ค่าความเชื่อมั่นของแบบทดสอบทั้งฉบับที่คำนวณโดยใช้สูตรแบ่งครึ่งฉบับ
χ^2	แทน ค่าไคสแควร์
Z	แทน ค่านัยสำคัญของการทดสอบด้วยสถิติ Z-test
Z_r	แทน คะแนนมาตรฐานจากค่าสัมประสิทธิ์ของค่าความเชื่อมั่นด้วยตาราง Fisher-Z
*	แทน มีระดับนัยสำคัญทางสถิติที่ระดับ .05
**	แทน มีระดับนัยสำคัญทางสถิติที่ระดับ .01

การนำเสนอผลการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้ ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลตามลำดับต่อไปนี้

1. ค่าความลำเอียงของข้อสอบและจำนวนข้อที่ลำเอียงจากแบบทดสอบวัดความเข้าใจทางภาษา จำแนกตามวิธีวิเคราะห์ความลำเอียง และลักษณะกลุ่มผู้สอบ
2. การเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียงด้วยวิธีต่างกัน เมื่อมีขนาดกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบเดียวกัน
3. การเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบต่างกัน เมื่อวิเคราะห์ความลำเอียงด้วยวิธีเดียวกัน
4. การเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วระหว่างวิธีวิเคราะห์ความลำเอียงด้วยวิธีต่างกัน เมื่อมีขนาดกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบเดียวกัน
5. การเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบต่างกัน เมื่อวิเคราะห์ความลำเอียงด้วยวิธีเดียวกัน

ผลการวิเคราะห์

1. ค่าความลำเอียงของข้อสอบและจำนวนข้อที่ลำเอียงจากแบบทดสอบวัดความเข้าใจทางภาษา จำแนกตามวิธีวิเคราะห์ความลำเอียง และลักษณะกลุ่มผู้สอบ

ผู้วิจัยได้นำแบบทดสอบไปทดสอบกับกลุ่มผู้สอบจำนวน 1,000 คน แบ่งกลุ่มผู้สอบเพศหญิงซึ่งเป็นกลุ่มอ้างอิงหรือกลุ่มที่ได้เปรียบในการทำแบบทดสอบ และแบ่งกลุ่มผู้สอบเพศชายซึ่งเป็นกลุ่มเปรียบเทียบหรือกลุ่มที่เสียเปรียบในการทำแบบทดสอบ เป็นอย่างละ 2 กลุ่ม คือ กลุ่มจำนวน 250 และ 500 คน ตามลำดับ จากนั้นได้ทำการวิเคราะห์หาค่าดัชนีความลำเอียงของข้อสอบด้วยวิธีวิเคราะห์ความลำเอียงทั้ง 3 วิธี คือ HW3 วิธีไบซีเรียล และวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด ตามลักษณะกลุ่มผู้สอบ 3 รูปแบบ คือกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:250 , กลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 500 500 และกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:500 ได้ผลการวิเคราะห์ดังตาราง 4 ดังนี้

ตาราง 4 ค่าความลำเอียง จำนวนข้อสอบที่มีความลำเอียงและค่าความเชื่อมั่นหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วของแบบทดสอบวัดความเข้าใจทางภาษา จำแนกตามวิธีวิเคราะห์ความลำเอียงและลักษณะผู้สอบทั้ง 3 ลักษณะ

กลุ่มผู้สอบ $N_R : N_F$	K	HW3			r_{bis}			r_{xy}		
		k_{D1}	Z_{HW3}	r_{tt}	k_{D2}	$Z_{r_{bis}}$	r_{tt}	k_{D3}	$Z_{r_{xy}}$	r_{tt}
250 : 250	40	2	1.669 - 3.215	0.7103	16	2.011-10.294	0.6216	3	2.056-2.853	0.5679
500 : 500	40	7	1.709 - 9.667	0.6990	22	2.158-6.008	0.4163	11	2.095-3.857	0.6506
250 : 500	40	5	1.711 - 3.880	0.6406	10	2.160-5.626	0.7230	4	2.610-3.298	0.4956

จากตาราง 4 แสดงให้เห็นว่า เมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:250 พบว่า มีค่าความลำเอียงของข้อสอบที่วิเคราะห์ความลำเอียง เรียงตามวิธี HW3 , วิธีไบซีเรียล และ วิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด มีค่าเท่ากับ 1.669-3.215 , 2.011-10.294 และ 2.056-2.853 ตามลำดับ ส่วนจำนวนข้อสอบที่มีความลำเอียง เรียงตามวิธีวิเคราะห์ความลำเอียง ได้จำนวน 2,16 และ 3 ข้อ ตามลำดับ และค่าความเชื่อมั่นหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว เรียงตามวิธีวิเคราะห์ความลำเอียงเป็น 0.7103 , 0.6216และ0.5679 ตามลำดับ และเมื่อวิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 500:500 พบว่า มีค่าความลำเอียงของข้อสอบที่วิเคราะห์ความลำเอียง เรียงตามวิธีวิเคราะห์ มีค่าเท่ากับ 1.709-9.667 , 2.158-6.008 และ 2.095-3.857 ตามลำดับ ส่วนจำนวนข้อสอบที่มีความลำเอียงเรียงตามวิธีวิเคราะห์ความลำเอียง ได้จำนวน 7,22 และ 11 ข้อ ตามลำดับ และค่าความเชื่อมั่นหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว เรียงตามวิธีวิเคราะห์ความลำเอียงเป็น 0.990 , 0.4163 และ 0.6505 ตามลำดับ ส่วนเมื่อวิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:500 พบว่ามีค่าความลำเอียงของข้อสอบที่วิเคราะห์ความลำเอียง เรียงตามวิธีวิเคราะห์ มีค่าเท่ากับ 1.711-3.880 , 2.160-5.626 และ 2.610 -3.298 ตามลำดับ ส่วนจำนวนข้อสอบที่มีความลำเอียง เรียงตามวิธีวิเคราะห์ความลำเอียง ได้จำนวน 5 , 10 และ 4 ข้อ ตามลำดับ และค่าความเชื่อมั่นหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว เรียงตามวิธีวิเคราะห์ความลำเอียงเป็น 0.6406 , 0.7230 และ 0.4956 ตามลำดับ

2. การเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิง และกลุ่มเปรียบเทียบที่มีรูปแบบเดียวกัน เมื่อวิเคราะห์ความลำเอียงด้วยวิธีต่างกัน

ผลการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีวิเคราะห์ความลำเอียงด้วยวิธีวิเคราะห์ต่างกัน เมื่อมีกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีขนาดเดียวกัน พบว่า ได้จำนวนข้อสอบที่มีความลำเอียงแตกต่างกัน ดังนั้นผู้วิจัยจึงทำการเปรียบเทียบความแตกต่างอย่างมีนัยสำคัญของจำนวนข้อสอบที่มีความลำเอียง โดยใช้สถิติ Z-test ซึ่งผลการทดสอบนัยสำคัญทางสถิติ ดังแสดงในตาราง 5

ตาราง 5 ผลการเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียงด้วยวิธีต่างกัน เมื่อมีขนาดกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบเดียวกัน

กลุ่มผู้สอบ $N_R : N_F$	วิธีวิเคราะห์ความลำเอียง		
	HW3 จำนวนข้อ(z-test)	r_{bis} จำนวนข้อ(z-test)	r_{xy} จำนวนข้อ(z-test)
HW3 ; 2 ข้อ 250 : 250 r_{bi} ; 16 ข้อ r_{xy} ; 3 ข้อ	-	14 (3.8417)**	1 (0.4646) 13 (3.4287)** -
HW3 ; 7 ข้อ 500 : 500 r_{bis} ; 22 ข้อ r_{xy} ; 11 ข้อ	-	15 (3.4981)** -	4 (1.0729) 11 (2.5000)* -
HW3 ; 5 ข้อ 250 : 500 r_{bis} ; 10 ข้อ r_{xy} ; 4 ข้อ	-	5 (1.4357) -	1 (0.3571) 6 (1.7688) -

* $Z_{.05} = 1.96$; ** $Z_{.01} = 2.58$

จากตาราง 5 แสดงให้เห็นว่า เมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบ เป็น 250:250 คน พบว่าวิธีHW3กับวิธีไบซีเรียล และวิธีไบซีเรียลกับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด มีจำนวนข้อที่ความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ส่วนวิธีHW3กับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด มีจำนวนข้อแตกต่างกันอย่างไม่

นัยสำคัญทางสถิติ และเมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 500:500 พบว่า วิธีHW3กับวิธีไบซีเรียล ,วิธีไบซีเรียลกับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด มีจำนวนข้อที่ความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และ .05 ตามลำดับ ส่วนเมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:500 พบว่ามีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทุกวิธี

3. การเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบต่างกัน เมื่อมีวิเคราะห์ความลำเอียงด้วยวิธีเดียวกัน

ผลการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน เมื่อมีขนาดกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีขนาดต่างกัน พบว่า ได้จำนวนข้อสอบที่มีความลำเอียงแตกต่างกัน ดังนั้นผู้วิจัยจึงทำการเปรียบเทียบความแตกต่างอย่างมีนัยสำคัญของจำนวนข้อสอบที่มีความลำเอียง โดยใช้สถิติ Z-test ซึ่งผลการทดสอบนัยสำคัญทางสถิติ ดังแสดงในตาราง 6

ตาราง 6 ผลการเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียง ระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงด้วยวิธีเดียวกัน

วิธีวิเคราะห์ ความลำเอียง	กลุ่มผู้สอบ ($N_R : N_F$)		
	250:250 จำนวนข้อ (z-test)	500:500 จำนวนข้อ (z-test)	250:500 จำนวนข้อ (z-test)
250:250; 2 ข้อ	-	5 (1.7705)	3 (0.1060)
HW3 500:500; 7 ข้อ		-	2 (1.5822)
250:500; 5 ข้อ			-
250:250; 16 ข้อ	-	6 (1.3453)	6 (1.4360)
r_{bis} 500:500; 22 ข้อ		-	12 (2.5000)*
250:500; 10ข้อ			-
250:250; 3 ข้อ	-	8 (2.3557)*	1 (0.4006)
r_{xy} 500:500; 11ข้อ		-	7 (2.0090)*
250:500; 4 ข้อ			-

* $Z_{.05} = 1.96$

จากตาราง 6 แสดงให้เห็นว่า ในวิธีไบซีเรียลและวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด เมื่อวิเคราะห์ระหว่างกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 500:500 กับ 250:500 พบว่ามีจำนวนข้อที่ความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 และเมื่อวิเคราะห์ระหว่างกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:250 กับ 500:500 พบว่า วิธีค่าความสัมพันธ์ของข้อรายข้อกับคะแนนรวมทั้งหมด มีจำนวนข้อที่ความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทุกส่วนวิธีHW3 พบว่ามีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทุกขนาดกลุ่มผู้สอบ

4.การเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วระหว่างวิธีวิเคราะห์ความลำเอียงด้วยวิธีต่างกัน เมื่อมีขนาดกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบเดียวกัน

หลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว จากวิธีวิเคราะห์ความลำเอียงทั้ง 3 วิธี แบ่งตามกลุ่มผู้สอบจำนวน 3 กลุ่ม ฉะนั้นจึงได้แบบทดสอบจำนวน 9 ฉบับ นำไปทดสอบกลับกลุ่มตัวอย่างจำนวนฉบับละ 200 คน รวมเป็น 1,800 คน แล้วนำผลคะแนนการสอบมาคำนวณหาค่าความเชื่อมั่นด้วยสูตรแบบแบ่งครึ่งฉบับและทำการทดสอบความแตกต่างของค่าความเชื่อมั่นด้วยสถิติไคสแควร์ ได้ผลดังตาราง 7

ตาราง 7 จำนวนข้อสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว ค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับ ค่าพีชเชอร์-ซี และการทดสอบด้วยสถิติไคสแควร์

กลุ่มผู้สอบ $N_R : N_F$	วิธีวิเคราะห์ความ ลำเอียง	K	r_{tt}	Z_r	χ^2
250 : 250	HW3	38	0.7103	0.887	6.402*
	r_{bis}	24	0.6216	0.725	
	r_{xy}	37	0.5679	0.640	
500 : 500	HW3	33	0.6990	0.758	13.872*
	r_{bis}	18	0.4163	0.442	
	r_{xy}	29	0.6505	0.775	
250 : 500	HW3	35	0.6406	0.858	15.130*
	r_{bis}	30	0.7230	0.908	
	r_{xy}	36	0.4956	0.543	

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

จากตาราง 7 แสดงให้เห็นว่า เมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:250 พบว่า มีจำนวนข้อสอบที่ไม่มีความลำเอียง เรียงตามวิธี HW3 , วิธีไบซีเรียล และ วิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด เป็น 38 , 24 และ 37 ข้อ ตามลำดับ และมีค่าความเชื่อมั่นหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว เรียงตามวิธีวิเคราะห์ความลำเอียงเป็น 0.7103 , 0.6216 และ 0.5679 ตามลำดับ และเมื่อวิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 500:500 พบว่า มีจำนวนข้อสอบที่ไม่มีความลำเอียง เรียงตามวิธีวิเคราะห์ เป็น 33, 18 และ 29 ข้อ ตามลำดับ และค่าความเชื่อมั่นหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว เรียงตามวิธีวิเคราะห์ความลำเอียงเป็น 0.6990, 0.4163 และ 0.6505 ตามลำดับ ส่วนการวิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:500 พบว่า มีจำนวนข้อสอบที่ไม่มีความลำเอียง เรียงตามวิธีวิเคราะห์ เป็น 35, 30 และ 36 ตามลำดับ และค่าความเชื่อมั่นหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว เรียงตามวิธีวิเคราะห์ความลำเอียงเป็น 0.6406, 0.7230 และ 0.4956 ตามลำดับ และเมื่อทำการทดสอบค่าความเชื่อมั่นด้วยสถิติไคสแควร์ พบว่ามีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ในทุกขนาดกลุ่มผู้สอบ แสดงว่ามีค่าความเชื่อมั่นอย่างน้อย 1 คู่ แตกต่างกัน ดังนั้นผู้วิจัยจึงทำการทดสอบความแตกต่างเป็นรายคู่ โดยใช้สถิติ z-test ดังแสดงผลการทดสอบในตาราง 8 และ 9

ตาราง 8 ผลการเปรียบเทียบค่าความเชื่อมั่นในรูปค่าพิชเชอร์-ซี ของแบบทดสอบหลังคัดเลือก
ข้อที่มีความลำเอียงออกแล้วระหว่างวิธีวิเคราะห์ความลำเอียงด้วยวิธีต่างกัน เมื่อมีขนาด
กลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบเดียวกัน

กลุ่มผู้สอบ $N_R : N_F$	วิธีวิเคราะห์ความลำเอียง		
	HW3	r_{bis}	r_{xy}
	ค่าพิชเชอร์-ซี (z-test)	ค่าพิชเชอร์-ซี (z-test)	ค่าพิชเชอร์-ซี (z-test)
HW3 ; .887 250 : 250 r_{bis} ; .725 r_{xy} ; .640	-	.162 (1.6135)	.247 (2.4601)* .085 (0.8466) -
HW3 ; .758 500 : 500 r_{bis} ; .442 r_{xy} ; .775	-	.316 (3.1474)** -	.017 (0.1693) .333 (3.3167)** -
HW3 ; .858 250 : 500 r_{bis} ; .908 r_{xy} ; .543	-	.050 (0.000) -	.315 (3.1374)** .365 (3.6354)** -

*.Z_{.05} = 1.96 ; ** Z_{.01} = 2.58

จากตาราง 8 แสดงให้เห็นว่า เมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น250:250 พบว่าวิธี HW3กับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด มีค่าความเชื่อมั่นแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีHW3 กับวิธีไบซีเรียล และวิธีไบซีเรียลกับวิธีค่าความสัมพันธ์ของข้อสอบกับคะแนนรวมทั้งหมด มีค่าความเชื่อมั่นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และเมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น500:500 พบว่า วิธีHW3กับวิธีไบซีเรียล,วิธีไบซีเรียลกับวิธีค่าความสัมพันธ์ของข้อสอบกับคะแนนรวมทั้งหมด มีค่าความเชื่อมั่นแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 นอกนั้นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ส่วนเมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น250:500 พบว่าวิธีHW3กับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมดและวิธีไบซีเรียลกับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด มีค่าความเชื่อมั่นแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 นอกนั้นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

5. การเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วระหว่างจากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน

ผู้วิจัยได้ทำการเปรียบเทียบค่าความเชื่อมั่นในรูปค่าพีชเชอร์-ซีของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว เมื่อมีลักษณะกลุ่มผู้สอบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน โดยใช้สถิติ Z-test ผลการทดสอบนัยสำคัญดังแสดงในตาราง 9

ตาราง 9 ผลการเปรียบเทียบค่าความเชื่อมั่นในรูปค่าพีชเชอร์-ซีของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วระหว่างจากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน

วิธีวิเคราะห์ ความลำเอียง	กลุ่มผู้สอบ ($N_R : N_F$)		
	250:250 ค่าพีชเชอร์-ซี(z-test)	500:500 ค่าพีชเชอร์-ซี (z-test)	250:500 ค่าพีชเชอร์-ซี (z-test)
250:250 ; .887 HW3 500:500 ; .758 250:500 ; .858	-	.129 (1.72848)	.029 (0.2888) - .100 (0.9960)
250:250 ; .725 r_{bis} 500:500 ; .442 250:500 ; .908	-	.283 (2.8087)** -	.183 (1.8227) .466 (4.6414)** -
250:250 ; .640 r_{xy} 500:500 ; .775 250:500 ; .543	-	.135 (1.3446) -	.097 (0.9661) .232 (2.3107)* -

* $Z_{.05} = 1.96$; ** $Z_{.01} = 2.58$

จากตาราง 9 แสดงให้เห็นว่า ในวิธีไบซีเรียลและวิธีค่าความสัมพันธ์ของข้อสอบ รายข้อกับคะแนนรวมทั้งหมด เมื่อวิเคราะห์ระหว่างกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 500:500 กับ 250:500 พบว่ามีค่าความเชื่อมั่นแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01และ.05 ตามลำดับ และเมื่อวิเคราะห์ระหว่างกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:250 กับ 500:500 พบว่า วิธีไบซีเรียล มีค่าความเชื่อมั่นแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01 นอกนั้นมีค่าความเชื่อมั่นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ส่วน วิธีHW3 พบว่ามีค่าความเชื่อมั่นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ทุกขนาดกลุ่มผู้สอบ

บทที่ 5

สรุปผล อภิปรายผล และข้อเสนอแนะ

สังเขปจุดมุ่งหมาย สมมติฐานและวิธีการดำเนินการศึกษา

จุดมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียงด้วยวิธีต่างกัน เมื่อมีขนาดกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบเดียวกัน
2. เพื่อเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบต่างกัน เมื่อวิเคราะห์ความลำเอียงด้วยวิธีเดียวกัน
3. เพื่อเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วระหว่างวิธีวิเคราะห์ความลำเอียงด้วยวิธีต่างกัน เมื่อมีขนาดกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบเดียวกัน
4. เพื่อเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบต่างกัน เมื่อวิเคราะห์ความลำเอียงด้วยวิธีเดียวกัน

สมมติฐานในการวิจัย

1. จำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียงด้วยวิธีต่างกัน เมื่อมีขนาดกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบเดียวกัน จะมีจำนวนแตกต่างกัน
2. จำนวนข้อสอบที่มีความลำเอียงระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบต่างกัน เมื่อมีวิเคราะห์ความลำเอียงด้วยวิธีเดียวกัน จะมีจำนวนแตกต่างกัน
3. ค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วระหว่างวิธีวิเคราะห์ความลำเอียงด้วยวิธีต่างกัน เมื่อมีขนาดกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบเดียวกัน จะมีค่าแตกต่างกัน
4. ค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วระหว่างที่วิเคราะห์จากกลุ่มผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่มีรูปแบบต่างกัน เมื่อวิเคราะห์ความลำเอียงด้วยวิธีเดียวกัน จะมีค่าแตกต่างกัน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนที่มีเพศต่างกัน ชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2543 ซึ่งเลือกมาโดยการสุ่มแบบสองขั้นตอน (Two-stage Random Sampling) โดยมีขนาดโรงเรียนเป็นชั้น (Strata) โรงเรียนเป็นหน่วยการสุ่ม (Sampling Unit) ได้จำนวนโรงเรียน 7 มีจำนวนห้อง 56 ห้องเรียน มีจำนวนนักเรียน 2,800 คน กลุ่มตัวอย่างแบ่งเป็น 2 กลุ่ม คือ

1.กลุ่มตัวอย่างที่ใช้ในการศึกษาผลการวิเคราะห์ความลำเอียง ด้านจำนวนข้อสอบที่มีความลำเอียง จำนวน 1,000 คน เป็นนักเรียนที่มีเพศแตกต่างกัน คือ นักเรียนเพศชาย จำนวน 500 คน และเป็นนักเรียนเพศหญิงจำนวน 500 คน

2.กลุ่มตัวอย่างที่ใช้ในการศึกษาผลวิเคราะห์ความลำเอียงด้านความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว จำนวน 1,800 คน

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้ คือ แบบทดสอบวัดความเข้าใจทางภาษาสำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 2 เป็นแบบทดสอบเลือกตอบแบบตัวเลือกถูกผิด ชนิด 5 ตัวเลือก จำนวน 40 ข้อ ที่ผู้วิจัยสร้างเอง

วิธีการเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลเพื่อทำการวิจัยในครั้งนี้ ผู้วิจัยได้ดำเนินการดังนี้

1. นำหนังสือจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ติดต่อขออนุญาตจากผู้บริหารโรงเรียนที่ใช้เป็นกลุ่มตัวอย่างและประสานงานกับอาจารย์ผู้เกี่ยวข้องในการดำเนินการสอบเพื่อแจ้งวัตถุประสงค์พร้อมทั้งกำหนดวัน-เวลา คือ อยู่ในระหว่างเดือนธันวาคม 2543 – มกราคม 2544 และชั้นที่จะสอบ คือ ชั้นมัธยมศึกษาปีที่ 2

2. จัดเตรียมอุปกรณ์ ได้แก่ แบบทดสอบ กระดาษคำตอบ ให้มากกว่าจำนวนผู้เข้าสอบ

3. ไปถึงสถานที่สอบในวันเวลาที่ทำการสอบ ก่อนกำหนดเวลาสอบ

4. ชี้แจงวัตถุประสงค์ของการสอบและพูดโน้มน้าวจิตใจผู้สอบให้มีความกระตือรือร้นที่จะทำข้อสอบด้วยความตั้งใจอย่างเต็มความสามารถและอธิบายคำชี้แจงในการทำข้อสอบให้ผู้สอบทุกคนเข้าใจและทำความเข้าใจกับผู้ช่วยดำเนินการสอบให้เข้าใจตรงกัน

5. นำแบบทดสอบไปทดสอบกับกลุ่มตัวอย่างครั้งที่ 1 จำนวน 458 คน เพื่อนำมาวิเคราะห์ค่าความยากและค่าอำนาจจำแนก ด้วยวิธีวิทนีและซาเบอร์ และทำการคัดเลือกข้อสอบที่อยู่ในเกณฑ์ที่ตั้งไว้

6. นำแบบทดสอบไปทดสอบกับกลุ่มตัวอย่างครั้งที่ 2 กับกลุ่มตัวอย่างจำนวน 500 คน แล้วทำการวิเคราะห์หาค่าความเชื่อมั่นของแบบทดสอบด้วยวิธีของฮอยท์

7. นำแบบทดสอบที่มีคุณภาพจำนวน 40 ข้อ ไปทดสอบกับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 2 ที่เป็นกลุ่มตัวอย่าง จำนวน 1,000 คน แบ่งเป็นนักเรียนเพศชาย 500 คน เพศหญิง 500 คน

8. ตรวจสอบให้คะแนนของแบบทดสอบวัดความเข้าใจทางภาษา โดยตรวจสอบให้คะแนนใช้วิธีการตรวจแบบให้คะแนนทุกตัวเลือก

9. ทำการสุ่มกลุ่มผู้สอบในกลุ่มอ้างอิง (เพศชาย) และกลุ่มเปรียบเทียบ (เพศหญิง) แล้วทำการแบ่งผู้สอบออกเป็น 4 กลุ่ม คือ กลุ่มผู้สอบที่เป็นเพศชาย 2 กลุ่ม คือ จำนวน 250 , 500 คน ตามลำดับ และกลุ่มผู้สอบเพศหญิง 2 กลุ่ม คือ จำนวน 250 , 500 คน ตามลำดับ

10. ทำการวิเคราะห์ความลำเอียงตามวิธีการวิเคราะห์ความลำเอียงทั้ง 3 วิธี เพื่อศึกษาจำนวนข้อสอบที่มีความลำเอียง จำแนกในแต่ละวิธีวิเคราะห์และลักษณะกลุ่มผู้สอบ เพื่อตอบคำถามในจุดมุ่งหมายข้อ 1 และ 2 แล้วทำการทดสอบสมมติฐานข้อที่ 1 และ 2

11. นำผลการวิเคราะห์ความลำเอียงตามข้อ 10 มาคัดเลือกข้อที่ลำเอียงออกแล้ว ทำการหาค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว ในแต่ละวิธีวิเคราะห์ความลำเอียง ตามขนาดกลุ่มผู้สอบทั้ง 3 กลุ่มละ 200 คน รวมจำนวนทั้งหมด 1,800 คน โดยใช้วิธีการหาค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับ (Split-Half) เพื่อตอบคำถามในจุดมุ่งหมายข้อ 3 และ 4 แล้วทำการทดสอบสมมติฐานข้อที่ 3 และ 4

สรุปผลการศึกษาค้นคว้า

1. จำนวนข้อสอบที่มีความลำเอียงเมื่อวิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:250 เมื่อวิเคราะห์ด้วยวิธี HW3 ,วิธีไบซีเรียลและวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด มีจำนวนข้อที่ลำเอียงเป็น 2,16 และ 3 ข้อ ตามลำดับ และจากการวิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 500:500 มีจำนวนข้อที่ลำเอียงเรียงตามวิธีวิเคราะห์ความลำเอียงเป็น 7, 22และ11 ข้อ ตามลำดับ ส่วนเมื่อวิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:500 มีจำนวนข้อที่ลำเอียงเรียงตามวิธีวิเคราะห์ความลำเอียงเป็น 5,10 และ 4 ข้อ ตามลำดับ แสดงว่าวิธีไบซีเรียลพบข้อสอบที่มีความลำเอียงมากที่สุด รองลงมาคือวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด และวิธีHW3 พบข้อสอบที่มีความลำเอียงน้อยที่สุด

2. จำนวนข้อสอบที่มีความลำเอียงระหว่างที่วิเคราะห์ด้วยวิธีHW3กับวิธีไบชีเรียล และวิธีไบชีเรียลกับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด เมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:250 และ 500:500 มีจำนวนข้อที่ลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01และ.05ตามลำดับ ส่วนจำนวนข้อที่ลำเอียงระหว่างวิธีHW3กับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด เมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:500 มีจำนวนข้อแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

3. จำนวนข้อสอบที่มีความลำเอียงเมื่อวิเคราะห์ด้วยวิธีไบชีเรียลและวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด ระหว่างที่วิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 500:500 กับ 250:500 มีจำนวนข้อที่ลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 และวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด ระหว่างที่วิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:250 กับ 500:500 มีจำนวนข้อที่ลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 และเมื่อวิเคราะห์ด้วยวิธีHW3 มีจำนวนข้อที่ลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทุกขนาดกลุ่มผู้สอบ

4. ค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อที่ลำเอียงออกแล้ว ระหว่างที่วิเคราะห์ด้วยวิธีHW3กับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด เมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:250และ500:500 มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05และ.01ตามลำดับ และจากการวิเคราะห์ระหว่างวิธีHW3กับวิธีไบชีเรียล และวิธีไบชีเรียลกับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด เมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:500 มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01

5. ค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว เมื่อวิเคราะห์ด้วยวิธีไบชีเรียล ระหว่างที่วิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 500:500 กับ 250:500 มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01 และเมื่อวิเคราะห์ด้วยวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด ระหว่างที่วิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น250:250 กับ 500:500 มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 ส่วนวิธีHW3 มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทุกขนาดกลุ่มผู้สอบ

อภิปรายผล

จากการศึกษาค้นคว้าครั้งนี้ สามารถอภิปรายผลได้ดังนี้

1. จำนวนข้อสอบที่มีความลำเอียงระหว่างที่วิเคราะห์ด้วยวิธีHW3กับวิธีไบซีเรียล และวิธีไบซีเรียลกับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด เมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:250 และ 500:500 มีจำนวนข้อที่ลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01และ.05ตามลำดับ ซึ่งสอดคล้องกับสมมติฐานการวิจัยที่ตั้งไว้ ส่วนจำนวนข้อที่ลำเอียงระหว่างวิธีHW3กับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด เมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:500 มีจำนวนข้อแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติซึ่งไม่สอดคล้องกับสมมติฐานการวิจัยที่ตั้งไว้ ทั้งนี้เนื่องจากมีจำนวนข้อสอบที่มีความลำเอียงจากการวิเคราะห์แต่ละวิธีมีจำนวนข้อแตกต่างกันน้อยมาก กล่าวคือวิธีHW3 กับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด มีจำนวนข้อที่ลำเอียง 5 และ 4 ข้อ แต่สอดคล้องกับผลงานวิจัยของแวงและเลนนี่ (Wang and Lane.1996:175-199) ที่ทำการศึกษาความลำเอียงโดยใช้วิธีการวิเคราะห์ความลำเอียงแตกต่างกัน 3 วิธี คือวิธีLogistic Regression (LOG) วิธี Logistic discriminant function analysis (LDFD) และวิธีHW3 พบว่ามีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยทางสถิติ และยังสอดคล้องกับงานวิจัยของคมศักดิ์ ชื่นชม (2539:75) ได้ทำการการศึกษาค่าผลการวิเคราะห์ความลำเอียงที่ใช้วิธีต่างกันของแบบทดสอบวัดจริยธรรมด้านความซื่อสัตย์ ตามทฤษฎีคลาสสิกอลที่ใช้ต่างกัน 3 วิธี คือวิธีค่าอำนาจจำแนกของข้อสอบ วิธีวิเคราะห์องค์ประกอบ และวิธีวิเคราะห์ถดถอย พบว่าการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีค่าอำนาจจำแนกของข้อสอบ วิธีวิเคราะห์องค์ประกอบและวิธีวิเคราะห์การถดถอยมีจำนวนข้อที่มีความลำเอียงน้อยและแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

2.จำนวนข้อสอบที่มีความลำเอียงเมื่อวิเคราะห์ด้วยวิธีไบซีเรียลและวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด ระหว่างที่วิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 500:500 กับ 250:500 มีจำนวนข้อที่ลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 และวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด ระหว่างที่วิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:250 กับ 500:500 มีจำนวนข้อที่ลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05 ซึ่งสอดคล้องกับสมมติฐานการวิจัยที่ตั้งไว้ และเมื่อวิเคราะห์ด้วยวิธีHW3 มีจำนวนข้อที่ลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทุกขนาดกลุ่มผู้สอบ ซึ่งไม่สอดคล้องกับสมมติฐานในการวิจัยที่ตั้งไว้ ทั้งนี้เนื่องจากมีจำนวนข้อสอบที่มีความลำเอียงจากการวิเคราะห์ความลำเอียงในแต่ละขนาดกลุ่มผู้สอบมีจำนวนข้อแตกต่างกันน้อยมาก กล่าวคือเมื่อทำการวิเคราะห์ตามกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:250,500:500 และ250:500 มีจำนวนข้อที่ความลำเอียง เป็น 2,7 และ 5 ข้อ ตามลำดับ และเมื่อนำจำนวนข้อที่มีความลำเอียงมาทดสอบ

ความแตกต่างแล้ว จึงพบว่า มีจำนวนข้อที่ลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ แต่สอดคล้องกับงานวิจัยของฮูเวอร์และโคลน (Hoover and Kolan.1984:173-181) ที่ทำการศึกษาเกี่ยวกับความเชื่อมั่นของดัชนีความลำเอียงทางเพศ ซึ่งใช้วิธีการวิเคราะห์ความลำเอียง 6 วิธี คือ วิธีแปลงค่าความยาก 2 วิธี, วิธีไบซีเรียล, วิธีพอยต์ไบซีเรียล, วิธีไคสแควร์ และ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ ผลการศึกษาพบว่า มีจำนวนข้อแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และยังสอดคล้องกับงานวิจัยของเวลช์และฮูเวอร์ (Wing and Lane.1996:183-184; citing Wech and Hoover.1993) ที่ทำการเปรียบเทียบวิธีการวิเคราะห์ความลำเอียงสองวิธีคือวิธีHW3กับวิธีไคสแควร์แบบแมนเทิล ซึ่งผลการศึกษาพบว่า มีจำนวนข้อแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

สำหรับวิธี HW3 ที่ผู้วิจัยใช้ในการศึกษาค้นคว้าครั้งนี้ให้ผลการศึกษาค้นคว้าได้สอดคล้องกับ เวลช์และฮูเวอร์ (Wing and Lane.1996:183-184 ; citing Wech and Hoover.1993) ที่กล่าวไว้ว่า วิธีHW 3 มีแนวโน้มที่จะค้นหาความลำเอียงได้ดี เมื่อขนาดกลุ่มผู้สอบเท่ากับหรือมากกว่า 500 คน ขึ้นไป กล่าวคือ วิธี HW3 เมื่อทำการวิเคราะห์ตามกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 250:250 , 250:500 และ 500:500 มีจำนวนข้อที่ลำเอียง เป็น 2,5 และ 7 ข้อ ตามลำดับ และวิธีการวิเคราะห์ความลำเอียงของข้อสอบที่ผู้วิจัยนำมาศึกษาครั้งนี้ทั้ง 3 วิธี เป็นวิธีวิเคราะห์ความลำเอียงโดยอาศัยหลักทฤษฎีการวัดผลแบบดั้งเดิม (Classical Test Theory) ซึ่งมีข้อจำกัดคือผลการวิเคราะห์จะมีค่าไม่คงที่ ทั้งนี้ขึ้นอยู่กับความสามารถของผู้สอบแต่ละกลุ่ม (Hambleton and Swaminathan.1985:25) กล่าวคือ เมื่อวิเคราะห์กับกลุ่มผู้สอบที่มีจำนวนมาก ก็จะตรวจพบข้อสอบที่มีความลำเอียงมากขึ้นด้วย

3. ค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อที่ลำเอียงออกแล้ว ระหว่างที่วิเคราะห์ด้วยวิธีHW3กับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด เมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น250:250 และ 500:500 มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05และ.01ตามลำดับ และจากการวิเคราะห์ระหว่างวิธีHW3กับวิธีไบซีเรียลและวิธีไบซีเรียลกับวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด เมื่อมีกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น250:500 มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01 ซึ่งสอดคล้องกับสมมติฐานในการวิจัยที่ตั้งและยังสอดคล้องกับงานวิจัยของฮูเวอร์และโคลน (Hoover and Kolan.1984:173-181) ได้ทำการศึกษาเกี่ยวกับความเชื่อมั่นของดัชนีความลำเอียงทางเพศ ซึ่งใช้วิธีการวิเคราะห์ความลำเอียง 6 วิธี คือวิธีแปลงค่าความยาก 2 วิธี, วิธีไบซีเรียล,วิธีพอยต์ไบซีเรียล,วิธีไคสแควร์ และวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ ผลการศึกษาพบว่าค่าความเชื่อมั่นที่ได้จากการวิเคราะห์ความลำเอียงแต่ละวิธีมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.05

4.ค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว เมื่อวิเคราะห์ด้วยวิธีไบซีเรียล ระหว่างที่วิเคราะห์จากกลุ่มอ้างอิง:กลุ่มเปรียบเทียบเป็น 500:500 กับ 250:500 มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ.01 และเมื่อวิเคราะห์ด้วยวิธี

ค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด ระหว่างที่วิเคราะห์จากกลุ่มอ้างอิง: กลุ่มเปรียบเทียบเป็น 250:250 กับ 500:500 มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งสอดคล้องกับสมมติฐานที่ตั้งไว้ ส่วนวิธี HW3 มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทุกขนาดกลุ่มผู้สอบ ซึ่งไม่สอดคล้องกับสมมติฐานในการวิจัยที่ตั้งไว้ ทั้งนี้เนื่องจากมีค่าความเชื่อมั่นแตกต่างกันน้อยมากจากการวิเคราะห์ในแต่ละขนาดกลุ่มผู้สอบกล่าวคือ เมื่อทำการวิเคราะห์ตามกลุ่มอ้างอิง: กลุ่มเปรียบเทียบเป็น 250:250, 500:500 และ 250:500 มีค่าความเชื่อมั่นเป็น 7103, 6990 และ 6406 ตามลำดับ และเมื่อนำค่าความเชื่อมั่นมาทำการทดสอบความแตกต่างกันแล้ว จึงพบว่ามีความเชื่อมั่นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ แต่สอดคล้องกับงานวิจัยของคมศักดิ์ ชื่นชม (2539:76-77) ได้ทำการการศึกษาผลการวิเคราะห์ลำเอียงที่ใช้วิธีต่างกันของแบบทดสอบวัดจริยธรรมด้านความซื่อสัตย์ ตามทฤษฎีคลาสสิกอลที่ใช้ต่างกัน 3 วิธี คือวิธีค่าอำนาจจำแนกของข้อสอบ วิธีวิเคราะห์องค์ประกอบ และวิธีวิเคราะห์ถดถอย ผลการวิจัยพบว่า มีความเชื่อมั่นของแบบทดสอบวัดจริยธรรมด้านความซื่อสัตย์หลังคัดเลือกข้อที่มีความลำเอียงออกแล้ว มีค่าไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

ผลจากการศึกษาครั้งนี้ แสดงว่าแบบทดสอบวัดความเข้าใจทางภาษามีค่าความเชื่อมั่นหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วจากวิธีการวิเคราะห์ความลำเอียงด้วยวิธี HW3, วิธีไบซีเรียล และวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด เมื่อนำผลการสอบมาคำนวณหาความเชื่อมั่นโดยสูตรการคำนวณแบบแบ่งครึ่งฉบับ พบว่าถ้าหากวิเคราะห์กับกลุ่มผู้สอบที่มีจำนวนมากและมีข้อสอบจำนวนมาก จะมีค่าความเชื่อมั่นเพิ่มขึ้นซึ่งสอดคล้องกับแอลเลนและเยน (Allen and Yen, 1979:91-92) ที่กล่าวว่าค่าความเชื่อมั่นของแบบทดสอบที่คำนวณจากสูตรสเปียร์แมน-บราวน์ (Spearman-Brown) จะมีแนวโน้มต่ำกว่าเมื่อยาวของแบบทดสอบสั้นกว่าและค่าความเชื่อมั่นของแบบทดสอบจะมีแนวโน้มสูงขึ้น เมื่อความยาวของแบบทดสอบเพิ่มขึ้น

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 ควรมีการวิเคราะห์ความลำเอียงของข้อสอบในแบบทดสอบก่อนนำไปใช้ โดยเลือกใช้วิธีไบซีเรียล เนื่องจากการศึกษาค้นคว้าครั้งนี้ พบว่าวิธีไบซีเรียลให้ผลในการค้นหาข้อสอบที่มีความลำเอียงมากกว่าวิธี HW3 และวิธีค่าความสัมพันธ์ของข้อสอบรายข้อกับคะแนนรวมทั้งหมด ในแต่ละขนาดกลุ่มผู้สอบทั้ง 3 กลุ่ม

1.2 ควรใช้ขนาดกลุ่มผู้สอบที่จะนำมาวิเคราะห์ตั้งแต่กลุ่มละ 500 คนขึ้นไป ทั้งนี้จากการศึกษาค้นคว้า พบว่า จำนวนผู้สอบมากจะได้ผลการค้นหาข้อสอบที่มีความลำเอียงได้มากเช่นกัน

1.3 ในการสร้างแบบทดสอบวัดความเข้าใจทางภาษา ควรมีการคัดเลือกข้อสอบตามสถานการณ์ อย่างน้อยสถานการณ์ละ 2 ข้อ ไม่ว่าจะเป็นการหาคุณภาพเบื้องต้นของแบบทดสอบหรือการทดสอบตามจุดมุ่งหมายของการศึกษาก็ตาม

2.ข้อเสนอแนะในการทำวิจัยต่อไป

2.1 ควรมีการศึกษาเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบแบบที่มีการตรวจให้คะแนนที่ไม่เป็น 0-1 เช่น แบบทดสอบปลายเปิด แบบทดสอบวัดเจตคติ แบบทดสอบแบบอัตนัย แบบทดสอบเรียงความ แบบทดสอบวัดการปฏิบัติจริง เป็นต้น

2.2 ควรมีการศึกษาเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบกับตัวแปรอื่นๆ เช่น ตัวแปรด้านจิตพิสัย ได้แก่ ระดับความสนใจ ระดับเจตคติต่อเรื่องที่จะศึกษาว่าจะมีผลสอดคล้องกับตัวแปรทางกายภาพ ได้แก่ เพศ ฐานะทางเศรษฐกิจ หรือไม่

2.3 ควรมีการศึกษาเปรียบเทียบความลำเอียงของข้อสอบจากแบบทดสอบที่สร้างขึ้น โดยวิธีไบซรีแยลซึ่งเป็นวิธีที่อาศัยทฤษฎีแบบดั้งเดิม กับวิธีอื่นๆที่อาศัยทฤษฎีการตอบข้อคำถาม เช่น วิธีGMH วิธีLogistic discriminant function analysis (LDFD) และ วิธี Logistic Regression (LOG) เป็นต้น ว่าให้ผลการวิเคราะห์สอดคล้องกันหรือไม่

2.4 ควรมีการศึกษาถึงแหล่งที่ก่อให้เกิดความลำเอียงของข้อสอบในข้อสอบที่มีความลำเอียง หลังจากที่ใช้วิธีวิธีการทางสถิติในการวิเคราะห์หาข้อสอบที่มีความลำเอียงแล้ว โดยให้ผู้วิจัยพิจารณาก่อนแล้วจึงให้ผู้เชี่ยวชาญพิจารณาอีกครั้งหนึ่ง เพื่อที่จะทราบถึงสาเหตุของความลำเอียงของข้อสอบอย่างแท้จริง

บรรณานุกรม

บรรณานุกรม

- กาญจนา วัฒนสุนทร. (2538) . การพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ. วิทยานิพนธ์ ค.ด . (การวัดและประเมินผลการศึกษา). กรุงเทพฯ:บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- จิตรสุตา ธราพร . (2539) . การศึกษาเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบโจทย์ปัญหาทางคณิตศาสตร์และแบบทดสอบวัดความเข้าใจทางภาษา โดยใช้วิธีการวิเคราะห์ต่างกัน . ปรินญาณินพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ ฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร. ถ่ายเอกสาร.
- คมศักดิ์ ชื่นชม . (2539) . การศึกษาผลการวิเคราะห์ความลำเอียงที่ใช้วิธีต่างกันของแบบทดสอบวัดจริยธรรมด้านความซื่อสัตย์ ตามทฤษฎีคลาสสิกอล . ปรินญาณินพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ ฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร. ถ่ายเอกสาร.
- ชัชชัย เผ่าพงศ์. (2527) . การวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์และภาษา ระดับชั้นมัธยมศึกษาตอนต้น. วิทยานิพนธ์ ศศ.ม. (การวัดผลและประเมินการศึกษา). ขอนแก่น : บัณฑิตวิทยาลัย มหาวิทยาลัยขอนแก่น. ถ่ายเอกสาร.
- ทัศนีย์ พิรมนตรี. (2530) . การวิเคราะห์ความลำเอียงของแบบทดสอบวิชาคณิตศาสตร์ โครงการตรวจสอบคุณภาพทางการศึกษา ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526. วิทยานิพนธ์ ค.ม. (การวัดผลและประเมินการศึกษา). กรุงเทพฯ ฯ : จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- นิรมล ชัยชาลิต. (2537). การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของแบบทดสอบความเข้าใจในการอ่านภาษาไทย ตามทฤษฎีคลาสสิกอลที่ใช้วิธีวิเคราะห์ต่างกัน. ปรินญาณินพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ ฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.ถ่ายเอกสาร.
- บุญชม ศรีสะอาด. (2535). การวิจัยทางการวัดผลและประเมินผล.มหาสารคาม : มหาวิทยาลัยศรีนครินทรวิโรฒ มหาสารคาม.
- พรณี จินตมาศ . (2540) . การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของแบบทดสอบ โดยใช้ขนาดของกลุ่มผู้สอบและวิธีวิเคราะห์ความลำเอียงต่างกัน. ปรินญาณินพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ ฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร. ถ่ายเอกสาร.
- พัชรี ปิยภักดิ์. (2531). การวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน ชั้นประถมศึกษาปีที่ 6. ปรินญาณินพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ ฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร. ถ่ายเอกสาร.

- ไพฑูรย์ เวทการ. (2536) . การประเมินผลและการสร้างแบบทดสอบ . ลำปาง : ภาควิชาทดสอบและวิจัยทางการศึกษา วิทยาลัยครูลำปาง.
- เรวดี อินทสระ. (2539). ผลการตรวจสอบความลำเอียงของข้อสอบต่อการศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่คิดคะแนนต่างกัน. ปรินญานิพนธ์ กศ.ด. (การทดสอบและการวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร. ถ่ายเอกสาร.
- ล้วน สายยศและอังคณา สายยศ. (2536). เทคนิคการวิจัยทางการศึกษา. พิมพ์ครั้งที่ 3 . กรุงเทพฯ : ศูนย์ส่งเสริมวิชาการ.
- .(2539).เทคนิคการวัดผลการเรียนรู้ กรุงเทพฯ : ชมรมเด็ก.
- .(2540). สถิติทางการศึกษา. พิมพ์ครั้งที่ 3 . กรุงเทพฯ : ศูนย์ส่งเสริมวิชาการ.
- สมศักดิ์ จันทอง. (2541). “ จากความลำเอียงของข้อสอบ(Item Bias) มาเป็นการทำหน้าที่ต่างกันของข้อสอบ (DIF),” วารสารศึกษาศาสตร์ มหาวิทยาลัยนเรศวร. 3(1) : 81-93.
พฤษภาคม – สิงหาคม.
- สุรศักดิ์ อมรัตน์ศักดิ์. (2531). การเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่แตกต่างกัน 4 วิธี. วิทยานิพนธ์ ค.ด. (การวัดและประเมินผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย.ถ่ายเอกสาร.
- อนันต์ ศรีโสภา. (2520). การวัดผลและประเมินผลการศึกษา. กรุงเทพฯ : ไทยวัฒนาพานิช.
- อเนก เพียรอนุกุลบุตร . (2524). การวัดผลและประเมินผลการศึกษา. กรุงเทพฯ : วิดีเตอร์การพิมพ์.
- อังคณา สายยศ. (2539) . การศึกษาผลการวิเคราะห์ความลำเอียงของแบบทดสอบความเข้าใจในการอ่าน ตามลักษณะการใช้ภาษาพูดในครอบครัว. กรุงเทพฯ : ภาควิชาการวัดผลและวิจัยทางการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร .
- อังคณา สายยศ. (2538). “ ข้อสอบเลือกตอบแบบถูกผิด(Multiple Ture – False : MTF),” วารสารการวัดผลการศึกษา. 17(50) : 42 – 50 .
- Allen,M.J. and W.M.Yen. (1979). *Introduction to Measurement Theory* . Monterey,Calif Brook/Cole.
- Allalof,Avi ,Ronald K.Hambleton and Stephen G.Sireci. (1999). “ Identifying the Causes of DIF in Translated Verbal Items,” *Journal of Educational Measurement*. 36(3) : 185-198.
- Angoff,W.H. (1982). “ Use of Difficulty and Discrimination Indices of Detecing Item Bias,” *In Handbook of Methods for Detecting test Bias* . Baltimore and London : The Johns Hopkins University Press.

- Angoff,W.H. and Susan F.Ford. (1973). " Item – Race Interaction on a test of Scholastic Aptitude," *Journal of Educational Measurement* . 10(2):95-105.
- Anastasi , Anne. (1976) . *Differential Psychology : individual and group differences in Behavior*. 3rd ed. New York : Macmillan.
- Baghi,H. and S.F.Ferrara. (1990) . *Detecting Differential Item Functioning Using IRT and Mantel Haenszel Techniques : Implementing Procedure and Comparing Results*. 14 – 17 .
- Berk,R.A. (Ed.)(1982). *Handbook of Methods for Detecting test Bias* . Baltimore and London : The Johns Hopkins University Press.
- Burrill,L.E. (1982) . " Comparative Studies of Item Bias Method," In *Handbook of Methods for Detecting test Bias* . Baltimore and London : The Johns Hopkins University Press.
- Camill,Gregory and Lorrie A.Shepard.(1994) . *Method for Identifying Biased Test Items*_. Thousand , Oaks , London , New Delhi : SAGE Publication.
- Cang,Hua-Hua , John Mazzeo and Louis Rouussos. (1996). " Detecting DIF for Polytomously Scord Items : An Adaptation of the SIBTES Procedure,"*Journal of Educational Measurement*_,33(3):333-335.Fall.
- Clauser,Bria E.,Ronald J.Nungester and H.Swaminathan. (1996). " Improving the Matching for DIF Analysis by Conditioning on Both Test Score and an Educational Background Variable," *Journal of Educational Measurement*. 33(4):453-464.
- Cohen,D.S. et al. (1996). " An Investigation of the Likelihood Ratio Test For Detection of Differential Item Functioning," *Applied Psychological Measurement*. 20(1):15-26.
- Cronbach, L. J. (1970). *Essentials of Psychological Testing* . New York : Harper & Row Publishers.
- Doolittle,Allen E. and Cleary T. Anne. (1987). " Gender – Based Differential Item Performance in Mathematics Achivement Item," *Journal of Educational Measurement*_, 24(2):157-166.
- Garrett,Hanry E. and R.S.Woodworth. (1966). *Statistics in Psychology and Education*. Bomby : Vakils,Feffer and Simon PVT.LTD.
- Ferguson,G.A. (1981). *Statistical Analysis in Psychology and Education*. New York : Mc Graw-Hill.

- Haebara,T.A. (1979). *Method for Investigating Item Bias Using Brimbuan's Three-Parameter Logistic Model* . Iowa Testing Programs Occasional Paper, The University of Iowa.
- Hambleton,Ronald K. and Roger H.Jane. (1989) . " Detecting potentially biased test item : Comparison of IRT area and Mantel-Haenszel methods," *Applied Measurement in Education*. 2 :313-334.
- Hambleton,Ronald K. and H.Swaminathan. (1985). *Item Response Theory Principle and Application*. Boston : Kluwer-Nijhoff Publishing.
- Hambleton,Ronald K. , H.Swaminathan and Roger H.Jane. (1991). *Foundamentals of Item Response Theory* . Newbury Park, California : Sage Publishing.
- Hill,J. (1990). " Screening for Potentially Biased Item in Testing Program," *Education Measurement : Issues and Practice*. 8 : 5–11.
- Holland,Paul W. and Howard Wainer. (1993). *Differential Item Functioning* . Hillsdale, New Jersey , Hove and London : Lawrence Erlbaum Associates Publishers.
- Hoover,H.D. and Michael J.Kolen. (1984). " The Reliability of Six Item Bias Indices," *Applied Psychological Measurement* . 8(2) :173-181.
- Hulin,Charles L. et al. (1983). *Item Response Theory Application to Psychological Measurement* . Illinois : Dow Jones Irwin.
- Intasuwan,P. (1979). *A Comparison of Three Approaches for Determining Item Bias in Gross-National Testing*. Doctor's Thesis The University of Pittsburgh.
- Ironson,Gail et al. (1984). " The Validity of Item Bias Techniques with math word Problems," *Applied Psychological Measurement* . 8 (4) :391–396.
- Laksana,S. (1979). *Application of Analysis of Variance Approach and ITBS Form 7* . The University of Iowa.
- Laksana,S. and W.E.Coffman. (1980). *A Comparison of and ANOVA Approach and ICC Approach for Assessing Item Bias in an Achievement Test* . Iowa Testing Programs Occasional Paper, The University of Iowa.
- Linn,Robert L. (1989). *Educational Measurement*. 3rd ed. New York,London.
- Lord,F.M. and M.R. Novick. (1968). *Statistical Theries of Mental Test Scores* . Reading MA : Addison-Wesley.Publishing Company.
- Lord,F.M. (1980). *Application of Item Response Theory to Practical Teseting Problems*. New Jersey : Lawrence Erlbaum Associates Publishers.

- Maller, S.J. (1994). "Validity and Item Bias of the WISC-III with Deaf Children," *Dissertation Abstracts International*. 55(5) : 1249-A.
- Mazor, Kathleen M. et al. (1991). "The Effect of sample size on the functioning of The Mantel-Haenszel Statistic," *Educational and Psychological Measurement* . 52 : 443-451.
- Mazor, Kathleen M. et al. (1995). "Using Logistic Regression and Mantel-Haenszel with Multiple Ability Estimates to Detect Differential Item Functioning," *Journal of Educational Measurement* . 32(2) : 131-144.
- Mc Canley, D. Cynthia and Jorge Mendoza. (1985). "A Simulation Study of Item Bias Using a Two-Parameter Item Response Model," *Applied Psychological Measurement* . 9(4) : 389-400.
- Mellbergh, G.J. (1989). "Item bias and item response theory," *International Journal of Education Research*. 13 : 127-143.
- Millsap, Roger E. and Howard T. Everson. (1993). "Methodology Review : Statistical Approaches for Assessing Measurement Bias," *Applied Psychological Measurement* . 17 (4) : 297-334.
- Narayanan, Pankaja and H. Swaminathan. (1994). "Performance of the Mantel-Haenszel and Simultaneous Item Bias Procedures for Detecting Item Function," *Applied Psychological Measurement* . 18(4) : 315-328.
- Osterind, S.L. (1983) . *Test Item Bias* . Beverly Hill : Sara McCune.
- Poham, W.J. (1990) . *Modern Educational Measurement : A Practitioner's Perspective*. 2nd ed. Prentice , Englewood Cliffs , New York.
- Potenza, Maria T. and Neil J. Dorans. (1995). "DIF Assessment for Polytomously Scored Items : A Framework for Classification and Evaluation," *Applied Psychological Measurement* . 19(1) : 23-37.
- Roger, H. Jane and H. Swaminathan. (1993). "A Comparison of Logistic Regression and Mantel-Haenszel Procedures for Detecting Differential Item Functioning," *Applied Psychological Measurement* . 17(2) : 105-116.
- Ruder, L.M., P.R. Geston and D.L. Knight. (1980). "A Monte Carlo Comparison of Seven Biased Item Detection Techniques," *Journal of Educational Measurement*. 17: 1-10.
- Scheuneman, Janice Dowd. (1987). An Experimental, Exploratory Study of Causes of Bias in test item," *Journal of Educational Measurement*. 14(2) : 97-118. Summer.

- Snedecor and Cochran. (1967) . *Statistical Method*. Iowa : Iowa State University Press.
- Subkovic, Micheal J. et al. (1984). " Empirical Comparison of Item Bias Detection Procedure with Bias Manipulation," *Journal of Educational Measurement*. 20(10): 49-59.Spring.
- Wainer,Howard and Henry I.Braun. (1988). *Test Validity*. New Jersey : Lawrence Erlbaum Associates.
- Wang,Ning and Suzanne Lane. (1996). " Detection of Gender-Related Differential Item Functioning in a Mathematics Performance Assessment," *Applied Measurement in Education* . 9(2) : 175-199.
- Welch,Catherine J. and Timothy R. Miller. (1995). " Assessing Differential Items Functioning in Direct Writing Assessment : Problems and an Example," *Journal of Educational Measurement*. 32(2) : 163-178.
- Wiersma,William and Stephen G.Jurs. (1990) . *Educational Measurement and Testing*. 2nd ed . Boston , London , Sydney Toronto : Allyn and Bacon.
- Wise,L.L. (1987). *Differential item difficulty indicators in small samples*. Paper Presented at the meeting of the American Educational Research Association , Washington, April.
- Zwick,Rebecca and Dorothy T.Thyer. (1996). " Evaluating the Magnitude of Differential Item Functioning in Polymous Items," *Journal of Educational and Behavioral Statistic*. 21(3) : 187-201.
- Zwick,Rebecca , Dorothy T.Thyer and John Mazzeo. (1997). " Descriptive and Inferential Procedures for Assessing Differential Item Function in Polytomous Items," *Applied Measurement in Education*. 10(4) : 321-344.
- Zwick,Rebecca,John R.Donoghue and Angela Grima. (1993). " Assessment of Differential Items Functioning for Performance Tasks," *Journal of Educational Measurement*. 30(3) : 233-251.

ภาคผนวก

ภาคผนวก ก
รายชื่อผู้เชี่ยวชาญ

ผู้ช่วยชาญในการพิจารณาคัดเลือกสถานการณ์และการเขียนข้อสอบที่ใช้วัดที่เหมาะสม
ในการออกข้อสอบของแบบทดสอบวัดความเข้าใจทางภาษา

- 1.รศ.วิญญา วิศาลาภรณ์ ภาควิชาวัดผลและวิจัยการศึกษา คณะศึกษาศาสตร์
มหาวิทยาลัยศรีนครินทรวิโรฒ
- 2.รศ.นิภา ศรีไพโรจน์ ภาควิชาวัดผลและวิจัยการศึกษา คณะศึกษาศาสตร์
มหาวิทยาลัยศรีนครินทรวิโรฒ
- 3.อาจารย์วาณี โชติพัฒนกุล หมวดภาษาไทย ร.ร.กุนนทีรุชารามวิทยาคม
ครูตัวอย่างหมวดวิชาภาษาไทย ของโรงเรียนมัธยม
ส่วนกลาง กลุ่ม 4
- 4.อาจารย์ประทุมพร จันทศิริ หมวดภาษาไทย ร.ร.ประชากรราษฎร์อุปถัมภ์
ครูต้นแบบเรื่องการนำภูมิปัญญาชาวบ้านมาใช้ในการ
จัดกิจกรรมการเรียนการสอน
- 5.อาจารย์ศิวรัตน์ กาญจนพฤษย์ หมวดภาษาไทย ร.ร.สุรศักดิ์มนตรี

ภาคผนวก ข

ค่าความยาก (p) และค่าอำนาจจำแนก (r)
ของแบบทดสอบวัดความเข้าใจทางภาษา

ตาราง 10 ค่าความยาก (p) และค่าอำนาจจำแนก (r) รายข้อของแบบทดสอบวัดความเข้าใจ
ทางภาษา จากการทดสอบครั้งที่ 1 ฉบับที่ 1

ข้อที่	p	r	ข้อที่	p	r
1*	0.52	0.24	22*	0.29	0.25
2	0.57	0.01	23*	0.35	0.24
3	0.49	0.16	24	0.33	0.14
4*	0.43	0.36	25*	0.23	0.22
5*	0.40	0.27	26	0.64	0.20
6	0.32	0.03	27	0.50	0.12
7	0.25	0.02	28*	0.44	0.34
8*	0.44	0.24	29	0.32	0.05
9*	0.36	0.24	30	0.51	0.16
10	0.42	0.14	31*	0.48	0.22
11	0.38	0.09	32*	0.44	0.23
12*	0.49	0.34	33	0.44	0.17
13	0.37	0.19	34*	0.37	0.26
14	0.42	0.08	35*	0.40	0.26
15	0.20	0.09	36	0.35	0.19
16*	0.40	0.34	37*	0.33	0.20
17	0.30	0.06	38	0.16	0.12
18*	0.45	0.24	39	0.29	0.17
19*	0.46	0.24	40*	0.36	0.24
20	0.46	0.13	41	0.50	0.14
21	0.46	0.19	42*	0.39	0.37

*เป็นข้อสอบที่เลือกไว้ใช้ในการวิจัย

ตาราง 11 ค่าความยาก (p) และค่าอำนาจจำแนก (r) รายข้อของแบบทดสอบวัดความเข้าใจ
ทางภาษา จากการทดสอบครั้งที่ 1 ฉบับที่ 2

ข้อที่	p	r	ข้อที่	p	r
1*	0.20	0.22	23	0.26	0.12
2*	0.30	0.25	24*	0.42	0.36
3*	0.25	0.27	25	0.41	0.17
4*	0.52	0.35	26*	0.28	0.28
5	0.35	0.29	27*	0.35	0.34
6*	0.43	0.42	28*	0.43	0.38
7	0.42	0.00	29	0.35	0.19
8*	0.52	0.36	30	0.42	0.11
9*	0.44	0.25	31	0.28	0.38
10	0.50	0.17	32	0.32	0.05
11*	0.51	0.28	33*	0.35	0.52
12*	0.39	0.21	34*	0.45	0.25
13	0.40	0.12	35*	0.34	0.53
14	0.48	0.08	36*	0.37	0.31
15	0.23	0.07	37*	0.34	0.26
16	0.36	0.18	38	0.34	0.12
17	0.32	0.01	39	0.32	0.04
18	0.41	0.23	40	0.28	0.05
19*	0.37	0.21	41	0.38	0.31
20*	0.43	0.31	42	0.17	0.22
21	0.40	0.17	43	0.25	0.15
22*	0.39	0.28	44	0.38	0.16
			45	0.28	0.32

*เป็นข้อสอบที่เลือกไว้ใช้ในการวิจัย

ภาคผนวก ค

ค่าความลำเอียงของแบบทดสอบวัดความเข้าใจทางภาษาจากการ
วิเคราะห์ด้วยและวิธีค่าความสัมพันธ์ของข้อสอบรายข้อ
กับคะแนนรวมทั้งหมด , วิธีไบซีเรียลและวิธี HW3

ตาราง 12 ค่าความลำเอียงจากการวิเคราะห์ความลำเอียงของข้อสอบ ตามวิธีค่าความสัมพันธ์ของ
ข้อสอบรายข้อกับคะแนนรวมทั้งหมด จำแนกตามลักษณะกลุ่มผู้สอบต่างกัน

ข้อที่	$N_R250:N_F250$ (ค่า Z)	$N_R500:N_F500$ (ค่า Z)	$N_R250:N_F500$ (ค่า Z)
1	-0.977	-2.095*	-0.792
2	0.595	0.825	1.090
3	-0.067	0.873	0.493
4	1.146	0.190	0.085
5	-0.573	0.253	0.805
6	1.382	0	1.129
7	0	0.952	0.077
8	-0.876	0.634	0.519
9	1.348	1.809	0.480
10	0.303	0.095	-0.142
11	1.955	2.317*	1.389
12	0.561	-0.428	0.064
13	-0.056	0.873	0
14	0	0.174	-0.545
15	1.483	0.619	1.506
16	0.831	-0.253	-0.753
17	0.382	-0.666	-1.311
18	1.764	-2.222*	2.610*
19	0.820	1.412	1.506
20	-0.067	0.476	-0.467
21	1.680	2.333*	-1.571
22	-1.089	1.730	-1.415
23	-2.853*	2.158*	2.779*
24	-0.808	-2.047*	1.597
25	0.539	-0.285	-0.857
26	-1.000	1.936	-1.090
27	-1.606	1.492	-1.363
28	-0.067	0.095	0.389
29	-1.696	-2.364*	1.779
30	-2.494*	-3.857*	-3.298*

ตาราง 12 (ต่อ)

ข้อที่	$N_R250:N_F250$ (ค่า Z)	$N_R500:N_F500$ (ค่า Z)	$N_R250:N_F500$ (ค่า Z)
31	-1.617	-0.920	-0.987
32	-0.123	-0.349	-0.025
33	-2.056*	-2.301*	-2.883*
34	0.471	-2.174*	-0.948
35	-0.651	2.111*	-1.532
36	1.505	1.015	1.545
37	-0.224	0.317	0.194
38	-0.224	-0.476	-1.181
39	1.516	0.095	1.246
40	1.157	0.841	0.688

* ข้อสอบที่มีความลำเอียง

ตาราง 13 ค่าความลำเอียงจากการวิเคราะห์ความลำเอียงของข้อสอบ ด้วยวิธีไบซีเรียล จำแนกตามลักษณะกลุ่มผู้สอบต่างกัน

ข้อที่	$N_R250:N_F250$ (ค่า Z)	$N_R500:N_F500$ (ค่า Z)	$N_R250:N_F500$ (ค่า Z)
1	1.747	-1.233	-0.746
2	-0.182	-0.508	-0.273
3	-0.178	0.333	1.186
4	0.917	-5.733*	1.573
5	0.547	-2.308*	-0.153
6	-1.082	-2.600*	1.126
7	0.288	1.733	-0.360
8	1.417	0.216	0.993
9	4.170*	2.933*	2.700*
10	-0.964	0.433	0.373
11	-2.011*	1.691	-2.160*
12	0	1.733	0.306
13	-1.123	-0.250	0.080
14	0.094	0.933	0.780
15	-0.647	-1.741	0.333
16	1.070	3.333*	3.353*
17	-3.670*	4.966*	-3.633*
18	1.641	-2.900*	1.506
19	-10.294*	-5.916*	-2.640*
20	-2.676*	-5.300*	1.233
21	-1.594	-2.158*	-1.140
22	-3.552*	-1.650	-0.540
23	-2.858*	-0.550	-0.660
24	-4.829*	-5.158*	-3.060*
25	-1.394	-4.683*	0.673
26	1.929	-1.141	0.126
27	-1.505	-2.658*	0.700
28	-1.047	2.850*	1.740
29	-3.552*	-5.341*	-2.420*
30	-7.523*	-6.008*	-4.726

ตาราง 13 (ต่อ)

ข้อที่	$N_R250:N_F250$ (ค่า Z)	$N_R500:N_F500$ (ค่า Z)	$N_R250:N_F500$ (ค่า Z)
31	-3.388*	-2.950*	1.233
32	1.017	-3.758*	-1.166
33	-2.164*	-2.791*	-0.840
34	-4.570*	-3.166*	-2.346*
35	-4.311*	-2.491*	-1.740
36	4.535*	1.838	5.626*
37	0.264	-0.333	0.100
38	0.064	-0.475	0.500
39	-3.052*	-3.550*	0.260
40	0.688	0.941	1.366

* ข้อสอบที่มีความลำเอียง

ตาราง 14 ค่าความลำเอียงจากการวิเคราะห์ความลำเอียงของข้อสอบ ด้วยวิธีHW3 จำแนก
ตามลักษณะกลุ่มผู้สอบต่างกัน

ข้อที่	$N_R250:N_F250$ (ค่า Z)	$N_R500:N_F500$ (ค่า Z)	$N_R250:N_F500$ (ค่า Z)
1	0.333	-1.033	1.330
2	0.746	-0.823	3.833*
3	0.629	0.377	0.220
4	0.156	-0.806	-1.570
5	-1.511	2.333*	-1.125
6	0.103	-2.203*	-0.428
7	-0.764	-9.667*	-2.570*
8	-0.284	-0.516	-0.500
9	-0.862	-2.032*	1.625
10	1.022	0.534	1.000
11	-0.123	0	0.285
12	0.126	0.016	0.375
13	0.022	-0.016	-0.125
14	0.659	1.596	1.625
15	-0.454	0.060	-1.140
16	-0.367	0.114	0.570
17	0.167	1.056	0.040
18	0.783	0.109	0.857
19	1.010	-1.606	1.711*
20	1.389	0.193	-0.660
21	0.727	0.026	1.139
22	-0.679	0.698	-0.867
23	-1.057	1.548	-3.880*
24	0.681	-1.029	-0.307
25	-0.356	0.140	1.038
26	-0.356	-1.049	-0.512
27	0.056	1.774*	-0.043
28	0.102	1.229	-0.558
29	-1.696*	1.246	-0.776
30	0.022	1.709*	-0.750

ตาราง 14 (ต่อ)

ข้อที่	$N_R250:N_F250$ (ค่า Z)	$N_R500:N_F500$ (ค่า Z)	$N_R250:N_F500$ (ค่า Z)
31	-3.388*	-2.950*	1.233
32	1.017	-3.758*	-1.166
33	-2.164*	-2.791*	-0.840
34	-4.570*	-3.166*	-2.346*
35	-4.311*	-2.491*	-1.740
36	4.535*	1.838	5.626*
37	0.264	-0.333	0.100
38	0.064	-0.475	0.500
39	-3.052*	-3.550*	0.260
40	0.688	0.941	1.366

* ข้อสอบที่มีความลำเอียง

ภาคผนวก ง
แบบทดสอบวัดความเข้าใจทางภาษา

คำชี้แจงวิธีทำแบบทดสอบ

๑. ข้อสอบนี้มีทั้งหมด ๔๐ ข้อ ใช้เวลาในการทำข้อสอบ ๖๐ นาที
๒. แบบทดสอบนี้เป็นแบบทดสอบเลือกตอบแบบถูกผิดชนิด ๕ ตัวเลือก วัดความเข้าใจทางภาษา โดยในแต่ละข้อจะกำหนดสถานการณ์มาให้ นักเรียนอ่านก่อน นักเรียนจะต้องพิจารณาจับ

เลขศูนย์ใจความหรือข้อความให้ตี ซึ่งตัวเลือกในแต่ละข้อนั้นมีตัวเลือกถูกและตัวเลือกผิดหลายตัว ดังนั้นนักเรียนจะต้องพิจารณาหาตัวเลือกถูกและตัวเลือกผิดให้ตี แล้วไปตอบในกระดาษคำตอบ โดยทำเครื่องหมายถูก (✓) ในช่องที่ตรงกับตัวเลือกที่เห็นว่าถูก และทำเครื่องหมายกากบาท (✕) ในช่องที่ตรงกับตัวเลือกที่เห็นว่าผิด ดังตัวอย่างข้อ (๐) ดังนี้

ให้นักเรียนอ่านข้อความต่อไปนี้ แล้วตอบคำถามข้อ ๐

“ไม่ขอรับทรัพย์สินจากใครหมด ทั้งยศศักดิ์สละลาภไพศาล
ไม่หวังให้ใครหนุนบุญบันดาล ชอบผลงานทุกอย่างสร้างด้วยตัว”

(๐). คำพูดดังกล่าวตรงกับสำนวนใด

- ก. ชักแม่น้ำทั้งห้า
- ข. กระซอกโฮกฮาก
- ค. กำปั้นทุบดิน
- ง. ขวานผ่าซาก
- จ. กลับกลอก

ในตัวอย่างข้อ ๐ จะเห็นว่าคำพูดดังกล่าวตรงกับสำนวน ในตัวเลือกข้อ ค และ ง ดังนั้นข้อ ค และ ง จึงเป็นคำตอบที่ถูกต้องส่วนตัวเลือกข้อ ก , ข และ จ เป็นคำตอบที่ผิด จึงไปตอบในกระดาษคำตอบ ดังนี้

(๐)

ก	ข	ค	ง	จ
✕	✕	✓	✓	✕

๓. อย่าเสียเวลาในการทำข้อใดข้อหนึ่งมากเกินไปเพราะจะทำให้เสียเวลา ถ้าคิดไม่ออกให้เว้นไว้แล้วข้ามไปทำข้ออื่นก่อน เพราะอาจจะมีข้อง่ายอยู่ด้านหลัง แล้วค่อยกลับมาทำข้อที่เว้นไว้ทีหลัง

๔. กรุณาอย่าขีดฆ่าหรือทำเครื่องหมายใด ๆ ลงในข้อสอบ ถ้ามีข้อสงสัยประการใดให้ถามผู้ดำเนินการสอบ

ขอให้นักเรียนตั้งใจทำให้ดี และ ครูขอสัญญาว่าจะเก็บข้อมูลเกี่ยวกับ ชื่อ-นามสกุล และโรงเรียนของนักเรียนเป็นความลับ

ให้นักเรียนอ่านข้อความต่อไปนี้ แล้วตอบคำถามข้อ ๑ - ๓

“ นักเตะเมืองลอดช่องถูกนักเตะพลังโสมยิงพุน ๕ ประตู ต่อ ๐ สิ้นสาย และเป็นไปตามฟอร์ม ที่นักวิจารณ์ส่วนใหญ่คาดหมายไว้ ไม่มีพลิกลือค ”

๑.คำว่า "พลิกลือค" มีความหมายตรงกับข้อใด

- ก. แก้วไข
- ข. ผิดคาด
- ค. ตรงข้าม
- ง. กลับกลาย
- จ. เปลี่ยนแปลง

๒.จากข้อความข้างต้นเป็นกีฬาประเภทใด

- ก. ตะกร้อ
- ข. ฟุตบอล
- ค. แฮร์บอล
- ง. วอลเลย์บอล
- จ. บาสเก็ตบอล

๓.คำว่า “ เป็นไปตามฟอร์ม ” มีความหมายตรงกับข้อใด

- ก. เป็นไปตามคาด
- ข. เป็นไปตามที่หวัง
- ค. เป็นไปตามที่ปรารถนา
- ง. เป็นไปตามที่วิจารณ์ไว้
- จ. เป็นไปตามที่ประกาศไว้

ให้นักเรียนอ่านบทความนี้ แล้วตอบคำถามข้อที่ ๔ - ๕

“ แม่ค้าที่เป็นคนจีนจะแตกต่างจากแม่ค้าที่เป็นคนไทย แม่ค้าจีนมักจะมีอารมณ์เย็นกว่าแม่ค้าไทย อย่างเช่น ลูกค้าต้องรอราคามากไป แม่ค้าจีนก็เพียงส่ายหน้าและพูดว่าขายไม่ได้ เพราะต้นทุนแพง จะซื้อหรือไม่ซื้อก็แล้วกันไป แต่ในขณะที่เดียวกันถ้าเป็นแม่ค้าไทยจะค่อนข้างกลัวแล้วจิบปากจิบคอจ้องดูใส่ลูกค้าว่าอยากกินแต่ไม่มีปัญญจะซื้อให้รอชาติหน้า ถ้าบังเอิญไปเจอลูกค้าปากจัดเข้าชาวบ้านก็พลอยได้ฟังลำตัดฟรีๆหนึ่งรอบ เข้าตำราขิงก็ราขากก็แรงนั่นเอง ที่ว่าอย่างนี้มีใช้กล่าวหาว่าแม่ค้าไทยหยาบคายหรือร้ายกาจ ยกย่องแม่ค้าจีนเพียงแต่ที่ฉันเห็นมาในตลาดเท่านั้น ”

๔.คำว่า “ ลำตัด ” ในเรื่องนี้ มีความหมายตรงตามข้อใด

- ก. การประชิดประชันกัน
- ข. การต่อล้อต่อเถียงกัน
- ค. การละเล่นพื้นเมือง
- ง. การปะทะคารมกัน
- จ. การทะเลาะกัน

๕.ผู้แต่งมีวัตถุประสงค์อย่างไร

- ก. ยกย่องแม่ค้าจีน
- ข. ตำหนิติเตียนแม่ค้าไทย
- ค. แสดงให้เห็นความแตกต่างของแม่ค้าไทย
- ง. แสดงให้เห็นถึงความเป็นนักการค้าของคนจีน
- จ. เปรียบเทียบคุณลักษณะระหว่างแม่ค้าจีนกับแม่ค้าไทย

ให้นักเรียนอ่านข้อความต่อไปนี้ แล้วตอบคำถามข้อ ๖ - ๘

“ การออกกำลังกายเป็นประจำจนกลายเป็นนิสัย ให้เหมือนกับการล้างหน้าแปรงฟัน เป็นสิ่งจำเป็นอย่างยิ่ง ท่านจะมองเห็นความสำคัญในการล้างหน้า แปรงฟัน ได้ชัดเจนกว่า ถ้าท่านไม่ปฏิบัติตามภารกิจนั้นแล้วใบหน้าจะสกปรกเป็นสิ่วฝ้า กลิ่นปากเหม็น ท่านมองเห็นผลลัพธ์ภายในวันเดียวเท่านั้น แต่การออกกำลังกายให้ผลระยะยาวนาน คนจึงมองไม่ค่อยเห็นถึงความสำคัญของการออกกำลังกายที่แท้จริง โดยการออกกำลังกายช่วยกระตุ้นอวัยวะทุกระบบให้ทำงานมีประสิทธิภาพ จึงส่งผลให้ผู้นั้นแข็งแรง สุขภาพดี ทรวดทรงงดงาม ”

๖. เรื่องนี้ให้ข้อคิดในเรื่องใด

- ข. ปัจจุบันคนเราไม่แข็งแรง
- ข. ประโยชน์ในการออกกำลังกาย
- ค. สถานที่ในการออกกำลังกาย
- ง. คนไม่เห็นความสำคัญของการออกกำลังกาย
- จ. การออกกำลังกายเป็นรักษาสุขภาพอย่างหนึ่ง

๗. เหตุผลใดที่ตรงกับคำที่ว่า “ การออกกำลังกายทำให้สุขภาพและทรวดทรงดี ”

- ก. ทำให้นอนหลับ
- ข. ทำให้กล้ามเนื้อแข็งแรง
- ค. ช่วยกระตุ้นอวัยวะทุกระบบ
- ง. ทำให้รับประทานอาหารได้มาก
- จ. ทำให้ร่างกายพร้อม ที่จะรับงานหนัก

๘. ผู้เขียนมีจุดประสงค์ใดจึงเขียนข้อความนี้ขึ้นมา

- ข. แนะนำวิธีการออกกำลังกาย
- ข. ชักชวนให้ผู้อ่านออกกำลังกาย
- ค. โน้มน้าวให้สละเวลาออกกำลังกาย
- ง. ชี้แจงเรื่องการดูแลสุขภาพกายและจิต
- จ. เปรียบเทียบให้เห็นความสำคัญของการออกกำลังกาย

ให้นักเรียนอ่านข้อความต่อไปนี้ แล้วตอบคำถามข้อ ๙ - ๑๑

“ พระบวชใหม่เม้มริมฝีปากแล้วมองตามเธอ จนออกไปพ้นประตูโบสถ์ ผ้าเหลืองไม่ใช่เครื่องชำระกิเลส เพียงแต่ห่มผ้าเหลืองไม่ได้ความหมายว่ากิเลสหมดสิ้นไปแล้ว หากแต่ต้องบำเพ็ญเพียรชำระจิตไปวันละเล็กละน้อย จนกว่าจะทุเลาเบาบางหรือหมดสิ้นไป ”

๙. ข้อความข้างต้นกล่าวถึงเรื่องใด

- ก. การบวชเป็นการชำระกิเลสได้ดีที่สุด
- ข. การหมั่นปฏิบัติรักษาศีลทำให้ละกิเลสได้
- ค. การบวชเป็นหนทางช่วยให้มนุษย์พ้นทุกข์
- ง. คนที่ห่มผ้าเหลืองมิใช่เป็นผู้ที่ละกิเลสได้เสมอไป
- จ. พระคือปุถุชนที่ต้องการบำเพ็ญเพียรเพื่อชำระจิตใจ

๑๐. เครื่องชำระกิเลสในที่นี้หมายถึงสิ่งใด

- ก. พระภิกษุ
- ข. พระธรรม
- ค. ความอดทน
- ง. การปฏิบัติตน
- จ. การรู้ข้อปฏิบัติ

๑๑. หากผู้อ่านนำข้อคิดนี้ไปใช้ในชีวิตประจำวัน ผู้อ่านจะเป็นบุคคลเช่นไร

- ก. รักษาศีล
- ข. หมั่นให้ทาน
- ค. ไม่ให้น้องสาวเข้าวัด
- ง. ศึกษาพระธรรมวินัย
- จ. หมั่นทำใจให้บริสุทธิ์

ให้นักเรียนอ่านข้อความนี้ แล้วตอบคำถามข้อ ๑๒ - ๑๓

“ เราต้องจำไว้ว่าชาวต่างประเทศนั้น อาจจะเต็มใจอยู่กับเรา กินข้าวกับเรา ค้าขายกับเรา และเป็นสหายกับเราได้ แต่เราจะหวังว่าเขาจะตายกับเราด้วยนั้นไม่ได้ ”

๑๒. ข้อความนี้ตรงกับสำนวนใด

- ก. รู้หน้าไม่รู้ใจ
- ข. เสียชีวิตอย่าเสียสัตย์
- ค. คบคนพาลพาลไปหาผิด
- ง. ไม่ตั้งอยู่ในความประมาท
- จ. เพื่อนกินหาง่ายเพื่อนตายหายาก

๑๓. ข้อความข้างต้นนี้กล่าวในลักษณะใด

- ก. ให้สติ
- ข. ข้อคิด
- ค. ตักเตือน
- ง. เสนอแนะ
- จ. ประชดประชัน

ให้นักเรียนอ่านข้อความต่อไปนี้ แล้วตอบคำถามข้อ ๑๔ - ๑๗

“ มีข่าวประเทศไทยติดอันดับสองของโลกในเรื่องความสกปรก ฉันได้ยินแล้วใจหาย คงเป็นจริงอย่างที่ว่า เพราะเวลาฉันเดินไปโรงเรียนเห็นเศษกระดาษ เศษขยะอยู่ทั่วไป แม้กระทั่งในโรงเรียนของฉันเอง เขาน่าจะออกกฎหมายแก้ไขกันได้แล้ว ”

๑๔. ข้อใดเป็นสาเหตุทำให้ผู้พูดว่า “ ใจหาย ”

- ก. รู้สึกอับอายคนต่างชาติ
- ข. ความเห็นแก่ตัวของคนไทย
- ค. พบเห็นขยะมากมายตามถนน
- ง. ความสกปรกลุกลามมาถึงโรงเรียน
- จ. ประเทศไทยสกปรกเป็นอันดับต้นๆ ของโลก

๑๕. “เขา” ในที่นี้หมายถึงใคร

- ก. ครู
- ข. ตำรวจ
- ค. รัฐบาล
- ง. นักการเมือง
- จ. ฝ่ายบริหาร

๑๖. ผู้พูดมีความคิดที่จะแก้ปัญหาอย่างไร

- ก. กำหนดให้มีข้อบังคับ
- ข. กำหนดให้มีบทลงโทษ
- ค. กำหนดให้มีการจับปรับ
- ง. กำหนดให้มีการดำเนินคดี
- จ. กำหนดให้มีการรณรงค์เผยแพร่

๑๗. จากข้อความนี้ นักเรียนมีความเห็นสอดคล้องกับผู้พูดในการออกกฎหมายหรือไม่

- ก. เห็นด้วยเพราะบ้านเมืองเราสกปรกมาก
- ข. คัดค้านเพราะประเทศไทยไม่ได้สกปรกทุกที่ไป
- ค. เห็นด้วยเพราะเพื่อแก้ไขนิสัยมักง่ายของคนไทย
- ง. คัดค้านเพราะประเทศมีพื้นที่เล็กน้อยกว่าประเทศอื่น
- จ. เห็นด้วยเพราะมีการดำเนินการเพื่อรักษาความสะอาด

ให้นักเรียนอ่านคำประพันธ์ต่อไปนี้ แล้วตอบคำถามข้อ ๑๘ - ๑๙

“ ก่อนนี้โลกสวยด้วยป่า บัดนี้ป่าไม้ถูกโรมรัน ครุฑาป่าของหนูก้อยู่ที่ไหน หรือหนูจะได้ชมเพียงแต่	สดชื่นเย็นตาพาสุขสันต์ พรั่งนี้คงมีวันโลกสิ้นใจ ต้นไม้และใบอย่างไหนแท้ กาบแกใบแห้งแสลงว่าจริง ”
---	--

๑๘. คำประพันธ์นี้กล่าวถึงเรื่องใด

- ก. ป่าไม้
- ข. สัตว์ป่า
- ค. ธรรมชาติ
- ง. ใบไม้แห้ง
- จ. แม่น้ำลำธาร

๑๙. ใจความสำคัญของคำประพันธ์นี้คือข้อใด

- ก. เด็กอยากอยู่ในป่า
- ข. ป่าในปัจจุบันเหลือน้อยเต็มที
- ค. ถ้าไม่มีต้นไม้โลกเราก็คงอยู่ไม่ได้
- ง. ป่าไม้มีคุณค่าควรรักษาให้คงอยู่ต่อไป
- จ. ภัยธรรมชาติเป็นเหตุให้ป่าไม้ถูกทำลาย

ให้นักเรียนอ่านข้อความต่อไปนี้ แล้วตอบคำถามข้อ ๒๐ - ๒๑

“ บ้านเราเป็นบ้านเมืองที่ไม่เคยเอาจริงเอาจังกับคนเมาขับรถ ซึ่งเป็นเหตุให้ผู้อื่นถึงตาย และพิกลพิการมานักต่อนัก ขณะที่ประเทศที่เจริญทางกฎหมายนั้น มีบทลงโทษหนักจนคนเมาไม่กล้าขับรถ การซื้อขายก็เช่นกัน ยังทำกันได้อย่างเสรี ไม่มีกฎระเบียบที่จะดูแลอย่างเคร่งครัดนอกจากยาไม่ก็ประเภทเท่านั้น ”

๒๐. ข้อความนี้แสดงความรู้สึกเช่นไร

- ก. ท้อใจ
- ข. ช้ำใจ
- ค. สิ้นเวร
- ง. ผิดหวัง
- จ. เบื่อหน่าย

๒๑. คำว่า “ ไม่เคยเอาจริงเอาจัง ” ตรงกับสำนวนตามข้อใด

- ก. หางแถว
- ข. ลดราวาศอก
- ค. เอาหูไปนาเอาตาไปไร่
- ง. ไม้หลักปักเลน
- จ. สองฝักสองฝาย

ให้นักเรียนอ่านข้อความต่อไปนี้ แล้วตอบคำถามข้อ ๒๒ - ๒๓

“ คนที่ตามใจตนเองอยู่เสมอคือ คนที่ไม่มีความอดทนที่จะรักษาระเบียบ วินัย หรือ ประเพณีอันดีงามได้ ถ้าคนไทยส่วนมากไม่มีความอดทนก็กลายเป็นชนชาติที่ไม่มีวินัย ”

๒๒.สาเหตุที่คนไทยไม่มีระเบียบ วินัย มาจาก ข้อใด

- ก. ค่านิยม
- ข. การกระทำ
- ค. ประเพณีนิยม
- ง. ตามใจตนเอง
- จ. ไม่มีความอดทน

๒๓.เหตุการณ์ใดตรงกับข้อความข้างต้น

- ก. มานีขอแทรกแถวเพื่อนซื้ออาหาร
- ข. วีระชอบแกล้งเพื่อนในชั่วโมงเรียน
- ค. สมใจส่งงานอาจารย์หลังหมดกำหนดส่ง
- ง. เมื่อเห็นคนแก่ขึ้นรถเมล์ ปิดิแกล้งนั่งหลับ
- จ. พิมพ็ใจไม่สนใจเรียนในวิชาที่ตนไม่ชอบ

ให้นักเรียนอ่านข้อความต่อไปนี้ แล้วตอบคำถามข้อ ๒๔ - ๒๕

“ สรรพสมบัติอะไรจะประเสริฐเท่าปัญญาไม่มี เพราะปัญญาเป็นสมบัติติดตัว จะซื้อขาย จ่ายแจกมิได้ แม้ใครจะแย่งชิงเอาไปก็ไม่ได้ การเรียนรู้ทำให้เกิดปัญญาและส่งให้ประพัตติ ความประพัตติทำให้ได้รับตำแหน่งที่ดี ตำแหน่งที่ดีช่วยให้เกิดทรัพย์ ทรัพย์หนุนให้ ประพัตติธรรมสะดวก ถ้าไม่มีในผู้ใด ผู้นั้นก็ไม่ผิดอะไรกับคนตาบอด ”

๒๔.ถ้าขาดสิ่งใดต่อไปนี้ แล้วทำให้เป็นคนตาบอด

- ก. เงินทอง
- ข. สติปัญญา
- ค. อำนาจหน้าที่
- ง. ความประพัตติ
- จ. ประสบการณ์การเรียนรู้

๒๕. “ คนตาบอด ” ในข้อความข้างต้นหมายถึงคน เช่นไร

- ก. คนที่ไม่มีสติปัญญา
- ข. คนที่มองไม่เห็น
- ค. คนไม่ได้เรียน
- ง. คนไม่มีความรู้
- จ. คนที่มีกิเลส

ให้นักเรียนอ่านข้อความต่อไปนี้ แล้วตอบคำถามข้อ ๒๖ – ๒๘

“ หมู่บ้านของเราแห่งเดียวที่ยังล้าหลังเขาอยู่ ทั้งนี้เพราะความไม่ปรองดองกันระหว่างครอบครัวของผมกับครอบครัวของชาโตรุ แต่ละฝ่ายต้องการจะเป็นผู้นำของหมู่บ้าน ดังนั้นจึงไม่มีใครทำอะไรกันขึ้นเลยเป็นที่น่าเสียดาย ไม่เพียงแต่จะทำให้สองตระกูลจะเสียประโยชน์ แต่ยังเป็นผลร้ายแก่หมู่บ้านอีกด้วย ในขณะที่เราบาดหมางไม่ลงรอยกัน คนอื่น ๆ ก็ต้องมาองง้อขอให้เราช่วย แล้วก็ไม่มีใครทำอะไรกัน หมู่บ้านก็ไม่เจริญรุ่งเรืองขึ้นเลย ผมเชื่อว่าพวกเราสองตระกูลน่าจะหันหน้าเข้าหากันได้ ร่วมใจกันทำประโยชน์ให้กับหมู่บ้าน วิธีที่ทำกันอยู่ในปัจจุบันนี้เป็นเรื่องโงเขลามาก เมื่อเราโตเป็นหนุ่มเราจะไม่ทำอย่างนี้ เราจะคิดถึงประโยชน์ของหมู่บ้านเราให้มากกว่านี้สักหน่อย ”

๒๖. ผู้กล่าวข้อความนี้ มีลักษณะนิสัยอย่างไร

- ก. ตรงไปตรงมา
- ข. รักวงศ์ตระกูล
- ค. รักความยุติธรรม
- ง. รักความก้าวหน้า
- จ. ไม่ยึดติดกับความหลัง

๒๗. ในอนาคต สภาพของหมู่บ้านนี้น่าจะเป็นอย่างไร

- ก. แร้นแค้นยิ่งขึ้น
- ข. ท้นสมัยมากขึ้น
- ค. เสื่อมโทรมมากขึ้น
- ง. ได้รับการพัฒนามากขึ้น
- จ. มีสภาพเศรษฐกิจที่ดีขึ้น

๒๘. “ วิธีที่ทำกันอยู่ในปัจจุบันนี้เป็นเรื่องโงเขลามาก ” ในที่นี้หมายถึงการกระทำในข้อใด

- ก. การร่วมมือกันทำประโยชน์
- ข. การทะเลาะวิวาทไม่ลงรอยกัน
- ค. การไม่ทำประโยชน์ให้กับส่วนรวม
- ง. การเสียประโยชน์ของสองตระกูล
- จ. การแย่งชิงความเป็นผู้นำของหมู่บ้าน

ให้นักเรียนอ่านข้อความต่อไปนี้ แล้วตอบคำถามข้อ ๒๙ - ๓๓

“ ในสมัยก่อนผู้อ่านออกเขียนหนังสือได้มีจำนวนน้อยมาก ส่วนใหญ่เป็นข้าราชการ ส่วนชาวบ้านหรือคนพื้น ๆ หายาก เพราะไม่ค่อยจะมีโรงเรียนให้เรียน ต้องอาศัยเรียนกับพระตามวัดซึ่งเป็นโรงเรียนของเด็กผู้ชายเท่านั้น จึงไม่นิยมส่งลูกหลานให้เรียนหนังสือ ยิ่งลูกผู้หญิงแล้ว พ่อแม่ เห็นว่าไม่จำเป็นเลยสมควรอยู่ทำงานบ้านดีกว่า แล้วถ้าให้เรียนหนังสือแล้วทำให้ใจแตก เขียนจดหมายถึงผู้ชายหรือเรียกว่า “ เพลงยาว ” ผู้หญิงจึงไม่ได้เล่าเรียนอย่างผู้ชาย อย่างดีก็เรียนกันเองพอเขียนชื่อสกุลตัวเองเป็นเท่านั้น ”

๒๙. ข้อความนี้กล่าวถึงเรื่องใด

- ก. สถานที่เรียนในสมัยก่อน
- ข. การเรียนหนังสือสมัยก่อน
- ค. ความเชื่อของคนสมัยก่อน
- ง. ความเป็นอยู่ของคนสมัยก่อน
- จ. โอกาสทางการเรียนของผู้หญิง

๓๐. “ เพลงยาว ” หมายถึงข้อใด

- ก. จดหมายรัก
- ข. จดหมายส่วนตัว
- ค. เพลงที่มีเนื้อเรื่องยาวๆ
- ง. เพลงพื้นเมืองสมัยก่อน
- จ. จดหมายที่เขียนโต้ตอบกันระหว่างชายกับหญิง

๓๑. การเรียนหนังสือสมัยก่อนต่างจากการเรียนหนังสือในปัจจุบันอย่างไร

- ก. สถานที่เรียน
- ข. สิทธิในการเรียน
- ค. คุณภาพของการเรียนการสอน
- ง. ความสามารถในการสอนของครู
- จ. อุปกรณ์ในการแสวงหาความรู้ของผู้เรียน

๓๒. ผู้หญิงสมัยก่อนที่กล่าวถึงในเรื่องนี้มีลักษณะอย่างไร

- ก. เชื้อฟ่ง
- ข. หัวอ่อน
- ค. ใจเขลา
- ง. ไม่ทันคน
- จ. วานอนสอนง่าย

๓๓. ข้อความนี้ให้ข้อคิดเกี่ยวกับผู้หญิงของคนสมัยก่อนอย่างไร

- ก. ผู้หญิงเป็นช้างเท้าหลัง
- ข. ผู้หญิงไม่ควรรับราชการ
- ค. ผู้หญิงเหยียบขี่ไก่ไม่ฝ่อ
- ง. ผู้หญิงควรมีหน้าที่ดูแลครอบครัว
- จ. ผู้หญิงมักทำความเสื่อมเสียมาสู่ครอบครัว

ให้นักเรียนอ่านข้อความต่อไปนี้ แล้วตอบคำถามข้อ ๓๔ - ๓๖

“ไม่ว่าสิ่งใด มิไกลเกินฝัน หากเริ่มเก็บเล็กผสมน้อยตั้งแต่วันนี้ สิ่งที่ย่อมเป็นจริงได้ ”

๓๔. ข้อความนี้เป็นข้อความประเภทใด

- ก. เสียดสี
- ข. ขอร้อง
- ค. ตักเตือน
- ง. ให้กำลังใจ
- จ. ให้คำแนะนำ

๓๕. ผู้ที่ปฏิบัติตามข้อความข้างต้นจะเป็นคน
เช่นไร

- ก. ยากจน
- ข. มีผู้นับถือ
- ค. ไม่อดอยาก
- ง. มีฐานะมั่นคง
- จ. มีความร่ำรวย

๓๖. จากข้อความข้างต้นมีความหมาย ตรงข้าม กับสำนวนใด

- ก. กระเซอกันรัว
- ข. ใช้เงินเป็นเบี้ย
- ค. สิบเบี้ยใกล้มือ
- ง. รู้จักเก็บรู้จักเขี่ย
- จ. มีสิ่งพึงบรรจบให้ครบบาท

ให้นักเรียนอ่านข้อความต่อไปนี้ แล้วตอบคำถามข้อ ๓๗ - ๓๘

“โบราณกล่าวว่า “ดักลอบต้องหมั่นกู่” คำพังเพยนี้มักพูดไปในทำนองหนุ่มเจ้าชู้ แต่เดิมเขาคงผูกคำนี้ขึ้นเพื่อเตือนใจให้ขยันทำมาหากิน ไปวางลอบที่ไหนก็ต้องหมั่นไปดู เพราะแต่ก่อนเมืองไทยอุดมสมบูรณ์ เรียกว่าในน้ำมีปลาในนามีข้าว แหล่งน้ำที่เอาจลอบไปวางดักปลาต้องเป็นทางน้ำไหล ก่อนวางก็หาเศษอาหารเน่าๆ คาวๆ ใส่ไว้ข้างในเพื่อล่อปลา ปลาที่เข้ามาจะเข้าได้อย่างเดียวไม่สามารถออกไปได้ ”

๓๗. “ดักลอบต้องหมั่นกู่” มีความหมาย
ตรงกับสำนวนใด

- ก. พายเรือทวนน้ำ
- ข. ฝนทั่งให้เป็นเข็ม
- ค. งมเข็มในมหาสมุทร
- ง. ปากเป็นเอกเลขเป็นโท
- จ. ความพยายามอยู่ที่ไหน
ความสำเร็จอยู่ที่นั่น

๓๘. ข้อความข้างต้น ไม่ได้ กล่าวถึงสิ่งใด

- ก. วิธีการวางลอบ
- ข. วัสดุในการทำลอบ
- ค. สถานที่ในการวางลอบ
- ง. ส่วนประกอบของลอบ
- จ. ขั้นตอนในการทำลอบ

ให้นักเรียนอ่านข้อความนี้ แล้วตอบคำถามข้อ ๓๙ - ๔๐

“ การที่เราทำตัวเป็นคนโง่ฟุ้งคนอื่นเขาแสดงความฉลาดแล้วเก็บความรู้จากคนอื่น นั่นแหละเป็นการกระทำของคนฉลาด เพราะคนอื่นเขารู้เองว่าเราโง่หรือฉลาดแค่ไหน ฉะนั้นจึงทำตัวเป็นคนโง่ไว้เสมอ อย่าคุยโวโอ้อวดว่าเราเป็นคนฉลาดเป็นอันขาดถึงแม้เราจะมี ความฉลาดสักเพียงใดก็ตาม โดยเฉพาะเมื่ออยู่ในสังคมคนหมู่มาก ”

๓๙. ข้อความข้างต้นนี้มีความหมายตรงกับสำนวนใด

- ก. คมในฝัก
- ข. น้ำนิ่งไหลลึก
- ค. เสือซ่อนเล็บ
- ง. ฉลาดแกมโกง
- จ. เขี้ยบบเรื่อสองแคม

๔๐. " การคุยโวโอ้อวดว่าเราเป็นคนฉลาด " ตรงกับสำนวนใด

- ก. ยกตนข่มท่าน
- ข. เปิดขันประชันไก่
- ค. หมาขี้ไม่มีใครยกหาง
- ง. สอนหนังสือสั่งฆราวาส
- จ. ชักแม่น้ำทั้งห้า

ประวัติย่อผู้วิจัย

ชื่อ	นายพิรุฬพงษ์ แนวทอง
วันเดือนปีเกิด	3 มีนาคม 2519
สถานที่เกิด	กิ่งอำเภอนมดงรัก จังหวัดสุรินทร์
สถานที่อยู่ปัจจุบัน	2000 / 22 ถนนประชาสงเคราะห์ แขวงดินแดง เขตดินแดง กรุงเทพมหานคร 10400
ตำแหน่งหน้าที่การงาน	อาจารย์ 1 ระดับ 4
สถานที่ทำงานปัจจุบัน	โรงเรียนสามเสนนอก(ประชาราษฎร์อนุกุล) เขตดินแดง กรุงเทพมหานคร 10400
ประวัติการศึกษา	
พ.ศ.2533	มัธยมศึกษาปีที่ 3 จากโรงเรียนวัดชัยศรีสะอาด
พ.ศ.2536	มัธยมศึกษาตอนปลาย(สอบเทียบ)จังหวัดสุรินทร์
พ.ศ.2540	ค.บ. (การวัดผลการศึกษา) จากสถาบันราชภัฏสุรินทร์
พ.ศ.2544	กศ.ม. (การวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ