

317.26093

พ. 2721

5-3

การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบ โดยใช้ขนาดกลุ่มผู้สอบ
และวิธีวิเคราะห์ต่างกัน

ปริญญาานิพนธ์

ของ

พรรณี จินตมาศ

15 ส.ค. 2541

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา

ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา

กันยายน 2540

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบ โดยใช้ขนาดกลุ่มผู้สอบ
และวิธีวิเคราะห์ต่างกัน

บทคัดย่อ
ของ
พรรณิ จินตมาศ

เสนอต่อมหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา
กันยายน 2540

บทที่ ๑๐

การวิจัยครั้งนี้มีจุดมุ่งหมายเพื่อศึกษาผลการวิเคราะห์ความลำเอียงต่อเพศของข้อสอบ จากแบบทดสอบคณิตศาสตร์โจทย์ปัญหาโดยจำแนกในแต่ละวิธีวิเคราะห์ความลำเอียง 3 วิธี คือ วิธีแปลงค่าความยาก วิธีแมนเทิล-แฮนส์เซลและวิธีชิปเทสท์ และในแต่ละขนาดกลุ่มผู้สอบ คือ ขนาดกลุ่มผู้สอบ 500 คนและขนาดกลุ่มผู้สอบ 1,000 คน โดยเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงและเปรียบเทียบค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว ในการศึกษาครั้งนี้มีกลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2539 ของโรงเรียนสังกัดกรมสามัญศึกษา ส่วนกลาง จำนวน 2,200 คน ซึ่งเลือกมาโดยการสุ่มแบบแบ่งชั้น มีขนาดโรงเรียนเป็นชั้นและโรงเรียนเป็นหน่วยการสุ่ม

ผลการวิจัยพบว่า เมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 500 คนวิธีชิปเทสท์พบข้อสอบที่มีความลำเอียงมากที่สุด และวิธีแปลงค่าความยากพบข้อสอบที่มีความลำเอียงน้อยที่สุด โดยจำนวนข้อสอบที่ลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทุกวิธีวิเคราะห์ และเมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 1,000 คน วิธีแมนเทิล-แฮนส์เซลพบข้อสอบที่มีความลำเอียงมากที่สุด วิธีแปลงค่าความยากไม่พบข้อสอบที่ลำเอียง โดยจำนวนข้อสอบที่ลำเอียงจากการวิเคราะห์ด้วยวิธีแปลงค่าความยากกับวิธีแมนเทิล-แฮนส์เซลและวิธีแปลงค่าความยากกับวิธีชิปเทสท์แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติสำหรับจำนวนข้อสอบที่ลำเอียงจากการวิเคราะห์ด้วยวิธีแปลงค่าความยากระหว่างกลุ่มผู้สอบขนาด 500 คนและกลุ่มผู้สอบขนาด 1,000 คน จะมีจำนวนข้อสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

ค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกเมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 500 คน จากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากกับวิธี ชิปเทสท์มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และเมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 1,000 คน ค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่ลำเอียงออกแล้วจากการวิเคราะห์ความลำเอียง 3 วิธี มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ สำหรับค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่ลำเอียงออกจากการวิเคราะห์ด้วยวิธีชิปเทสท์ระหว่างกลุ่มผู้สอบ 500 คนและกลุ่มผู้สอบ 1,000 คนมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 นอกนั้นมีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

**A COMPARISON OF THE ITEM BIAS ANALYSIS USING DIFFERENT
SAMPLE SIZES AND APPROACHES**

**AN ABSTRACT
BY
PHANNEE CHINTAMAS**

**Present in partial of the requirements for the Master of
Education degree in Education Measurement
at Srinakharinwirot University
September 1997**

The purpose of this research was to compare the biased items of sex to the Mathematical Problem Solving using three different approaches ; Transformed Item Difficulty (TID), Mantel-Haenszel (MH), and SIBTEST (SIB) and two different sample sizes, 500 and 1,000 students . The comparison of the item bias was to determine numbers of the biased items and Split-Half reliabilities of the test after the deleted biased items. The sample of the study was 2,200 of students Mathayom Suksa 1, 2ND semester academic year 1996 from secondary school under Department Of General Education selected by stratified random sampling, taking school's size as a strata as well as considering schools as sampling units.

The results of the study showed that from the 500 students sample size SIBTEST approach (SIB) gave the largest number of biased items while Transformed Item Difficulty approach (TID) gave the least. They were found that no significantly of each approach. When the sample size was 1,000 students Mantel-Haenszel approach (MH) gave the largest number of biased items while Transformed Item Difficulty approach (TID) was not found the biased item. The number of biased item between Transformed Item Difficulty (TID) and Mantel-Haenszel (MH) , Transformed Item Difficulty (TID) and SIBTEST (SIB) were found to differ significantly at .05 level while the other was found to have no significance . In addition Transformed Item Difficulty (TID) between the different sample sizes was found to differ significantly at .05 level while the others were found to have no significance.

The Split-Half reliabilities of the test without biased items in the 500 students sample size between Transformed Item Difficulty approach (TID) and SIBTEST approach (SIB) were found to differ significantly at .05 level while the others were found to have no significance. In the 1,000 students sample size, the Split-Half reliabilities of the test without biased items were found no significance of each approach. In addition comparing the different sample sizes, the Split-Half reliabilities of the test without biased item, using SIBTEST approach (SIB) were found to differ significantly at .05 level while the others were found to have no significance.

คณะกรรมการควบคุมและคณะกรรมการสอบได้พิจารณาปริญญาบัตรฉบับนี้แล้ว เห็นสมควรรับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอก การวัดผลการศึกษา ของมหาวิทยาลัยศรีนครินทรวิโรฒได้

คณะกรรมการควบคุม

..... ประธาน
(รศ. อังคณา สายยศ)
..... กรรมการ
(รศ. อาวุธ วัฒนสิน)

คณะกรรมการสอบ

..... ประธาน
(รศ. อังคณา สายยศ)
..... กรรมการ
(รศ. อาวุธ วัฒนสิน)
..... กรรมการที่แต่งตั้งเพิ่มเติม
(รศ. ชูศรี วงศ์รัตน์)
..... กรรมการที่แต่งตั้งเพิ่มเติม
(ผศ. เขาวานา ขวลิขิตอำรง)

บัณฑิตวิทยาลัยอนุมัติให้รับปริญญาบัตรฉบับนี้ เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร ปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา ของมหาวิทยาลัยศรีนครินทรวิโรฒ

..... คณบดีบัณฑิตวิทยาลัย
(ดร. ศิริยุภา พูลสุวรรณ)

วันที่ 23 เดือน กันยายน พ.ศ. 2540

ขอขอบพระคุณ
มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร
ที่พิจารณาให้
“ทุนพระพรหมคุณาภรณ์”
เพื่อสนับสนุนในการทำปริญญาโท

ประกาศคุณูปการ

ปริญญาานิพนธ์ฉบับนี้สำเร็จลุล่วงไปได้ด้วยความช่วยเหลือจากบุคคลที่เกี่ยวข้องหลายท่าน ผู้วิจัยขอกราบขอบพระคุณมา ณ ที่นี้ ดังต่อไปนี้

รองศาสตราจารย์อังคณา สายยศ ได้กรุณาช่วยชี้แนะแนวทางในการแก้ปัญหา เป็นที่ปรึกษาและให้กำลังใจแก่ผู้วิจัย ทั้งในฐานะอาจารย์ที่ปรึกษาและประธานที่ปรึกษาปริญญาานิพนธ์ ด้วยความเอาใจใส่อย่างดีเยี่ยมตลอดมา

รองศาสตราจารย์อาวุธ วัฒนสิน ได้กรุณาแนะนำให้ทั้งความรู้และข้อคิดเห็น เพื่อนำไปแก้ไข ปรับปรุงข้อบกพร่องต่าง ๆ แก่ผู้วิจัย ในฐานะที่ปรึกษาปริญญาานิพนธ์

รองศาสตราจารย์ชูศรี วงศ์รัตนะ และผู้ช่วยศาสตราจารย์เขาวนา ชวลิตธำรง ได้ช่วยเหลือแนะนำให้ข้อคิดเห็น และแก้ข้อบกพร่องต่าง ๆ ของปริญญาานิพนธ์

ผู้ช่วยศาสตราจารย์เขาวนา ชวลิตธำรง อาจารย์ระวีวรรณ พันธุ์พานิช อาจารย์ชวลิต รวยหาจิดน อาจารย์ตรรชนี บุญทริก และอาจารย์แสงระวี มิ่งขวัญ ที่กรุณาตรวจสอบเครื่องมือในการวิจัย ตลอดจนให้คำแนะนำและข้อคิดเห็นต่าง ๆ ในการแก้ไขและปรับปรุงเครื่องมือที่ใช้ในการศึกษาครั้งนี้

คณาจารย์ภาควิชาวัดผลและวิจัยทางการศึกษาทุกท่านที่ได้อบรมสั่งสอนวิชาความรู้ในศาสตร์เกี่ยวกับการวัดผลการศึกษา

ผู้บริหารโรงเรียน คณะครูโรงเรียนที่เป็นกลุ่มตัวอย่าง และขอขอบใจนักเรียนทุกคนที่ให้ความร่วมมือในการเก็บข้อมูลเป็นอย่างดี

พี่ ๆ เพื่อน ๆ และน้อง ๆ นิสิตสาขาการวัดผลการศึกษาทุกท่านที่คอยช่วยเหลือเป็นกำลังใจในการเรียน และการทำวิจัยในครั้งนี้เป็นอย่างดีตลอดมา

สุดท้ายนี้ ผู้วิจัยขอกราบขอบพระคุณคุณพ่อ คุณแม่ และขอขอบคุณ พี่ ๆ น้อง ๆ โดยเฉพาะคุณรุจ จินตมาศ ที่สนับสนุนทั้งด้านกำลังใจ กำลังทรัพย์และคอยดูแลช่วยเหลือในทุกสิ่งทุกอย่างแก่ผู้วิจัยอย่างดีเสมอมา

พรรณณี จินตมาศ

สารบัญ

| บทที่ | หน้า |
|---|------|
| 1 บทนำ..... | 1 |
| ภูมิหลัง..... | 1 |
| ✓ จุดมุ่งหมายของการศึกษาค้นคว้า..... | 4 |
| ความสำคัญของการศึกษาค้นคว้า..... | 5 |
| ✓ ขอบเขตการศึกษาค้นคว้า..... | 5 |
| คำนิยามศัพท์เฉพาะ..... | 7 |
| 2 เอกสารและงานวิจัยที่เกี่ยวข้อง..... | 10 |
| ความหมายของความลำเอียงของข้อสอบ..... | 10 |
| ทฤษฎีที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบ..... | 11 |
| หลักการวิเคราะห์ความลำเอียงของข้อสอบ..... | 34 |
| งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบ..... | 36 |
| <i>(สุ่ม) ขอบเขตวิจัย</i> | 50 |
| 3 วิธีดำเนินการศึกษาค้นคว้า..... | 51 |
| ประชากร..... | 51 |
| ✓ กลุ่มตัวอย่าง..... | 51 |
| เครื่องมือที่ใช้ในการศึกษาค้นคว้า..... | 54 |
| วิธีดำเนินการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า..... | 54 |
| วิธีดำเนินการเก็บรวบรวมข้อมูล..... | 59 |
| ✓ สถิติที่ใช้ในการวิเคราะห์ข้อมูล..... | 60 |

| บทที่ | หน้า |
|---|------|
| 4 ผลการวิเคราะห์ข้อมูล..... | 67 |
| สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล..... | 67 |
| การเสนอผลการวิเคราะห์ข้อมูล..... | 68 |
| ผลการวิเคราะห์ข้อมูล..... | 69 |
| 5 สรุปผล อภิปรายผล และข้อเสนอแนะ..... | 88 |
| จุดมุ่งหมายของการศึกษาค้นคว้า..... | 88 |
| กลุ่มตัวอย่าง..... | 88 |
| เครื่องมือที่ใช้ในการศึกษาค้นคว้า..... | 88 |
| วิธีดำเนินการเก็บรวบรวมข้อมูล..... | 89 |
| การเสนอผลการวิเคราะห์ข้อมูล..... | 90 |
| สรุปผลการวิเคราะห์ข้อมูล..... | 90 |
| อภิปรายผล..... | 92 |
| ข้อเสนอแนะในการวิจัย..... | 95 |
| บรรณานุกรม..... | 97 |
| ภาคผนวก..... | 105 |
| ประวัติย่อของผู้วิจัย..... | 128 |

บัญชีตาราง

| ตาราง | หน้า |
|---|------|
| 1 ข้อมูลผลการสอบของกลุ่มตัวอย่าง กลุ่ม R และ F ที่มีคะแนนรวมอยู่ในช่วง คะแนน j ของข้อสอบข้อที่ i | 28 |
| 2 จำนวนนักเรียนในกลุ่มตัวอย่างที่ใช้ในการวิจัย จำแนกตามขนาดของโรงเรียน... | 53 |
| 3 ค่าดัชนีความลำเอียงของข้อสอบที่วิเคราะห์ด้วยวิธีแปลงค่าความยาก จำแนก ตามขนาดกลุ่มผู้สอบ..... | 70 |
| 4 ค่าดัชนีความลำเอียงของข้อสอบที่วิเคราะห์ด้วยวิธีแมนเทิล-แฮนส์เชล จำแนก ตามขนาดกลุ่มผู้สอบ..... | 73 |
| 5 ค่าดัชนีความลำเอียงของข้อสอบที่วิเคราะห์ด้วยวิธีชิปเทสท์ จำแนกตามขนาด กลุ่มผู้สอบ..... | 77 |
| 6 สรุปจำนวนข้อสอบที่มีความลำเอียงในแต่ละวิธีวิเคราะห์ความลำเอียง จำแนกตามขนาดกลุ่มผู้สอบ..... | 80 |
| 7 ผลการเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีวิเคราะห์ความ ลำเอียง 3 วิธี เมื่อขนาดกลุ่มผู้สอบเท่ากัน..... | 82 |
| 8 ผลการเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงเมื่อขนาดกลุ่มผู้สอบต่าง กันเมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน..... | 83 |
| 9 ค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับของแบบทดสอบ ค่าFisher-Z และค่าการ ทดสอบนัยสำคัญของค่าความเชื่อมั่นของแบบทดสอบที่คัดเลือก ข้อสอบที่มีความลำเอียงออกแล้ว ระหว่างวิธีวิเคราะห์ความลำเอียงที่ แตกต่างกันเมื่อขนาดกลุ่มผู้สอบเท่ากัน..... | 84 |
| 10 ผลการเปรียบเทียบค่าความเชื่อมั่นรายคู่ของแบบทดสอบหลังจากคัดเลือก ข้อสอบที่มีความลำเอียงออกแล้วด้วยวิธีวิเคราะห์ความลำเอียงต่างกัน เมื่อมีขนาดกลุ่มผู้สอบ 500 คน..... | 86 |

| | | |
|----|--|-----|
| 11 | ผลการเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว เมื่อขนาดกลุ่มผู้สอบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน..... | 87 |
| 12 | ค่าความยาก (p) ค่าอำนาจจำแนก (r) เฉพาะตัวถูกของแบบทดสอบคณิตศาสตร์โจทย์ปัญหา จากการทดสอบครั้งที่ 1..... | 109 |
| 13 | ค่าเฉลี่ยของข้อสอบแต่ละข้อระหว่างกลุ่มผู้สอบเพศชายและเพศหญิง จำแนกตามขนาดกลุ่มผู้สอบ..... | 112 |
| 14 | ค่าดัชนีเบต้า (β) ของข้อสอบแต่ละข้อจำแนกตามขนาดกลุ่มผู้สอบ..... | 117 |

บัญชีภาพประกอบ

| ภาพประกอบ | หน้า |
|---|------|
| 1 กราฟไบนารีเอทของค่าเฉลี่ยของคนมีวดีกลุ่ม 1 และคนมีวดีกลุ่ม 2 จากแบบทดสอบวัดความถนัดทางการเรียนด้านภาษา..... | 17 |
| 2 กราฟไบนารีเอทของค่าเฉลี่ยของคนมีวดีกลุ่ม 5 และคนมีวดีกลุ่ม 6 จากแบบทดสอบวัดความถนัดทางการเรียนด้านภาษา..... | 18 |
| 3 กราฟไบนารีเอทของค่าเฉลี่ยของคนมีวดีกลุ่ม 5 และคนมีวดีกลุ่ม 1 จาก แบบทดสอบวัดความถนัดทางการเรียนด้านภาษา..... | 19 |
| 4 ความหมายของค่าพารามิเตอร์ข้อสอบตามโมเดลโลจิสติกที่มีพารามิเตอร์ 1 ตัว..... | 23 |
| 5 ความหมายของค่าพารามิเตอร์ข้อสอบตามโมเดลโลจิสติกที่มีพารามิเตอร์ 2 ตัว..... | 25 |
| 6 ความหมายของค่าพารามิเตอร์ข้อสอบตามโมเดลโลจิสติกที่มีพารามิเตอร์ 3 ตัว..... | 26 |
| 7 ลำดับชั้นในการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า..... | 55 |
| 8 การกระจายของคู่อันดับค่าเฉลี่ยของข้อสอบแต่ละข้อระหว่างผู้สอบเพศชาย และผู้สอบเพศหญิง เมื่อวิเคราะห์จากขนาดกลุ่มผู้สอบ 500 คน..... | 114 |
| 9 การกระจายของคู่อันดับค่าเฉลี่ยของข้อสอบแต่ละข้อระหว่างผู้สอบเพศชาย และผู้สอบเพศหญิง เมื่อวิเคราะห์จากขนาดกลุ่มผู้สอบ 1,000 คน..... | 115 |

บทที่ 1

บทนำ

ภูมิหลัง

ในการทดสอบนั้น มีปรัชญาที่เกี่ยวกับการทดสอบว่าการทดสอบเป็นการค้นและพัฒนาสมรรถภาพของมนุษย์ (ชวาล แพรัตกุล. 2518 : 34 - 129) แสดงว่าการทดสอบมีความมุ่งหมายอยู่ 2 ประการ คือ เพื่อตรวจดูความรู้ความสามารถของผู้เรียน และเพื่อกระตุ้นสมรรถภาพต่าง ๆ ของนักเรียน หรือเพื่อพัฒนานักเรียนให้เกิดการเรียนรู้สูงขึ้น ดังนั้นการทดสอบจึงเป็นวิธีการวัดผลวิธีหนึ่งในหลายวิธี และเป็นวิธีที่นิยมใช้กันมากที่สุดในวงการศึกษา โดยเฉพาะอย่างยิ่งถ้าใช้กับเด็กจำนวนมาก ๆ จะนิยมใช้แบบทดสอบปรนัย ซึ่งแบบทดสอบที่จะนำมาใช้ควรจะเป็นแบบทดสอบที่มีคุณภาพ เพื่อให้ผลที่ได้จากการวัดเป็นที่เชื่อถือได้ มีประสิทธิภาพ และสามารถพยากรณ์ หรืออธิบายความสามารถของผู้สอบได้อย่างแม่นยำ สำหรับคุณภาพของแบบทดสอบนั้นจะเน้นคุณภาพในด้านความเที่ยงตรง ความเชื่อมั่น ความง่ายที่พอเหมาะ ไม่ง่ายจนเกินไปหรือยากจนเกินไป ความสามารถของผู้สอบ และสามารถจำแนกผู้สอบที่มีความสามารถต่างกันออกจากกันได้ นอกจากคุณภาพของแบบทดสอบแล้วยังมีคุณลักษณะประการอื่นที่ผู้สร้างจะต้องคำนึงถึงเพื่อให้แบบทดสอบที่สร้างขึ้นทำหน้าที่ได้อย่างสมบูรณ์และบรรลุจุดมุ่งหมายของการทดสอบ ซึ่งคุณลักษณะเหล่านั้นคือ ข้อคำถามในแบบทดสอบจะต้องถามลึก ถามเพื่อวัดพฤติกรรมสมองที่สูงกว่าพฤติกรรมสมองด้านความรู้ความจำ กระตุ้นให้ผู้สอบต้องการตอบข้อคำถามข้อต่อ ๆ ไปเพื่อให้เกิดการพัฒนาด้านความคิด ข้อคำถามที่จำเพาะเจาะจง มีประสิทธิภาพ และมีความเป็นปรนัย นอกจากคุณลักษณะที่กล่าวมานี้แล้วยังมีคุณลักษณะที่สำคัญอีกประการ คือ คุณลักษณะของข้อคำถามที่ให้ความเสมอภาค ความยุติธรรมกับผู้สอบ ไม่เกิดการได้เปรียบเสียเปรียบกับกลุ่มผู้สอบกลุ่มใดกลุ่มหนึ่งโดยเฉพาะ ไม่ว่าจะผู้สอบจะมีความแตกต่างกันในด้านเพศ ภาษา เชื้อชาติ วัฒนธรรม ศาสนา ซึ่งคุณลักษณะดังกล่าวนี้ เป็นคุณลักษณะของข้อคำถามที่ปราศจากความลำเอียง ข้อคำถามที่ดีจึงควรเป็นข้อคำถามที่มีทั้งคุณภาพ และความเสมอภาคปราศจากความลำเอียง แต่จากงานวิจัยของกรีน (Intasuwan. 1979 : 1 ; citing Green.1972) พบว่า ข้อคำถามบางข้อของแบบทดสอบวัดผลสัมฤทธิ์ของแคลิฟอร์เนีย (California Achievement Test) ซึ่งเป็นข้อคำถามที่มีคุณภาพ เมื่อนำ

ไปทดสอบกับกลุ่มผู้สอบที่มีความแตกต่างกันทางเชื้อชาติ—ภาษา—สภาพเศรษฐกิจ—และสภาพภูมิอากาศ จะให้ผลการสอบที่แตกต่างกัน ซึ่งอาจมีสาเหตุมาจากวัฒนธรรม และสิ่งแวดล้อมที่ต่างกันมีผลต่อความเข้าใจภาษาที่ใช้ในข้อคำถามต่างกัน และเรียกข้อคำถามประเภทนี้ว่า ข้อคำถามลำเอียงหรือข้อสอบลำเอียง (Item Bias or Differential Item Function : DIF) นอกจากนี้ยังมีผลงานวิจัยของ สงบ ลักษณะ(Burrill. 1982 : 168 ; citing Laksana . 1979) ที่ศึกษาความลำเอียงที่เกิดจากเพศและผิวต่างกันเมื่อทำแบบทดสอบด้านภาษา และคณิตศาสตร์ พบว่า ข้อคำถามในแบบทดสอบด้านภาษาและคณิตศาสตร์บางข้อมีความลำเอียงระหว่างกลุ่มผู้สอบเพศชายและเพศหญิง และกลุ่มผู้สอบผิวดำและผิวขาว ทั้งนี้อาจมีสาเหตุมาจากความสามารถในการอ่านโจทย์ปัญหาระหว่างเพศชายและเพศหญิงต่างกันหรือเกิดจากการใช้ภาษาในแบบทดสอบที่เอื้อต่อเพศหนึ่งแต่ไม่เอื้อต่ออีกเพศหนึ่งก็เป็นได้ซึ่งสอดคล้องกับผลงานวิจัยของพัชรี ปิยภัณฑ (2531 : 65-67) ชัยชัย เผ่าพงษ์ (2527 : 73 - 76) นิรมล ชัยชวลิต (2537 : 57-59) ดูลิทเทิล และเคลียร์ (Doolittle and Cleary. 1987 : 157) เชปาร์ด (Shepard. 1982 : 11) และปอปแฮม (Popham. 1981 : 181) ที่ระบุว่าภาษาและเพศเป็นแหล่งที่ก่อให้เกิดความลำเอียงของข้อสอบ นอกจากนี้ยังมีผลงานวิจัยของไฟรแมน (Friedman. 1995 : 22 - 50) และโดแรนกับลิวิงตัน (Dorans and Livingston. 1987 : 65 - 71) ที่ระบุว่า นักเรียนหญิงจะใช้ทักษะด้านภาษาในการเรียนคณิตศาสตร์มากกว่า นักเรียนชาย เพราะมีทักษะด้านภาษาสูงกว่านักเรียนชาย

จากที่กล่าวมาข้างต้น จะเห็นว่า ความแตกต่างของเชื้อชาติ ศาสนา วัฒนธรรม สภาพภูมิศาสตร์ สภาพทางสังคม เป็นแหล่งที่ทำให้ข้อคำถามบางข้อที่มีคุณภาพเกิดความลำเอียงกับกลุ่มผู้สอบบางกลุ่ม นอกจากนี้ ภาษาและเพศก็เป็นแหล่งที่ก่อให้เกิดความลำเอียงของข้อคำถามด้วยเช่นกัน ฉะนั้นในการสร้างและพัฒนาแบบทดสอบ โดยเฉพาะอย่างยิ่งแบบทดสอบที่มีภาษาเข้ามาเกี่ยวข้อง เช่น แบบทดสอบคณิตศาสตร์โจทย์ปัญหาที่ผู้สอบจะต้องใช้ความเข้าใจทางภาษาเพื่อแปลความหมายของโจทย์ปัญหาแล้วใช้ทักษะทางการคิดคำนวณเพื่อหาคำตอบ จึงควรศึกษาทั้งคุณภาพและความลำเอียงของข้อสอบระหว่างกลุ่มผู้สอบที่มีเพศต่างกัน เพื่อให้แน่ใจว่าผลที่ได้จากการวัดมีความถูกต้องแม่นยำเป็นการวัดความสามารถที่แท้จริงของผู้สอบโดยไม่เกิดการได้เปรียบเสียเปรียบต่อกลุ่มผู้สอบเพศใดเพศหนึ่งโดยเฉพาะ

สำหรับวิธีการวิเคราะห์ความลำเอียงของข้อสอบนั้น มีวิธีการวิเคราะห์หลายวิธี ดังนี้ วิธีที่ใช้หลักของทฤษฎีการวัดผลแบบดั้งเดิม (Classical Test Theory) เช่น วิธีแปลงค่าความยาก (Transformed Item Difficulty) การวิเคราะห์ความแปรปรวน (Analysis of Variance)

การวิเคราะห์ด้วยไคสแควร์ (Chi-Square) วิเคราะห์องค์ประกอบ (Factor Analysis) วิเคราะห์การถดถอย (Regression Analysis) วิธิล็อกลิเนียร์ (Log-Linear) และวิธีที่ใช้หลักของทฤษฎีการตอบข้อคำถาม (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 แต่การใช้ทฤษฎีการตอบข้อคำถามในการวิเคราะห์ความลำเอียงของข้อสอบเป็นวิธีการที่ใช้การคำนวณยุ่งยากซับซ้อน ใช้กลุ่มตัวอย่างขนาดใหญ่ และสิ้นเปลืองค่าใช้จ่ายมาก (Narayanan and Swaminathan. 1994 : 315) ทำให้เกิดความยุ่งยากในทางปฏิบัติ จากเหตุผลที่กล่าวมาข้างต้น จึงมีการศึกษาวิธีการวิเคราะห์ความลำเอียงของข้อสอบ โดยหลีกเลี่ยงการใช้ทฤษฎีการตอบข้อคำถาม สำหรับวิธีที่มีผู้เสนอ คือ วิธีการที่ใช้สถิติแบบนอน - พารามิเตอร์ (Non - Parametric) ในการวิเคราะห์ความลำเอียงของข้อสอบ เช่น วิธีแมนเทล - ฮานส์เซล (Mantel - Haenszel : MH) และ วิธีซิปเทสท์ (SIBTEST : SIB) ซึ่งวิธีการทั้ง 2 เป็นวิธีการคำนวณที่ไม่ยุ่งยากซับซ้อน และสิ้นเปลืองค่าใช้จ่ายไม่มากนักเหมาะที่จะนำมาใช้ในทางปฏิบัติจึงเป็นที่สนใจและนิยมใช้ในปัจจุบัน

นอกจากนี้ยังมีงานวิจัยที่ศึกษาผลกระทบของขนาดกลุ่มตัวอย่างในการวิเคราะห์ความลำเอียงของข้อสอบ แบบ MH เช่น งานวิจัยของ ไวส (Wise. 1987) โดยศึกษาผลการวิเคราะห์ความลำเอียงของข้อสอบในขนาดกลุ่มตัวอย่าง 400 คน และ 800 คน พบว่า วิธี MH จะวิเคราะห์ข้อสอบที่มีความลำเอียงได้เพิ่มขึ้นเมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้น และยังพบว่า วิธี MH จะให้ผลการวิเคราะห์ที่ถูกต้องที่สุดเมื่อข้อสอบมีค่าอำนาจจำแนกเพิ่มขึ้นเป็น 1.5 ($a = 1.5$) แต่จะให้ผลตรงข้ามกันเมื่อค่าอำนาจจำแนกของ ข้อสอบต่ำกว่า 1.0 หรือสูงกว่า 2.0 และจากงานวิจัยของ โรเจอร์ (Rogers.1993 : 105) ที่ศึกษา การเปรียบเทียบวิธีการวิเคราะห์ข้อสอบแบบ MH กับวิธีการถดถอยโลจิสติก (Logistic Regression) โดยศึกษาผลกระทบของขนาดกลุ่มตัวอย่างด้วยกลุ่มตัวอย่างที่ศึกษา ขนาด 250 คน และ 500 คน ได้ผลการศึกษา เหมือนของ ไวส (Wise. 1987) คือ วิธี MH จะวิเคราะห์ข้อสอบที่มีความลำเอียงได้เพิ่มขึ้นเมื่อขนาดกลุ่มตัวอย่างเพิ่มขึ้น ส่วนของ ฮิล (Hill.1990 : 5) ได้เสนอว่า วิธี MH เป็นวิธีที่ให้ผลการวิเคราะห์ถูกต้องเมื่อกลุ่มตัวอย่างมีขนาดตั้งแต่ 100 ถึง 300 คน ในแต่ละกลุ่ม และงานวิจัยของนารายานาน และสวามินาทาน (Narayanan and Swaminathan.1994 : 315) ที่ศึกษาการวิเคราะห์ความลำเอียง โดยวิธี MH และ วิธี SIB พบว่า วิธีการทั้ง 2 ใช้วิเคราะห์ข้อสอบที่มีความลำเอียง เข้าสู่กลุ่มใดเพียงกลุ่มเดียว

(Uniform DIF) เมื่อกลุ่มตัวอย่างมีการแจกแจงความสามารถเท่าเทียมกัน แต่วิธี SIB จะให้ผลการวิเคราะห์ที่ถูกต้องกว่าวิธี MH เมื่อกลุ่มตัวอย่างมีการแจกแจงความสามารถไม่เท่ากัน และวิธีวิเคราะห์ความลำเอียงทั้ง 2 วิธีจะใช้ได้ดีมีประสิทธิภาพเมื่อกลุ่มตัวอย่างแต่ละกลุ่มมีขนาด 300 คน

จากที่กล่าวมาข้างต้นและจากการศึกษารายงานการวิจัยที่ผ่านมา จะเห็นว่า วิธีที่ใช้ในการวิเคราะห์ความลำเอียงของข้อสอบมีหลายวิธี เช่น วิธีแปลงค่าความยาก(Transformed Item Difficulty) วิธีวิเคราะห์ความแปรปรวน(Analysis of Variance) วิธีวิเคราะห์ด้วยไคสแควร์ (Chi-Square) วิธีโค้งลักษณะข้อสอบ(Item Characteristic Curve) วิธีแมนเทล - แฮนส์เซล (Mantel - Haenszel) และวิธีซิปเทสต์(SIBTEST) แต่ยังไม่มียานวิจัยที่ศึกษาเปรียบเทียบความลำเอียงของข้อสอบเมื่อขนาดกลุ่มผู้สอบต่างกันโดยวิธีแปลงค่าความยาก(Transformed Item Difficulty) วิธีแมนเทล - แฮนส์เซล(Mantel - Haenszel) และวิธีซิปเทสต์(SIBTEST) ซึ่งเป็นวิธีที่ได้ดีในทางปฏิบัติ ง่ายต่อการคำนวณและการแปลผล

ด้วยเหตุผลดังกล่าวข้างต้น ผู้วิจัยจึงสนใจที่จะศึกษาเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบต่อกลุ่มผู้สอบที่มีเพศต่างกัน โดยใช้วิธีวิเคราะห์ 3 วิธี คือ วิธีแปลงค่าความยาก(Transformed Item Difficulty) วิธีแมนเทล-แฮนส์เซล(Mantel - Haenszel) วิธีซิปเทสต์(SIBTEST) และขนาดกลุ่มผู้สอบที่ต่างกัน โดยศึกษาจำนวนข้อสอบที่มีความลำเอียงและค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดข้อสอบที่มีความลำเอียงออกแล้ว อันจะเป็นแนวทางในการเลือกวิธีการวิเคราะห์ความลำเอียงมาใช้ในทางปฏิบัติ เพื่อปรับปรุงและพัฒนาแบบทดสอบให้มีความยุติธรรมต่อไป

① จุดมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อศึกษาดัชนีความลำเอียงของข้อสอบแต่ละข้อในแบบทดสอบวิชาคณิตศาสตร์ ใจทย์ปัญหา จำแนกในแต่ละวิธีวิเคราะห์ความลำเอียงของข้อสอบและขนาดกลุ่มผู้สอบ
2. เพื่อเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียงที่แตกต่างกัน เมื่อขนาดกลุ่มผู้สอบเท่ากัน

3. เพื่อเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงเมื่อขนาดกลุ่มผู้สอบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน

4. เพื่อเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังการคัดข้อสอบที่มีความลำเอียงออกแล้วระหว่างวิธีวิเคราะห์ความลำเอียงที่แตกต่างกัน เมื่อขนาดกลุ่มผู้สอบเท่ากัน

5. เพื่อเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังการคัดข้อสอบที่มีความลำเอียงออกแล้ว เมื่อขนาดกลุ่มผู้สอบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน

ความสำคัญของการศึกษาค้นคว้า

ผลจากการศึกษาครั้งนี้ จะทำให้ทราบว่า วิธีวิเคราะห์ความลำเอียงข้อสอบวิธีใดจะพบจำนวนข้อสอบที่มีความลำเอียงต่อเพศได้มากที่สุด และในแต่ละวิธีวิเคราะห์ความลำเอียงจะคัดเลือกข้อสอบที่มีความลำเอียงได้มากที่สุดเมื่อกลุ่มผู้สอบมีขนาดเท่าใด รวมทั้งทราบถึงค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว ซึ่งจะเป็นแนวทางในการตัดสินใจเลือกวิธีวิเคราะห์ความลำเอียงที่เหมาะสมในทางปฏิบัติมาพัฒนาแบบทดสอบให้มีคุณภาพ และมีความยุติธรรมต่อไป

๒) ขอบเขตการศึกษาค้นคว้า

1. ประชากรที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2539 ของโรงเรียนสังกัดกรมสามัญศึกษา ส่วนกลาง กรุงเทพมหานคร จำนวน 113 โรงเรียน 1,332 ห้องเรียน มีจำนวนนักเรียนทั้งหมด 58,545 คน

2. กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นนักเรียนชายและนักเรียนหญิง ชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2539 ของโรงเรียนสังกัดกรมสามัญศึกษา ส่วนกลาง กรุงเทพมหานคร ซึ่งได้มาโดยการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) มีขนาดโรงเรียนเป็นชั้น (Strata) โรงเรียนเป็นหน่วยการสุ่ม (Sampling Unit) ได้จำนวนโรงเรียน 10 โรงเรียนที่มีจำนวนห้องเรียน 53 ห้องเรียน และมีจำนวนนักเรียน 2,200 คน

3. ตัวแปรที่ศึกษา

3.1 ตัวแปรอิสระ

3.1.1 วิธีวิเคราะห์ความลำเอียง ได้แก่

3.1.1.1 วิธีแปลงค่าความยาก (Transformed Item Difficulty)

3.1.1.2 วิธีแมนเทล-แฮนส์เซล (Mantel - Haenszel)

3.1.1.3 วิธีชิปเทสต์ (SIBTEST)

3.1.2 ขนาดของกลุ่มตัวอย่าง ได้แก่

3.1.2.1 กลุ่มผู้สอบจำนวน 500 คน

3.1.2.2 กลุ่มผู้สอบจำนวน 1,000 คน

3.2 ตัวแปรตาม ได้แก่

3.2.1 ดัชนีความลำเอียงของข้อสอบ

3.2.2 จำนวนข้อสอบที่ลำเอียง

3.2.3 ความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว

3.3 ตัวแปรที่ใช้ในการวิเคราะห์ความลำเอียง คือ เพศ ได้แก่ เพศชาย และ เพศหญิง

4. เครื่องมือที่ใช้ในการศึกษาค้นคว้า

4.1 เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นแบบทดสอบคณิตศาสตร์ใจทย์ปัญหา ชั้นมัธยมศึกษาปีที่ 1 เรื่อง สมการ อัตราส่วนและร้อยละ ตามหลักสูตรมัธยมศึกษาตอนต้น พุทธศักราช 2521 (ฉบับปรับปรุง 2533) เป็นแบบทดสอบเลือกตอบ ชนิด 5 ตัวเลือก จำนวน 40 ข้อ ที่ผู้วิจัยสร้างขึ้นเอง 1 ฉบับ

คำนิยามศัพท์เฉพาะ

1. ข้อสอบที่มีความลำเอียง หมายถึง ข้อสอบที่วัดความสามารถหรือคุณลักษณะเดียวกันแต่เมื่อนำไปสอบกับผู้สอบที่มีเพศต่างกัน 2 กลุ่ม แต่มีความสามารถเท่ากันแล้วได้ผลการสอบต่างกัน โดยศึกษาความลำเอียงของข้อสอบด้วยวิธีวิเคราะห์ความลำเอียง 3 วิธี

2. ดัชนีความลำเอียงของข้อสอบ หมายถึง ค่าสถิติที่ใช้แสดงความแตกต่างของผลการตอบข้อสอบระหว่างกลุ่มผู้สอบที่มีเพศต่างกันจากวิธีวิเคราะห์ความลำเอียง 3 วิธี ดังนี้

2.1 ดัชนีความลำเอียงของข้อสอบจากวิธีแปลงค่าความยาก หมายถึง ค่าระยะห่างตั้งฉากจากจุดศูนย์กลางค่าเฉลี่ยของข้อสอบแต่ละข้อไปยังเส้นแกนหลัก คือ ค่า d_i

2.2 ดัชนีความลำเอียงของข้อสอบจากวิธีแมนเทล - แฮนส์เซล หมายถึง ค่าสัดส่วนของการตอบของกลุ่มผู้สอบ 2 กลุ่ม คือ ค่า α_{MH}

2.3 ดัชนีความลำเอียงของข้อสอบจากวิธีชิปเทสท์ หมายถึง ค่าที่ได้จากการทดสอบความแตกต่างของการตอบถูกของกลุ่มผู้สอบแต่ละกลุ่ม คือ ค่า β

3. การวิเคราะห์ความลำเอียงของข้อสอบ หมายถึง การคำนวณค่าดัชนีความลำเอียงของข้อสอบ ซึ่งในที่นี้ใช้วิธีวิเคราะห์ความลำเอียง 3 วิธี ดังนี้

3.1 วิธีแปลงค่าความยาก (Transformed Item Difficulties Approach) หมายถึง การวิเคราะห์หาดัชนีความลำเอียงของข้อสอบโดยพิจารณาค่าระยะห่างตั้งฉากจากจุดศูนย์กลางค่าเฉลี่ยของข้อสอบแต่ละข้อไปยังเส้นแกนหลัก (d_i) ข้อสอบลำเอียงจะมีค่า d_i มากกว่า 0.75 หรือน้อยกว่า -0.75

3.2 วิธีแมนเทล - แฮนส์เซล(Mantel - Haenszel) หมายถึง การวิเคราะห์หาดัชนีความลำเอียงของข้อสอบ โดยพิจารณาเปรียบเทียบสัดส่วนของการตอบของกลุ่มผู้สอบ 2 กลุ่ม ข้อสอบที่มีความลำเอียงจะมีค่า α_{MH} ไม่เท่ากับ 1 และค่า χ^2_{MH} มากกว่าค่า χ^2 ที่ df เท่ากับ 1 อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

3.3 วิธีชิปเทสท์ (SIBTEST) หมายถึง การวิเคราะห์หาดัชนีความลำเอียงของข้อสอบโดยพิจารณาค่าความลำเอียง (β) ที่ได้จากการทดสอบความแตกต่างของการตอบถูกของกลุ่มผู้สอบแต่ละกลุ่ม ข้อสอบที่มีความลำเอียงจะมีค่า B_u หรือค่า Z_{SIB} แตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

4. ผลการวิเคราะห์ความลำเอียงของข้อสอบ หมายถึง ผลที่ได้จากการคำนวณด้วยวิธีวิเคราะห์ความลำเอียง 3 วิธี ซึ่งได้แก่ วิธีแปลงค่าความยาก (Transformed Item Difficulty) วิธีแมนเทล - แฮนส์เซล (Mantel - Haenszel) และวิธีชิปเทสท์ (SIBTEST) ซึ่งการวิจัยครั้งนี้หมายถึงผล ดังนี้

4.1 ความลำเอียงของข้อสอบแต่ละข้อ หมายถึง ข้อสอบในแบบทดสอบที่ผู้วิจัยสร้างขึ้น ให้ผลการสอบระหว่างกลุ่มผู้สอบเพศชาย และเพศหญิง แตกต่างกัน ซึ่งวัดได้จากค่าดัชนีความลำเอียงของข้อสอบที่วิเคราะห์ด้วยวิธีต่างกัน

4.2 จำนวนข้อสอบที่มีความลำเอียง หมายถึง จำนวนข้อสอบจากแบบทดสอบที่ผู้วิจัยสร้างขึ้น ให้ผลการสอบแตกต่างกันระหว่างกลุ่มผู้สอบเพศชายและเพศหญิง ที่วิเคราะห์ความลำเอียงด้วยวิธีวิเคราะห์ 3 วิธี จำแนกในแต่ละวิธีวิเคราะห์ความลำเอียงและขนาดกลุ่มผู้สอบ

4.3 ค่าความเชื่อมั่นของแบบทดสอบ หมายถึง คุณสมบัติของแบบทดสอบที่วัดความรู้ความสามารถของนักเรียนได้คงที่แน่นอน ซึ่งในการวิจัยครั้งนี้ศึกษาความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อที่มีความลำเอียงออกแล้ว โดยคำนวณค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับ (Split-half) ของแบบทดสอบ โดยใช้สูตรของ Spearman-Brown

5. ขนาดกลุ่มผู้สอบ หมายถึง จำนวนกลุ่มผู้สอบที่ตอบข้อสอบจากแบบทดสอบที่ผู้วิจัยสร้างขึ้น แบ่งเป็น 2 กลุ่ม ได้แก่ กลุ่มที่ 1 จำนวน 500 คน กลุ่มที่ 2 จำนวน 1,000 คน โดยการสุ่มกระดาษคำตอบด้วยวิธีสุ่มอย่างง่าย

6. แบบทดสอบคณิตศาสตร์ใจทย์ปัญหา หมายถึง แบบทดสอบที่มีลักษณะข้อคำถามเป็นประเภทใจทย์คณิตศาสตร์ที่มีภาษาเข้ามาเกี่ยวข้องกับและมุ่งแก้ปัญหาทางคณิตศาสตร์ซึ่งเป็นแบบทดสอบคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1 เรื่อง สมการ อัตราส่วนและร้อยละ จำนวน 40 ข้อ ที่ผู้วิจัยสร้างขึ้นเอง

7. ผู้เชี่ยวชาญ หมายถึง ครูผู้สอนวิชาคณิตศาสตร์ ระดับมัธยมศึกษาปีที่ 1 ที่มีประสบการณ์อย่างน้อย 5 ปี จำนวน 2 คน กับนักวัดผลการศึกษา จำนวน 3 คน

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องกับการศึกษาค้นคว้า ผู้วิจัยได้ศึกษาเนื้อหาต่าง ๆ ดังนี้

1. ความหมายของความลำเอียงของข้อสอบ
2. ทฤษฎีที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบ
3. หลักการวิเคราะห์ความลำเอียงของข้อสอบ
4. ผลงานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบ

ความหมายของความลำเอียงของข้อสอบ

มีผู้ให้ความหมายของคำว่า ความลำเอียงของข้อสอบไว้หลายท่านด้วยกัน เช่น เอเลน และเยน (Allen and Yen. 1979 : 129) ได้ให้ความหมายว่า ข้อสอบลำเอียงเป็นข้อสอบที่วัดความสามารถหรือคุณลักษณะทางจิตวิทยาของผู้สอบที่มีระดับความสามารถเท่ากัน แล้วได้ผลต่างกัน โดยพิจารณาจากโค้งลักษณะข้อสอบของผู้สอบในแต่ละกลุ่มจะต่างกัน เช่นเดียวกับ แฮมเบิลตัน และคณะ (Hambleton et al. 1991 : 110) แมคแคนลีย์ และเมนโดซา (Mc Canley and Mendoza. 1985 : 89) อังคณา สายยศ (2539 : 6) ที่ให้ความหมายของความลำเอียงของข้อสอบโดยพิจารณาจากผลการสอบที่แตกต่างกันจากผู้สอบสองกลุ่มที่มีความสามารถเท่ากัน สำหรับบุญชม ศรีสะอาด (2535 : 195) ให้ความหมายว่า ข้อสอบลำเอียงเป็นข้อสอบที่ให้ผลการสอบแตกต่างกันจากกลุ่มผู้สอบที่มีความสามารถไม่ต่างกัน และเสนอว่า ผลการสอบที่ต่างกันนั้นมาจากองค์ประกอบบางอย่าง เช่น ภาษา เพศ วัฒนธรรม ศาสนา เชื้อชาติ สภาพทางภูมิศาสตร์ สภาพทางเศรษฐกิจและสังคม ประสบการณ์ส่วนตัว เป็นต้น

ลอร์ด (Lord.1980 : 212) สับโคเวียต และคณะ (Subkoviak et al .1984 : 49) ฮูลิน, ดรอสโก และพาร์สัน (Hulin, Drasgow and Parsons .1983 : 152) ให้ความหมายไว้เหมือน

กันในรูปแบบของความน่าจะเป็นในการตอบข้อสอบได้ถูกต้องของผู้สอบที่มีความสามารถเท่ากัน แต่ อยู่ต่างกลุ่มกันจะไม่เท่ากัน

นอกจากนี้ ยังมีผู้ให้ความหมายของความลำเอียงที่แตกต่างออกไป เช่น ปอปแฮม (Popham.1981 : 78) ให้ความหมายของความลำเอียงของข้อสอบในรูปแบบการสอบที่ไม่ยุติธรรม เนื่องมาจากความชอบหรือความเอนเอียงของบุคคล ส่วนออสเตอร์ลายด์ (Osterlind. 1983 : 10) ได้ให้ความหมายว่า เป็นความคลาดเคลื่อนอย่างมีระบบทางการวัด และมีลล์แซปและอีเวอร์สัน (Millsap and Everson. 1993 : 297) กล่าวว่า ความลำเอียงของข้อสอบเกิดจากระบบการวัดที่ไม่แน่นอน

จากนิยามดังกล่าวข้างต้น พอสรุปได้ว่า ความลำเอียงของข้อสอบหมายถึง ข้อสอบที่วัดความสามารถ หรือคุณลักษณะทางจิตวิทยาเดียวกัน แต่เมื่อนำไปสอบกับกลุ่มผู้สอบสองกลุ่มที่มีความแตกต่างกันในด้าน เพศ ภาษา วัฒนธรรม ศาสนา เชื้อชาติ สภาพทางภูมิศาสตร์ สภาพทางเศรษฐกิจและสังคม ประสบการณ์ส่วนตัวแล้วได้ผลการวัดที่ต่างกัน

ทฤษฎีที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบ

ได้มีผู้เสนอแนะวิธีการวิเคราะห์ความลำเอียงของข้อสอบไว้หลายวิธี ซึ่งเมื่อนำมาแบ่งเป็นประเภทต่าง ๆ โดยยึดทฤษฎีทางการวัดผลเป็นเกณฑ์ในการแบ่ง จะได้เป็น 2 ประเภทใหญ่ ๆ คือ

1. วิธีวิเคราะห์ความลำเอียงโดยยึดทฤษฎีการวัดผลแบบดั้งเดิม (Classical Test Theory)
2. วิธีวิเคราะห์ความลำเอียง โดยยึดทฤษฎีการตอบข้อสอบ (Item Response Theory)

ทฤษฎีการวัดผลแบบดั้งเดิม (Classical Test Theory)

เป็นทฤษฎีวัดผลที่ใช้กันทั่วไป ทฤษฎีนี้อธิบายอิทธิพลของความคลาดเคลื่อนของการวัดที่มีต่อคะแนน โดยอาศัยข้อตกลงเบื้องต้น และข้อสรุปบางประการที่ได้จากการใช้หลักการทางสถิติและคณิตศาสตร์ในการพิสูจน์ และหาข้อสรุป โดยมีเงื่อนไขว่า ถ้าข้อตกลงเบื้องต้นสมเหตุสมผลแล้ว ข้อสรุปที่ได้จะสมเหตุสมผลด้วย แต่ถ้าข้อตกลงไม่สมเหตุสมผล จะทำให้ได้ข้อสรุปที่ผิดพลาดเช่นกัน สำหรับข้อตกลงเบื้องต้นและข้อสรุปที่ได้จะนำไปสู่การอธิบายคุณภาพของแบบทดสอบที่สร้างขึ้น

ข้อตกลงเบื้องต้นของทฤษฎีการวัดผลแบบดั้งเดิม มีดังนี้ (Allen and Yen. 1979 : 56-60)

$$1. X = T + E$$

คะแนนที่ได้จากการสอบ (คะแนนที่สังเกตได้) เป็นผลรวมของคะแนนจริงกับความคลาดเคลื่อน โดยที่คะแนนจริงของผู้สอบจะเป็นค่าคงที่ ในขณะที่คะแนนจากการสอบ และคะแนนความคลาดเคลื่อนจะเป็นค่าที่แปรเปลี่ยนไปในการสอบแต่ละครั้ง

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

ค่าคาดหวัง หรือค่าเฉลี่ยประชากร (μ_x) ของคะแนนจากการสอบของผู้สอบคนหนึ่ง เท่ากับคะแนนจริงของผู้สอบนั้น โดยทำการสอบซ้ำหลาย ๆ ครั้ง และการทดสอบซ้ำแต่ละครั้ง ต้องเป็นอิสระจากกัน

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

คะแนนความคลาดเคลื่อน และคะแนนจริงที่ได้จากประชากรผู้สอบจากแบบทดสอบ ฉบับหนึ่งไม่มีความสัมพันธ์กัน

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

คะแนนความคลาดเคลื่อนจากแบบทดสอบ 2 ฉบับจะไม่มีความสัมพันธ์กัน

$$5. \rho_{E_1T_2} = 0$$

คะแนนความคลาดเคลื่อนในการสอบแบบทดสอบฉบับที่หนึ่ง ไม่มีความสัมพันธ์กับคะแนนจริงในแบบทดสอบฉบับอื่น

6. ข้อตกลงนี้เป็นข้อตกลงที่กล่าวถึงระดับความคู่ขนานของแบบทดสอบ 2 ฉบับ โดยกล่าวว่า ถ้าแบบทดสอบทั้ง 2 ฉบับ มีความคู่ขนาน (Classical Parallel Test) แบบทดสอบทั้ง 2 ฉบับนั้นจะต้องมีความเป็นเอกพันธ์ในเนื้อหา (Homogeneous in Content) คะแนนจริงจากแบบทดสอบทั้ง 2 ฉบับจะมีค่าเท่ากัน ($T_1 = T_2$) ความแปรปรวนของคะแนนความคลาดเคลื่อนจากแบบทดสอบทั้ง 2 ฉบับจะมีค่าเท่ากัน ($\sigma_{E_1}^2 = \sigma_{E_2}^2$) คะแนนสอบเฉลี่ย และความแปรปรวนของคะแนนสอบจากแบบทดสอบทั้ง 2 ฉบับ จะมีค่าเท่ากัน ($\mu_1 = \mu_2$ และ $\sigma_1^2 = \sigma_2^2$) ความแปรปรวนร่วมของแบบทดสอบทั้งสองฉบับจะเท่ากัน และความแปรปรวนร่วมกับเกณฑ์ภายนอกของแบบทดสอบจะเท่ากันด้วย

7. ข้อตกลงนี้กล่าวถึง ระดับความคู่ขนานอีกระดับหนึ่งของแบบทดสอบสองฉบับ คือแบบทดสอบ 2 ฉบับจะมีความเท่าเทียมกัน (Tau -Equivalent) เมื่อแบบทดสอบทั้ง 2 ฉบับนั้น มีความเป็นเอกพันธ์ในเนื้อหา และคะแนนจริงจากแบบทดสอบฉบับที่หนึ่งมีค่าเท่ากับผลรวมของคะแนนจริงจากแบบทดสอบฉบับที่สองกับค่าคงที่ ($T_1 = T_2 + C_{12}$)

ข้อตกลงเบื้องต้นทั้ง 7 ข้อนี้ นำไปสู่ข้อสรุปอีกมากมายที่ใช้ในการอธิบายความสัมพันธ์ระหว่างคะแนนความคลาดเคลื่อน คะแนนจริง และคะแนนจากการสอบ และการแปลผลจากแบบทดสอบที่สร้างขึ้น เพื่อนำผลที่ได้มาอธิบายความสามารถของผู้สอบ โดยพยายามค้นหาคะแนนจริงของผู้สอบ และการหาความคลาดเคลื่อนมาตรฐานในการวัด นอกจากนี้ยังนำข้อตกลงและข้อสรุปมาประยุกต์หาคุณภาพของแบบทดสอบที่สร้างขึ้น ทั้งคุณภาพรายข้อ เช่น ค่าความยากง่าย ค่าอำนาจจำแนกรายข้อ และคุณภาพของแบบทดสอบ เช่น ค่าความเชื่อมั่น และค่าความเที่ยงตรงของแบบทดสอบ ซึ่งคุณภาพเหล่านี้จะมีความเกี่ยวพันอยู่กับค่าสหสัมพันธ์ของความแปรปรวน และความแปรปรวนร่วมของคะแนนจริง คะแนนที่ได้จากการสอบ และคะแนนความคลาดเคลื่อน

* วิธีการวิเคราะห์ความลำเอียงโดยใช้ทฤษฎีการวัดผลแบบดั้งเดิม

วิธีการวิเคราะห์ความลำเอียงโดยใช้ทฤษฎีการวัดผลแบบดั้งเดิม มีหลายวิธีที่มีผู้นิยมทำการศึกษาไว้ เช่น วิธีการวิเคราะห์ความแปรปรวน (Analysis Of Variance) วิธีวิเคราะห์องค์ประกอบ (Factor Analysis) วิธีการวิเคราะห์ด้วยไคสแควร์ (Chi-Square) และวิธีแปลงค่าความยาก (Transformed Item Difficulty) ซึ่งวิธีที่กล่าวมาทั้งหมด เป็นการวิเคราะห์ความลำเอียงโดยอาศัยข้อตกลงและข้อสรุปของทฤษฎีการวัดผลแบบดั้งเดิม

วิธีการวิเคราะห์ความแปรปรวน (Analysis Of Variance)

การวิเคราะห์ความลำเอียงโดยวิธีนี้ เป็นการทดสอบนัยสำคัญทางสถิติของผลของปฏิสัมพันธ์ (Interaction Effect) ทั้งหมดระหว่างข้อสอบกับกลุ่มผู้สอบ โดยมีข้อตกลงเบื้องต้นว่าแบบทดสอบจะต้องมีความเอกพันธ์ในเนื้อหา ความแปรปรวนของข้อสอบต้องเท่ากัน และความแปรปรวนร่วมของข้อสอบจะต้องเท่ากันด้วย สำหรับการทดสอบปฏิสัมพันธ์ระหว่างข้อสอบกับกลุ่มผู้สอบ มีรูปแบบการวิเคราะห์ ดังนี้ (Laksana. 1979 : 12)

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

| | | | |
|-------|----------------------|-----|--|
| เมื่อ | μ | แทน | ค่าเฉลี่ยทั้งหมด |
| | α_i | แทน | อิทธิพลจากข้อสอบข้อที่ i ($i = 1, 2, 3, \dots, I$) |
| | β_j | แทน | อิทธิพลจากกลุ่มผู้สอบ กลุ่มที่ j ($j = 1, 2, 3, \dots, J$) |
| | $(\alpha\beta)_{ij}$ | แทน | ปฏิสัมพันธ์ระหว่างข้อสอบกับกลุ่มผู้สอบ |
| | E_{ij} | แทน | ความคลาดเคลื่อนแบบสุ่มภายในเซลล์ i, j |

ขั้นตอนการวิเคราะห์ความลำเอียง

1. จัดข้อสอบและกลุ่มผู้สอบ ลงตามแนวสดมภ์ (Column) และแนวแถว (Row)
2. คำนวณค่าสัดส่วนในการตอบข้อสอบถูกในแต่ละข้อระหว่างกลุ่มผู้สอบ

3. ทดสอบปฏิสัมพันธ์ทั้งหมด ระหว่างข้อสอบกับกลุ่มผู้สอบ หากพบว่าปฏิสัมพันธ์อย่างมีนัยสำคัญทางสถิติ แสดงว่า แบบทดสอบฉบับนั้นมีความลำเอียง

4. หากผลการทดสอบในข้อ 3 มีปฏิสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติแล้วทำการทดสอบรายข้อว่า มีข้อใดบ้างที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ และถือว่าข้อสอบข้อนั้นเป็นข้อสอบที่มีความลำเอียง

จากขั้นตอนการวิเคราะห์โดยวิธีนี้ จะเห็นว่า ข้อสอบที่ถูกคัดเลือกกว่าเป็นข้อสอบลำเอียงนั้น จะเป็นข้อสอบที่ให้ผลการวัดแตกต่างกันใน 2 กลุ่ม คือ อาจจะเป็นข้อสอบที่ง่ายสำหรับกลุ่มหนึ่ง แต่ยากสำหรับกลุ่มอื่น หรือข้อสอบข้อนั้น อาจไม่ได้วัดความสามารถเดียวกันในผู้สอบแต่ละกลุ่ม

วิธีวิเคราะห์องค์ประกอบ (Factor Analysis)

เป็นวิธีวิเคราะห์ความลำเอียง โดยทำการวิเคราะห์องค์ประกอบ(Laksana and Coffman.1980:2) เพื่อหาน้ำหนักองค์ประกอบ (Factor Loading) จากแต่ละกลุ่มแล้วทำการทดสอบความแตกต่างของน้ำหนักองค์ประกอบระหว่างกลุ่มผู้สอบ 2 กลุ่ม ถ้าทดสอบแล้วพบว่ามีความแตกต่างอย่างมีนัยสำคัญทางสถิติ ถือว่าข้อสอบนั้นมีความลำเอียง

วิธีการวิเคราะห์ความลำเอียงโดยวิธีนี้ จะเห็นว่า ข้อสอบที่ลำเอียงจะเป็นข้อสอบที่ไม่ได้วัดสิ่งเดียวกันในผู้สอบต่างกลุ่มกัน

วิธีวิเคราะห์ด้วยไคสแควร์ (Chi-Square : χ^2)

ผู้ที่เสนอการวิเคราะห์ความลำเอียงวิธีนี้ คือ ชูเนแมน (Intasuwan.1979 : 26 ; citing Scheuneman. 1975) โดยให้นิยามของข้อสอบที่ไม่ลำเอียงว่า เป็นข้อสอบที่ผู้สอบที่มีความสามารถเท่ากัน มีโอกาสตอบถูกเหมือนกัน โดยไม่คำนึงว่าจะเป็นกลุ่มที่มีเชื้อชาติแตกต่างกันหรือไม่ สำหรับความสามารถของผู้สอบจะวัดได้โดยใช้คะแนนที่ได้จากการสอบ คะแนนรวมที่ต่างกัน จะถูกแบ่งออกเป็นช่วง ๆ แล้วคำนวณหาสัดส่วนการตอบข้อสอบแต่ละข้อถูกและความถี่ที่คาดหวัง เพื่อคำนวณค่าไคสแควร์ ข้อสอบที่ลำเอียงจะมีค่าไคสแควร์ที่คำนวณได้ สูงกว่าค่าวิกฤตที่เปิดตาราง

ขั้นตอนการวิเคราะห์ความลำเอียงด้วยวิธีไคสแควร์ มีดังนี้

1. แบ่งคะแนนรวม ซึ่งแทนความสามารถออกเป็น 3-5 ช่วง
2. คำนวณค่าไคสแควร์ แล้วตรวจสอบความมีนัยสำคัญทางสถิติ

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

เมื่อ $df = (I - 1)(J - 1)$

χ^2 แทน Scheunemam's Chi-Square

O_{1j}, O_{2j} แทน ความถี่ที่สังเกตได้ของการตอบถูกของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบ ในแต่ละชั้นคะแนน j

E_{1j}, E_{2j} แทน ความถี่ที่คาดหวังของการตอบถูกของกลุ่มอ้างอิง และกลุ่มเปรียบเทียบในแต่ละชั้นคะแนน j

I แทน จำนวนกลุ่มที่ทำการวิเคราะห์

J แทน จำนวนชั้นคะแนน

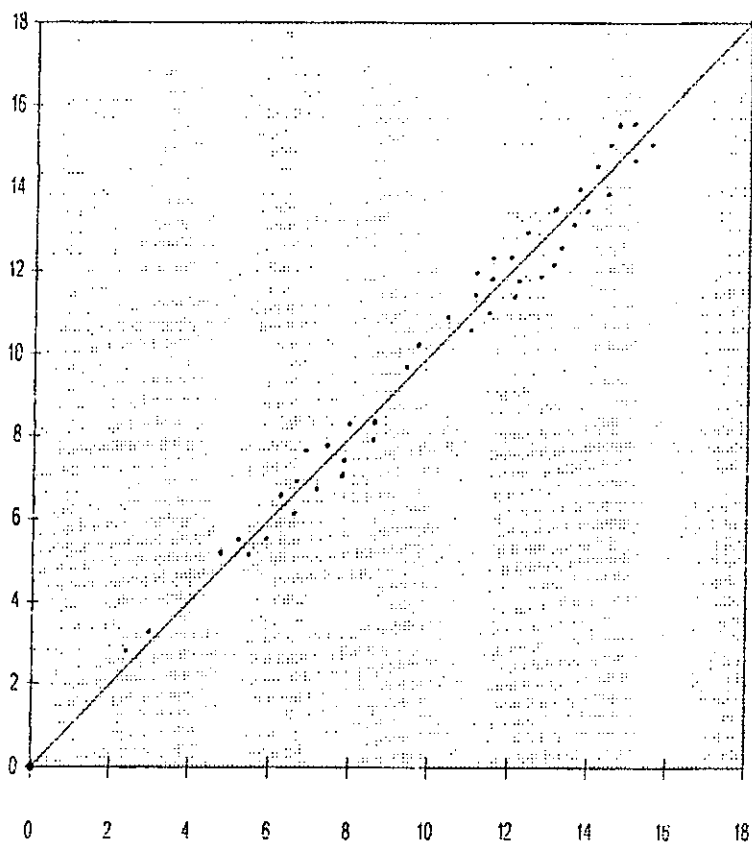
df แทน Degree of Freedom

วิธีแปลงค่าความยาก (Transformed Item Difficulty)

ผู้เสนอการวิเคราะห์ความลำเอียงวิธีนี้ คือ แองกอฟฟ์ (Berk. 1982 : 97 ; citing Angoff and Ford. 1973) โดยนิยามข้อสอบที่มีความลำเอียงว่าเป็นข้อสอบที่เบี่ยงเบนไปจากเส้นแกนหลัก วิธีนี้ใช้ค่าความยาก (p) ของข้อสอบแต่ละข้อ จากกลุ่มผู้สอบแต่ละกลุ่ม (ดังนั้นค่าความยากของข้อสอบจะมี 2 ค่า) มาแปลงให้เป็นคะแนนมาตรฐานที่ แล้วแปลงคะแนนมาตรฐานที่ ให้เป็นค่าความยากมาตรฐานหรือค่าเดลต้า (Δ) โดยใช้สมการ $\Delta = 13 + 4z$ นำค่าเดลต้า (Delta) แต่ละคู่มาลงจุดคู่อันดับบนกราฟไบวาเรียต (Bivariate Graph) จุดต่าง ๆ บนกราฟไบวาเรียตซึ่งแทนค่าเดลต้าจะอยู่ในรูปวงรีแล้วทำการคำนวณค่าระยะห่างตั้งฉากจากจุดคู่อันดับของค่าเดลต้าของข้อสอบแต่ละข้อไปยังเส้นแกนหลัก (Major Axis) ข้อที่มีระยะห่างตั้งฉากมากกว่า หรือเท่า

กับ 0.75 หรือน้อยกว่า หรือเท่ากับ -0.75 ลงไป ถือว่าเป็นข้อสอบที่มีความลำเอียง ดังแสดงตามภาพต่อไปนี้

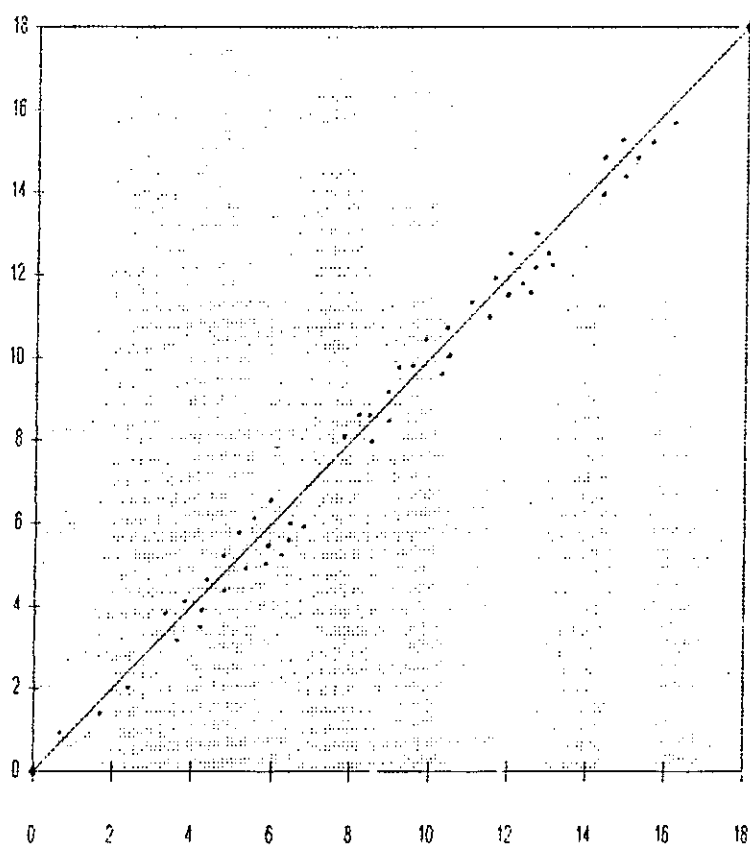
เดลต้าของคนผิวดำกลุ่ม 2



เดลต้าของคนผิวดำกลุ่ม 1

ภาพประกอบ 1 แสดงกราฟไคววรีเอทของค่าเดลต้าของคนผิวดำกลุ่ม 1 และคนผิวดำกลุ่ม 2 จากแบบทดสอบวัดความถนัดทางการเรียนด้านภาษา (Preliminary Scholastic Aptitude (PSAT)-Verble) (Berk. 1982 : 99 ; citing Angoff and Ford. 1973)

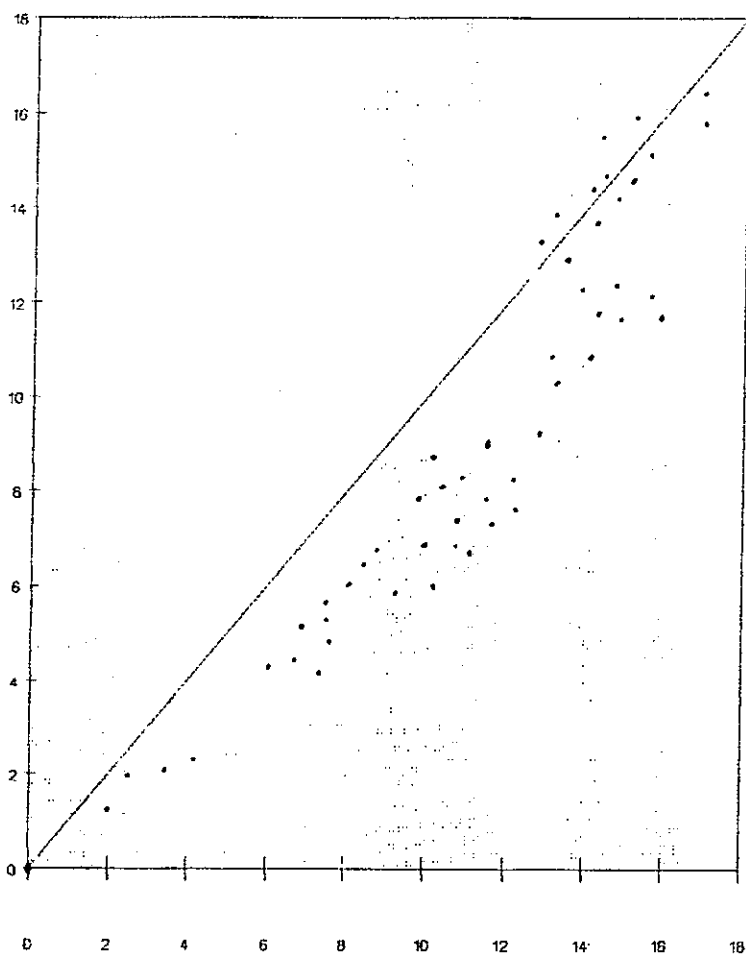
เดลด้าของคนผิขาวกลุ่ม 6



เดลด้าของคนผิขาวกลุ่ม 5

ภาพประกอบ 2 แสดงกราฟไขว้กันของค่าเดลด้าของคนผิขาวกลุ่ม 5 และคนผิขาวกลุ่ม 6 จากแบบทดสอบวัดความถนัดทางการเรียนด้านภาษา (Preliminary Scholastic Aptitude (PSAT)-Verbal) (Berk. 1982 : 100 ; citing Angoff and Ford. 1973)

เดลต้าของคนผิวขาวกลุ่ม 5



เดลต้าของคนผิวดำกลุ่ม 1

ภาพประกอบ 3 แสดงกราฟไบวาริเอท ของค่าเดลต้าของคนผิวขาวกลุ่ม 5 และคนผิวดำ กลุ่ม 1 จากแบบทดสอบวัดความถนัดทางการเรียนด้านภาษา(Preliminary Scholastic Aptitude) (PSAT)- Verbal (Berk. 1982 : 101 ; citing Angoff and Ford. 1973)

เมื่อลงจุดคู่อันดับ ค่าเดลต้า ของข้อสอบแต่ละข้อ ของผู้เข้าสอบแต่ละกลุ่มข้อสอบที่ไม่มี ความลำเอียงจะได้กราฟไบวาริเอท ดังภาพประกอบ 1 และภาพประกอบ 2 ซึ่งภาพประกอบ 1 เป็นกราฟไบวาริเอท ของค่าเดลต้าจากกลุ่มตัวอย่างคนผิวดำ 2 กลุ่ม โดยใช้แบบทดสอบวัดความ ถนัดทางการเรียนด้านภาษา (Berk. 1982 : 99 ; citing Angoff and Ford. 1973) ภาพประกอบ 2 เป็นกราฟไบวาริเอทของค่าเดลต้า ที่ได้จากการทดสอบกลุ่มตัวอย่างคนผิวขาว 2 กลุ่ม โดยใช้

แบบทดสอบชุดเดียวกัน จะเห็นว่าจุดต่าง ๆ ของค่าเฉลี่ยของกลุ่มผู้เข้าสอบ 2 กลุ่มจะกระจายอยู่ในแนวเส้นแกนหลัก คือ ไม่มีการเบี่ยงเบนไปจากเส้นแกนหลัก

ส่วนภาพประกอบ 3 เป็นกราฟไปวาริเอทของค่าเฉลี่ย ที่ได้จากการทดสอบกลุ่มตัวอย่างคนผิวดำกับคนผิวขาว โดยใช้แบบทดสอบชุดเดียวกันจะเห็นว่าจุดต่าง ๆ ของค่าเฉลี่ยของผู้เข้าสอบ 2 กลุ่มนี้จะกระจายห่างจากเส้นแกนหลัก คือ มีการเบี่ยงเบนไปจากเส้นแกนหลัก ข้อสอบเหล่านี้ยากต่อกลุ่มผู้สอบคนผิวดำ ดังนั้นแบบทดสอบชุดนี้มีปฏิสัมพันธ์ระหว่างข้อสอบกับเชื้อชาติ (Item Race Interaction) และเรียกได้ว่าเป็นข้อสอบที่มีความลำเอียงต่อเชื้อชาติ ซึ่งคำนวณได้ดังนี้

จากสมการเส้นแกนหลัก $y = ax + b$

เมื่อ a คือ ความชันของเส้นแกนหลัก

b คือ ค่าคงที่ของการตัดแกน Y

ซึ่งคำนวณค่า a และ b ได้จากสูตร

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

$$b = M_y - aM_x$$

เมื่อ x คือ ค่าเฉลี่ยของกลุ่มผู้สอบผิวขาว

y คือ ค่าเฉลี่ยของกลุ่มผู้สอบผิวดำ

M_x คือ ค่าเฉลี่ยของค่าเฉลี่ย ของกลุ่มผู้สอบผิวขาว

M_y คือ ค่าเฉลี่ยของค่าเฉลี่ย ของกลุ่มผู้สอบผิวดำ

S_x^2 คือ ค่าความแปรปรวนของค่าเฉลี่ยกลุ่มผู้สอบผิวขาว

S_y^2 คือ ค่าความแปรปรวนของค่าเฉลี่ยกลุ่มผู้สอบผิวดำ

S_x คือ ค่าความเบี่ยงเบนมาตรฐานของค่าเฉลี่ยกลุ่มผู้สอบผิวขาว

S_y คือ ค่าความเบี่ยงเบนมาตรฐานของค่าเฉลี่ยของกลุ่มผู้สอบผิดค่า

r_{xy} คือ สหสัมพันธ์ของค่าเฉลี่ยของกลุ่มผู้สอบผิดวิชา และกลุ่มผู้สอบผิดค่า

แองกอฟฟ์ และฟอร์ด (Angoff and Ford. 1973 : 95-105) เสนอการคำนวณระยะทางตั้งฉากจากจุดศูนย์กลางค่าเฉลี่ยของข้อสอบ แต่ละข้อไปยังเส้นแกนหลักคือ

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

เมื่อ d_i คือ ระยะทางตั้งฉากจากจุดศูนย์กลางค่าเฉลี่ยของข้อสอบแต่ละข้อไปยังเส้นแกนหลัก

วิธีการวิเคราะห์ความลำเอียงของข้อสอบ โดยใช้ทฤษฎีการวัดผลแบบดั้งเดิม เป็นวิธีที่ง่ายต่อการคำนวณ และการแปลผล ไม่ซับซ้อน ใช้ได้ในทางปฏิบัติ ไม่ต้องการข้อมูลจำนวนมาก การวิเคราะห์ข้อสอบที่มีความลำเอียง อาศัยค่าสถิติที่ใช้ทดสอบนัยสำคัญ เป็นค่าดัชนีความลำเอียงของข้อสอบ แต่วิธีเหล่านี้มีข้อจำกัดในทางทฤษฎี เนื่องจากผลการวิเคราะห์จะมีค่าไม่คงที่ ขึ้นอยู่กับความสามารถของผู้สอบแต่ละกลุ่ม

ทฤษฎีการตอบข้อคำถาม (Item Response Theory)

ทฤษฎีตอบข้อคำถาม เป็นทฤษฎีที่อธิบายถึงความสัมพันธ์ระหว่างลักษณะหรือความสามารถที่มีอยู่ในตัวบุคคล กับพฤติกรรมการตอบสนองต่อข้อสอบของบุคคลนั้น โดยทฤษฎีนี้มีความเชื่อว่าพฤติกรรมการตอบสนองต่อข้อสอบของบุคคล จะถูกกำหนดโดยลักษณะหรือความสามารถที่มีอยู่ในตัวบุคคล (Lord and Novick. 1968 : 358) ซึ่งไม่สามารถจะสังเกตได้ ทฤษฎีนี้พยายามที่จะอธิบายความสัมพันธ์ระหว่างลักษณะภายในตัวบุคคลกับพฤติกรรมที่บุคคลตอบสนองต่อข้อสอบ โดยใช้แบบจำลองทางคณิตศาสตร์ดังนี้

ให้ P แทน ผลการสอบ (Performance)

θ แทน ความสามารถ (Ability or Trait)

f แทน ฟังก์ชัน (Function) แล้วจะได้ว่า

$$P = f(\theta)$$

ข้อตกลงเบื้องต้นของทฤษฎีการตอบข้อคำถาม

1. ข้อตกลงเกี่ยวกับ Dimensionality

ทฤษฎีนี้ มีข้อตกลงว่า ความสามารถของมนุษย์มีอยู่ทั้งหมด k อย่าง แต่จะมีความสามารถเดียว (Only one trait) อธิบายการตอบข้อสอบของแบบทดสอบได้ หรือคะแนนจากผู้สอบสามารถอธิบายได้เพียงความสามารถเดียว ดังนั้นแบบทดสอบที่ใช้จึงควรเป็นแบบทดสอบที่มีมิติเดียว (Unidimension Test) หมายถึง แบบทดสอบที่นำไปสอบกับกลุ่มตัวอย่าง r กลุ่ม ที่มีความสามารถเท่ากับ θ นำคะแนนการสอบทั้ง r กลุ่มมาแจกแจงความถี่แล้ว คะแนนมีลักษณะการแจกแจงเป็นอย่างเดียวกัน การสร้างแบบทดสอบให้เป็น Unidimension ทำได้โดยการสร้างคำถามที่วัดความสามารถเดียว แล้วทำการวิเคราะห์หองค์ประกอบเพื่อรวมคำถามที่มีลักษณะองค์ประกอบเดียวกันไว้ด้วยกัน

2. ข้อตกลงเกี่ยวกับ Local Independence

ข้อตกลงนี้ กล่าวว่า โอกาสที่ผู้สอบตอบคำถามข้อหนึ่งถูกต้องนั้นไม่เกี่ยวกับการตอบคำถามข้ออื่น ๆ ในแบบทดสอบนั้น

3. ข้อตกลงเกี่ยวกับลักษณะโค้งจากผลการสอบแบบทดสอบและข้อสอบ

โอกาสของการตอบข้อสอบถูก ไม่ขึ้นอยู่กับการแจกแจงความสามารถของกลุ่มประชากรนั้นคือรูปร่างของเส้นโค้งของข้อคำถาม ไม่ได้ขึ้นอยู่กับการกระจายของความสามารถของกลุ่มผู้สอบ จึงเห็นได้ว่า ลักษณะของข้อคำถาม และคุณสมบัติของคำถามนี้ ไม่ขึ้นอยู่กับการลักษณะของกลุ่มผู้สอบ จึงเป็นข้อคำถามที่มีคุณสมบัติคงที่ตลอดเวลา

จากข้อตกลงเบื้องต้น 3 ประการนี้ สามารถนำมาใช้เป็นแนวทางของการวิเคราะห์ความลำเอียงของข้อสอบได้หลายวิธี ที่นิยมใช้ คือ วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์, วิธีโค้งลักษณะ 2 พารามิเตอร์ วิธีโค้งลักษณะ 3 พารามิเตอร์

วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ หรือ การวิเคราะห์ความลำเอียงตามแบบจำลองของราสซ (Rasch Model)

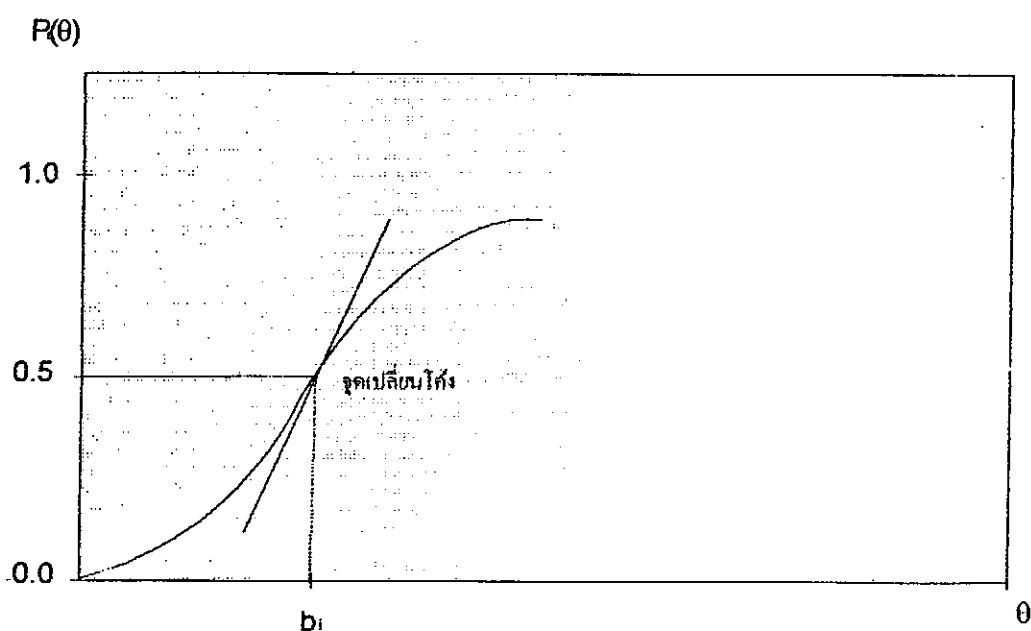
วิธีการนี้ กำหนดให้ข้อสอบทุกข้อมีค่าอำนาจจำแนก (a) เป็น 1.00 ทุกข้อ และค่าการเดา (c) มีค่าเป็น 0 ทุกข้อ แล้วทำการคำนวณค่าความยาก (b) ของข้อสอบจากโค้งลักษณะข้อสอบแต่ละกลุ่ม โดยมีรูปแบบทางคณิตศาสตร์ ดังนี้ (Hambleton, Swaminathan and Rogers, 1991 : 12)

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

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

θ แทน ความสามารถของผู้สอบ

b_i แทน ค่าความยากของข้อสอบข้อที่ i



ภาพประกอบ 4 แสดงความหมายของค่าพารามิเตอร์ข้อสอบตามโมเดลโลจิสติกที่มีพารามิเตอร์ 1

ตัว

จากนั้น นำค่าความยากของแต่ละกลุ่มมาทดสอบนัยสำคัญ โดยใช้ค่าสถิติ t จากสูตร

$$t = \frac{d_1 - d_2}{\sqrt{SE_1^2 + SE_2^2}}$$

เมื่อ d_1, d_2 แทน ค่าความยากของข้อสอบที่ได้จากการวิเคราะห์ ในกลุ่มที่ 1 และกลุ่มที่ 2 ตามลำดับ

SE_1^2, SE_2^2 แทน ค่าความคลาดเคลื่อนมาตรฐานของข้อสอบที่ได้จากการวิเคราะห์

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

วิธีโค้งลักษณะ 2 พารามิเตอร์

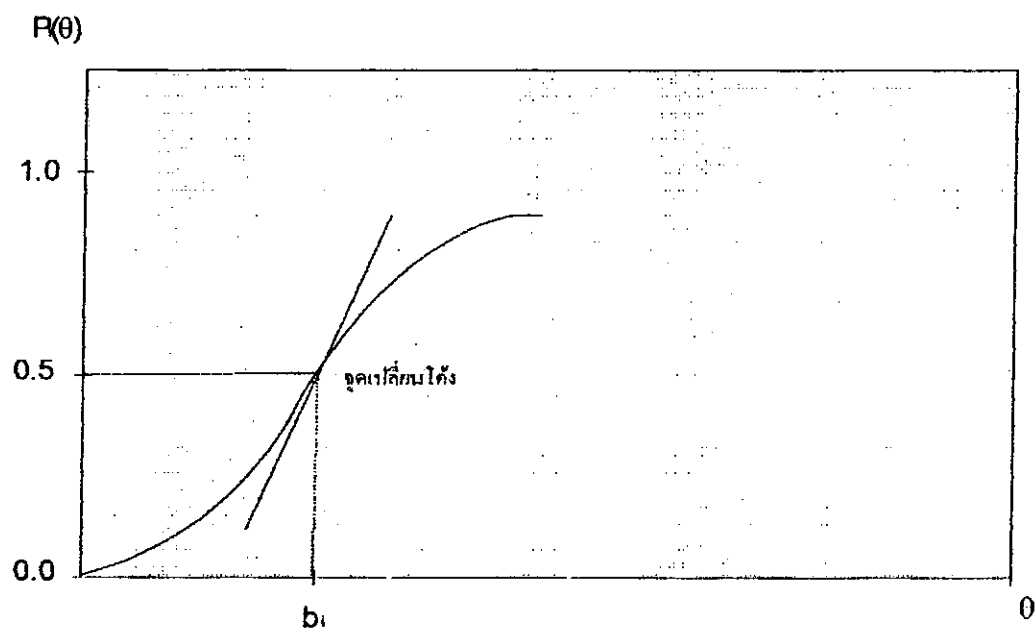
วิธีนี้จะกำหนดให้ค่าการเดา (c) ของข้อสอบทุกข้อเป็น 0 ส่วนค่าอำนาจจำแนกและค่าความยากจะแปรผันไปตามกลุ่มย่อยแต่ละกลุ่ม ซึ่งมีรูปแบบทางคณิตศาสตร์ดังนี้ (Hambleton ,Ronald and Swaminathan. 1985 : 36)

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

เมื่อ b_i แทน ค่าความยากที่ระดับความสามารถที่แท้จริงที่จุดโค้งชันที่สุด

a_i แทน ค่าอำนาจจำแนก ซึ่งมีค่าเป็นสัดส่วนโดยตรงกับค่าความชันของโค้ง ณ จุดเปลี่ยนโค้ง

D แทน ค่า Scaling Factor ซึ่งมีค่า 1.7



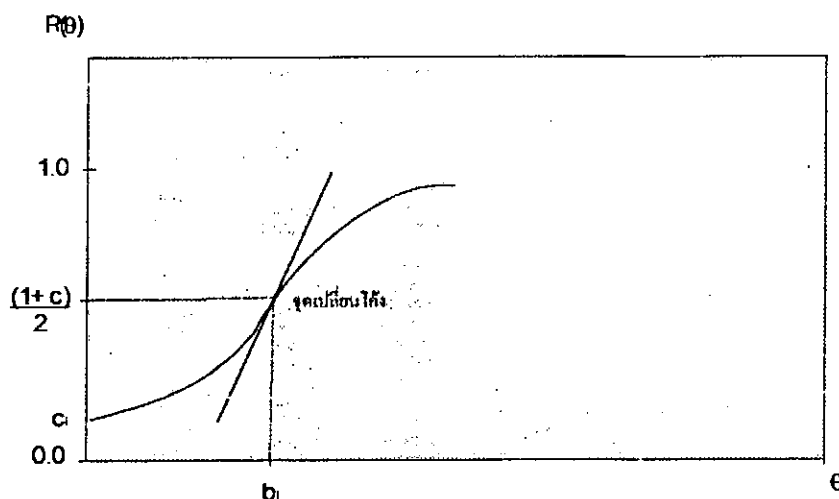
ภาพประกอบ 5 แสดงความหมายของค่าพารามิเตอร์ข้อสอบตามโมเดลโลจิสติกที่มีพารามิเตอร์ 2 ตัว

จากนั้นนำลักษณะโค้งของข้อสอบ แต่ละกลุ่มที่ศึกษาไปทดสอบด้วยสถิติการแจกแจง F ถ้า F มีค่าอย่างมีนัยสำคัญทางสถิติ แสดงว่าข้อสอบข้อนั้นมีความลำเอียง

วิธีโค้งลักษณะ 3 พารามิเตอร์

วิธีนี้ กำหนดให้ ค่าอำนาจจำแนก ค่าความยาก และค่าการเดาแปรผันไปตามกลุ่มย่อย แต่ละกลุ่ม 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 6 แสดงความหมายของค่าพารามิเตอร์ข้อสอบตามโมเดลโลจิสติกที่มีพารามิเตอร์ 3 ตัว

วิธีการวิเคราะห์ความลำเอียงของข้อสอบโดยใช้โค้งลักษณะข้อสอบนั้น จะถือว่าข้อสอบมีความลำเอียง เมื่อผู้สอบที่มีระดับความสามารถเท่ากัน แต่อยู่ต่างกลุ่มกัน มีโอกาสจะตอบถูกไม่เท่ากัน ซึ่งความลำเอียงของข้อสอบจะแสดงให้เห็นด้วยโค้งลักษณะข้อสอบที่แตกต่างกันของข้อสอบข้อเดียวกัน ในทำนองเดียวกัน หากโค้งลักษณะข้อสอบข้อเดียวกันจากผู้สอบต่างกลุ่มกัน มีลักษณะโค้งเหมือนกัน ย่อมแสดงว่า ข้อสอบนั้นไม่ลำเอียง ดังนั้น ค่าดัชนีความลำเอียงของข้อสอบจึงคำนวณโดยการหาความแตกต่างของพื้นที่ที่คำนวณจากโค้งลักษณะข้อสอบจากผู้สอบแต่ละกลุ่ม สำหรับการประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบ จะใช้โปรแกรมสำเร็จรูปในการประมาณค่า

การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้ทฤษฎีการตอบข้อคำถาม จะให้ค่าพารามิเตอร์ของข้อสอบที่เป็นจริง และเหมาะสมกับข้อมูลเนื่องจากการไม่แปรเปลี่ยนของค่า

พารามิเตอร์ ซึ่งเป็นลักษณะเด่นของทฤษฎีนี้ จึงเป็นวิธีวิเคราะห์ที่ใช้ได้ดีตามทฤษฎี แต่ในทางปฏิบัติ นั้น จะต้องใช้จำนวนกลุ่มตัวอย่างขนาดใหญ่มาก และการวิเคราะห์เป็นวิธีการที่ยุ่งยาก ซับซ้อน จำเป็นต้องใช้เครื่องคอมพิวเตอร์ในการคำนวณ ซึ่งทำให้ค่าใช้จ่ายในการวิเคราะห์สูง

วิธีของแมนเทลและแฮนส์เซล (Mantel - Haenszel : MH)

เทคนิคแมนเทล-แฮนส์เซล (MH) (Millsap and Everson, 1993 : 302-304) เป็นเทคนิคที่แมนเทลและแฮนส์เซล (Mantel and Haenszel) ได้เสนอขึ้นใช้ตั้งแต่ปี 1959 แต่ ฮอลแลนด์และเทเยอร์ (Holland and Thayer) เพิ่งนำเสนอมาใช้เพื่อการตรวจสอบความลำเอียงของข้อสอบในปี 1988 และในปีนั้นก็ได้มีการตรวจสอบความลำเอียงของข้อสอบด้วยเทคนิค MH โดยโดแวน, แฮมเบิลตัน, โรเจอร์, ฮอลแลนด์ และไรท์

เทคนิค MH เป็นเทคนิคที่มีความคล้ายคลึงกับวิธีโคสควาร์ที่เสนอโดยชูเนแมน (Scheuneman, 1979) มาราสคูโล และสล็อตเตอร์ (Marascuilo and Slaughter, 1981) และ เมลเลนเบิร์ก (Mellenberg, 1982) โดยเป็นการเปรียบเทียบผลการตอบข้อสอบของกลุ่มผู้สอบ 2 กลุ่ม หรือที่ฮอลแลนด์เรียกว่า กลุ่มอ้างอิงและกลุ่มเปรียบเทียบ (Reference and Focal Group) โดยที่กลุ่มแรกจะให้อ้างถึงกลุ่มที่คาดว่าจะได้ประโยชน์จากข้อสอบและกลุ่มหลังเป็นกลุ่มเสียประโยชน์จากข้อสอบในกรณีนี้ที่ข้อสอบมีความลำเอียงโดยมีการตรวจสอบทุก ๆ ระดับคะแนนรวมจากการสอบ ข้อสอบใดที่ผู้สอบทั้ง 2 กลุ่ม ทำได้เท่า ๆ กัน จะเป็นข้อสอบที่ถือได้ว่าไม่มีความลำเอียงต่อกลุ่มใดกลุ่มหนึ่ง วิธีการในการตรวจสอบมีดังนี้

ถ้าให้ N_{R_j} และ N_{F_j} แทนจำนวนผู้สอบในกลุ่มอ้างอิง และกลุ่มเปรียบเทียบที่มีคะแนนรวมในช่วงคะแนน j และให้ N_j ซึ่ง $= N_{R_j} + N_{F_j}$ เป็นจำนวนผู้สอบรวมทั้ง 2 กลุ่ม ที่ได้คะแนนรวมจากการสอบข้อสอบข้อที่ i ในช่วงคะแนน j แล้ว นำมาเขียนในรูปตาราง 2 ทาง แสดงผลการตอบถูก (1) และผิด (0) ดังตารางที่ 1 (Narayanan and Swaminathan, 1994 : 315-317)

ตาราง 1 แสดงข้อมูลผลการสอบของกลุ่มตัวอย่างกลุ่ม R และ F ที่มีคะแนนรวมอยู่ในช่วงคะแนน j ของข้อสอบข้อที่ 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) และ ผิด (0) ในช่วงคะแนน j

จากตาราง 1 จะได้ว่า ในช่วงระดับคะแนน j A_j และ B_j คือจำนวนผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบข้อที่ถูกต้อง และไม่ถูกต้องตามลำดับ ในขณะที่ C_j และ D_j คือจำนวนผู้สอบกลุ่มเปรียบเทียบที่ตอบข้อสอบข้อที่ถูกต้องและไม่ถูกต้องตามลำดับ เช่นกัน

ตามเทคนิควิธี MH ที่เสนอโดยฮอลแลนด์ และเทเยอร์ (Holland and Thayer) นั้น ข้อมูลการตอบข้อสอบแต่ละข้อจะถูกนำมาแสดงในรูปตาราง 2 ทางนี้ เป็นจำนวนตารางเท่ากับคะแนนรวมที่มีผู้สอบทำได้บวกด้วย 1 จากนั้นจะคำนวณค่าความน่าจะเป็นในรูปของค่าสัดส่วนการตอบข้อสอบถูก-ผิด ระหว่างกลุ่มโดยกำหนดให้เป็นค่า α ดังนี้

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

ค่า α_{MH} จะมีค่าระหว่าง 0 และ ∞ โดยที่ $\alpha_{MH} = 1$ จะแสดงถึงสมมติฐานที่เป็นกลาง ที่ว่าไม่มีความแตกต่างหรือไม่มีความลำเอียง ซึ่งแสดงว่ากลุ่มตัวอย่างทั้ง 2 กลุ่ม ทำข้อสอบได้ผลเช่นเดียวกันในแต่ละระดับคะแนนรวม ค่า $A_j D_j / N_j$ และ $B_j C_j / N_j$ มีค่าเท่ากันทั้ง 2 กลุ่ม ทำให้ $\alpha_{MH} = 1$ แต่ถ้า α_{MH} มีค่ามากกว่า 1 แสดงว่ากลุ่มอ้างอิงจะทำข้อสอบข้อนั้นได้ถูกมากกว่ากลุ่มเปรียบเทียบ และถ้า α_{MH} มีค่าน้อยกว่า 1 แสดงว่ากลุ่มเปรียบเทียบจะทำข้อสอบข้อนั้นได้ถูกมากกว่ากลุ่มอ้างอิง สำหรับสมมติฐานที่เป็นกลางที่แสดงว่าข้อสอบข้อนั้นไม่ลำเอียงเป็นดังนี้

$$H_0 : \left[\pi_{R_j} / (1 - \pi_{R_j}) \right] = \left[\pi_{F_j} / (1 - \pi_{F_j}) \right]$$

$$\text{และ } H_1 : \left[\pi_{R_j} / (1 - \pi_{R_j}) \right] = \alpha \left[\pi_{F_j} / (1 - \pi_{F_j}) \right]$$

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

และ π_{R_j} เป็น ค่าความน่าจะเป็นที่กลุ่มอ้างอิงตอบข้อสอบข้อนั้นถูกในระดับช่วงคะแนน j

π_{F_j} เป็น ค่าความน่าจะเป็นที่กลุ่มเปรียบเทียบตอบข้อสอบข้อนั้นถูกในระดับช่วงคะแนน j

ซึ่งสมมติฐานทั้ง 2 นี้ ใช้ในการตรวจสอบความลำเอียงของข้อสอบแบบยูนิฟอร์ม (Uniform DIF)

อย่างไรก็ตามในทางปฏิบัติแล้ว แม้ว่าจะมีการควบคุมให้ผู้สอบ 2 กลุ่ม มีความสามารถในสิ่งที่ต้องการวัดเท่ากัน แต่เป็นไปได้ที่ผลการตอบข้อสอบของผู้สอบทั้ง 2 กลุ่ม จะเป็นเช่นเดียวกันดังกล่าว แมนเทล และแฮนส์เซล (Mantel and Haenszel, 1959) จึงเสนอค่าสถิติไคสแควร์เพื่อทดสอบค่าที่ได้ว่าจะมีความแตกต่างจาก 1.00 อย่างมีนัยสำคัญหรือไม่ที่ระดับชั้นแห่งความเป็นอิสระ = 1 ดังนี้

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

โดยที่ค่าผลรวมเป็นค่าที่ได้จากการรวมทุกชั้นของคะแนนรวม และค่า $E(A_j)$, $\text{Var}(A_j)$ มีค่าดังนี้

$$E(A_j) = \frac{N_{R_j} N_{1j}}{N_j}$$

$$\text{Var}(A_j) = \frac{N_{R_j} N_{F_j} N_{1j} N_{0j}}{(N_j)^2 (N_j - 1)}$$

เมื่อ $E(A_j)$ และ $\text{Var}(A_j)$ เป็นค่าคาดหวังและความแปรปรวนของ A ที่ระดับคะแนน j

นอกจากนั้น Education Testing Service 's ได้เสนอให้แปลงค่า α_{MH} ให้เป็นคะแนนมาตรฐานในรูปเดลต้าที่มีค่าเฉลี่ยเป็น 0 โดแรน และฮอลแลนด์ (Dorans and Holland. 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 และใช้เป็นตัวชี้แสดงความลำเอียงข้อสอบที่ไม่ลำเอียงจะมีค่า $\alpha_{MH} = 1$ หรือ $\Delta_{MH} = 0$

เท่าที่ผ่านมามีผู้นำวิธี MH มาศึกษา เช่น คลอสเซอร์ และคณะ (Clauser et al. 1991 : 3-7) ใช้ MH วิเคราะห์ความลำเอียงแบบยูนิฟอร์ม (Uniform DIF) พบว่า MH จะใช้ได้ดีในกรณีข้อสอบมีค่าอำนาจจำแนกสูง แต่จะไม่สามารถวิเคราะห์ข้อสอบที่มีค่าความยากสูงมากได้ ซึ่งสอดคล้องกับข้อค้นพบของเมเซอร์ และคณะ (Mazor et al. 1991 : 443-451) จูดวิกส์ และ ทอลแมน (Sudweeks and Tolman.1990 : 8-11) ที่พบว่า ข้อสอบที่ลำเอียงมักเป็นข้อสอบยากส่วนมาก และเฟอราธา (Baghi and Ferrara. 1980 : 14-17) พบว่า เมื่อใช้ขนาดกลุ่มตัวอย่าง 750 คนขึ้นไป MH ใช้แทน IRT-3 ได้ สวามินาธาน และโรเจอร์ (Swaminathan and Rogers. 1990 : 361-370) ซึ่งศึกษาจากข้อมูลจำลองพบว่า MH วิเคราะห์ได้ดีกว่าการถดถอยแบบโลจิสต์เล็กน้อยโดยตรวจค้นได้ถูกต้อง ร้อยละ 75 ในกรณีใช้กลุ่มตัวอย่าง 250 คน และตรวจค้นได้ ร้อยละ 100 กรณีกลุ่มตัวอย่าง 500 คน กรณีที่ลำเอียงแบบยูนิฟอร์มและกรณีที่ลำเอียงแบบนอนยูนิฟอร์มที่ตัดกันที่ปลายข้างใดข้างหนึ่ง แต่ MH มีค่าใช้จ่ายน้อยกว่า โลจิสต์ประมาณ 3-4 เท่า แฮมเบิลตัน และคณะ (Hambleton et al. 1989 : 313) พบว่า MH ให้ค่าใกล้เคียงกับ IRT ทั้งที่ใช้ค่าความแตกต่างของค่าเฉลี่ยกำลังสองและการตรวจสอบความแตกต่างของพื้นที่รวมได้ไค้ แต่ MH มีค่าใช้จ่ายต่ำกว่าและใช้เวลาน้อยกว่า เพริลแมน และคณะ (Periman et al. 1988 : 5-9) พบว่า MH มีปัญหาด้านความเที่ยงเมื่อจำนวนกลุ่มตัวอย่างน้อยกว่า 660 คน ธิสเซน,สไตน์เบอร์ก และวายนอร์ (Thissen ,Steinberg and Wainer. 1988 : 147-170) พบว่า MH ให้ผลการวิเคราะห์คล้าย ITR-LR และอาจใช้วิเคราะห์ความลำเอียงก่อนใช้ IRT-LR ส่วนมิลแซป (Millsap.1992 : 389-402) เสนอว่าความยาวแบบทดสอบที่น้อยกว่า 20 ข้อ วิธี MH จะไม่สามารถตรวจค้นข้อที่มีความลำเอียงได้

วิธีชิปทดสอบ (SIBTEST : SIB)

การทดสอบ SIB เป็นวิธีการที่มีแนวคิดบนการตรวจสอบความลำเอียงของแบบสอบชนิดพหุมิติ (Multidimensional) โดยมีพื้นฐานบนทฤษฎีการตอบข้อสอบแบบพหุมิติ แต่การทดสอบ SIB แตกต่างจาก IRT ตรงที่เป็นการใช้การทดสอบค่าสถิตินอนพารามิเตอร์ (Non-Parametric) และมีการคำนวณที่ง่ายไม่ซับซ้อน ใช้ได้ดีสำหรับการตรวจสอบความลำเอียงอย่างสม่ำเสมอและมีทิศทางเดียว (Unidirectional)

การทดสอบ SIB เป็นวิธีการตรวจสอบความลำเอียงของข้อสอบที่เสนอโดยซิลลีและสเตาท์ (Shealy and Stout, 1993 : 159) และรูปแบบอยู่ที่การวิเคราะห์ข้อมูลที่มีข้อตกลงว่ามีมิติการวัด 2 มิติ มิติหนึ่งเป็นความสามารถหรือลักษณะแฝงเป้าหมายที่ต้องการวัด และอีกมิติเป็นลักษณะแฝงแทรกซ้อนที่ไม่ต้องการวัด

ซิลลีและสเตาท์ (Shealy and Stout, 1993 : 159-194) เริ่มจากกลุ่มผู้สอบ 2 กลุ่ม คือ กลุ่มอ้างอิงและกลุ่มเปรียบเทียบ การตอบข้อสอบของผู้สอบที่สุ่มมาจากกลุ่มผู้สอบคือ U โดยที่ $U = (U_1, \dots, U_n)$ เมื่อ U_i มีค่าเป็น 0 ถ้าตอบข้อสอบข้อ i ผิดและเป็น 1 เมื่อตอบถูก

สำหรับโมเดล IRT แล้ว โดยทั่ว ๆ ไป U จะเกิดได้จาก 2 องค์ประกอบ คือ (1) พารามิเตอร์ความสามารถของผู้สอบจำนวน d มิติ และ (2) ฟังก์ชันการตอบข้อสอบแต่ละข้อ ซึ่งเป็นตัวกำหนดความน่าจะเป็นของการตอบข้อสอบได้ถูก

ในที่นี้ d คือ 2 เพราะเราจะพิจารณาความสามารถแทรกซ้อนเพิ่มเข้ามาจากความสามารถเป้าหมายที่ต้องการวัดเวกเตอร์ความสามารถในที่นี้ คือ (θ, η) เมื่อ θ คือความสามารถเป้าหมายและ η คือ ความสามารถสอดแทรก ส่วนฟังก์ชันการตอบข้อสอบสำหรับข้อ i คือ $P_i(\theta, \eta)$ โดยที่ข้อสอบทุกข้อจะประกอบด้วย θ ในขณะที่บางข้อจะมี η อยู่ด้วย สำหรับข้อที่ i ประกอบด้วย θ อย่างเดียว ฟังก์ชันการตอบข้อสอบ คือ $P_i(\theta)$ ค่า $P_i(\theta, \eta)$ จะสูงขึ้นใน θ และใน η เมื่อข้อสอบข้อที่ i ต้องการความสามารถทั้ง 2 อย่าง และจะสูงขึ้นใน θ ถ้าต้องการเฉพาะความสามารถ θ

นิยามของความลำเอียงของข้อสอบในการทดสอบ SIB คือ "ความลำเอียงของข้อสอบ จะเกิดขึ้นต่อกลุ่มเปรียบเทียบที่ระดับความสามารถ θ ถ้าฟังก์ชันการตอบข้อสอบสำหรับความสามารถเป้าหมาย (Target Ability) ของกลุ่มอ้างอิงมีค่ามากกว่าฟังก์ชัน การตอบข้อสอบของ กลุ่มเปรียบเทียบ" (Stout and Shealy, 1993 : 159-194) ซึ่งแสดงได้ในรูปของ

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

$$\text{เมื่อ } T_R(\theta) = E_R[P(\theta, \eta)/\theta]$$

โดยที่ R คือ กลุ่มอ้างอิง F คือกลุ่มเปรียบเทียบ

θ ความสามารถเป้าหมาย

η ความสามารถสอดแทรกซึ่งอาจเป็นพหุมิติก็ได้

$T_R(\theta)$ ฟังก์ชันการตอบข้อสอบของกลุ่มอ้างอิง

$T_F(\theta)$ ฟังก์ชันการตอบข้อสอบของกลุ่มเปรียบเทียบ

ความลำเอียงของข้อสอบจะเกิดขึ้นเมื่อ

$$\eta_F/\theta < \eta_R/\theta$$

นิยามความลำเอียงในแง่ของแบบทดสอบทั้งฉบับ เมื่อให้ U เป็นการตอบข้อสอบแบบ สุ่มของผู้สอบคนหนึ่ง $h(U)$ คือ คะแนนจากแบบทดสอบ โดยที่ $h(U) = \sum U_i$ แล้วจะได้ว่า ความลำเอียงของข้อสอบจะเกิดต่อกลุ่มเปรียบเทียบในระดับความสามารถ θ ถ้า

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

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

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

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

สมมุติฐานเพื่อการทดสอบความลำเอียงของข้อสอบ หรือ แบบทดสอบของ SIBTEST

$$H_0 : \beta_U = 0$$

$$H_1 : \beta_U > 0$$

การทดสอบทำโดยแบ่งแบบทดสอบออกเป็นแบบทดสอบฉบับย่อย 2 ฉบับคือแบบทดสอบที่มีความเที่ยงตรงซึ่งเป็นแบบทดสอบที่ถือว่าไม่ลำเอียงกับแบบทดสอบที่ต้องการศึกษา และคะแนนของแต่ละแบบทดสอบย่อย คือ

$$X = \sum_{i=1}^n U_i$$

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

เมื่อ X เป็นคะแนนรวมของแบบทดสอบที่มีความเที่ยงตรง ซึ่ง $X = k$ ($k = 0, 1, 2, \dots, n$)

U_i เป็นการตอบข้อสอบข้อที่ i มีค่าเป็น 0 หรือ 1

Y เป็นคะแนนรวมจากแบบทดสอบที่ต้องการศึกษา

$\overline{Y_{R_k}}$ $\overline{Y_{F_k}}$ เป็น คะแนนเฉลี่ยที่ได้จากแบบทดสอบย่อยที่ต้องการศึกษาสำหรับผู้เข้าสอบทุกคนของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบตามลำดับที่มีคะแนนรวม ในแบบทดสอบย่อยที่ถือว่าเป็นแบบทดสอบที่มีความเที่ยงตรง (X) = k และความลำเอียงวัดได้จาก $\overline{Y_{R_k}} - \overline{Y_{F_k}}$ ทั้งนี้เพราะ $\overline{Y_{R_k}} - \overline{Y_{F_k}}$ เป็นความแตกต่างของการตอบของกลุ่มย่อยที่มีความสามารถระดับเดียวกัน (=k) บนความสามารถเป้าหมายและกรณีข้อสอบไม่มีความลำเอียงค่า $\overline{Y_{R_k}} - \overline{Y_{F_k}}$ จะมีค่าเป็น 0

สำหรับค่าสถิติที่ประมาณค่าความลำเอียง β_U คำนวณได้จาก

$$\hat{\beta}_u = \sum_{k=0}^n \hat{p}_k (\bar{Y}_{Rk} - \bar{Y}_{Fk})$$

เมื่อ \hat{p}_k = สัดส่วนของผู้สอบกลุ่มที่เปรียบเทียบที่ตอบแบบทดสอบย่อยที่ถือว่าไม่ลำเอียงได้ถูก k ข้อ

ค่า $\hat{\beta}_u$ จะเป็นค่าประมาณของ β_u ที่มีการปรับแก้โดยวิธีการใช้ค่าถดถอยตามทฤษฎีการตอบข้อคำถามเมื่อมีความแตกต่างระหว่างความสามารถเป้าหมายซึ่งทำให้ค่าสถิติ β_u สูงกว่าปกติ ทำให้เกิดความคลาดเคลื่อนแบบที่ 1 ขึ้น

สถิติทดสอบของ SIBTEST คือ

$$B_u = \hat{\beta}_u / \hat{\sigma}(\hat{\beta}_u)$$

$$\hat{\sigma}(\hat{\beta}_u) = \left\{ \sum_{k=0}^n \hat{p}_k^2 \left[\frac{1}{J_{Rk}} \hat{\sigma}^2(Y/R, k) + \frac{1}{J_{Fk}} \hat{\sigma}^2(Y/F, k) \right] \right\}^{\frac{1}{2}}$$

เมื่อ $\hat{\sigma}(\hat{\beta}_u)$ เป็น การประมาณค่าความคลาดเคลื่อนมาตรฐานของ β_u

J_{Rk}, J_{Fk} เป็น จำนวนผู้สอบในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบที่ทำคะแนนได้เท่ากันในรูปแบบทดสอบย่อยที่เที่ยงตรง ($X = k$)

สำหรับการทดสอบนัยสำคัญทางสถิติของค่า B_u ที่คำนวณได้จากสูตรถ้ามีค่ามากกว่าค่า Z ที่ $P[N(0, 1)]$ จะปฏิเสธสมมติฐานที่เป็นกลางและยอมรับสมมติฐานอื่นที่แสดงว่าข้อสอบนั้นมีความลำเอียง

หลักการวิเคราะห์ความลำเอียงของข้อสอบ

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

การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้เกณฑ์ภายนอก

การวิเคราะห์ความลำเอียงของข้อสอบ โดยใช้เกณฑ์ภายนอกนี้ สามารถวิเคราะห์ได้ทั้งรายข้อและรายฉบับ วิธีการวิเคราะห์จะเขียนกราฟแสดงความสัมพันธ์ ระหว่างคะแนนจากตัวแปรเกณฑ์ภายนอกกับตัวแปรทำนาย วิธีนี้มีจุดมุ่งหมายที่จะวิเคราะห์ความถดถอยของตัวแปรทั้งสอง แล้วทำการเปรียบเทียบค่าความชัน ค่าการตัดแกนของเส้นกราฟ ในการวิเคราะห์ความลำเอียงของข้อสอบทั้งฉบับ จะใช้คะแนนรวมเป็นตัวแปรทำนาย แต่ถ้าเป็นการวิเคราะห์รายข้อจะใช้ค่าความยาก (p) ของข้อสอบแต่ละข้อเป็นตัวแปรทำนาย ส่วนตัวแปรเกณฑ์ภายนอกจะใช้คะแนนรวม หรือเกรดเกี่ยวกับผลสัมฤทธิ์ของงานบางอย่างที่ให้ทำ (Cronbach. 1970 ; Anastssi. 1976)

การวิเคราะห์ความลำเอียงของข้อสอบด้วยหลักการนี้ จึงเป็นการเปรียบเทียบเส้นกราฟระหว่างกลุ่มผู้สอบ โดยถ้าเส้นกราฟดังกล่าวมีค่าความชันและค่าการตัดแกนแตกต่างกันในแต่ละกลุ่มแล้ว ข้อสอบข้อนั้นหรือแบบทดสอบฉบับนั้นก็จะมี ความลำเอียงต่อกลุ่มผู้สอบที่มีค่าการตัดแกนและค่าความชันมากกว่า (Haebara . 1979)

การวิเคราะห์ความลำเอียงของข้อสอบตามหลักการที่กล่าวมานี้ มีจุดอ่อนตรงที่ ในทางปฏิบัติแล้ว เป็นการยากมากที่จะหาตัวแปรเกณฑ์ภายนอกที่มีความเที่ยงตรงเชิงพยากรณ์ และมีความยุติธรรม ซึ่งหากตัวแปรเกณฑ์ภายนอกขาดคุณสมบัติดังกล่าว จะทำให้ผลการวิเคราะห์ความลำเอียงของข้อสอบขาดความสมบูรณ์ และถูกต้องเท่าที่ควร

การวิเคราะห์ความลำเอียงของข้อสอบโดยใช้เกณฑ์ภายใน

การวิเคราะห์ความลำเอียงของข้อสอบ โดยใช้เกณฑ์ภายในนี้ ยึดหลักการที่ว่า ข้อสอบที่ผู้สอบที่มีระดับความสามารถเท่ากัน แต่อยู่ต่างกลุ่มกัน มีคะแนนจริงผลการสอบแตกต่างกันเป็นข้อสอบที่มีความลำเอียง การวิเคราะห์ความลำเอียงโดยหลักการนี้อาจแบ่งได้เป็น 3 หลักใหญ่ ๆ (Laksana and Coffman . 1980) ดังนี้

1. ใช้หลักการวัดความเบี่ยงเบนสัมพัทธ์ (Relative Deviation) ของข้อสอบแต่ละข้อ จากแนวโน้มเข้าสู่ส่วนกลาง วิธีนี้มีข้อตกลงเบื้องต้นว่า ข้อสอบส่วนใหญ่ในแบบทดสอบมีความ

เป็นเอกพันธ์ในการวัดความสามารถใดความสามารถหนึ่งในผู้สอบต่างกลุ่มกัน ข้อสอบใดที่เบี่ยงเบนไปจากส่วนกลางมากกว่าที่คาดหวังไว้ ก็อาจตั้งข้อสงสัยได้ว่าเป็นข้อสอบที่มีความลำเอียง วิธีการหนึ่งที่ใช้กันบ่อย ๆ ก็คือ การวิเคราะห์ความแปรปรวน (ANOVA) ข้อสอบที่ลำเอียงคือ ข้อสอบที่มีปฏิสัมพันธ์ระหว่างข้อกับกลุ่มผู้สอบมีนัยสำคัญทางสถิติ ส่วนอีกวิธีหนึ่งได้แก่ การเขียนกราฟไบวาริเคท วิธีการนี้ค่าพารามิเตอร์เกี่ยวกับค่าความยาก เช่น ค่า p ค่าอาร์คไซน์ ค่าเดลต้า ค่า b จากกลุ่มหนึ่งจะลงจุดคู่อันดับ คู่กับค่าพารามิเตอร์จากอีกกลุ่มหนึ่ง ข้อสอบข้อใดที่เบี่ยงเบนไปจากเส้นแกนหลักมาก ถือว่าเป็นข้อสอบที่มีความลำเอียง

2. ใช้หลักการประเมินความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) ของแบบทดสอบวิธีการหนึ่งที่ใช้กันมาก คือการวิเคราะห์องค์ประกอบ ดัชนีความลำเอียงของข้อสอบตามวิธีการนี้ก็คือ ความแตกต่าง (Discrepancy) ของน้ำหนักองค์ประกอบ (Factor loading) จากผู้สอบต่างกลุ่มกัน หรือความแตกต่างอย่างมีนัยสำคัญ ระหว่างค่าเฉลี่ยของคะแนนตัวประกอบ (Factor scores) จากแต่ละกลุ่มที่นำมาเปรียบเทียบกัน ดัชนีความลำเอียงทั้งสองที่กล่าวมาเป็นตัวชี้ว่า ข้อสอบไม่ได้วัดสิ่งเดียวกันในผู้สอบต่างกลุ่มกัน

3. ใช้หลักการประมาณค่าโอกาส ในการตอบข้อสอบแต่ละข้อถูก วิธีการที่ใช้กันทั่วไป คือวิธีไคสแควร์ (Chi - Square) และวิธีโค้งลักษณะข้อสอบ (Item Characteristic Curve : ICC) วิธีการทั้งสองนี้มีความคล้ายคลึงกันในแง่การใช้โอกาสในการตอบถูกที่แตกต่างกันจากผู้สอบ 2 กลุ่ม หรือ มากกว่า เป็นดัชนีความลำเอียงของข้อสอบ ส่วนข้อแตกต่างระหว่างวิธีการทั้งสอง ก็คือ วิธี ไคสแควร์จะประมาณความสามารถของผู้สอบ โดยใช้คะแนนดิบ ส่วนวิธีโค้งลักษณะข้อสอบจะประมาณความสามารถของผู้สอบจากคุณลักษณะแฝง

งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบ

งานวิจัยที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบนั้น แบ่งเป็นงานวิจัยภายในประเทศ และงานวิจัยต่างประเทศ ดังนี้

งานวิจัยในประเทศ

จากการศึกษารายงานการวิจัยภายในประเทศที่ผ่านมา มีผู้ทำการวิจัยเกี่ยวกับการวิเคราะห์ความลำเอียงของข้อสอบไว้หลายท่าน โดยศึกษาในรูปการเปรียบเทียบผลการวิเคราะห์

ความลำเอียงด้วยวิธีต่าง ๆ กับกลุ่มย่อยที่มีลักษณะแตกต่างกันในด้านภาษา, เพศ และ สภากฎุมิตศาสตร์ รวมทั้งการเปรียบเทียบคุณภาพของแบบทดสอบหลังคิดข้อสอบที่ลำเอียงออก แล้ว นอกจากนี้ยังมีการพัฒนาเกณฑ์ที่ใช้ในการตัดสินความลำเอียงทางเพศ ซึ่งผู้วิจัยจะได้นำ เสนอ ดังนี้

รัชชัย เผ่าพงศ์ (2527 : 73-76) ได้ทำการวิเคราะห์ความลำเอียงของข้อสอบมาตรฐาน วัดความถนัดทางการเรียน ด้านคณิตศาสตร์และภาษา ในระดับมัธยมศึกษาตอนต้น ซึ่งพัฒนา โดยสำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร ระหว่างกลุ่มผู้สอบเพศชายและเพศหญิง โดยการเปรียบเทียบโค้งลักษณะข้อสอบด้วยวิธีแบบ จำลองโลจิสติก แบบ 3 พารามิเตอร์ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ปี การศึกษา 2524 เพศชายและหญิง ซึ่งประกอบด้วยทุกภาคภูมิศาสตร์ทั่วประเทศ 5 ภาค คือ ภาคเหนือ ภาคกลาง ภาคใต้ ภาคตะวันออก และภาคตะวันออกเฉียงเหนือ กลุ่มซึ่งทำการ วิเคราะห์ด้วยแบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์เป็นนักเรียนชาย 1,610 คน และนักเรียนหญิง 1,337 คน ส่วนอีกกลุ่มหนึ่งเป็นนักเรียนชาย 1,316 คน นักเรียนหญิง 985 คน ทำการวิเคราะห์ด้วยแบบทดสอบวัดความถนัดทางการเรียนด้านภาษา ผลการวิเคราะห์พบว่าจาก แบบทดสอบวัดความถนัดทางการเรียนด้านคณิตศาสตร์ จำนวน 30 ข้อ มีข้อสอบที่ลำเอียงต่อ กลุ่มใดกลุ่มหนึ่งโดยเฉพาะ 9 ข้อ คือ ลำเอียงต่อกลุ่มนักเรียนชาย 7 ข้อ และลำเอียงต่อกลุ่ม นักเรียนหญิง 2 ข้อ ข้อสอบที่ลำเอียงต่อกลุ่มนักเรียนชายและหญิงในระดับปานกลางขึ้นไปมี จำนวน 5 ข้อ ซึ่งวัดในเรื่องร้อยละ การหาปริมาตรและการหาความยาวของเส้นในด้านของรูป สามเหลี่ยมจำนวนเรื่องละ 1 ข้อ และเรื่องเกี่ยวกับโจทย์ปัญหา จำนวน 2 ข้อ จากข้อสอบจำนวน 5 ข้อ ดังกล่าวนี้เป็นข้อสอบที่มีความลำเอียงต่อกลุ่มนักเรียนหญิงในช่วงความสามารถแรกและ ลำเอียงต่อกลุ่มนักเรียนชายในช่วงความสามารถต่อมาจำนวน 1 ข้อ ส่วนแบบทดสอบวัดความ ถนัดทางการเรียนด้านภาษาจำนวน 40 ข้อ มีข้อสอบที่ลำเอียงต่อนักเรียนชายโดยเฉพาะ 3 ข้อ และลำเอียงต่อนักเรียนหญิงโดยเฉพาะมี 8 ข้อ ข้อสอบที่มีความลำเอียงในระดับปานกลางขึ้นไป มีจำนวน 9 ข้อ ซึ่งวัดความเข้าใจเกี่ยวกับการอ่านคำประพันธ์ บทร้อยกรองอย่างละ 1 ข้อ และ วัดความเข้าใจเกี่ยวกับการอ่านข้อความจำนวน 7 ข้อ จากข้อสอบจำนวน 9 ข้อนี้เป็นข้อที่มีความ ลำเอียงต่อกลุ่มนักเรียนชายโดยเฉพาะจำนวน 1 ข้อ และเป็นข้อสอบที่มีความลำเอียงต่อกลุ่ม นักเรียนหญิงโดยเฉพาะจำนวน 6 ข้อ และมีข้อสอบจำนวน 2 ข้อ มีความลำเอียงทางการทดสอบ มากกว่า 1 ช่วงความสามารถ

ทัศนีย์ พิรมนตรี (2530) : 77 ได้ทำการวิเคราะห์ความลำเอียงของแบบทดสอบวิชาคณิตศาสตร์ โครงการตรวจสอบคุณภาพการศึกษาของสำนักงานทดสอบทางการศึกษา กรมวิชาการ กระทรวงศึกษาธิการ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526 จำนวน 7,036 คน โดยเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างกลุ่มนักเรียนในกรุงเทพมหานครกับกลุ่มนักเรียนในภาคภูมิศาสตร์ทั้ง 5 ภาค คือ ภาคกลาง ภาคใต้ ภาคเหนือ ภาคตะวันออกเฉียงเหนือ และภาคตะวันออก ด้วยวิธีวิเคราะห์ 3 วิธี คือ วิธีกำหนดจุดค่าเดลต้า วิธีทดสอบความแตกต่างระหว่างกลุ่มด้วยสถิติไคสแควร์ ในโมเดลลอกลีเนียร์ 2 โมเดล คือโมเดลที่ไม่มีพารามิเตอร์ผลรวมระหว่างระดับคะแนนกับกลุ่ม และโมเดลที่ไม่มีพารามิเตอร์ของผลหลักที่เกิดจากกลุ่มและวิธีโค้งลักษณะข้อสอบที่ใช้พารามิเตอร์ 3 ตัว พบข้อสอบที่มีความลำเอียงจากวิธีโค้งลักษณะข้อสอบที่ใช้พารามิเตอร์ 3 ตัวมากที่สุด และในแต่ละวิธีวิเคราะห์พบว่ามีข้อสอบที่ลำเอียงซ้ำกันระหว่างกลุ่มนักเรียนในกรุงเทพมหานครกับกลุ่มนักเรียนในทุกภาค

สุรศักดิ์ อมรรัตนศักดิ์ (2531) : 200-207 ได้ทำการศึกษาเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่แตกต่างกัน 4 วิธี คือ วิธีวิเคราะห์ความแปรปรวน วิธีแปลงค่าความยากของข้อสอบ วิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 1 ตัว และวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว โดยหาสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียง 4 วิธี และเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อนและหลังการศึกษาความลำเอียงของข้อสอบตามวิธีการคิดคะแนนรวมที่แตกต่างกัน 6 วิธี ในด้านจำนวนผู้ได้รับการคัดเลือก สัดส่วนของชาย / หญิง ที่ได้รับการคัดเลือกและความเชื่อมั่นของแบบทดสอบโดยกลุ่มตัวอย่าง คือ ผู้เข้าสอบแข่งขันเพื่อบรรจุเข้ารับราชการเป็นข้าราชการครู ในปี พ.ศ. 2529 จำนวน 2,340 คน เป็นชาย 1,170 คน หญิง 1,170 คน เครื่องมือที่ใช้ในการรวบรวมข้อมูลได้แก่แบบทดสอบที่ใช้สอบแข่งขันเพื่อบรรจุเข้ารับราชการเป็นข้าราชการครูในปี พ.ศ. 2539 ซึ่งมี 4 ฉบับ คือ ฉบับที่ 1 แบบทดสอบวิชาเอก(ความรู้เกี่ยวกับการประถมศึกษา) ฉบับที่ 2 แบบทดสอบวิชาการศึกษาและกฎหมาย ฉบับที่ 3 แบบทดสอบภาษาไทย ฉบับที่ 4 แบบทดสอบความสามารถทั่วไป ผลการศึกษาพบว่า วิธีวิเคราะห์ความลำเอียงแต่ละวิธีพบจำนวนข้อที่ลำเอียงแตกต่างกัน โดยวิธีโค้งลักษณะข้อสอบที่มีพารามิเตอร์ 3 ตัว พบจำนวนข้อสอบที่ลำเอียงมากที่สุด รองลงมาได้แก่ วิธีวิเคราะห์ความแปรปรวน และวิธีแปลงค่าความยากของข้อสอบพบจำนวนข้อที่ลำเอียงน้อยที่สุด ส่วนค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียง 4 วิธี พบว่ามีค่าอยู่ระหว่าง .7535 ถึง .9921 ในด้านจำนวนผู้ได้รับการคัดเลือกเมื่อเปรียบเทียบความแตกต่างของผลการคัดเลือกก่อนและหลังการศึกษาความลำเอียงของ

ข้อสอบ ตามวิธีคิดคะแนนดิบกับวิธีรวมคะแนนแบบอื่น ๆ อีก 5 วิธี มีจำนวนผู้ได้รับการคัดเลือกแตกต่างกันประมาณ 4% ถึง 24% ส่วนวิธีแปลงคะแนนดิบเป็นคะแนนมาตรฐานที่ปกติกับวิธีรวมคะแนนแบบอื่น ๆ อีก 4 วิธี มีจำนวนผู้ได้รับการคัดเลือกแตกต่างกันประมาณ 6% ถึง 23% ส่วนในเรื่องความเชื่อมั่นของแบบทดสอบ พบว่าค่าความเชื่อมั่นของแบบทดสอบลดลงกว่าแบบทดสอบเดิมเล็กน้อย

พิชรี ปิยภักดิ์ (2531) : 64-69 ได้ทำการวิเคราะห์ความลำเอียงของข้อสอบวิชาคณิตศาสตร์ เรื่องสมการ ชั้นประถมศึกษาปีที่ 6 จำนวน 45 ข้อ ที่ผู้วิจัยสร้างขึ้น ระหว่างกลุ่มผู้สอบที่มีเพศ และสภาพภูมิศาสตร์ต่างกัน ด้วยวิธีวิเคราะห์ 3 วิธี คือ วิธีแปลงค่าความยากของข้อสอบ วิธีไคสแควร์ และวิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ โดยทำการเปรียบเทียบจำนวนข้อสอบที่ลำเอียง การวิเคราะห์หาสัมประสิทธิ์สหสัมพันธ์ของวิธีวิเคราะห์ทั้ง 3 วิธี และศึกษาเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบก่อนและหลังคัดข้อสอบที่ลำเอียงออกแล้ว กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชาย หญิง ชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2530 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ ในกรุงเทพมหานคร และจังหวัดสมุทรสาคร จำนวน 1,422 คน ซึ่งเลือกโดยวิธีสุ่มแบบแบ่งชั้น ผลการวิเคราะห์พบว่า วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ พบจำนวนข้อสอบที่ลำเอียงต่อเพศมากที่สุด จำนวน 7 ข้อ รองลงมาคือวิธีไคสแควร์ และวิธีแปลงค่าความยากของข้อสอบ โดยค้นพบได้จำนวนข้อ 3 และ 2 ข้อ ตามลำดับ สำหรับความแตกต่างในด้านสภาพภูมิศาสตร์ ได้ผลการค้นพบเช่นเดียวกัน คือ วิธีโค้งลักษณะข้อสอบ 1 พารามิเตอร์ พบจำนวนข้อสอบที่ "ลำเอียงมากที่สุด" จำนวน 18 ข้อ รองลงมาคือ วิธีไคสแควร์ และวิธีแปลงค่าความยากของข้อสอบ โดยค้นพบได้จำนวนข้อ 3 ข้อ และ 1 ข้อ ตามลำดับ ค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงต่อเพศ ระหว่างวิธีวิเคราะห์ทั้ง 3 วิธี มีค่าระหว่าง .1713 ถึง .5618 และค่าสัมประสิทธิ์สหสัมพันธ์ของดัชนีความลำเอียงต่อกลุ่มนักเรียนกรุงเทพมหานคร กับกลุ่มนักเรียนสมุทรสาคร มีค่าระหว่าง .1868 ถึง .6009 สำหรับค่าความเชื่อมั่นของแบบทดสอบก่อนและหลังการคัดข้อสอบที่ลำเอียงออกแล้ว พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

นิรมล ชัยขวลิต (2537) : 57-58 ได้ทำการเปรียบเทียบผลการวิเคราะห์ความลำเอียงของแบบทดสอบความเข้าใจในการอ่านภาษาไทย จำนวน 50 ข้อ ที่ผู้วิจัยสร้างขึ้นเอง ระหว่างกลุ่มผู้สอบที่มีเพศ ต่างกัน ด้วยวิธีวิเคราะห์ความลำเอียงของข้อสอบตามทฤษฎีคลาสสิกอล 3 วิธี คือ วิธีแปลงค่าความยาก วิธีไคสแควร์ และวิธีวิเคราะห์ความแปรปรวน โดยศึกษาเปรียบเทียบ

จำนวนข้อสอบที่มีความลำเอียงและศึกษาเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบ หลังคัดข้อที่ลำเอียงออกแล้ว ที่คำนวณด้วยสูตร KR-20 และแบบแบ่งครึ่งฉบับ กลุ่มตัวอย่างที่ใช้เป็นนักเรียนชาย-หญิง ชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2536 ของโรงเรียนในสังกัดกรมสามัญศึกษา ส่วนกลาง กลุ่มที่ 5 กรุงเทพมหานคร จำนวน 1,006 คน ซึ่งได้มาโดยการสุ่มอย่างง่าย ผลการวิเคราะห์พบว่า การวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากกับวิธีไคสแควร์ และวิธีแปลงค่าความยากกับวิธีวิเคราะห์ความแปรปรวนมีจำนวนข้อที่มีความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีไคสแควร์กับวิธีวิเคราะห์ความแปรปรวน พบจำนวนข้อที่มีความลำเอียง แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับค่าความเชื่อมั่นของแบบทดสอบหลังคัดข้อที่ลำเอียงออกแล้ว ที่คำนวณด้วยสูตร KR-20 และแบบแบ่งครึ่งฉบับที่วิเคราะห์ด้วยวิธีทั้ง 3 พบว่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติที่ระดับ .05

(กาญจนา วัฒนสุนทร (2538) : 195-207) ได้ศึกษาการพัฒนาเกณฑ์ตัดสินข้อสอบความลำเอียงทางเพศ ด้วยข้อมูลเชิงประจักษ์ โดยใช้ดัชนีความลำเอียง 4 ตัว คือ พื้นที่ระหว่างโค้งการตอบข้อสอบชนิดคิดเครื่องหมาย (SA) และไม่คิดเครื่องหมาย (UA) จากทฤษฎีการตอบข้อคำถามโมเดล 2 พารามิเตอร์ ดัชนีแอลฟา (α_{MH}) จากวิธีแมนเทิล-แฮนส์เซล และเบต้า จากวิธี SIBTEST ใช้ข้อมูลการตอบข้อสอบคัดเลือกบุคคลเข้าศึกษาในสถาบันอุดมศึกษาของทบวงมหาวิทยาลัย ปีการศึกษา 2535 ความยาวของแบบทดสอบเป็น 20, 30 และ 40 ข้อ สำหรับวิชาคณิตศาสตร์ และ 50, 60, 70 และ 80 ข้อ สำหรับวิชาภาษาอังกฤษ ขนาดกลุ่มผู้สอบมี 6 ขนาด คือ 100, 200, 400, 600, 800 และ 1,000 คน การพัฒนาเกณฑ์กระทำโดยคำนวณค่าดัชนีทั้ง 4 ตัว จากข้อมูลการตอบข้อสอบของผู้สอบเพศเดียวกัน เพศละ 50 ค่าสำหรับแต่ละความยาวแบบทดสอบและขนาดกลุ่มผู้สอบ จากนั้นนำค่าดัชนีที่ได้ทั้งหมดมาวิเคราะห์หาค่าเฉลี่ย และกำหนดเกณฑ์จากค่าเฉลี่ย 2 ลักษณะ คือ เกณฑ์ที่กำหนดจากค่าเฉลี่ย ซึ่งรวมค่าดัชนีทุกข้อโดยไม่พิจารณาความแตกต่างในด้านความยาวของแบบทดสอบ และขนาดกลุ่มผู้สอบด้วย จากนั้นนำเกณฑ์ที่กำหนดไว้ไปตัดสินค่าดัชนีที่ได้จากการวิเคราะห์ระหว่างผู้สอบเพศหญิงและชาย พบว่า ความสอดคล้องของการตัดสินภายในดัชนีเดียวกัน มีความไม่คงที่ข้ามขนาดผู้สอบ แต่ความสอดคล้องนี้จะมีแนวโน้มสูงขึ้น ที่ขนาดผู้สอบตั้งแต่ 600 คนขึ้นไป

นอกจากนี้ยังค้นพบว่า เกณฑ์ที่ใช้ในการตัดสินความลำเอียงของข้อสอบระหว่างผู้สอบเพศหญิงและชาย คือ $ISAI > .80$ และ $UA > .50$ สำหรับแบบทดสอบที่สั้นกว่า 50 ข้อ และ $ISAI > .80$ และ $UA > 1.20$ สำหรับแบบทดสอบที่มีความยาวตั้งแต่ 50 ข้อขึ้นไป และ

.06 > α_{MH} > 1.40 สำหรับทุกความยาวแบบทดสอบและทุกขนาดกลุ่มผู้สอบ โดย ผู้วิจัยสรุปว่า ควรใช้ขนาดผู้สอบอย่างน้อย 800 คน สำหรับดัชนี SA และ UA และ 600 คน สำหรับวิธี แมนเทล-แฮนส์เซล และวิธี SIBTEST การตรวจค้นข้อสอบลำเอียงทางเพศมีความไม่คงที่ข้ามขนาดกลุ่มผู้สอบและความยาวของแบบทดสอบ ความสอดคล้องในการตรวจค้นข้อสอบที่ลำเอียงภายในวิธีเดียวกัน ข้ามขนาดผู้สอบค่อนข้างต่ำ แต่จะสูงขึ้นที่ขนาดผู้สอบตั้งแต่ 600 คนขึ้นไป และข้อสอบที่ลำเอียงในวิชาคณิตศาสตร์ส่วนใหญ่จะลำเอียงเข้าสู่ผู้สอบเพศชาย แต่วิชาภาษาอังกฤษจะลำเอียงเข้าสู่เพศหญิง เมื่อใช้ดัชนี SA และ α_{MH} แต่ดัชนี β_{SIB} ให้ผลตรงข้าม

เรวัตี อินทสระ (2539): 248-252) ศึกษาความเที่ยงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่วิเคราะห์ความลำเอียงต่อเพศ ด้วยวิธีทฤษฎีตอบข้อคำถาม (IRT) วิธีแมนเทล-แฮนส์เซล (MH) และวิธี SIBTEST การตัดสินผลการสอบที่คิดคะแนนมาตรฐานที่ปกติ และคะแนนนำหน้าความสามารถและสาเหตุความลำเอียงของข้อสอบ โดยศึกษาความลำเอียงของข้อสอบคัดเลือกเข้าศึกษาในชั้นปีที่ 1 ประเภทรับตรง ปีการศึกษา 2538 ของมหาวิทยาลัยสงขลานครินทร์ ในวิชาภาษาไทย ก วิชาสังคมศึกษา ก วิชาภาษาอังกฤษ กข วิชาละ 8,127 คน (ชาย 2,722 คน หญิง 5,405 คน) วิชาภาษาไทย กข วิชาสังคมศึกษา กข และวิชาภาษาอังกฤษ กขค วิชาละ 5,415 คน (ชาย 1,454 คน หญิง 3,961 คน) ความเที่ยงตรงเชิงพยากรณ์ศึกษาจากคะแนนสอบคัดเลือกกับระดับผลการเรียนภาคเรียนที่ 1 ปีการศึกษา 2538 ของนักศึกษาที่ได้รับการคัดเลือกจากประเภทรับตรง สายวิทยาศาสตร์ 763 คน และสายศิลปศาสตร์ 281 คน และสาเหตุความลำเอียงของข้อสอบจากการระบุสาเหตุของนักวัดผลการศึกษานี้หรืออาจารย์ผู้สอน จำนวน 50 คน และนิสิตนักศึกษาที่เรียนในสาขาวิชานั้น ๆ วิชาละ 50 คน

ผลการศึกษาพบว่า วิธีตรวจสอบความลำเอียงทั้ง 3 วิธี ตัดสินจำนวนข้อสอบที่ลำเอียงแตกต่างกันในวิชาภาษาไทย ก ฉบับที่ 2 และวิชาสังคมศึกษา ก ฉบับที่ 1 ที่ระดับนัยสำคัญทางสถิติที่ .05 นอกนั้นต่างกันที่ระดับนัยสำคัญทางสถิติที่ .01 โดยวิธีใช้ทฤษฎีการตอบข้อคำถามตัดสินจำนวนข้อสอบที่ลำเอียงได้มากที่สุด ความสัมพันธ์ของลำดับที่ของการสอบไม่ว่าจะคิดคะแนนมาตรฐานที่ปกติ หรือ คิดคะแนนนำหน้าความสามารถและใช้ข้อสอบจำนวนทั้งหมด หรือ ใช้เฉพาะข้อสอบที่ปราศจากความลำเอียงต่างมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

ประสิทธิภาพในการทำนายผลสัมฤทธิ์ทางการเรียนในสายวิทยาศาสตร์ การคิดคะแนนนำหน้าความสามารถทั้งใช้ข้อสอบจำนวนทั้งหมด และใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง

มีประสิทธิภาพในการทำนายสูงกว่าการคิดคะแนนมาตรฐานที่ปกติ ส่วนสายศิลปศาสตร์ การคิดคะแนนมาตรฐานที่-ปกติทั้งใช้ข้อสอบจำนวนทั้งหมด และใช้เฉพาะข้อสอบที่ปราศจากความลำเอียง มีประสิทธิภาพในการทำนายสูงกว่าการคิดคะแนนน้ำหนักความสามารถ และสาเหตุของความลำเอียงของข้อสอบต่อเพศทั้งชายและหญิง เกิดจากข้อสอบเป็นข้อคำถามที่ผู้สอบได้รับการฝึกฝนเฉพาะ จะมีโอกาสตอบถูกมากกว่า เป็นเรื่องราวที่กลุ่มนั้น ๆ สนใจ และเป็นข้อสอบที่ถามความจำ

(อังคณา สายยศ (2539) : 24-26) ได้ศึกษาผลการวิเคราะห์ความลำเอียงของแบบทดสอบความเข้าใจในการอ่านตามลักษณะการใช้ภาษาพูดในครอบครัวเป็นภาษาลาว เขมร และภาษาส่วย โดยศึกษาเปรียบเทียบจำนวนข้อที่ลำเอียง ความเชื่อมั่นของแบบทดสอบหลังคิดข้อที่ลำเอียงออกแล้ว และคะแนนที่ได้จากการสอบหลังคิดข้อที่ลำเอียงออกแล้ว กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 ในจังหวัดสุรินทร์ ได้มาโดยการสุ่มแบบแบ่งชั้น อำเภอเป็นชั้นและโรงเรียนเป็นหน่วยการสุ่ม ได้จำนวนโรงเรียนทั้งสิ้น 29 โรงเรียน นักเรียน 860 คน ทำการวิเคราะห์ความลำเอียงด้วยวิธีแมนเทิล - แชนส์เซล ผลการวิจัยพบว่า จำนวนข้อสอบที่ลำเอียงของแบบทดสอบความเข้าใจในการอ่านระหว่างกลุ่มผู้สอบที่ใช้ภาษาพูดเป็นภาษาลาว ภาษาเขมร และภาษาส่วยแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ สำหรับความเชื่อมั่นของแบบทดสอบหลังคิดข้อลำเอียงออกแล้วพบว่า ค่าความเชื่อมั่นระหว่างกลุ่มผู้สอบที่ใช้ภาษาพูดเป็นภาษาลาว ภาษาเขมร และภาษาส่วยแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ส่วนคะแนนที่ได้จากการสอบด้วยแบบทดสอบที่คิดข้อลำเอียงออกแล้วมีคะแนนระหว่างกลุ่มผู้สอบที่พูดภาษาลาวกับภาษาส่วย ภาษาเขมรกับภาษาส่วยมีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ส่วนภาษาลาวกับภาษาเขมรมีคะแนน แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

งานวิจัยต่างประเทศ

สำหรับงานวิจัยต่างประเทศที่เกี่ยวข้องกับการวิเคราะห์ความลำเอียงของข้อสอบนั้น มีผู้วิจัยทำวิจัยไว้หลายท่าน ซึ่งผู้วิจัยได้รวบรวมและสรุปไว้ดังนี้

(ผจงจิต อินทสุวรรณ (Intasuwarn 1979) : 92) ศึกษาวิธีวิเคราะห์ความลำเอียงที่ต่างกัน 3 วิธี คือ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ วิธีราสซ และวิธีไคสแควร์ ของแบบทดสอบ IEA (The International Association for The Evaluation of Education Achivement) กลุ่มตัวอย่างที่ใช้ศึกษาเป็นนักเรียนอายุ 14 ปี ชาวอังกฤษ อเมริกัน และนิวซีแลนด์ จำนวน 8,601 คน

ผลการวิจัยพบว่า สหสัมพันธ์ระหว่างวิธีทั้ง 3 มีค่าอยู่ระหว่าง .51 - .98 และวิธีไคสแควร์ กับวิธีราสซ มีสหสัมพันธ์กันสูงถึง .98

สงบ ลักษณะ (Burrill. 1982 : 168 ; citing Laksana. 1979) ได้วิเคราะห์ความ
ลำเอียงต่อเพศ และผิวของข้อสอบ ในแบบทดสอบ ITBS (IOWA Tests of Basic Skills) ฉบับ
ภาษาด้านคำศัพท์และฉบับคณิตศาสตร์ ด้วยวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ และวิธี
วิเคราะห์ความแปรปรวน ผลการวิจัยพบว่า วิธีวิเคราะห์ความแปรปรวนเป็นวิธีที่ใช้ได้ดีในทาง
ปฏิบัติ แต่เป็นวิธีที่ไม่ครอบคลุม ส่วนวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ เป็นวิธีที่ครอบคลุม
กว่า แต่ต้องใช้กลุ่มตัวอย่างขนาดใหญ่ และการคำนวณยุ่งยากซับซ้อน

รัตเนอร์ และคณะ (Subkoviak. 1984) : 51; citing Ruder et al. 1980) ได้ทำการ
วิเคราะห์ความลำเอียงของข้อสอบ 7 วิธี คือวิธีแปลงค่าความยากของข้อสอบ 2 วิธี และวิธีโค้ง
ลักษณะข้อสอบ 3 วิธี กับข้อมูลที่จำลองขึ้น (Simulate data) ผลการวิจัยพบว่า วิธี โค้งลักษณะ
ข้อสอบ 3 พารามิเตอร์ เป็นวิธีที่วิเคราะห์ความลำเอียงของข้อสอบได้ดีที่สุด รองลงมาได้แก่ หนึ่งใน
สองของวิธีวิเคราะห์ไคสแควร์ ส่วนวิธีแปลงค่าความยากของข้อสอบ ผู้วิจัยแนะนำให้ใช้ได้
ทางปฏิบัติ

สับโคเวียต และคณะ (Subkoviak. 1984) : 49-59) ศึกษาวิธีวิเคราะห์ความลำเอียงของ
ข้อสอบ 3 วิธี คือ วิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ วิธีไคสแควร์ และวิธี แปลงค่าความยาก
ของแบบทดสอบวัดความรู้เกี่ยวกับคำศัพท์ 50 ข้อ เป็นคำศัพท์ภาษาอังกฤษมาตรฐาน จำนวน 40
ข้อ ศัพท์แสงลงข่าวมีคำจำนวน 10 ข้อ กลุ่มตัวอย่างที่ใช้ในการศึกษาเป็นนักเรียนผิวขาว 1,021
คน นักเรียนผิวดำจำนวน 1,008 คน ผลการวิจัยพบว่า วิธีโค้งลักษณะ ข้อสอบ 3 พารามิเตอร์ เป็น
วิธีที่มีประสิทธิภาพมากที่สุด รองลงมาคือวิธีไคสแควร์ ส่วนวิธีแปลง ค่าความยากเป็นวิธีที่นำไปใช้
ในทางปฏิบัติได้ ในกรณีที่ไม่มีเครื่องคอมพิวเตอร์

ดูลิทเทิล และเคลียร์ (Doolittle and Cleary. 1987 : 157-166) ได้วิเคราะห์ความลำเอียง
ต่อเพศของข้อสอบในแบบทดสอบ ACT (Assessment Mathematic Test) ฉบับการคำนวณทาง
คณิตศาสตร์ เรขาคณิต และคณิตศาสตร์การใช้เหตุผล จำนวน 8 ฉบับ กลุ่มตัวอย่างที่ใช้เป็นเพศ
ชายและหญิง จำนวน 1,300 - 1,400 คน โดยเป็นเพศหญิงประมาณร้อยละ 55 วิเคราะห์ข้อมูล
ด้วยดัชนีของ ลินน์ และฮาร์นิสซ์ (Linn and Harnisch. 1981) ซึ่งเป็นโมเดลโลจิสติก แบบ 3
พารามิเตอร์ ถ้าดัชนี Z ที่ได้เป็นบวกแสดงว่าง่าย สำหรับกลุ่มเปรียบเทียบ ถ้าเป็นลบแสดงว่า

ยากสำหรับกลุ่มเปรียบเทียบ ผลการวิจัยพบว่า ค่าดัชนี Z มีค่าเป็นลบสำหรับข้อสอบที่วัดด้าน
เรขาคณิต การใช้เหตุผลเชิงพีชคณิต และเลขคณิตในทุกฉบับ แสดงว่าง่ายสำหรับเพศชาย

— สวามินาธาน และโรเจอร์ (Swaminathan and Roger. 1990 : 361-370) ศึกษาการ
วิเคราะห์ความลำเอียงของข้อสอบ โดยวิธีการถดถอยแบบโลจิสติก (Logistic Regression) กับวิธี
แมนเทิล-แฮนส์เชล (MH) โดยศึกษาผลการตรวจสอบความลำเอียงแบบยูนิฟอร์ม (Uniform DIF)
กับแบบนอนยูนิฟอร์ม (Nonuniform DIF) ด้วยวิธีการวิเคราะห์ความลำเอียงทั้ง 2 วิธี จากกลุ่ม
ตัวอย่างที่จำลองผลการตอบข้อสอบด้วยโปรแกรม DATAGEN แบบ 3 พารามิเตอร์ กลุ่มตัวอย่าง
ที่ใช้มีขนาด 250 และ 500 คนต่อกลุ่ม ความยาวแบบทดสอบมี 3 ขนาด คือ 40 ข้อ 60 ข้อ และ 80
ข้อ โดยแต่ละฉบับจะมีข้อสอบที่ลำเอียง ฉบับละ 20% เป็นข้อสอบที่ลำเอียงแบบยูนิฟอร์ม
(Uniform DIF) จำนวน 10% และแบบนอนยูนิฟอร์ม (Nonuniform DIF) อีก 10% ผลการ
วิเคราะห์พบว่าวิธีวิเคราะห์ความลำเอียงทั้ง 2 วิธีให้ผลใกล้เคียงกัน โดยวิธี MH จะดีกว่าเล็กน้อย
ในขนาดกลุ่มตัวอย่าง 250 คน MH จะตรวจสอบได้ถูกต้องร้อยละ 75 และในขนาดกลุ่มตัวอย่าง
500 คน จะตรวจสอบได้ถูกต้องร้อยละ 100 ในทุกความยาวของแบบทดสอบ เมื่อข้อสอบมีความ
ลำเอียงแบบยูนิฟอร์ม แต่ไม่สามารถตรวจสอบได้ ในข้อสอบที่มีความลำเอียงแบบนอนยูนิฟอร์ม
สำหรับวิธีการถดถอยแบบโลจิสติกจะตรวจสอบได้ถูกต้อง ประมาณร้อยละ 50 เมื่อกลุ่มตัวอย่าง
ขนาดเล็ก และความยาวของแบบทดสอบสั้น แต่จะตรวจสอบได้ถูกต้องร้อยละ 75 ในแบบทดสอบ
ที่ยาวและกลุ่มตัวอย่างมีขนาดใหญ่ ในด้านความคลาดเคลื่อนแบบที่ 1 (การระบุว่าข้อสอบ
ลำเอียงทั้งๆที่ไม่ลำเอียง) วิธี MH จะตรวจค้นผิดประมาณ ร้อยละ 1 ส่วนวิธีการถดถอยแบบ
โลจิสติก จะตรวจค้นผิดประมาณร้อยละ 1 ถึงร้อยละ 6

สำหรับนิยามของความลำเอียงแบบยูนิฟอร์ม (Uniform DIF) และความลำเอียงแบบ
นอนยูนิฟอร์ม (Nonuniform DIF) เมลเลนเบิร์ก (Mellenbergh. 1989 : 127-143) ได้นิยามว่า
ข้อสอบที่ลำเอียงแบบยูนิฟอร์ม คือ ข้อสอบที่ไม่มีปฏิสัมพันธ์ระหว่างระดับความสามารถของ
ผู้สอบกับกลุ่มผู้สอบซึ่งจะมีผลให้โอกาสในการตอบข้อสอบถูกจะสูงสำหรับกลุ่มหนึ่งมากกว่ากลุ่ม
อื่นในทุกระดับความสามารถ และข้อสอบที่ลำเอียงแบบนอนยูนิฟอร์ม คือ ข้อสอบที่มีปฏิสัมพันธ์
ระหว่างระดับความสามารถของผู้สอบกับกลุ่มผู้สอบ ทำให้ความแตกต่างของโอกาสในการตอบ
ข้อสอบถูกสำหรับ 2 กลุ่มต่างกันในทุกระดับความสามารถ

ฟรีเดิล และคอสติน (Freedle and Kostin. 1990 : 329-343) ได้ศึกษาวิธีแปลงค่าความ
ยากของข้อสอบเป็นดัชนีชี้บ่งความลำเอียงต่อสีผิวของแบบทดสอบทางภาษาของ Graduate

Record Examination (GRE) จำนวน 13 ฉบับ 988 ข้อ และแบบทดสอบ Scholastic Aptitude Test (SAT) จำนวน 11 ฉบับ 935 ข้อ ซึ่งเป็นแบบทดสอบวัดความถนัดด้านภาษา 4 ด้านคือ ด้านอุปมาอุปมัย ด้านคำตรงข้าม การเติมคำในประโยคให้สมบูรณ์ และด้านความเข้าใจในการอ่าน กลุ่มตัวอย่างเป็นกลุ่มผู้สอบนิเวศน์ที่ตอบว่าใช้ภาษาอังกฤษเป็นภาษาหลักในแบบทดสอบ GRE และแบบทดสอบ SAT จำนวน 1,688 คน และ 6,332 คน ตามลำดับเป็นกลุ่มเปรียบเทียบ กลุ่มผู้สอบนิเวศน์ที่ตอบว่าใช้ภาษาอังกฤษเป็นภาษาหลักในแบบทดสอบทั้ง 2 ฉบับจำนวน 23,375 คน และ 34,208 คน ตามลำดับเป็นกลุ่มอ้างอิง สำหรับวิธีวิเคราะห์ความลำเอียงใช้วิธีแปลงค่าความยาก (เปลี่ยนค่า p ให้เป็นค่า Z และแปลงเป็นค่าเดลด้า) และวิธี Standardization Approach ที่เสนอโดยคูแวน และคูลิค (Freedle and Kostin. 1990 : 329 ; citing Dorans and Kulick.1986) เพื่อหาขนาดของความลำเอียงและใช้การปรับแก้ค่าร้อยละของการตอบถูกต้องด้วยวิธีต่าง ๆ 3 แบบ จากนั้นนำค่าความลำเอียงจากทั้ง 2 วิธีมาหาความสัมพันธ์เฉลี่ยในแต่ละฉบับของแบบทดสอบ GRE และ SAT แล้วทำการทดสอบนัยสำคัญของค่าสหสัมพันธ์เฉลี่ยที่คำนวณได้ ผลการวิจัยพบว่า ค่าความยากของข้อสอบสามารถนำมาใช้เป็นดัชนีที่บ่งชี้ขนาดความลำเอียงได้ในแบบทดสอบ GRE และ SAT กลุ่มผู้สอบนิเวศน์จะทำข้อสอบที่มีความยากสูงได้ดีกว่ากลุ่มผู้สอบนิเวศน์แต่จะทำข้อสอบที่ง่ายได้ดีน้อยกว่าในระดับความสามารถเดียวกัน เนื่องจากในข้อสอบที่ง่ายมักใช้คำที่มีความหมายหลายอย่าง จึงมีผลต่อค่าความยากและเป็นแหล่งที่ก่อให้เกิดความลำเอียงของข้อสอบ สำหรับการปรับแก้ 3 แบบในการคำนวณหาขนาดความลำเอียงกับการคำนวณหาขนาดความลำเอียงโดยไม่ใช้วิธีปรับแก้ให้ผลไม่แตกต่างกัน

เมเซอร์ คลอเซอร์ และแฮมเบิลตัน (Mazor ,Clouser and Hambleton. 1992 : 443-451) ศึกษาผลกระทบของขนาดกลุ่มตัวอย่างที่มีต่อการวิเคราะห์ความลำเอียงด้วยวิธี MH โดยใช้ข้อมูลที่จำลองด้วยโปรแกรม DATAGEN แบบ 3 พารามิเตอร์ ขนาดกลุ่มตัวอย่างที่ศึกษามี 5 ขนาด คือ 2,000 คน 1,000 คน 500 คน 200 คน และ 100 คน ความยาวแบบทดสอบชุดละ 75 ข้อ ผลการวิจัยพบว่า เมื่อใช้ขนาดกลุ่มตัวอย่าง 2,000 คน วิธี MH จะตรวจค้นความลำเอียงได้ผิดพลาดร้อยละ 30 และเมื่อใช้จำนวนผู้สอบน้อยกว่า 500 คน พบว่าจะตรวจค้นความลำเอียงผิดพลาดร้อยละ 50 และ ข้อสอบที่ไม่สามารถตรวจค้นความลำเอียงได้เป็นข้อสอบที่ยากมาก ข้อสอบที่มีค่าอำนาจจำแนกต่ำ และ ข้อสอบที่ค่าความยากต่างกันเล็กน้อยสำหรับ 2 กลุ่ม

มิลเลอร์ และโอชิม่า (Miller and Oshima. 1992 : 381-388) ได้ศึกษาผลของขนาดกลุ่มตัวอย่าง จำนวนข้อสอบที่ลำเอียง และขนาดของความลำเอียงที่ประมาณค่าความลำเอียงด้วยวิธี

สองชั้นตอน ข้อมูลที่ใช้เป็นข้อมูลที่จำลองชั้นแบบ 2 พารามิเตอร์ จำนวน 40 ข้อ มีค่าอำนาจจำแนกเฉลี่ยเป็น 1.13 ซึ่งอยู่ระหว่าง .41 ถึง 3.82 ส่วนเบี่ยงเบนมาตรฐานเป็น .63 ค่าความยากเฉลี่ยเป็น .25 มีค่าอยู่ระหว่าง -1.99 ถึง 2.10 ส่วนเบี่ยงเบนมาตรฐานเป็น 1.05 เงื่อนไขที่ทำการศึกษามี 24 เงื่อนไข ดังนี้ พารามิเตอร์ของข้อสอบ แบบ 2 พารามิเตอร์ ขนาดความลำเอียง 3 ขนาด คือ น้อย ปานกลาง และแบบผสมโดยที่เงื่อนไขขนาดความลำเอียงแบบน้อยจะมีค่าความยากเพิ่มขึ้น .20 ในขนาดความลำเอียงแบบปานกลางค่าความยากเพิ่มขึ้น .35 และขนาดความลำเอียงแบบผสมเป็นแบบผสมระหว่าง 2 แบบ และเงื่อนไขจำนวนข้อที่ลำเอียง 4 แบบ คือ ร้อยละ 5 ร้อยละ 10 ร้อยละ 20 และร้อยละ 40 ของแบบทดสอบ ($2 \times 3 \times 4$) ดังนั้นความลำเอียงคำนวณจากวิธีที่ใช้ทฤษฎีตอบข้อคำถาม 5 วิธี คือ วิธี SA, UA , SSOS , USOS , WSSOS และวิธี MH ชั้นที่ 2 ของวิธีการเป็นการหาจำนวนข้อสอบที่ผิดพลาดทางบวก (วิเคราะห์ข้อสอบที่ไม่ลำเอียงว่าเป็นข้อสอบที่ลำเอียง) ในชั้นตอนนี้จะมีประโยชน์เมื่อขนาดความลำเอียงมีมากและจำนวนข้อสอบที่ลำเอียงมีมาก (ร้อยละ 20 หรือ ร้อยละ 40 ของแบบทดสอบ) สถิติของ MH จะตรวจสอบความผิดพลาดทางบวกได้น้อย และวิธี MH จะวิเคราะห์ความลำเอียงได้ดีพอ ๆ กับวิธีทฤษฎีตอบข้อคำถาม

โรเจอร์ และสวามินาทาน (Roger and Swaminathan. 1993 : 105-116) ได้ศึกษาเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีถดถอยแบบโลจิสติก (วิธี LR) กับวิธีแมนเทิล - ฮานส์เชล (วิธี MH) โดยใช้ข้อมูลที่จำลองชั้น เงื่อนไขการศึกษาเป็น 4 แบบ คือ ขนาดกลุ่มตัวอย่าง 2 กลุ่ม ขนาด 250 คน และ 500 คนต่อกลุ่ม ระดับความเหมาะสมของโมเดลข้อมูลเป็น 2 ระดับ คือ ระดับที่เหมาะสมใช้รูปแบบ 2 พารามิเตอร์ ระดับที่ไม่เหมาะสมใช้รูปแบบ 3 พารามิเตอร์โดยค่าการเดา (c) จะถูกกำหนดที่ .2 เครื่องมือที่ใช้เป็นข้อสอบจำนวน 40 ข้อที่จำลองชั้นมา 7 แบบ ดังนี้ แบบที่ 1 ค่าความยากและค่าอำนาจจำแนกต่ำ แบบที่ 2 ค่าความยากสูงค่าอำนาจจำแนกต่ำ แบบที่ 3 ค่าความยากปานกลางค่าอำนาจจำแนกต่ำ แบบที่ 4 ค่าความยากปานกลางค่าอำนาจจำแนกสูง แบบที่ 5 ค่าความยากสูงค่าอำนาจจำแนกต่ำ แบบที่ 6 ค่าความยากและค่าอำนาจจำแนกสูง แบบที่ 7 เป็นแบบผสมคือค่าอำนาจจำแนกและค่าความยากต่างกันสำหรับ 2 กลุ่ม โดยวิเคราะห์ร้อยละของความลำเอียงที่ตรวจค้นพบในความลำเอียงแบบยูนิฟอร์มและนอนยูนิฟอร์ม ผลการวิจัยพบว่า ร้อยละของความลำเอียงที่ตรวจค้นพบในความลำเอียงแบบยูนิฟอร์มสูงกว่าแบบนอนยูนิฟอร์มทั้ง 2 วิธีและในทุกเงื่อนไข โดยวิธี LR จะตรวจพบร้อยละของความลำเอียงสูงที่ระดับความยากปานกลางค่าอำนาจจำแนกสูง สำหรับความลำเอียงแบบนอนยูนิฟอร์มพบว่า วิธี LR จะตรวจค้นความลำเอียงลดลงร้อยละ 14 ในโมเดลข้อมูล

ที่ไม่เหมาะสม ส่วนวิธี MH จะตรวจค้นพบลดลงร้อยละ 7 และในขนาดกลุ่มตัวอย่างที่เพิ่มขึ้น วิธี LR ตรวจค้นพบความลำเอียงได้เพิ่มขึ้นร้อยละ 19 ส่วนวิธี MH จะตรวจพบเพิ่มขึ้นร้อยละ 11 และที่ความยากต่ำวิธี MH ตรวจค้นพบได้น้อยกว่าวิธี LR แต่ที่ความยากสูงทั้ง 2 วิธีให้ผลใกล้เคียงกัน แต่วิธี MH ไม่สามารถตรวจสอบได้ในระดับความยากปานกลาง และวิธี LR จะตรวจค้นความลำเอียงได้ดีกว่าวิธี MH ในความลำเอียงทั้ง 2 แบบแต่ค่าใช้จ่ายที่ใช้ในการวิเคราะห์จะสูงกว่าวิธี MH

รูสโซ และสตาท์ (Roussos and Stout, 1996: 215-230) ได้ศึกษาผลการใช้จำนวนกลุ่มตัวอย่างขนาดเล็ก ความแตกต่างของค่าพารามิเตอร์ข้อสอบ รวมทั้งความแตกต่างในค่าเฉลี่ยของความสามารถของกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเพื่อศึกษาผลที่มีต่อความคลาดเคลื่อนแบบที่ 1 (Type I error) คือการยอมรับว่าข้อสอบลำเอียงทั้ง ๆ ที่ไม่ลำเอียงโดยวิธี MH และวิธี SIBTEST

การศึกษาที่ 1 ใช้กลุ่มตัวอย่างเป็นกลุ่มอ้างอิงและกลุ่มเปรียบเทียบกลุ่มละ 100, 200, 500 และ 1000 คน เลือกข้อสอบ 1 ข้อ จากแบบทดสอบ ASVAB เลือกข้อที่มีค่าพารามิเตอร์ของข้อสอบ คือ a , b และ c ใกล้เคียงกับค่าเฉลี่ยของแบบทดสอบมากที่สุด คือ 1.32, 0.03 และ 0.25 (ค่าเฉลี่ย 1.22, 0.09 และ 0.02) และเป็นข้อสอบที่ไม่ลำเอียงความสามารถ (θ) ของกลุ่มตัวอย่างทั้ง 2 กลุ่มถูกจำลองให้มีการกระจายเป็นปกติมีค่าความแปรปรวนเป็น 1.00 ความแตกต่างระหว่างค่าเฉลี่ยของการแจกแจงของ 2 กลุ่ม ถูกจำลองให้มีค่า 0.0, 0.5 และ 1.0 แล้วคำนวณค่าความสามารถเฉลี่ยของแต่ละกลุ่มให้จุดกลางของค่าเฉลี่ยอยู่บนระดับความยากเฉลี่ยของแบบทดสอบ จำลองการตอบ 400 ครั้ง และวิเคราะห์ 400 ครั้งในแต่ละสถานการณ์

ผลการใช้ SIBTEST และ MH ตรวจสอบความลำเอียงของข้อสอบที่เลือกขึ้นมา พบว่านอกจากกลุ่มตัวอย่างขนาด 100 คน และค่าความแตกต่างของความสามารถเป็น 0.0 แล้ว อัตราการเกิดความคลาดเคลื่อนแบบที่ 1 ของวิธีทั้ง 2 แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ แต่ในขนาดกลุ่มตัวอย่าง 100 คนและความแตกต่างของความสามารถเป็น 0.0 จะมีอัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 2 วิธีแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05 โดยที่วิธี SIB จะมีอัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 มากกว่าวิธี MH และในขนาดกลุ่มตัวอย่างไม่เกิน 1,000 คน จะมีแนวโน้มการเกิดความคลาดเคลื่อนสูงทั้ง 2 วิธี

การศึกษาที่ 2 ใช้แบบทดสอบชุดเดิม แต่กำหนดค่าความยากและค่าอำนาจจำแนกที่แตกต่างกัน ดังนี้

ค่าอำนาจจำแนกเป็น 0.4 , 1.0 และ 2.5 ค่าความยากเป็น -1.5 , -0.5 , 0.0 , 0.5 และ 1.5 ค่าการเดาเป็น 0.20 ทำการจำลองการแจกแจงความสามารถให้เป็น 2 แบบคือ 0.0 และ 1.0 นอกนั้นจำลองทุกอย่างเช่นเดียวกับการศึกษาที่ 1 แต่ใช้จำนวนผู้สอบขนาดใหญ่ คือ 500 , 1,000 และ 3,000 คน ในแต่ละสถานการณ์คำนวณซ้ำ 100 ครั้ง และตรวจสอบความลำเอียงโดยใช้ MH และ SIB ผลการวิจัยพบว่า เมื่อค่าอำนาจจำแนกสูงขึ้นและค่าความยากต่ำ อัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 จะสูงทั้ง 2 วิธีที่ระดับนัยสำคัญ .05 และวิธี MH อัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 สูงกว่าวิธี SIB และขนาดกลุ่มตัวอย่างที่เพิ่มขึ้น ขนาด 3,000 คน ที่ค่าความยากเป็น -1.5 ค่าอำนาจจำแนกเป็น 2.5 อัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 สูงถึงกว่าวิธี SIB เมื่อการแจกแจงความสามารถเป็น 1.0 ค่าการเดาเป็น 0.20 วิธี SIB มีอัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธี MH แต่เมื่อใช้การแจกแจงความสามารถที่ 0.0 ไม่มีการเดาแล้ว อัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 ของทั้ง 2 วิธีไม่แตกต่างกัน

สรุปได้ว่า เมื่อใช้กลุ่มตัวอย่างขนาดเล็ก ไม่เกิน 1,000 คน วิธีทั้ง 2 ให้ผลด้านความคลาดเคลื่อนประเภทที่ 1 ใกล้เคียงกัน แต่เมื่อใช้ขนาดกลุ่มตัวอย่างที่มากกว่า 1,000 คน และค่าเฉลี่ยการแจกแจงความสามารถของทั้ง 2 กลุ่มแตกต่างกัน รวมทั้งข้อสอบมีค่าความยากต่ำ ค่าอำนาจจำแนกสูง วิธี SIB มีอัตราการเกิดความคลาดเคลื่อนประเภทที่ 1 ต่ำกว่าวิธี MH

นารายานาน และสวามินาธาน (Narayanan and Swaminathan . 1994 : 315-328) ได้ศึกษาเปรียบเทียบผลการวิเคราะห์ความลำเอียงแบบยูนิฟอร์มด้วยวิธี MH และวิธี SIB โดยใช้ข้อมูลที่จำลองภายใต้เงื่อนไข 1,296 เงื่อนไข ($9 \times 3 \times 2 \times 4 \times 6$) แล้วทำการคำนวณซ้ำ 100 ครั้ง ในแต่ละเงื่อนไข เงื่อนไขที่ทำการศึกษาประกอบด้วย

1. ขนาดกลุ่มตัวอย่างที่ต่างกัน 3 ขนาด คือกลุ่มเปรียบเทียบใช้ขนาด 100, 200 และ 300 คน กลุ่มอ้างอิงใช้ขนาด 300, 500 และ 1,000 คน วางไขว้กันได้กลุ่มตัวอย่างทั้งหมด 9 กลุ่ม
2. ความแตกต่างในการแจกแจงความสามารถ 3 ระดับ ระดับที่ 1 ให้ทั้ง 2 กลุ่มมีค่าเฉลี่ยการแจกแจงความสามารถอยู่ที่ 0.0 ระดับที่ 2 ให้ค่าเฉลี่ยของความสามารถในกลุ่มอ้างอิงและกลุ่มเปรียบเทียบเป็น 0.0 และ 0.5 ตามลำดับ และระดับที่ 3 ค่าเฉลี่ยความสามารถในกลุ่ม

อ้างอิงและกลุ่มเปรียบเทียบเป็น 0.0 และ -1.0 ตามลำดับ ส่วนเบี่ยงเบนมาตรฐานของทั้ง 3 ระดับ เป็น 1.0

3. ร้อยละของข้อสอบที่ลำเอียงมี 2 ขนาด คือ ร้อยละ 10 และร้อยละ 20 จากแบบทดสอบ จำนวน 40 ข้อ

4. ขนาดของความลำเอียง 4 ระดับ คือ 0.4, 0.6, 0.8 และ 1.0

5. ชนิดของแบบทดสอบ 6 ชนิด คือ ชนิดของแบบทดสอบที่ประกอบด้วยค่าความยากและค่าอำนาจจำแนก 3 ระดับคือ สูง ปานกลาง ต่ำ วางไว้ด้วยกัน แต่คัดแบบทดสอบที่ประกอบด้วยค่าความยากและค่าอำนาจจำแนกสูง ปานกลาง ต่ำ ทั้ง 2 ค่าออก จำนวน 3 ฉบับ

ผลการวิจัยพบว่า วิธี MH และวิธี SIB สามารถตรวจค้นความลำเอียงได้ใกล้เคียงกันโดยจะตรวจค้นได้ดีในขนาดกลุ่มตัวอย่าง 600 คน (กลุ่มอ้างอิง และกลุ่มเปรียบเทียบกลุ่มละ 300 คน) ขนาดของความลำเอียงทั้ง 4 ขนาดมีผลต่อการตรวจค้นความลำเอียงทั้ง 2 วิธีอย่างมีนัยสำคัญทางสถิติ ความแตกต่างในการแจกแจงความสามารถไม่มีผลต่อวิธีการวิเคราะห์ความลำเอียงแบบ SIB แต่มีผลต่อวิธี MH ที่เป็นเช่นนี้อาจเป็นเพราะวิธี SIB ได้มีการปรับแก้การถดถอย ขนาดของแบบทดสอบและร้อยละของข้อสอบที่ลำเอียงไม่มีผลต่อวิธีทั้งสอง สำหรับการเกิดความคลาดเคลื่อนแบบที่ 1 วิธี MH จะมีอัตราการเกิดอยู่ในระดับปกติแต่วิธี SIB จะเพิ่มขึ้นเล็กน้อยในกลุ่มที่มีความสามารถเท่ากัน แต่ในกลุ่มที่มีความสามารถไม่เท่ากันการเกิดความคลาดเคลื่อนแบบที่ 1 จะมีอัตราการเกิดสูงขึ้นทั้ง 2 วิธี ซึ่งการเพิ่มความยาวของแบบทดสอบจะทำให้เกิดความเชื่อมั่นสูงขึ้นซึ่งมีผลให้ความคลาดเคลื่อนแบบที่ 1 นำจะมีอัตราการเกิดลดลง

จากการศึกษาผลการวิจัยข้างต้น จะเห็นว่าการศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบมีหลายวิธี บางวิธีเป็นวิธีที่ยึดทฤษฎีการวัดผลแบบดั้งเดิม เช่น วิธีแปลงค่าความยาก วิธีวิเคราะห์ความแปรปรวน วิธีไคสแควร์ เป็นต้น บางวิธียึดทฤษฎีการตอบข้อคำถาม เช่นวิธีโค้งลักษณะข้อสอบ วิธีแมนเทิล - ฮานส์เซล วิธีชิปเทสท์ ซึ่งเป็นวิธีที่ใช้สถิติแบบนอนพาราเมตริก จากวิธีต่าง ๆ ดังกล่าวบางวิธีเป็นวิธีที่ให้ผลการวิเคราะห์ได้ถูกต้อง แม่นยำแต่ไม่เหมาะที่จะนำมาใช้ในทางปฏิบัติเพราะต้องใช้กลุ่มตัวอย่างขนาดใหญ่ การคำนวณและการแปลผลซับซ้อน ยุ่งยาก ต้องเสียค่าใช้จ่ายสูง ในขณะที่บางวิธีเป็นวิธีที่ใช้กลุ่มตัวอย่างไม่มาก การคำนวณและการแปลผลง่าย ไม่ซับซ้อน แต่ให้ผลการวิเคราะห์ที่มีความถูกต้องน้อยกว่าวิธีอื่นและบางวิธีใช้กลุ่มตัวอย่างไม่มาก การคำนวณและการแปลผลง่าย เหมาะที่จะนำมาใช้ในทางปฏิบัติ

ได้ และให้ผลการวิเคราะห์ที่มีความถูกต้อง ซึ่งวิธีที่มีผู้แนะนำให้นำมาใช้เช่น วิธีแมนเทล - แอนส์เซล วิธีชิปเทสท์ และ วิธีแปลงค่าความยาก สำหรับตัวแปรความลำเอียงที่นิยมทำการศึกษา มีดังนี้ ภาษา เพศ วัฒนธรรม ศาสนา สภาพภูมิศาสตร์ เป็นต้น นอกจากนี้จากการศึกษางานวิจัย ยังพบว่า มีตัวแปรบางตัวที่ส่งผลต่อการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีต่าง ๆ เช่น ขนาดกลุ่มตัวอย่าง ความยาวแบบทดสอบ การแจกแจงความสามารถของผู้สอบในแต่ละกลุ่มย่อย เป็นต้น ซึ่งผลการวิจัยที่ผ่านมา ยังไม่มีผู้ศึกษาเปรียบเทียบการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยาก วิธีแมนเทล-แอนส์เซล และวิธีชิปเทสท์ที่ระบุว่า เป็นวิธีที่ใช้ได้ดีในทางปฏิบัติแต่ วิธีใดที่จะให้ผลการวิเคราะห์ได้ถูกต้องที่สุดและเหมาะสมที่สุดและในแต่ละวิธีควรใช้ขนาดกลุ่มตัวอย่างเท่าไรจึงจะเหมาะสมก็ยังไม่ปรากฏว่ามีผู้ศึกษาและจากการศึกษาเอกสารอ้างอิงดังกล่าว ผู้วิจัยจึงกำหนดสมมติฐานในการวิจัยแบบไม่มีทิศทางดังนี้

3) สมมติฐานในการวิจัย

1. จำนวนข้อสอบที่ลำเอียงจากการใช้วิธีวิเคราะห์ต่างกัน เมื่อมีขนาดกลุ่มผู้สอบเท่ากันจะมีจำนวนแตกต่างกัน
2. จำนวนข้อสอบที่ลำเอียงระหว่างกลุ่มผู้สอบขนาดต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกันจะมีจำนวนแตกต่างกัน
3. ความเชื่อมั่นของแบบทดสอบหลังตัดข้อลำเอียงออกแล้วจากการใช้วิธีวิเคราะห์ต่างกัน เมื่อมีขนาดกลุ่มผู้สอบเท่ากันจะมีค่าแตกต่างกัน
4. ความเชื่อมั่นของแบบทดสอบหลังตัดข้อลำเอียงออกแล้วระหว่างกลุ่มผู้สอบขนาดต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกันจะมีค่าแตกต่างกัน

บทที่ 3

วิธีดำเนินการศึกษาค้นคว้า

ประชากร

ประชากรที่ใช้ในการศึกษาค้นคว้าเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2539 ของโรงเรียนสังกัดกรมสามัญศึกษา กระทรวงศึกษาธิการ ในเขตกรุงเทพมหานคร ส่วนกลาง จำนวน 113 โรงเรียน จำนวนห้องเรียน 1,332 ห้องเรียน มีจำนวนนักเรียนทั้งหมด 58,545 คน จำแนกออกเป็น 3 ขนาดคือ โรงเรียนขนาดใหญ่พิเศษ จำนวน 50 โรงเรียน 667 ห้องเรียน จำนวนนักเรียน 31,978 คน โรงเรียนขนาดใหญ่ จำนวน 49 โรงเรียน 557 ห้องเรียน จำนวนนักเรียน 23,265 คน และโรงเรียนขนาดกลาง จำนวน 14 โรงเรียน 108 ห้องเรียน จำนวนนักเรียน 3,302 คน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าเป็นนักเรียนชายหญิง ชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2539 ของโรงเรียนสังกัดกรมสามัญศึกษา กระทรวงศึกษาธิการ ในเขตกรุงเทพมหานคร ส่วนกลาง จำนวน 2,200 คน ซึ่งมีความเชื่อมั่นได้ร้อยละ 95 ($\alpha = .05$) (ลิ้วน สายยศ และ อังคณา สายยศ. 2536 : 261 ; อ้างอิงมาจาก Yamane. 1967) โดยแบ่งเป็นกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์ข้อสอบที่มีความลำเอียงและกลุ่มตัวอย่างที่ใช้ในการคำนวณหาค่าความเชื่อมั่นของแบบทดสอบหลังคิดข้อลำเอียงออกแล้วโดยใช้การสุ่มกลุ่มตัวอย่างแบบสองขั้นตอน (Two-states Random Sampling) ซึ่งมีลำดับขั้นตอนในการสุ่ม ดังนี้

ขั้นที่ 1 สุ่มโรงเรียนจากแต่ละขนาดของโรงเรียนด้วยสัดส่วน 0.04 หรือร้อยละ 4 ด้วยวิธีสุ่มแบบแบ่งชั้น (Stratified Random Sampling) มีขนาดโรงเรียนเป็นชั้น (Strata) และมีโรงเรียนเป็นหน่วยการสุ่ม (Sampling unit) ได้โรงเรียนขนาดใหญ่พิเศษ จำนวน 2 โรงเรียน โรงเรียนขนาดใหญ่ จำนวน 2 โรงเรียน และโรงเรียนขนาดกลาง 1 โรงเรียน

ขั้นที่ 2 สุ่มห้องเรียนจากโรงเรียนที่สุ่มได้ในแต่ละขนาด โดยวิธีสุ่มอย่างง่าย (Simple Random Sampling) มีห้องเรียนเป็นหน่วยการสุ่ม ได้จำนวนห้องเรียน 53 ห้องเรียน แบ่งเป็นกลุ่มตัวอย่างที่ใช้ในการวิเคราะห์ข้อสอบที่มีความลำเอียง จำนวน 24 ห้องเรียน จำนวนนักเรียน 1,000

คน เป็นนักเรียนชาย 501 คน นักเรียนหญิง 499 คน และ กลุ่มตัวอย่างที่ใช้ในการหาค่าความเชื่อมั่นของแบบทดสอบหลังคัดข้อลำเอียงออกแล้ว จำนวน 29 ห้องเรียน จำนวนนักเรียน 1,200 คน ดังรายละเอียดที่เสนอในตาราง 2

ตาราง 2 จำนวนนักเรียนในกลุ่มตัวอย่างที่ใช้ในการวิจัย จำแนกตามขนาดของโรงเรียน

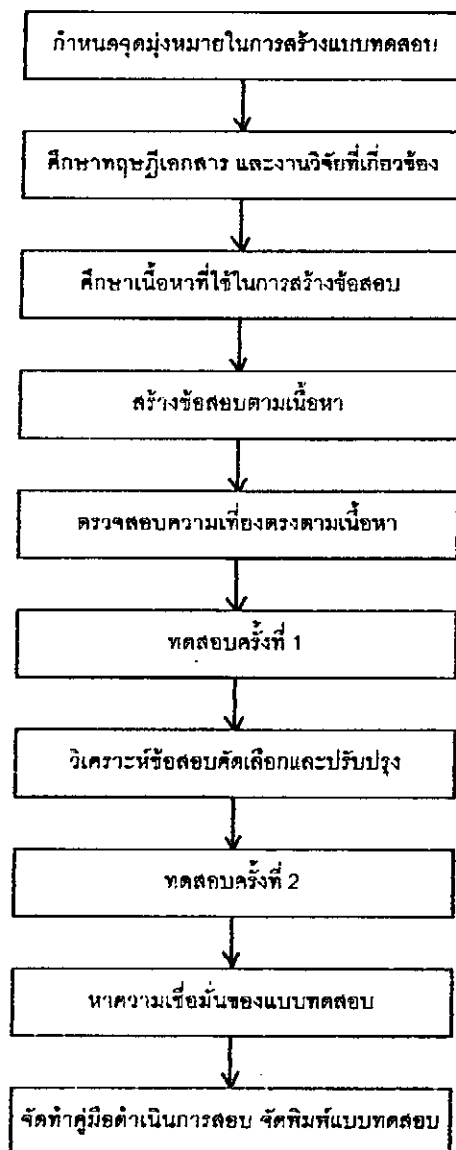
| ขนาด | โรงเรียน | จำนวนนักเรียนที่ใช้ ในการวิเคราะห์ข้อ คำถาม(ห้อง) | จำนวนนักเรียนที่ใช้ ในการหาความเชื่อมั่น หลังตัดข้อ คำถาม(ห้อง) |
|-----------|-----------------------------|---|--|
| ใหญ่พิเศษ | สามเสนวิทยาลัย | 294(6) | |
| | บดินทร์เดชา | 188(4) | |
| | สุรศักดิ์มนตรี | | 376(8) |
| | เทพศิรินทร์ร่มเกล้า | | 196(4) |
| ใหญ่ | กุนนทีรุทธาราม วิทยาคม | 234(6) | |
| | รัตนโกสินทร์สมโภช บางเขน | 175(5) | |
| | สตรีศรีนครินทร์ บ่อเป็ด | | 310(8) |
| | ประชากรศาสตร์อุปถัมภ์ | | 175(5) |
| | | | |
| กลาง | มักกะสันพิทยาศาสตร์ | 109(3) | |
| | วัดสระเกศ | | 143(4) |
| | รวม | 1,000(24) | 1,200(29) |

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นแบบทดสอบคณิตศาสตร์โจทย์ปัญหา ชั้นมัธยมศึกษาปีที่ 1 (ค.102) ในเรื่อง สมการ อัตราส่วนและร้อยละ ตามหลักสูตรมัธยมศึกษาตอนต้น พุทธศักราช 2521 (ฉบับปรับปรุง 2533) เป็นแบบเลือกตอบชนิด 5 ตัวเลือก จำนวน 40 ข้อ ที่ผู้วิจัยสร้างขึ้นเอง

วิธีดำเนินการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า

ในการสร้างเครื่องมือสำหรับการศึกษาค้นคว้าครั้งนี้ได้ดำเนินการสร้างตามขั้นตอนดังนี้



ภาพประกอบ 7 แสดงลำดับขั้นตอนการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า

รายละเอียดขั้นตอนการสร้างเครื่องมือ

จากแผนภูมิดำเนินการสร้างเครื่องมือ มีรายละเอียดดังนี้

1. กำหนดจุดมุ่งหมายในการสร้างแบบทดสอบเพื่อสร้างแบบทดสอบคณิตศาสตร์โจทย์ปัญหา ชั้นมัธยมศึกษาปีที่ 1 (ค.102) ในเรื่อง สมการ อัตราส่วนและร้อยละ เพื่อใช้ในการวิจัยเปรียบเทียบผลการวิเคราะห์ความลำเอียงของแบบทดสอบที่ใช้วิธีวิเคราะห์ต่างกัน

2. ศึกษา ทฤษฎี เอกสาร และงานวิจัยที่เกี่ยวข้องกับวิธีการสร้างแบบทดสอบปรนัยชนิดเลือกตอบคณิตศาสตร์โจทย์ปัญหาจากหนังสือการสร้างเครื่องมือวัดผลการเรียนรู้ และวารสารการวัดผล

3. ศึกษาเนื้อหาที่จะใช้ในการเขียนข้อสอบจากหลักสูตร คู่มือครู แบบเรียน และเอกสารที่เกี่ยวข้องในวิชาคณิตศาสตร์ (ค.102) ชั้นมัธยมศึกษาปีที่ 1 ในเรื่องสมการ อัตราส่วนและร้อยละ ได้ขอบเขตเนื้อหาวิชาในการสร้างแบบทดสอบที่ใช้ในการศึกษาวิจัย ดังนี้ การแก้โจทย์ปัญหาสมการที่มีตัวไม่ทราบค่าหนึ่งตัวและการแก้โจทย์ปัญหาอัตราส่วน สัดส่วนและร้อยละที่มีตัวไม่ทราบค่าหนึ่งตัว

4. สร้างแบบทดสอบคณิตศาสตร์โจทย์ปัญหา แบบเลือกตอบชนิด 5 ตัวเลือก ตามขอบเขตเนื้อหาที่กำหนดไว้ในข้อ 3 จำนวน 60 ข้อ โดยสร้างตามตารางวิเคราะห์รายละเอียด .

5. นำแบบทดสอบที่สร้างขึ้นจำนวน 60 ข้อ ไปให้ผู้เชี่ยวชาญซึ่งมีความรู้ในเนื้อหา และการเขียนข้อสอบวิชาคณิตศาสตร์ ที่มีประสบการณ์การสอนคณิตศาสตร์มาแล้วไม่ต่ำกว่า 5 ปี จำนวน 2 คน กับนักวัดผลการศึกษา จำนวน 3 คน เป็นผู้ตรวจสอบความเที่ยงตรงเชิงเนื้อหา (Content Validity) ของแบบทดสอบเป็นรายข้อ โดยพิจารณาความสอดคล้องระหว่างข้อสอบแต่ละข้อกับเนื้อหา และข้อสอบที่ถือว่าวัดได้ตรงตามเนื้อหาจะต้องได้รับความเห็นชอบจากผู้เชี่ยวชาญ 3 ใน 5 คน จึงถือว่าข้อสอบข้อนั้นใช้ได้ ในการสร้างเครื่องมือที่ใช้ในการวิจัยครั้งนี้ เมื่อให้ผู้เชี่ยวชาญพิจารณาและทำการแก้ไขปรับปรุงภาษาที่ใช้ในข้อสอบบางข้อแล้วนำไปให้ผู้เชี่ยวชาญพิจารณาอีกครั้ง ผลการพิจารณาปรากฏว่าข้อสอบที่สร้างขึ้นและทำการแก้ไขข้อบกพร่องแล้ว จำนวน 60 ข้อ เป็นข้อสอบที่ใช้ได้ มีความเที่ยงตรงเชิงเนื้อหาทุกข้อ

6. นำแบบทดสอบที่ผู้เชี่ยวชาญพิจารณาแล้วไปทดสอบครั้งที่ 1 กับกลุ่มตัวอย่างจำนวน 310 คน โดยสุ่มจากประชากรที่เหลือจากกลุ่มตัวอย่างที่ใช้ในการศึกษา โดยวิธีสุ่มอย่างง่าย (Simple Random Sampling) ใช้ห้องเรียนเป็นหน่วยการสุ่ม (Sampling Unit) ได้โรงเรียนโยธินบูรณะ จำนวน 6 ห้องเรียน มีนักเรียน 169 คน และโรงเรียนเตรียมอุดมศึกษาน้อมเกล้า จำนวน 3 ห้องเรียน มีนักเรียน 141 คน

7. นำแบบทดสอบมาตรวจให้คะแนน แล้วนำผลการสอบไปวิเคราะห์หาค่าความยากและค่าอำนาจจำแนกของข้อสอบเป็นรายข้อและรายตัวเลือก ทั้งตัวถูกและตัวลวง ด้วยโปรแกรม Item Analysis ผลการวิเคราะห์คุณภาพรายข้อของแบบทดสอบ 60 ข้อได้ค่าความยากรายข้อ

ของตัวถูก อยู่ระหว่าง 0.068 ถึง 0.829 และค่าอำนาจจำแนกรายข้อของตัวถูกอยู่ระหว่าง -0.039 ถึง 0.684 แล้วทำการพิจารณาคัดเลือกข้อสอบที่มีคุณภาพอยู่ในเกณฑ์ คือ ข้อสอบที่มีค่าความยากอยู่ระหว่าง 0.20 ถึง 0.80 และค่าอำนาจจำแนกตั้งแต่ 0.20 ขึ้นไป ได้จำนวน 40 ข้อ มีค่าความยากรายข้อของตัวถูกอยู่ระหว่าง 0.210 ถึง 0.610 และค่าอำนาจจำแนกรายข้อของตัวถูกอยู่ระหว่าง 0.325 ถึง 0.526 แล้วจัดเรียงข้อใหม่จากค่าความยาก (ค่า p) สูงไปค่าความยาก (ค่า p)ต่ำ เพื่อนำไปทดลองหาค่าความเชื่อมั่นต่อไป

8. นำแบบทดสอบที่คัดเลือกแล้ว จำนวน 40 ข้อ ไปทดสอบครั้งที่ 2 เพื่อหาค่าความเชื่อมั่นของเครื่องมือโดยใช้สูตร KR-20 ของคูเดอร์ ริชาร์ดสัน กับกลุ่มตัวอย่าง จำนวน 308 คน ซึ่งได้มาโดยการสุ่มอย่างง่าย (Simple Random Sampling) มีห้องเรียนเป็นหน่วยการสุ่ม (Sampling Unit) ได้ โรงเรียนเบญจมราชานุสรณ์ จำนวน 6 ห้องเรียน มีนักเรียน 227 คน และโรงเรียนบดินทรเดชา (สิงห์ สิงหเสนี) จำนวน 2 ห้องเรียน มีนักเรียน 81 คน ได้ความเชื่อมั่นของแบบทดสอบที่ใช้เป็นเครื่องมือในการวิจัยมีค่า 0.855

9. นำแบบทดสอบตามข้อ 8 จำนวน 40 ข้อ จัดพิมพ์เป็นรูปเล่ม และจัดทำคู่มือดำเนินการสอบ เพื่อทำการวิจัยต่อไป

ลักษณะของเครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นแบบทดสอบคณิตศาสตร์โจทย์ปัญหา ชั้นมัธยมศึกษาปีที่ 1 (ค.102) เรื่อง สมการ อัตราส่วน และ ร้อยละเป็นแบบเลือกตอบชนิด 5 ตัวเลือก จำนวน 40 ข้อ ที่ผู้วิจัยสร้างขึ้น ซึ่งมีลักษณะ ดังนี้

ตัวอย่างข้อสอบคณิตศาสตร์โจทย์ปัญหา

(0) โรงเรียนแห่งหนึ่ง มีนักเรียนชั้นมัธยมศึกษาปีที่ 1 จำนวน 660 คน เลือกแผนการเรียนคณิตศาสตร์คิดเป็น $\frac{1}{3}$ ของจำนวนนักเรียนทั้งหมด เลือกแผนการเรียนพาณิชยกรรม จำนวน 55 คน ที่เหลือเลือกแผนการเรียนภาษา คิดเป็น มีนักเรียนที่เรียนแผนภาษากี่คน ?

ก. 165 คน

ข. 220 คน

ค. 275 คน

ง. 385 คน

จ. 440 คน

(คำตอบที่ถูกต้อง คือ ข้อ ง.)

(00) แดงฝากเงิน 3,100 บาท กับธนาคารออมสิน ธนาคารคิดดอกเบี้ยเงินฝากทบต้นในอัตราร้อยละ 8.50 ต่อปี ครบ 3 ปี จึงปิดบัญชี แดงจะได้รับดอกเบี้ยเงินฝากทั้งหมดเท่าไร ?

ก. 263.50 บาท

ข. 790.50 บาท

ค. 859.50 บาท

ง. 3,363.50 บาท

จ. 3,890.50 บาท

(คำตอบที่ถูกต้อง คือ ข้อ ค.)

การตรวจให้คะแนน

การตรวจให้คะแนนของแบบทดสอบเลือกตอบวิชาคณิตศาสตร์ ชนิด 5 ตัวเลือก ใช้เกณฑ์การตรวจให้คะแนนดังนี้ คือ ถ้าตอบถูกให้ 1 คะแนน ตอบผิดหรือไม่ตอบ หรือตอบมากกว่า 1 แห่ง ในข้อนั้นๆ ให้ 0 คะแนน

วิธีดำเนินการเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลเพื่อทำการวิจัยในครั้งนี้ ผู้วิจัยได้ดำเนินการตามลำดับขั้นตอน ดังนี้

1. ติดต่อขอความร่วมมือกับโรงเรียนที่เป็นกลุ่มตัวอย่าง เพื่อกำหนด วัน เวลา ในการดำเนินการสอบ
2. จัดเตรียมแบบทดสอบให้เพียงพอกับจำนวนนักเรียนที่จะทำการสอบในแต่ละครั้ง
3. วางแผนดำเนินการสอบโดยผู้วิจัยเป็นผู้ดำเนินการสอบด้วยตนเอง
4. ชี้แจงให้นักเรียนที่เป็นกลุ่มตัวอย่างทราบวัตถุประสงค์ของการสอบเพื่อให้ได้ผลสอบตามความเป็นจริง
5. นำแบบทดสอบไปทดสอบกับกลุ่มตัวอย่างครั้งที่ 1 จำนวน 310 คน
6. นำผลการสอบในข้อ 5 มาวิเคราะห์หาค่าความยาก และค่าอำนาจจำแนกของข้อสอบเป็นรายข้อ แล้วคัดเลือกข้อสอบไว้ จำนวน 40 ข้อ
7. นำแบบทดสอบที่มีคุณภาพในข้อ 6 จำนวน 40 ข้อ ไปทดสอบกับกลุ่มตัวอย่างครั้งที่ 2 จำนวน 308 คน เพื่อหาค่าความเชื่อมั่นของแบบทดสอบทั้งฉบับ
8. นำแบบทดสอบที่มีคุณภาพตามข้อ 7 จำนวน 40 ข้อไปทดสอบกับกลุ่มตัวอย่างที่มีจำนวน 1,000 คน แล้วสุ่มกลุ่มตัวอย่างจำนวน 500 คน เพื่อศึกษาความลำเอียงของข้อสอบจำแนกในแต่ละวิธีวิเคราะห์และทดสอบสมมติฐานการวิจัย ข้อที่ 1 และ 2
9. นำผลการวิเคราะห์ความลำเอียงตามข้อ 8 มาคัดข้อที่ลำเอียงออกแล้วศึกษา ค่าความเชื่อมั่นของแบบทดสอบหลังคัดข้อลำเอียงออกกับกลุ่มตัวอย่างจำนวน 1,200 คนและทำการทดสอบสมมติฐานการวิจัยข้อที่ 3 และ 4

4) สถิติที่ใช้ในการวิเคราะห์ข้อมูล

1. สถิติที่ใช้ในการหาคุณภาพของเครื่องมือ

1.1 สถิติพื้นฐาน คือ ค่าเฉลี่ย และ ส่วนเบี่ยงเบนมาตรฐาน

1.2 ค่าความยากและค่าอำนาจจำแนกของแบบทดสอบทั้งตัวถูกและตัวลวง ด้วยโปรแกรม Item Analysis

1.3 ค่าความเชื่อมั่นของแบบทดสอบด้วยวิธีของคูเดอริชาร์ดสัน สูตรที่ 20 (KR-20) (ลิวน สายยศ และ อังคนา สายยศ. 2536 : 168)

$$r_u = \frac{n}{n-1} \left[1 - \frac{\sum pq}{s_i^2} \right]$$

| | | |
|-------------|-----|--|
| เมื่อ r_u | แทน | ความเชื่อมั่นของแบบทดสอบ |
| n | แทน | จำนวนข้อของแบบทดสอบ |
| p | แทน | สัดส่วนของผู้ตอบถูกในข้อหนึ่ง ๆ |
| q | แทน | สัดส่วนของผู้ตอบผิดในข้อหนึ่ง ๆ หรือ คือ $1-p$ |
| S_i^2 | แทน | ความแปรปรวนของแบบทดสอบทั้งฉบับ |

2. สถิติที่ใช้ในการวิจัย

2.1 สถิติที่ใช้ในการวิเคราะห์ความลำเอียงของข้อสอบ

2.1.1. วิธีแปลงค่าความยากของข้อสอบ (Transformed Item Difficulty)
มีวิธีดำเนินการดังนี้ (Angoff .1982 : 96)

2.1.1.1 คำนวณค่าความยาก(p) ของข้อสอบทุกข้อจากผู้สอบแต่ละกลุ่ม

2.1.1.2 แปลงค่า p เป็นคะแนนมาตรฐาน (Z) แล้วเปลี่ยนให้อยู่
ในรูปของเดลต้า Δ Values โดยใช้สมการ $\Delta = 13+4z$

2.1.1.3 คำนวณค่าระยะห่างตั้งฉากจากจุดคู่อันดับค่าเดลต้า
ของข้อสอบแต่ละข้อไปยังเส้นแกนหลัก จากสูตร

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

| | | | |
|-------|-------|-----|---|
| เมื่อ | d_i | แทน | ระยะห่างตั้งฉากจากจุดคู่อันดับค่าเดลต้าไปยังเส้นแกนหลัก |
| | x_i | แทน | ค่าเดลต้าของกลุ่มอ้างอิง สำหรับข้อที่ i |
| | y_i | แทน | ค่าเดลต้าของกลุ่มเปรียบเทียบ สำหรับข้อที่ i |
| | a | แทน | ความชันของเส้นแกนหลัก |
| | b | แทน | ค่าคงที่ของการตัดแกน y |

ซึ่ง a, b คำนวณจากสูตรดังนี้

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

$$b = M_y - aM_x$$

| | | | |
|-------|------------|-----|--|
| เมื่อ | s_x^2 | แทน | ความแปรปรวนของค่าเดลต้ากลุ่มอ้างอิง |
| | s_y^2 | แทน | ความแปรปรวนของค่าเดลต้ากลุ่มเปรียบเทียบ |
| | s_x | แทน | ความเบี่ยงเบนมาตรฐานของค่าเดลต้ากลุ่มอ้างอิง |
| | s_y | แทน | ความเบี่ยงเบนมาตรฐานของค่าเดลต้ากลุ่มเปรียบเทียบ |
| | r_{xy} | แทน | สหสัมพันธ์ระหว่างค่าเดลต้าของผู้สอบ 2 กลุ่ม |
| | M_x, M_y | แทน | ค่าเฉลี่ยของค่าเดลต้าของกลุ่มอ้างอิง และกลุ่มเปรียบเทียบ ตามลำดับ |

2.1.1.4 เกณฑ์ที่ใช้ในการตัดสินความลำเอียง คือระยะห่างจาก เส้นแกนหลัก $\geq \pm 0.75$ ถือว่าเป็นข้อสอบที่มีความลำเอียง (Berk. 1982 : 107 ; citing Angoff and Ford. 1973)

2.1.1.5 ระยะห่างที่คำนวณได้ในข้อ 2.1.1.3 คือดัชนีความลำเอียงของข้อสอบ

2.1.2 วิธีของแมนเทล - แฮนส์เชล โดยใช้สูตร ดังนี้ (Narayanan and Swaminathan. 1994 : 316)

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

| | | | |
|-------|---------------|-----|---|
| เมื่อ | A_j | แทน | จำนวนผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบถูก |
| | B_j | แทน | จำนวนผู้สอบกลุ่มอ้างอิงที่ตอบข้อสอบผิด |
| | C_j | แทน | ผู้สอบกลุ่มเปรียบเทียบที่ตอบข้อสอบถูก |
| | D_j | แทน | ผู้สอบกลุ่มเปรียบเทียบที่ตอบข้อสอบผิด |
| | N_j | แทน | ผู้สอบทั้งสองกลุ่มที่ได้รับคะแนนรวมจากการสอบอยู่ในช่วงคะแนน j |
| | α_{MH} | แทน | ค่าดัชนีความลำเอียงของข้อสอบ |

2.1.2.1 สถิติที่ใช้ทดสอบค่าความแตกต่างจาก 1.00 อย่างมีนัยสำคัญที่ระดับชั้นแห่งความเป็นอิสระ เท่ากับ 1 ใช้สูตรดังนี้ คือ

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

เมื่อ χ^2_{MH} เป็นค่าการทดสอบความลำเอียงอย่างมีนัยสำคัญของข้อสอบ

$E(A_j)$ และ $Var(A_j)$ เป็นค่าความคาดหวังและค่าความแปรปรวนของ A ที่ระดับคะแนน j

$$\text{เมื่อ } E(A_j) = \frac{N_{Rj} N_{1j}}{N_j}$$

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

2.1.3 วิธี SIBTEST โดยใช้สูตร (Narayanan and Swaminathan, 1994 : 317)

$$\hat{\beta}_u = \sum_{k=0}^n \hat{p}_k (\bar{Y}_{Rk} - \bar{Y}_{Fk})$$

| | | | |
|-------|----------------|-----|--|
| เมื่อ | \hat{p}_k | แทน | สัดส่วนของผู้สอบกลุ่มเปรียบเทียบที่ตอบแบบทดสอบย่อยที่ไม่ลำเอียง ได้ถูก k ข้อ |
| | \bar{Y}_{Rk} | แทน | คะแนนเฉลี่ยของผู้สอบทุกคนในกลุ่มอ้างอิงที่ตอบแบบทดสอบย่อย ที่ศึกษาได้ถูก k ข้อ |
| | \bar{Y}_{Fk} | แทน | คะแนนเฉลี่ยของผู้สอบทุกคนในกลุ่มเปรียบเทียบที่ตอบแบบทดสอบย่อย ที่ศึกษาได้ถูก k ข้อ |

2.1.3.1 สถิติที่ใช้ในการทดสอบ ค่า SIBTEST เมื่อค่า B_u หรือค่า Z_{SIB} ที่คำนวณจากสูตรแตกต่างจากค่า Z_u ที่ $N(0,1)$ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05

$$B_u = \hat{\beta}_u / \hat{\sigma}(\hat{\beta}_u)$$

$$\hat{\sigma}(\hat{\beta}_u) = \left\{ \sum_{k=0}^n \hat{p}_k^2 \left[\frac{1}{J_{Rk}} \hat{\sigma}^2(Y/R, k) + \frac{1}{J_{Fk}} \hat{\sigma}^2(Y/F, k) \right] \right\}^{\frac{1}{2}}$$

| | | | |
|-------|-------------------------------|-----|--|
| เมื่อ | $\hat{\sigma}(\hat{\beta}_u)$ | แทน | ค่าประมาณความคลาดเคลื่อนมาตรฐานของ β_u |
| | J_{Rk} | แทน | จำนวนผู้สอบในกลุ่มอ้างอิงในช่วงคะแนน k |
| | J_{Fk} | แทน | จำนวนผู้สอบในกลุ่มเปรียบเทียบในช่วงคะแนน k |

โดยที่กลุ่มอ้างอิงแทนกลุ่มที่ได้เปรียบในการตอบข้อสอบได้แก่ กลุ่มผู้สอบเพศชาย

กลุ่มเปรียบเทียบแทนกลุ่มที่เสียเปรียบในการตอบข้อสอบได้แก่ กลุ่มผู้สอบเพศหญิง

2.2 สถิติที่ใช้ในการทดสอบความแตกต่างของจำนวนข้อที่มีความลำเอียงจากการวิเคราะห์ความลำเอียงด้วยวิธีและขนาดกลุ่มผู้สอบต่างกัน ใช้สูตร (ลัวัน สายยศ และ อังคณา สายยศ. 2536 : 93)

$$Z = \frac{p_1 - p_2}{S_{p_1-p_2}}$$

เมื่อ Z แทน ค่าการแจกแจงของ Z

p_1 แทน สัดส่วนของข้อสอบที่ลำเอียงของกลุ่มที่ 1 คำนวณจากจำนวนข้อสอบที่ลำเอียงของกลุ่มที่ 1หารด้วยข้อสอบทั้งหมด

$$p_1 = \frac{f_1}{n_1}$$

p_2 แทน สัดส่วนของข้อสอบที่ลำเอียงของกลุ่มที่ 2 คำนวณจากจำนวนข้อสอบที่ลำเอียงของกลุ่มที่ 2 หารด้วยจำนวนข้อสอบทั้งหมด

$$p_2 = \frac{f_2}{n_2}$$

$S_{p_1-p_2}$ แทน ค่าความคลาดเคลื่อนมาตรฐานของความแตกต่างระหว่างสัดส่วนที่วิเคราะห์ความลำเอียงของ 2 กลุ่ม ซึ่งมีสูตรดังนี้

$$S_{p_1-p_2} = \sqrt{pq \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}$$

เมื่อ p แทน สัดส่วนของข้อสอบที่มีความลำเอียงที่วิเคราะห์ด้วยวิธีวิเคราะห์ทั้ง 2 วิธี ซึ่งคำนวณมาจากจำนวนข้อสอบที่ลำเอียง ที่วิเคราะห์วิธีที่ 1 และ วิธีที่ 2 หารด้วยจำนวนข้อสอบของทั้ง 2 วิธี

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

$$\therefore q = 1 - p$$

และ $q = 1 - p$

2.3 ค่าความเชื่อมั่นของแบบทดสอบหลังคิดข้อลำเอียงออกแล้วด้วยวิธีแบ่งครึ่งฉบับ (Split Half) (ล้วน สายยศ และอังคณา สายยศ. 2536 : 166)

$$r_{tt} = \frac{2r_{1/2, 1/2}}{1 + r_{1/2, 1/2}}$$

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

เมื่อ r_{tt} แทน ความเชื่อมั่นของแบบทดสอบทั้งฉบับ

$r_{1/2, 1/2}$ แทน ความเชื่อมั่นของแบบทดสอบครึ่งฉบับ คำนวณโดยใช้สูตร Pearson Product - Moment Correlation Coefficient

2.4 สถิติใช้ทดสอบความแตกต่างของค่าความเชื่อมั่นหลังคิดข้อลำเอียงออกแล้วจากการวิเคราะห์ความลำเอียงด้วยวิธีและขนาดกลุ่มผู้สอบต่างกัน ใช้ทดสอบด้วยไคสแควร์ (χ^2) ดังนี้ (Snedecor and Cochran. 1967 : 187)

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

เมื่อ χ^2 แทน ค่าไคสแควร์

Z_i แทน คะแนนมาตรฐานซึ่งแปลงจากสัมประสิทธิ์ความเชื่อมั่นด้วยตาราง Fisher - Z

n_i แทน จำนวนนักเรียนในแต่ละกลุ่ม

k แทน จำนวนแบบทดสอบ

df แทน degree of freedom มีค่า = $k - 1$

2.5 สถิติที่ใช้ทดสอบความแตกต่างของค่าความเชื่อมั่นรายคู่

(Ferguson. 1981 : 196)

$$Z = \frac{Z_{r_1} - Z_{r_2}}{\sqrt{\frac{1}{(N_1 - 3)} + \frac{1}{(N_2 - 3)}}$$

| | | | |
|-------|--------------------|-----|--|
| เมื่อ | Z | แทน | คะแนนมาตรฐานโค้งปกติ |
| | Z_{r_1}, Z_{r_2} | แทน | คะแนนมาตรฐานที่แปลงจากค่าสัมประสิทธิ์ความเชื่อมั่นด้วยตาราง Fisher-Z |
| | N_1, N_2 | แทน | จำนวนนักเรียนในกลุ่มย่อย 2 กลุ่มตามลำดับ |

บทที่ 4

ผลการวิเคราะห์ข้อมูล

ในการเสนอผลการวิเคราะห์ข้อมูล เพื่อความสะดวกและเข้าใจตรงกัน ผู้วิจัยจึงได้กำหนดสัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล ดังนี้

สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล

| | | |
|---------------|-----|--|
| n | แทน | จำนวนข้อของแบบทดสอบ |
| N | แทน | จำนวนนักเรียนในกลุ่มตัวอย่าง |
| p | แทน | ค่าความยากง่ายของข้อสอบที่ได้จากการวิเคราะห์โดยใช้ทฤษฎีดั้งเดิม |
| r | แทน | ค่าอำนาจจำแนกของข้อสอบที่ได้จากการวิเคราะห์โดยใช้ทฤษฎีดั้งเดิม |
| d_i | แทน | ค่าดัชนีความลำเอียงของข้อสอบที่ได้จากการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแปลงค่าความยาก |
| α_{MH} | แทน | ค่าดัชนีความลำเอียงของข้อสอบที่ได้จากการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแมนเทิล-แฮนส์เซล ที่ใช้ค่าแอลฟา |
| χ^2_{MH} | แทน | ค่าทดสอบความลำเอียงอย่างมีนัยสำคัญของข้อสอบที่ได้จากการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแมนเทิล-แฮนส์เซล |
| Z_{SIB} | แทน | ค่าทดสอบความลำเอียงอย่างมีนัยสำคัญของข้อสอบที่ได้จากการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีซิปเทสท์ |
| TID | แทน | การวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแปลงค่าความยาก |
| MH | แทน | การวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแมนเทิล-แฮนส์เซล |
| SIB | แทน | การวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีซิปเทสท์ |

- r_s แทน ค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับของแบบทดสอบที่คำนวณโดยใช้สูตร Spearman- Brown
- X^2 แทน ค่าไคสแควร์
- Z แทน คำนัยสำคัญของกาทดสอบด้วยสถิติ Z-test
- Z_r แทน คะแนนมาตรฐานจากค่าสัมประสิทธิ์ของความเชื่อมั่นด้วยตาราง Fisher-Z

การเสนอผลการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้ ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลตามลำดับ ดังต่อไปนี้

1. ค่าดัชนีความลำเอียงของข้อสอบแต่ละข้อจำแนกในแต่ละวิธีวิเคราะห์ความลำเอียง และขนาดกลุ่มผู้สอบ
2. เปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียงที่แตกต่างกันเมื่อขนาดกลุ่มผู้สอบเท่ากัน
3. เปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงเมื่อขนาดกลุ่มผู้สอบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน
4. เปรียบเทียบค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วระหว่างวิธีวิเคราะห์ความลำเอียงที่แตกต่างกันเมื่อขนาดกลุ่มผู้สอบเท่ากัน
5. เปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วเมื่อขนาดกลุ่มผู้สอบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน

5
ผลการวิเคราะห์ข้อมูล 1. ค่าดัชนีความลำเอียงของข้อสอบแต่ละข้อ จำแนกในแต่ละวิธีวิเคราะห์ความ
 * ←

1. ค่าดัชนีความลำเอียงของข้อสอบแต่ละข้อ จำแนกในแต่ละวิธีวิเคราะห์ความ
 ลำเอียง และขนาดกลุ่มผู้สอบ

ผู้วิจัยได้นำแบบทดสอบไปทดสอบกับกลุ่มตัวอย่างจำนวน 1,000 คน แบ่งเป็นกลุ่ม
 ตัวอย่างเพศชายซึ่งเป็นกลุ่มอ้างอิงหรือกลุ่มที่ได้เปรียบในการทำข้อสอบ จำนวน 500 คน และ
 กลุ่มตัวอย่างเพศหญิงซึ่งเป็นกลุ่มเปรียบเทียบกับหรือกลุ่มที่เสียเปรียบในการทำข้อสอบ จำนวน
 500 คน จากนั้นจึงนำผลการตอบแบบทดสอบมาวิเคราะห์หาค่าดัชนีความลำเอียงของข้อสอบ
 ด้วยวิธีวิเคราะห์ความลำเอียง 3 วิธี คือ วิธีแปลงค่าความยาก วิธีแมนเทิล-แฮนส์เชล และวิธี
 ชิปเทสท์ ได้ผลการวิเคราะห์ดังตาราง 3 ถึง ตาราง 6 ดังนี้

ตาราง 3 ค่าดัชนีความลำเอียงของข้อสอบที่วิเคราะห์ด้วยวิธีแปลงค่าความยาก จำแนกตามขนาดกลุ่มผู้สอบ

| ข้อที่ | ขนาดกลุ่มผู้สอบ 500 คน | ขนาดกลุ่มผู้สอบ 1,000 คน |
|--------|------------------------|--------------------------|
| | d_i | d_i |
| 1 | 0.369 | -0.051 |
| 2 | -0.384 | -0.397 |
| 3 | -0.183 | -0.051 |
| 4 | -0.611 | -0.696 |
| 5 | -0.357 | -0.583 |
| 6 | 0.146 | 0.063 |
| 7 | -0.420 | -0.547 |
| 8 | -0.670 | -0.244 |
| 9 | -0.137 | -0.203 |
| 10 | 0.109 | 0.069 |
| 11 | 0.126 | -0.032 |
| 12 | 0.446 | 0.025 |
| 13 | -0.058 | -0.115 |
| 14 | 0.546 | 0.668 |
| 15 | -0.279 | -0.091 |
| 16 | 1.041* | 0.552 |
| 17 | -0.324 | -0.206 |

ตาราง 3 (ต่อ)

| ข้อที่ | ขนาดกลุ่มผู้สอบ 500 คน | ขนาดกลุ่มผู้สอบ 1,000 คน |
|--------|------------------------|--------------------------|
| | d_i | d_i |
| 18 | -0.127 | -0.439 |
| 19 | 0.172 | 0.328 |
| 20 | -0.431 | -0.084 |
| 21 | -0.521 | -0.475 |
| 22 | 0.223 | 0.352 |
| 23 | -0.038 | -0.097 |
| 24 | 0.390 | 0.338 |
| 25 | 0.193 | 0.347 |
| 26 | 0.047 | 0.094 |
| 27 | 0.386 | 0.318 |
| 28 | 0.728 | 0.397 |
| 29 | 0.020 | -0.186 |
| 30 | 0.641 | 0.148 |
| 31 | 0.752* | 0.303 |
| 32 | -0.829* | -0.553 |
| 33 | 0.693 | 0.355 |
| 34 | -0.188 | -0.127 |

ตาราง 3 (ต่อ)

| ข้อที่ | ขนาดกลุ่มผู้สอบ 500 คน | ขนาดกลุ่มผู้สอบ 1,000 คน |
|--------|------------------------|--------------------------|
| | d_i | d_i |
| 35 | -0.013 | -0.094 |
| 36 | 0.014 | 0.154 |
| 37 | -1.056* | -0.435 |
| 38 | -0.036 | 0.069 |
| 39 | -0.301 | -0.237 |
| 40 | -0.406 | -0.678 |

*เป็นข้อสอบที่มีความลำเอียง

จากตาราง 3 แสดงค่าดัชนีความลำเอียงของข้อสอบ จำนวน 40 ข้อ ที่วิเคราะห์โดยวิธีแปลงค่าความยากจากกลุ่มผู้สอบ 500 คน มีค่าดัชนีความลำเอียงของข้อสอบ (d_i) อยู่ระหว่าง -1.056 ถึง 1.041 และเมื่อพิจารณาค่าดัชนีความลำเอียงตามเกณฑ์การพิจารณาความลำเอียงของข้อสอบที่ใช้ในการวิจัยนี้ พบว่า มีข้อสอบที่มีความลำเอียง 4 ข้อ คือ ข้อสอบข้อที่ 16, 31, 32 และข้อที่ 37 ซึ่งมีค่า d_i เท่ากับ 1.041, 0.752, -0.829 และ -1.056 ตามลำดับ โดยข้อสอบข้อที่ 16 และข้อที่ 31 เป็นข้อสอบที่ลำเอียงโดยกลุ่มผู้สอบเพศชาย(กลุ่มอ้างอิง)จะทำข้อสอบได้เปรียบมากกว่ากลุ่มผู้สอบเพศหญิง(กลุ่มเปรียบเทียบ) ส่วนข้อสอบข้อที่ 32 และข้อที่ 37 เป็นข้อสอบที่ลำเอียงโดยกลุ่มผู้สอบเพศหญิงจะทำข้อสอบได้เปรียบมากกว่ากลุ่มผู้สอบเพศชาย

สำหรับขนาดกลุ่มผู้สอบ 1,000 คน พบว่าค่าดัชนีความลำเอียงของข้อสอบ (d_i) มีค่าอยู่ระหว่าง -0.696 ถึง 0.668 แต่เมื่อพิจารณาค่าดัชนีความลำเอียงตามเกณฑ์การพิจารณาความลำเอียงที่ใช้ในการวิจัยครั้งนี้จะไม่พบข้อสอบที่มีความลำเอียง

ตาราง 4 ค่าดัชนีความลำเอียงของข้อสอบที่วิเคราะห์ด้วยวิธีแมนเทิล-แอนส์เซล จำแนกตามขนาดกลุ่มผู้สอบ

| ข้อที่ | ขนาดกลุ่มผู้สอบ 500 คน | | ขนาดกลุ่มผู้สอบ 1,000 คน | |
|--------|------------------------|---------------|--------------------------|---------------|
| | α_{MH} | χ^2_{MH} | α_{MH} | χ^2_{MH} |
| 1 | 0.70 | 1.28 | 1.15 | 0.34 |
| 2 | 1.26 | 0.95 | 1.32 | 3.12 |
| 3 | 1.14 | 0.23 | 1.07 | 0.17 |
| 4 | 1.55 | 3.51 | 1.73 | 12.87* |
| 5 | 1.37 | 1.38 | 1.45 | 4.07* |
| 6 | 0.85 | 0.56 | 0.93 | 0.27 |
| 7 | 1.17 | 0.51 | 1.35 | 4.87* |
| 8 | 1.69 | 5.58* | 1.26 | 2.18 |
| 9 | 1.08 | 0.08 | 1.19 | 1.16 |
| 10 | 0.84 | 0.55 | 0.92 | 0.29 |
| 11 | 0.88 | 0.27 | 1.07 | 0.15 |
| 12 | 0.71 | 2.21 | 0.94 | 0.12 |
| 13 | 1.11 | 0.15 | 1.04 | 0.03 |
| 14 | 0.67 | 3.76 | 0.65 | 9.28* |
| 15 | 1.23 | 0.99 | 1.06 | 0.12 |
| 16 | 0.48 | 9.97* | 0.66 | 7.59* |

ตาราง 4 (ต่อ)

| ข้อที่ | ขนาดกลุ่มผู้สอบ 500 คน | | ขนาดกลุ่มผู้สอบ 1,000 คน | |
|--------|------------------------|---------------|--------------------------|---------------|
| | α_{MH} | χ^2_{MH} | α_{MH} | χ^2_{MH} |
| 17 | 1.19 | 0.45 | 1.09 | 0.23 |
| 18 | 1.29 | 1.03 | 1.35 | 3.40 |
| 19 | 0.86 | 0.39 | 0.80 | 2.25 |
| 20 | 1.54 | 4.44* | 0.96 | 0.05 |
| 21 | 1.45 | 2.12 | 1.34 | 3.25 |
| 22 | 0.83 | 0.60 | 0.76 | 3.49 |
| 23 | 1.12 | 0.16 | 1.04 | 0.05 |
| 24 | 0.81 | 0.73 | 0.79 | 2.39 |
| 25 | 0.89 | 0.22 | 0.77 | 3.29 |
| 26 | 1.03 | 0.00 | 0.90 | 0.33 |
| 27 | 0.73 | 1.68 | 0.76 | 2.92 |
| 28 | 0.62 | 4.33* | 0.69 | 5.72* |
| 29 | 0.89 | 0.15 | 1.08 | 0.16 |
| 30 | 0.70 | 2.62 | 0.83 | 1.47 |
| 31 | 0.54 | 6.15* | 0.75 | 2.76 |
| 32 | 1.84 | 6.78* | 1.49 | 6.41* |

ตาราง 4 (ต่อ)

| ข้อที่ | ขนาดกลุ่มผู้สอบ 500 คน | | ขนาดกลุ่มผู้สอบ 1,000 คน | |
|--------|------------------------|---------------|--------------------------|---------------|
| | α_{MH} | χ^2_{MH} | α_{MH} | χ^2_{MH} |
| 33 | 0.71 | 2.13 | 0.76 | 3.25 |
| 34 | 1.10 | 0.10 | 0.96 | 0.04 |
| 35 | 0.94 | 0.02 | 1.04 | 0.02 |
| 36 | 0.97 | 0.00 | 0.83 | 0.98 |
| 37 | 1.71 | 5.76* | 1.24 | 1.98 |
| 38 | 1.00 | 0.01 | 0.90 | 0.43 |
| 39 | 1.30 | 0.95 | 1.10 | 0.26 |
| 40 | 1.20 | 0.54 | 1.48 | 6.32* |

*ข้อสอบที่มีความลำเอียงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ตาราง 4 แสดงค่าดัชนีความลำเอียงของข้อสอบ (α_{MH}) จำนวน 40 ข้อ ที่วิเคราะห์ด้วยวิธีแมนเทิล-แฮนส์เซลจากกลุ่มผู้สอบ 500 คน พบว่า ในขนาดกลุ่มผู้สอบ 500 คน ค่า α_{MH} ของข้อสอบมีค่าอยู่ระหว่าง 0.48 ถึง 1.84 และค่า χ^2_{MH} ของข้อสอบมีค่าอยู่ระหว่าง 0.00 ถึง 9.97 แต่เมื่อพิจารณาค่า χ^2_{MH} ที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 พบว่ามีข้อสอบที่มีความลำเอียง 7 ข้อ คือ ข้อสอบข้อที่ 8, 16, 20, 28, 31, 32 และข้อที่ 37 ซึ่งมีค่า α_{MH} อยู่ระหว่าง 0.48 ถึง 1.84 และค่า χ^2_{MH} อยู่ระหว่าง 4.33 ถึง 9.97 โดยข้อสอบข้อที่ 8, 20, 32 และข้อที่ 37 เป็นข้อสอบที่ลำเอียงโดยกลุ่มผู้สอบเพศชายจะทำข้อสอบได้เปรียบมากกว่ากลุ่มผู้สอบเพศหญิง และข้อสอบข้อที่ 16, 28 และข้อที่ 31 เป็นข้อสอบที่ลำเอียงโดยกลุ่มผู้สอบเพศหญิงจะทำข้อสอบได้เปรียบมากกว่ากลุ่มผู้สอบเพศชาย

สำหรับกลุ่มผู้สอบขนาด 1,000 คน ค่า α_{MH} ของข้อสอบมีค่าอยู่ระหว่าง 0.66 ถึง 1.35 และค่า χ^2_{MH} ของข้อสอบมีค่าอยู่ระหว่าง 0.02 ถึง 12.87 แต่เมื่อพิจารณาค่า χ^2_{MH} ที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 พบว่ามีข้อสอบที่มีความลำเอียง 8 ข้อ คือ ข้อสอบข้อที่ 4, 5, 7, 14, 16, 28, 32 และข้อที่ 40 ซึ่งมีค่า α_{MH} อยู่ระหว่าง 0.65 ถึง 1.73 และค่า χ^2_{MH} อยู่ระหว่าง 5.72 ถึง 12.87 โดยข้อสอบข้อที่ 4, 5, 7, 32 และข้อที่ 40 เป็นข้อสอบที่ลำเอียงโดยกลุ่มผู้สอบเพศชายจะทำข้อสอบได้เปรียบมากกว่ากลุ่มผู้สอบเพศหญิง และข้อสอบข้อที่ 14, 16 และข้อที่ 28 เป็นข้อสอบที่ลำเอียงโดยกลุ่มผู้สอบเพศหญิงจะทำข้อสอบได้เปรียบมากกว่ากลุ่มผู้สอบเพศชาย

ตาราง 5 ค่าดัชนีความลำเอียงของข้อสอบที่วิเคราะห์ด้วยวิธีชิปเทสต์ จำแนกตามขนาดกลุ่มผู้
สอบ

| ข้อที่ | ขนาดกลุ่มผู้สอบ 500 คน | ขนาดกลุ่มผู้สอบ 1,000 คน |
|--------|------------------------|--------------------------|
| | Z_{SIB} | Z_{SIB} |
| 1 | -1.316 | 0.530 |
| 2 | 1.045 | 1.892 |
| 3 | 0.697 | 0.832 |
| 4 | 2.279* | 3.844* |
| 5 | 0.939 | 1.459 |
| 6 | -0.203 | -0.247 |
| 7 | 1.102 | 2.422* |
| 8 | 2.456* | 1.368 |
| 9 | 0.389 | 1.558 |
| 10 | -0.336 | -0.060 |
| 11 | -0.360 | 0.746 |
| 12 | -1.972* | -0.353 |
| 13 | 0.580 | 0.602 |
| 14 | -1.538 | -2.635* |
| 15 | 1.103 | 0.462 |
| 16 | -3.426* | -2.839* |
| 17 | 0.948 | 0.965 |

ตาราง 5 (ต่อ)

| ข้อที่ | ขนาดกลุ่มผู้สอบ 500 คน | ขนาดกลุ่มผู้สอบ 1,000 คน |
|--------|------------------------|--------------------------|
| | Z_{SIB} | Z_{SIB} |
| 18 | 0.544 | 1.360 |
| 19 | -0.299 | -1.183 |
| 20 | 0.847 | -0.648 |
| 21 | 1.956* | 1.812 |
| 22 | -0.604 | -1.693 |
| 23 | 0.542 | 0.630 |
| 24 | -1.450 | -1.664 |
| 25 | -0.550 | -1.553 |
| 26 | 0.119 | -0.144 |
| 27 | -1.249 | -1.542 |
| 28 | -2.033* | -1.721 |
| 29 | 0.089 | 0.983 |
| 30 | -1.634 | -1.189 |
| 31 | -2.136* | -1.666 |
| 32 | 2.663* | 2.518* |
| 33 | -1.360 | -1.876 |
| 34 | 0.595 | -0.617 |

ตาราง 5 (ต่อ)

| ข้อที่ | ขนาดกลุ่มผู้สอบ 500 คน | ขนาดกลุ่มผู้สอบ 1,000 คน |
|--------|------------------------|--------------------------|
| | Z_{SIB} | Z_{SIB} |
| 35 | -0.042 | 0.452 |
| 36 | 0.135 | -0.869 |
| 37 | 2.817* | 1.900 |
| 38 | 0.378 | -0.701 |
| 39 | 0.885 | 1.497 |
| 40 | 1.331 | 2.357* |

* ข้อสอบที่มีความลำเอียงอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

ตาราง 5 แสดงค่าดัชนีความลำเอียงของข้อสอบ (Z_{SIB}) จำนวน 40 ข้อที่วิเคราะห์ด้วยวิธีชิปเทสต์จากกลุ่มผู้สอบ 500 คน ค่า Z_{SIB} ของข้อสอบมีค่าอยู่ระหว่าง -2.136 ถึง 2.817 และเมื่อพิจารณาค่า Z_{SIB} ที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 พบว่า มีข้อสอบที่มีความลำเอียง 9 ข้อ คือข้อสอบข้อที่ 4, 8, 12, 16, 21, 28, 31, 32 และข้อที่ 37 มีค่า Z_{SIB} อยู่ระหว่าง -3.426 ถึง 2.817 โดยข้อสอบข้อที่ 4, 8, 21, 32 และข้อที่ 37 เป็นข้อสอบที่ลำเอียงโดยกลุ่มผู้สอบเพศชายจะทำข้อสอบได้เปรียบมากกว่ากลุ่มผู้สอบเพศหญิง และข้อสอบข้อที่ 12, 16, 28 และข้อที่ 31 เป็นข้อสอบที่ลำเอียงโดยกลุ่มผู้สอบเพศหญิงจะทำข้อสอบได้เปรียบมากกว่ากลุ่มผู้สอบเพศชาย

สำหรับขนาดกลุ่มผู้สอบ 1,000 คน พบว่า ค่า Z_{SIB} ของข้อสอบมีค่าอยู่ระหว่าง -2.839 ถึง 3.844 และเมื่อพิจารณาค่า Z_{SIB} ที่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 พบว่ามีข้อสอบที่มีความลำเอียง 6 ข้อ คือ ข้อสอบข้อที่ 4, 8, 14, 16, 32 และข้อที่ 40 โดยข้อสอบข้อที่ 4, 7, 32 และข้อที่ 40 เป็นข้อสอบที่ลำเอียงโดยกลุ่มผู้สอบเพศชายจะทำข้อสอบได้เปรียบมากกว่ากลุ่มผู้สอบเพศหญิง และข้อสอบข้อที่ 14 และข้อที่ 16 เป็นข้อสอบที่ลำเอียงโดยกลุ่มผู้สอบเพศหญิงจะทำข้อสอบได้เปรียบมากกว่ากลุ่มผู้สอบเพศชาย

จากผลการวิเคราะห์ความลำเอียงต่อเพศของข้อสอบโดยใช้วิธีวิเคราะห์ความลำเอียง 3 วิธี คือ วิธีแปลงค่าความยาก วิธีแมนเทิล-แฮนส์เชล และวิธีชิปเทสท์ ในขนาดกลุ่มผู้สอบ 500 คน และในขนาดกลุ่มผู้สอบ 1,000 คน พบว่าในแต่ละวิธีวิเคราะห์ความลำเอียง และในแต่ละขนาดกลุ่มผู้สอบ จะพบจำนวนข้อสอบที่มีความลำเอียงไม่เท่ากัน และเมื่อพิจารณาในแต่ละขนาดกลุ่มผู้สอบ พบว่าวิธีวิเคราะห์ความลำเอียงบางวิธีสามารถตรวจพบข้อสอบที่มีความลำเอียงซ้ำกัน แต่ในบางวิธีวิเคราะห์ความลำเอียงก็สามารถตรวจพบข้อสอบที่มีความลำเอียงต่างกัน ดังนั้นผู้วิจัยจึงได้สรุปจำนวนข้อสอบที่มีความลำเอียงจากวิธีวิเคราะห์ความลำเอียงทั้ง 3 วิธี จำแนกตามขนาดกลุ่มผู้สอบเพื่อให้เห็นจำนวนข้อสอบที่มีความลำเอียงได้ชัดเจนขึ้น ดังแสดงในตาราง 6

ตาราง 6 สรุปจำนวนข้อสอบที่มีความลำเอียงในแต่ละวิธีวิเคราะห์ความลำเอียง จำแนกตามขนาดกลุ่มผู้สอบ

| วิธีวิเคราะห์ความลำเอียง | ข้อสอบที่ลำเอียงในขนาดกลุ่มผู้สอบ 500 คน | จำนวนข้อสอบที่มีความลำเอียง | ข้อสอบที่ลำเอียงในขนาดกลุ่มผู้สอบ 1,000 คน | จำนวนข้อสอบที่มีความลำเอียง |
|--------------------------|--|-----------------------------|--|-----------------------------|
| TID | 16,31,32,37 | 4 | - | - |
| MH | 8,16,20,28,31,32,37 | 7 | 4,5,7,14,16,28,32,40 | 8 |
| SIB | 4,8,12,16,21,28,31,32,37 | 9 | 4,7,14,16,32,40 | 6 |

จากตาราง 6 เมื่อพิจารณาผลการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีวิเคราะห์ความลำเอียง 3 วิธีในขนาดกลุ่มผู้สอบ 500 คน พบว่า วิธีชิปเทสท์จะพบจำนวนข้อสอบที่มีความลำเอียงมากที่สุดคือจำนวน 9 ข้อ รองลงมาคือวิธีแมนเทิล-แฮนส์เชล พบจำนวนข้อสอบที่มีความลำเอียง 7 ข้อ และวิธีแปลงค่าความยากพบจำนวนข้อสอบที่มีความลำเอียงได้น้อยที่สุดคือ จำนวน 4 ข้อ โดยที่วิธีแปลงค่าความยากจะตรวจพบข้อสอบที่ลำเอียงซ้ำกับวิธีแมนเทิล-แฮนส์เชลและวิธีชิปเทสท์ทุกข้อและวิธีแมนเทิล-แฮนส์เชลตรวจพบข้อสอบที่ลำเอียงซ้ำกับวิธีชิปเทสท์ด้วย

สำหรับขนาดกลุ่มผู้สอบ 1,000 คน วิธีแมนเทิล-แฮนส์เซล พบจำนวนข้อสอบที่มีความลำเอียงมากที่สุด จำนวน 8 ข้อ รองลงมาคือวิธีชิปเทสท์พบจำนวนข้อสอบที่มีความลำเอียง 6 ข้อ และวิธีแปลงค่าความยากไม่พบข้อสอบที่มีความลำเอียง โดยที่วิธีแมนเทิล-แฮนส์เซลจะตรวจพบข้อสอบที่ลำเอียงซ้ำกับวิธีชิปเทสท์ทุกข้อ

เมื่อพิจารณาผลการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีวิเคราะห์ความลำเอียงวิธีเดียวกันในขนาดกลุ่มผู้สอบต่างกันพบว่าในขนาดกลุ่มผู้สอบ 500 คน วิธีแปลงค่าความยากจะพบข้อสอบที่มีความลำเอียงจำนวน 4 ข้อ แต่ไม่พบข้อสอบที่มีความลำเอียงในขนาดกลุ่มผู้สอบ 1,000 คน วิธีแมนเทิล-แฮนส์เซล พบจำนวนข้อสอบที่มีความลำเอียงในขนาดกลุ่มผู้สอบ 500 คนน้อยกว่าในขนาดกลุ่มผู้สอบ 1,000 คน คือ ในขนาดกลุ่มผู้สอบ 500 คน วิธีแมนเทิล-แฮนส์เซลพบจำนวนข้อสอบที่มีความลำเอียง 7 ข้อ แต่ในขนาดกลุ่มผู้สอบ 1,000 คน จะพบจำนวนข้อสอบที่มีความลำเอียง 8 ข้อ และวิธีชิปเทสท์พบจำนวนข้อสอบที่มีความลำเอียงในขนาดกลุ่มผู้สอบ 500 คนมากกว่าในขนาดกลุ่มผู้สอบ 1,000 คน คือในขนาดกลุ่มผู้สอบ 500 คน วิธีชิปเทสท์พบจำนวนข้อสอบที่มีความลำเอียง 9 ข้อ แต่ในขนาดกลุ่มผู้สอบ 1,000 คน พบจำนวนข้อสอบที่มีความลำเอียง 6 ข้อ

2.เปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียงที่แตกต่างกันเมื่อขนาดกลุ่มผู้สอบเท่ากัน

จากผลการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีวิเคราะห์ความลำเอียงแต่ละวิธีและแต่ละขนาดกลุ่มผู้สอบ พบว่าได้จำนวนข้อสอบที่มีความลำเอียงไม่เท่ากัน ดังนั้นผู้วิจัยจึงทำการเปรียบเทียบความแตกต่างของจำนวนข้อสอบที่มีความลำเอียงโดยใช้สถิติ Z-test เพื่อทดสอบความแตกต่างอย่างมีนัยสำคัญทางสถิติของจำนวนข้อสอบที่มีความลำเอียง ซึ่งผลการทดสอบนัยสำคัญ ดังแสดงในตาราง 7

ตาราง 7 ผลการเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียง 3 วิธี เมื่อขนาดกลุ่มผู้สอบเท่ากัน

| วิธีวิเคราะห์ความลำเอียง | N = 500 คน | | | N = 1,000 คน | | |
|--------------------------|------------|-------|-------|--------------|--------|--------|
| | TID | MH | SIB | TID | MH | SIB |
| N = 500 คน | | | | | | |
| TID | - | 0.974 | 1.515 | - | - | - |
| MH | | - | 0.559 | - | - | - |
| SIB | | | - | - | - | - |
| N = 1,000 คน | | | | | | |
| TID | - | - | - | - | 2.982* | 2.547* |
| MH | - | - | - | - | - | 0.588 |
| SIB | - | - | - | - | - | - |

*มีนัยสำคัญทางสถิติที่ระดับ .05

$$Z_{.05} = 1.96$$

จากตาราง 7 แสดงให้เห็นว่า เมื่อขนาดกลุ่มผู้สอบ 500 คนจำนวนข้อสอบที่มีความลำเอียงจากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยาก วิธีแมนเทิล-แฮนส์เชล และวิธีชิปเทสท์ แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ซึ่งแสดงว่า การวิเคราะห์ความลำเอียงทั้ง 3 วิธีพบจำนวนข้อสอบที่มีความลำเอียงมีจำนวนข้อใกล้เคียงกัน

สำหรับขนาดกลุ่มผู้สอบ 1,000 คน พบว่าจำนวนข้อสอบที่มีความลำเอียงจากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากกับวิธีแมนเทิล-แฮนส์เชล และวิธีแปลงค่าความยากกับวิธีชิปเทสท์ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนวิธีแมนเทิล-

แฮนส์เซลกับวิธีซิปเทสที่พบจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

3. เปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงเมื่อขนาดกลุ่มผู้สอบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน

ผู้วิจัยได้ทำการเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียง เมื่อขนาดกลุ่มผู้สอบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน โดยใช้สถิติ Z-test ผลการเปรียบเทียบ ดังแสดงในตาราง 8

ตาราง 8 ผลการเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียง เมื่อขนาดกลุ่มผู้สอบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน

| วิธีวิเคราะห์ความลำเอียง | N = 500 คน | | |
|--------------------------|------------|-------|-------|
| | TID | MH | SIB |
| N = 1,000 คน | 2.052* | 0.286 | 0.859 |

*มีนัยสำคัญทางสถิติที่ระดับ .05

$$Z_{.05} = 1.96$$

จากตาราง 8 แสดงให้เห็นว่า เมื่อขนาดกลุ่มผู้สอบต่างกัน จำนวนข้อสอบที่มีความลำเอียงจากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยาก แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับการวิเคราะห์ความลำเอียงด้วยวิธีแมนเทล-แฮนส์เซลและวิธีซิปเทส จะพบจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

4. ค่าความเชื่อมั่นและเปรียบเทียบความเชื่อมั่นแบบแบ่งครึ่งฉบับของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว ระหว่างวิธีวิเคราะห์ความลำเอียงที่แตกต่างกัน เมื่อขนาดกลุ่มผู้สอบเท่ากัน

เมื่อทำการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีวิเคราะห์ความลำเอียงที่แตกต่างกันในแต่ละขนาดกลุ่มผู้สอบแล้ว ผู้วิจัยได้นำแบบทดสอบที่ทำการคัดเลือกข้อสอบที่มี

ความลำเอียงออกแล้ว จำนวน 6 ฉบับ ซึ่งแต่ละฉบับประกอบด้วยจำนวนข้อสอบไม่เท่ากันดังนี้ แบบทดสอบฉบับที่ 1 มีจำนวนข้อสอบ 36 ข้อ, แบบทดสอบฉบับที่ 2 มีจำนวนข้อสอบ 33 ข้อ, แบบทดสอบฉบับที่ 3 มีจำนวนข้อสอบ 31 ข้อ, แบบทดสอบฉบับที่ 4 มีจำนวนข้อสอบ 40 ข้อ, แบบทดสอบฉบับที่ 5 มีจำนวนข้อสอบ 32 ข้อ และแบบทดสอบฉบับที่ 6 มีจำนวนข้อสอบ 34 ข้อ ไปทดสอบกับกลุ่มตัวอย่างจำนวน 200 คนในแต่ละฉบับเพื่อคำนวณค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับของแบบทดสอบและทำการเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว ด้วยสถิติไคสแควร์ ได้ผลการวิเคราะห์ดังตาราง 9 ถึง ตาราง 10 ดังนี้

ตาราง 9 ค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับของแบบทดสอบ ค่า Fisher-Z และค่าการทดสอบนัยสำคัญของค่าความเชื่อมั่นของแบบทดสอบที่คัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว ระหว่างวิเคราะห์ความลำเอียงที่แตกต่างกัน เมื่อขนาดกลุ่มผู้สอบเท่ากัน

| N | วิธีวิเคราะห์ความลำเอียง | K | r_n | Z_r | χ^2 |
|----------|--------------------------|----|--------|-------|----------|
| 500 คน | TID | 36 | 0.7789 | 1.045 | 6.5067* |
| | MH | 33 | 0.7591 | 0.996 | |
| | SIB | 31 | 0.6673 | 0.802 | |
| 1,000 คน | TID | 40 | 0.7934 | 1.085 | 1.2377 |
| | MH | 32 | 0.7506 | 0.973 | |
| | SIB | 34 | 0.7742 | 1.033 | |

* มีนัยสำคัญทางสถิติที่ระดับ .05

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

จากตาราง 9 แสดงให้เห็นว่า ค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว เมื่อวิเคราะห์จากขนาดกลุ่มผู้สอบ 500 คน มีค่าอยู่ระหว่าง 0.6675 ถึง 0.7789 โดยแบบทดสอบที่วิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแปลงค่าความยาก จะมีค่าความเชื่อมั่นสูงสุดเป็น 0.7789 รองลงมาคือ แบบทดสอบที่

ตาราง 10 ผลการเปรียบเทียบค่าความเชื่อมั่นรายคู่ของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วด้วยวิธีวิเคราะห์ความลำเอียงต่างกันเมื่อมีขนาดกลุ่มผู้สอบ 500 คน

| วิธีวิเคราะห์ความ ลำเอียง | TID | MH | SIB |
|------------------------------|-----|--------|---------|
| TID | - | 0.4863 | 2.4117* |
| MH | | - | 1.9254 |
| SIB | | | - |

* มีนัยสำคัญทางสถิติที่ระดับ .05

$$Z_{.05} = 1.96$$

จากตาราง 10 แสดงให้เห็นว่าเมื่อวิเคราะห์จากขนาดกลุ่มผู้สอบ 500 คน ค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วที่วิเคราะห์ความลำเอียงของข้อสอบโดยวิธีแปลงค่าความยาก กับวิธีชิปเทสท์ มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วที่วิเคราะห์ความลำเอียงโดยวิธีแปลงค่าความยากกับวิธีแมนเทล-แฮนส์เชล และวิธีแมนเทล-แฮนส์เชลกับวิธีชิปเทสท์ มีค่าแตกต่างกันอย่างไรไม่มีนัยสำคัญทางสถิติ

5. เปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังการคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว เมื่อขนาดกลุ่มผู้สอบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน

ผู้วิจัยได้ทำการเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว เมื่อขนาดกลุ่มผู้สอบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน โดยใช้สถิติ Z-test ผลการทดสอบนัยสำคัญดังแสดงในตาราง 11

วิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแมนเทิล-แฮนส์เซล มีค่าความเชื่อมั่นเป็น 0.7591 และแบบทดสอบที่มีค่าความเชื่อมั่นต่ำสุดคือ แบบทดสอบที่วิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีชิปเทสท์ ซึ่งมีค่าความเชื่อมั่นเป็น 0.6673 และเมื่อเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบด้วย สถิติไคสแควร์ พบว่า มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งแสดงว่ามีค่าความเชื่อมั่นของแบบทดสอบอย่างน้อย 1 คู่ที่มีความแตกต่างกัน

สำหรับขนาดกลุ่มผู้สอบ 1,000 คน ค่าความเชื่อมั่นของแบบทดสอบแบบแบ่งครึ่งฉบับของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว มีค่าอยู่ระหว่าง 0.7506 ถึง 0.7934 โดยแบบทดสอบที่วิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแปลงค่าความยาก จะมีค่าความเชื่อมั่นสูงสุดเป็น 0.7934 รองลงมาคือแบบทดสอบที่วิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีชิปเทสท์ มีค่าความเชื่อมั่นเป็น 0.7742 และแบบทดสอบที่มีค่าความเชื่อมั่นต่ำสุด คือแบบทดสอบที่วิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแมนเทิล-แฮนส์เซล ซึ่งมีค่าความเชื่อมั่นเป็น 0.7506 และเมื่อทำการเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบด้วยสถิติไคสแควร์ พบว่า มีค่าแตกต่างกันอย่างไรไม่มีนัยสำคัญทางสถิติ

จากการเปรียบเทียบค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับของแบบทดสอบที่ทำการวิเคราะห์ความลำเอียงด้วยวิธีวิเคราะห์ความลำเอียงที่แตกต่างกันเมื่อวิเคราะห์จากขนาดกลุ่มผู้สอบ 500 คน พบว่ามีค่าความเชื่อมั่นอย่างน้อย 1 คู่ที่แตกต่างกัน ผู้วิจัยจึงทำการทดสอบค่าความเชื่อมั่นของแบบทดสอบเป็นรายคู่เพื่อให้ทราบว่าค่าความเชื่อมั่นของแบบทดสอบคู่ใดบ้างที่แตกต่างกันโดยใช้สถิติ Z-test ในการทดสอบความแตกต่างเป็นรายคู่ ดังแสดงผลในตาราง 10

ตาราง 11 ผลการเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วเมื่อขนาดกลุ่มผู้สอบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน

| วิธีวิเคราะห์ความลำเอียง | N = 500 คน | | |
|--------------------------|------------|--------|---------|
| | TID | MH | SIB |
| N = 1,000 คน | 0.3970 | 0.2283 | 2.2926* |

* มีนัยสำคัญทางสถิติที่ระดับ .05

$$Z_{.05} = 1.96$$

จากตาราง 11 แสดงให้เห็นว่า ค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วที่วิเคราะห์ความลำเอียงโดยใช้วิธีแปลงค่าความยาก และวิธีแมนเทิล-แฮนส์เซล มีค่าความเชื่อมั่นระหว่างที่วิเคราะห์จากขนาดกลุ่มผู้สอบ 500 คน กับ 1,000 คน มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ สำหรับค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วที่วิเคราะห์ความลำเอียงโดยวิธีชิปเทสท์ มีค่าความเชื่อมั่นของแบบทดสอบระหว่างที่วิเคราะห์จากขนาดกลุ่มผู้สอบ 500 คน และ 1,000 คน มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

บทที่ 5

สรุปผล อภิปรายผล และข้อเสนอแนะ

จุดมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อศึกษาดัชนี ความลำเอียงของข้อสอบแต่ละข้อในแบบทดสอบวิชาคณิตศาสตร์โจทย์ปัญหา จำแนกในแต่ละวิธีวิเคราะห์ความลำเอียงของข้อสอบและขนาดกลุ่มผู้สอบ
2. เพื่อเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียงที่แตกต่างกัน เมื่อขนาดกลุ่มผู้สอบเท่ากัน
3. เพื่อเปรียบเทียบจำนวนข้อสอบที่มีความลำเอียง เมื่อขนาดกลุ่มผู้สอบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน
4. เพื่อเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังการคัดข้อสอบที่มีความลำเอียงออกแล้ว ระหว่างวิธีวิเคราะห์ความลำเอียงที่แตกต่างกัน เมื่อขนาดกลุ่มผู้สอบเท่ากัน
5. เพื่อเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังการคัดข้อสอบที่มีความลำเอียงออกแล้ว เมื่อขนาดกลุ่มผู้สอบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2539 ของโรงเรียนสังกัดกรมสามัญศึกษา ส่วนกลาง กรุงเทพมหานคร ซึ่งได้มาโดยการสุ่มแบบแบ่งชั้น (Stratified Random Sampling) มีขนาดโรงเรียนเป็นชั้น (Strata) โรงเรียนเป็นหน่วยการสุ่ม (Sampling Unit) ได้จำนวน 10 โรงเรียน ที่มีจำนวนห้องเรียน 53 ห้องเรียน และมีจำนวนนักเรียน 2,200 คน

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นแบบทดสอบคณิตศาสตร์โจทย์ปัญหา ชั้นมัธยมศึกษาปีที่ 1 (ค.102) ในเรื่อง สมการ อัตราส่วนและร้อยละ ตามหลักสูตรมัธยมศึกษา

ตอนต้น พุทธศักราช 2521 (ฉบับปรับปรุง 2533) เป็นแบบเลือกตอบชนิด 5 ตัวเลือก จำนวน 40 ข้อ ที่ผู้วิจัยสร้างขึ้นเอง

วิธีดำเนินการเก็บรวบรวมข้อมูล

การเก็บรวบรวมข้อมูลเพื่อทำการวิจัยในครั้งนี้ ผู้วิจัยได้ดำเนินการตามลำดับขั้น ดังนี้

1. ติดต่อขอความร่วมมือจากโรงเรียนที่เป็นกลุ่มตัวอย่าง เพื่อกำหนด วัน เวลา ในการดำเนินการสอบ
2. จัดเตรียมแบบทดสอบให้เพียงพอกับจำนวนนักเรียนที่จะทำการสอบในแต่ละครั้ง
3. วางแผนดำเนินการสอบโดยผู้วิจัยดำเนินการสอบด้วยตนเอง
4. ชี้แจงให้นักเรียนที่เป็นกลุ่มตัวอย่างทราบวัตถุประสงค์ของการสอบ เพื่อให้ได้ผลสอบตามความเป็นจริง
5. นำแบบทดสอบไปทดสอบกับกลุ่มตัวอย่างครั้งที่ 1 จำนวน 310 คน
6. นำผลการสอบในข้อ 5 มาวิเคราะห์หาค่าความยาก และค่าอำนาจจำแนกของข้อสอบเป็นรายข้อ แล้วคัดเลือกข้อสอบไว้จำนวน 40 ข้อ
7. นำแบบทดสอบที่มีคุณภาพในข้อ 6 จำนวน 40 ข้อ ไปทดสอบกับกลุ่มตัวอย่างครั้งที่ 2 จำนวน 308 คน เพื่อหาค่าความเชื่อมั่นของแบบทดสอบทั้งฉบับ
8. นำแบบทดสอบที่มีคุณภาพตามข้อ 7 จำนวน 40 ข้อ ไปทดสอบกับกลุ่มตัวอย่างจำนวน 1,000 คน แล้วทำการสุ่มกลุ่มตัวอย่างจำนวน 500 คน เพื่อศึกษาความลำเอียงของข้อสอบ จำแนกในแต่ละวิธีวิเคราะห์และทดสอบสมมติฐานการวิจัย ข้อที่ 1 และ 2
9. นำผลการวิเคราะห์ความลำเอียงตามข้อ 8 มาคัดข้อที่ลำเอียงออกแล้วศึกษาความเชื่อมั่นของแบบทดสอบหลังคัดข้อลำเอียงออกกับกลุ่มตัวอย่างจำนวน 1,200 คน และทำการทดสอบสมมติฐานการวิจัยข้อที่ 3 และ 4

การเสนอผลการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้ ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลตามลำดับ ต่อไปนี้

1. ค่าดัชนีความลำเอียงของข้อสอบแต่ละข้อ จำแนกในแต่ละวิธีวิเคราะห์ความลำเอียงและขนาดกลุ่มผู้สอบ
2. เปรียบเทียบจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีวิเคราะห์ความลำเอียงที่แตกต่างกัน เมื่อขนาดกลุ่มผู้สอบเท่ากัน
3. เปรียบเทียบจำนวนข้อสอบที่มีความลำเอียง เมื่อขนาดกลุ่มผู้สอบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน
4. เปรียบเทียบค่าความเชื่อมั่นแบบแบ่งครึ่งฉบับของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว ระหว่างวิธีวิเคราะห์ความลำเอียงที่แตกต่างกัน เมื่อขนาดกลุ่มผู้สอบเท่ากัน
5. เปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว เมื่อขนาดกลุ่มผู้สอบต่างกัน เมื่อมีวิธีวิเคราะห์ความลำเอียงวิธีเดียวกัน

๖) สรุปผลการวิเคราะห์ข้อมูล

ผู้วิจัยได้สรุปผลการวิเคราะห์ข้อมูล ดังนี้

1. จำนวนข้อสอบที่ลำเอียง เมื่อวิเคราะห์จากกลุ่มผู้สอบ 500 คน และวิเคราะห์ด้วยวิธีแปลงค่าความยากพบข้อสอบที่มีความลำเอียง 4 ข้อ วิธีแมนเทิล-แฮนส์เซลพบข้อสอบที่ลำเอียง 7 ข้อ และวิธีชิปเทสท์พบข้อสอบที่ลำเอียง 9 ข้อ แสดงว่าวิธีชิปเทสท์พบข้อสอบที่มีความลำเอียงมากที่สุด รองลงมาคือวิธีแมนเทิล-แฮนส์เซล และวิธีแปลงค่าความยากพบข้อสอบที่มีความลำเอียงน้อยที่สุด โดยที่วิธีแปลงค่าความยากจะตรวจพบข้อสอบที่ลำเอียงช้ากับวิธีแมนเทิล-แฮนส์เซลและวิธีชิปเทสท์ทุกข้อและวิธีแมนเทิล-แฮนส์เซลตรวจพบข้อสอบที่ลำเอียงช้ากับวิธีชิปเทสท์ด้วย เมื่อวิเคราะห์จากขนาดกลุ่มผู้สอบ 1,000 คน และวิเคราะห์ด้วยวิธีแปลงค่าความยากไม่พบข้อสอบที่ลำเอียง วิธีแมนเทิล-แฮนส์เซลพบข้อสอบที่ลำเอียง 8 ข้อ และวิธีชิปเทสท์พบข้อสอบที่มีความลำเอียง 6 ข้อ แสดงว่าวิธีแมนเทิล-แฮนส์เซลพบข้อสอบที่มี

ความลำเอียงมากที่สุด รองลงมาคือวิธีชิปเทสท์และวิธีแปลงค่าความยากไม่พบข้อสอบที่มีความลำเอียง โดยที่วิธีชิปเทสท์จะตรวจพบข้อสอบที่ลำเอียงซ้ำกับวิธีแมนเทล-แฮนส์เซลทุกข้อ

2. จำนวนข้อสอบที่มีความลำเอียงเมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 500 คนพบว่า จำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีแปลงค่าความยาก วิธีแมนเทล - แฮนส์เซล และวิธีชิปเทสท์มีจำนวนข้อสอบที่ลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทุกวิธี เมื่อวิเคราะห์จากกลุ่มผู้สอบขนาด 1,000 คน พบว่าจำนวนข้อสอบที่มีความลำเอียงระหว่างวิธีแปลงค่าความยากกับวิธีแมนเทล-แฮนส์เซล และวิธีแปลงค่าความยากกับวิธีชิปเทสท์ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับจำนวนข้อสอบที่วิเคราะห์ความลำเอียงด้วยวิธีแมนเทล - แฮนส์เซลกับวิธีชิปเทสท์ แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

3. จำนวนข้อสอบที่มีความลำเอียง จากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากระหว่างกลุ่มผู้สอบขนาด 500 คนและกลุ่มผู้สอบขนาด 1,000 คน พบว่าจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สำหรับวิธีแมนเทล-แฮนส์เซลและวิธีชิปเทสท์ระหว่างกลุ่มผู้สอบขนาด 500 คนและกลุ่มผู้สอบขนาด 1,000 คน พบว่าจำนวนข้อสอบที่มีความลำเอียง แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

4. ค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว เมื่อวิเคราะห์จากกลุ่มผู้สอบ 500 คน พบว่าค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยาก วิธีแมนเทล-แฮนส์เซลและวิธีชิปเทสท์ มีค่าเป็น 0.7789, 0.7591 และ 0.6673 ตามลำดับ โดยที่ค่าความเชื่อมั่นของแบบทดสอบที่วิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแปลงค่าความยากกับวิธีชิปเทสท์มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบที่วิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแปลงค่าความยากกับวิธีแมนเทล-แฮนส์เซล และแบบทดสอบที่วิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแมนเทล-แฮนส์เซลกับวิธีชิปเทสท์มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ สำหรับกลุ่มผู้สอบขนาด 1,000 คน พบว่าค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกด้วยวิธีแปลงค่าความยาก วิธีแมนเทล-แฮนส์เซล และวิธีชิปเทสท์ มีค่าความเชื่อมั่นของแบบทดสอบเป็น 0.7934, 0.7506 และ 0.7742 ตามลำดับโดยที่ค่าความเชื่อมั่นของแบบทดสอบทั้ง 3 ฉบับแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

5. ค่าความเชื่อมั่นของแบบทดสอบหลังจากคัดเลือกข้อสอบที่มีความลำเอียงออกแล้ว ที่วิเคราะห์ความลำเอียงของข้อสอบโดยวิธีแปลงค่าความยากและวิธีแมนเทิล-แฮนส์เซล ระหว่างกลุ่มผู้สอบขนาด 500 คนและกลุ่มผู้สอบขนาด 1,000 คนพบว่ามีความเชื่อมั่นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ สำหรับค่าความเชื่อมั่นของแบบทดสอบที่วิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีชิปเทสท์ระหว่างกลุ่มผู้สอบขนาด 500 คนและกลุ่มผู้สอบขนาด 1,000 คนมีความเชื่อมั่นแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อภิปรายผล

จากการศึกษาค้นคว้าครั้งนี้ สามารถอภิปรายผลได้ ดังนี้

1. จำนวนข้อสอบที่มีความลำเอียงจากการใช้วิธีวิเคราะห์ความลำเอียง 3 วิธี คือ วิธีแปลงค่าความยาก วิธีแมนเทิล-แฮนส์เซลและวิธีชิปเทสท์ เมื่อมีขนาดกลุ่มผู้สอบ 500 คน จะพบจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ซึ่งผลที่ได้ไม่สอดคล้องกับสมมติฐานการวิจัย ทั้งนี้อาจเนื่องมาจาก วิธีวิเคราะห์ความลำเอียงทั้ง 3 วิธี พบจำนวนข้อสอบที่มีความลำเอียงใกล้เคียงกัน คือ วิธีแปลงค่าความยาก พบจำนวนข้อสอบที่มีความลำเอียง 4 ข้อ วิธีแมนเทิล-แฮนส์เซล พบจำนวนข้อสอบที่มีความลำเอียง 7 ข้อ และวิธีชิปเทสท์ พบจำนวนข้อสอบที่มีความลำเอียง 9 ข้อ เมื่อนำจำนวนข้อสอบที่มีความลำเอียงมาทดสอบนัยสำคัญทางสถิติจึงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และเมื่อพิจารณาข้อสอบที่ลำเอียงจากการวิเคราะห์ด้วยวิธีแปลงค่าความยากจะซ้ำกับข้อสอบที่ลำเอียงเมื่อวิเคราะห์ด้วยวิธีแมนเทิล-แฮนส์เซลและวิธีชิปเทสท์ทุกข้อ และข้อสอบที่ลำเอียงจากการวิเคราะห์ด้วยวิธีแมนเทิล-แฮนส์เซลจะซ้ำกับข้อสอบที่ลำเอียงเมื่อวิเคราะห์ด้วยวิธีชิปเทสท์ด้วย ซึ่งผลที่ได้แสดงว่าวิธีชิปเทสท์มีความไวในการตรวจพบข้อสอบที่มีความลำเอียงได้ดีกว่าวิธีแมนเทิล-แฮนส์เซลและวิธีแปลงค่าความยาก

สำหรับขนาดกลุ่มผู้สอบ 1,000 คน พบว่า การวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากกับวิธีแมนเทิล-แฮนส์เซล และวิธีแปลงค่าความยากกับวิธีชิปเทสท์ พบจำนวนข้อสอบที่มีความลำเอียง แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งผลที่ได้สอดคล้องกับสมมติฐานการวิจัย และสอดคล้องกับงานวิจัยของจิตสุดา ธราพร (2539 : 128) ที่ศึกษาวิธีการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแปลงค่าความยากกับวิธีแมนเทิล - แฮนส์เซลจากแบบทดสอบคณิตศาสตร์ใจทย์ปัญหา พบว่า จำนวนข้อสอบที่มีความลำเอียง

จากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากกับวิธีแมนเทิล-แฮนส์เซล มีจำนวนข้อสอบที่มีความลำเอียงแตกต่างกัน และเมื่อพิจารณาจำนวนข้อสอบที่มีความลำเอียงจากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากกับวิธีชิปเทสท์ พบว่า วิธีแปลงค่าความยากไม่พบจำนวนข้อสอบที่มีความลำเอียง ส่วนวิธีชิปเทสท์พบจำนวนข้อสอบที่มีความลำเอียง 6 ข้อ เมื่อนำจำนวนข้อสอบที่ลำเอียงจากวิธีวิเคราะห์ความลำเอียงทั้ง 2 วิธี มาทดสอบนัยสำคัญทางสถิติจึงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และเมื่อพิจารณาข้อสอบที่มีความลำเอียงจากการวิเคราะห์ด้วยวิธีชิปเทสท์จะซ้ำกับวิธีแมนเทิล-แฮนส์เซลทุกข้อ ซึ่งผลที่ได้แสดงว่าวิธีแมนเทิล-แฮนส์เซลมีความไวในการตรวจพบข้อสอบที่มีความลำเอียงได้ดีกว่าวิธีแปลงค่าความยากและวิธีชิปเทสท์

สำหรับวิธีแมนเทิล-แฮนส์เซลกับวิธีชิปเทสท์ พบจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ซึ่งผลที่ได้ไม่สอดคล้องกับสมมติฐานการวิจัย แต่สอดคล้องกับงานวิจัยของนารายานาน และสวามินาธาน (Narayanan and Swaminathan. 1994 : 315-328) ที่ศึกษาเปรียบเทียบผลการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแมนเทิล - แฮนส์เซลและวิธีชิปเทสท์โดยใช้ข้อมูลที่จำลองขึ้นพบว่า วิธีแมนเทิล-แฮนส์เซลและวิธีชิปเทสท์สามารถตรวจพบความลำเอียงของข้อสอบได้ใกล้เคียงกัน

2. จำนวนข้อสอบที่มีความลำเอียงจากการใช้วิธีแปลงค่าความยากในการวิเคราะห์ความลำเอียงของข้อสอบระหว่างกลุ่มผู้สอบขนาด 500 คน และขนาดกลุ่มผู้สอบ 1,000 คน พบว่ามีจำนวนข้อที่ลำเอียงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งสอดคล้องกับสมมติฐานการวิจัย และสอดคล้องกับงานวิจัยของรัตเนอร์ และคณะ (Subkoviak. 1984 : 51; citing Rudner et al. 1980) และงานวิจัยของสับโคเวียค และคณะ (Subkoviak et al. 1984 : 49-59) ที่ศึกษาวิธีวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแปลงค่าความยาก วิธีโคแอสควร์ และวิธีโค้งลักษณะข้อสอบ 3 พารามิเตอร์ จากข้อมูลที่จำลองขึ้นและจากข้อมูลจริงตามลำดับ พบว่า วิธีแปลงค่าความยากเป็นวิธีที่มีข้อจำกัด สามารถนำไปใช้ได้ดีในทางปฏิบัติ ในกรณีที่มีข้อจำกัดเกี่ยวกับคอมพิวเตอร์ กลุ่มตัวอย่างมีขนาดเล็ก และเมื่อพิจารณาจำนวนข้อสอบที่มีความลำเอียง เมื่อวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยาก จะพบว่ามีจำนวนข้อสอบที่มีความลำเอียงน้อย อาจเนื่องมาจากเกณฑ์ที่ใช้ในการตัดสินความลำเอียงของข้อสอบเป็นเกณฑ์ที่ค่อนข้างสูง จึงทำให้พบจำนวนข้อสอบที่มีความลำเอียงน้อย

สำหรับวิธีแมนเทิล-แฮนส์เซลและวิธีชิปเทสท์ ระหว่างกลุ่มผู้สอบขนาด 500 คน และกลุ่มผู้สอบขนาด 1,000 คน พบจำนวนข้อสอบที่มีความลำเอียงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ซึ่งผลที่ได้ไม่สอดคล้องกับสมมติฐานการวิจัย ทั้งนี้อาจเนื่องมาจากการวิเคราะห์ความลำเอียงด้วยวิธีแมนเทิล-แฮนส์เซลและวิธีชิปเทสท์ ได้จำนวนข้อสอบที่มีความลำเอียงใกล้เคียงกัน คือ เมื่อวิเคราะห์จากขนาดกลุ่มผู้สอบ 500 คน วิธีแมนเทิล-แฮนส์เซล พบข้อสอบที่มีความลำเอียง 7 ข้อ และเมื่อวิเคราะห์จากขนาดกลุ่มผู้สอบ 1,000 คนพบข้อสอบที่มีความลำเอียง 8 ข้อ สำหรับวิธีชิปเทสท์ เมื่อวิเคราะห์จากกลุ่มผู้สอบ 500 คนพบข้อสอบที่มีความลำเอียง 9 ข้อ และเมื่อวิเคราะห์จากกลุ่มผู้สอบ 1,000 คนพบข้อสอบที่มีความลำเอียง 6 ข้อ เมื่อนำจำนวนข้อสอบที่มีความลำเอียงจากการวิเคราะห์ความลำเอียงด้วยวิธีแมนเทิล-แฮนส์เซลและวิธีชิปเทสท์ ระหว่างกลุ่มผู้สอบขนาด 500 คน และกลุ่มผู้สอบขนาด 1,000 คน มาทดสอบนัยสำคัญทางสถิติจึงแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และอาจเนื่องมาจากแบบทดสอบที่ใช้ในการศึกษาค้นคว้ามีข้อสอบส่วนใหญ่เป็นข้อสอบที่ค่อนข้างยาก และผู้สอบทั้งสองกลุ่ม คือ กลุ่มผู้สอบเพศชาย และกลุ่มผู้สอบเพศหญิง มีความสามารถใกล้เคียงกัน เมื่อนำผลการตอบแบบทดสอบมาวิเคราะห์ความลำเอียงด้วยวิธีแมนเทิล-แฮนส์เซลและวิธีชิปเทสท์ ซึ่งทั้งสองวิธีดังกล่าว เป็นวิธีการวิเคราะห์ความลำเอียงที่ใช้ผลจากคะแนนการตอบแบบทดสอบหรือคะแนนดิบของผู้สอบมาใช้ในการแบ่งความสามารถของผู้สอบ จึงทำให้การตรวจสอบความลำเอียงของข้อสอบระหว่างกลุ่มผู้สอบต่างกันไม่ว่าจะใช้วิธีแมนเทิล-แฮนส์เซลหรือวิธีชิปเทสท์จึงได้จำนวนข้อสอบที่มีความลำเอียงใกล้เคียงกัน

3. ค่าความเชื่อมั่นแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วจากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากกับวิธีชิปเทสท์ เมื่อวิเคราะห์จากขนาดกลุ่มผู้สอบ 500 คน มีค่าความเชื่อมั่นของแบบทดสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งสอดคล้องกับสมมติฐานการวิจัย ส่วนแบบทดสอบที่คัดเลือกข้อสอบที่มีความลำเอียงออกจากการวิเคราะห์ความลำเอียงด้วยวิธีแปลงค่าความยากกับวิธีแมนเทิล-แฮนส์เซล และแบบทดสอบที่คัดเลือกข้อสอบที่มีความลำเอียงออกจากการวิเคราะห์ความลำเอียงด้วยวิธีแมนเทิล - แฮนส์เซลกับวิธีชิปเทสท์ มีค่าความเชื่อมั่นของแบบทดสอบแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติซึ่งไม่สอดคล้องกับสมมติฐานการวิจัย แต่สอดคล้องกับงานวิจัยของจิตสุตารุทธพร (2359 : 130) ที่ศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแปลงค่าความยากกับวิธีแมนเทิล-แฮนส์เซล พบว่าค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วมีค่าไม่แตกต่างกัน

สำหรับขนาดกลุ่มผู้สอบ 1,000 คน ค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วจากการวิเคราะห์ความลำเอียง 3 วิธีคือ วิธีแปลงค่าความยาก วิธีแมนเทิล-แฮนส์เชล และวิธีชิปเทสท์ พบว่าแบบทดสอบแต่ละฉบับมีค่าความเชื่อมั่นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติซึ่งไม่สอดคล้องกับสมมติฐานการวิจัยแต่สอดคล้องกับงานวิจัยของจิตสุดา ธรพร (2539 : 130) ที่ศึกษาเปรียบเทียบวิธีวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีแปลงค่าความยากกับวิธีแมนเทิล-แฮนส์เชล พบว่าค่าความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วมีค่าไม่แตกต่างกัน

4. ความเชื่อมั่นของแบบทดสอบหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วระหว่างกลุ่มผู้สอบขนาด 500 คน และกลุ่มผู้สอบขนาด 1,000 คน เมื่อใช้วิธีแปลงค่าความยาก และวิธีแมนเทิล-แฮนส์เชล ในการวิเคราะห์ความลำเอียงของข้อสอบ มีค่าความเชื่อมั่นแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ซึ่งไม่สอดคล้องกับสมมติฐานการวิจัย ทั้งนี้อาจเนื่องมาจากจำนวนข้อสอบของแบบทดสอบแต่ละฉบับหลังคัดเลือกข้อสอบที่มีความลำเอียงออกแล้วมีจำนวนข้อใกล้เคียงกันจึงมีผลให้ค่าความเชื่อมั่นของแบบทดสอบแต่ละฉบับไม่แตกต่างกัน ส่วนค่าความเชื่อมั่นของแบบทดสอบที่คัดเลือกข้อสอบที่มีความลำเอียงออกด้วยวิธีชิปเทสท์ เมื่อวิเคราะห์จากขนาดกลุ่มผู้สอบ 500 คนและขนาดกลุ่มผู้สอบ 1,000 คน มีค่าความเชื่อมั่นแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ซึ่งสอดคล้องกับสมมติฐานการวิจัยและอาจเนื่องมาจากแบบทดสอบฉบับที่วิเคราะห์จากขนาดกลุ่มผู้สอบ 500 คนมีจำนวนข้อสอบน้อยกว่าแบบทดสอบฉบับที่วิเคราะห์จากขนาดกลุ่มผู้สอบ 1,000 คนจึงทำให้ค่าความเชื่อมั่นของแบบทดสอบทั้ง 2 ฉบับมีค่าแตกต่างกันซึ่งสอดคล้องกับเฮเลน และเยน (Allen and Yen. 1979 : 91-92) ที่กล่าวว่า ค่าความเชื่อมั่นของแบบทดสอบที่คำนวณจากสูตรสเปียร์แมน-บราวน์ (Spearman-Brown) จะมีแนวโน้มต่ำกว่าเมื่อความยาวของแบบทดสอบสั้นกว่าและค่าความเชื่อมั่นของแบบทดสอบจะมีแนวโน้มสูงขึ้นเมื่อความยาวของแบบทดสอบเพิ่มขึ้น

ข้อเสนอแนะในการวิจัย

1. ควรมีการวิเคราะห์ความลำเอียงของข้อสอบในแบบทดสอบก่อนนำไปใช้ โดยเลือกใช้วิธีแมนเทิล - แฮนส์เชลหรือวิธีชิปเทสท์ เนื่องจากการศึกษาค้นคว้าพบว่า วิธีแมนเทิล - แฮนส์เชลและวิธีชิปเทสท์สามารถตรวจพบจำนวนข้อสอบที่มีความลำเอียงได้ใกล้เคียงกัน และตรวจพบจำนวนข้อสอบที่มีความลำเอียงได้มากกว่าวิธีแปลงค่าความยาก

2. ควรมีการศึกษาความลำเอียงของข้อสอบจากแบบทดสอบฉบับเดียวกันโดยใช้ตัวแปรอื่น ๆ ในการแบ่งกลุ่มเช่น กลุ่มนักเรียนในโรงเรียนสังกัดกรมสามัญศึกษาส่วนกลางกับส่วนภูมิภาค หรือกลุ่มนักเรียนในโรงเรียนสังกัดกรมสามัญศึกษากับโรงเรียนขยายโอกาสทางการศึกษา เพื่อให้แบบทดสอบที่สร้างขึ้นมีความยุติธรรมกับผู้สอบหลายกลุ่ม

3. ควรมีการศึกษาการวิเคราะห์ความลำเอียงของข้อสอบด้วยวิธีวิเคราะห์เดียวกันเมื่อมีขนาดกลุ่มผู้สอบต่าง ๆ กัน เช่น ขนาดกลุ่มผู้สอบเป็น 300 คน, 600 คน และ 900 คน เพื่อศึกษาเปรียบเทียบกับการศึกษาครั้งนี้

บรรณานุกรม

บรรณานุกรม

กาญจนา วัฒนสุนทร. การพัฒนาเกณฑ์ตัดสินข้อสอบลำเอียงทางเพศ. วิทยานิพนธ์ ค.ศ. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2538. อัดสำเนา.

ทัชชัย เผ่าพงศ์. การวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบวัดความถนัดทางการเรียน ด้านคณิตศาสตร์และภาษา ระดับชั้นมัธยมศึกษาตอนต้น. วิทยานิพนธ์ ศศ.ม. ขอนแก่น : มหาวิทยาลัยขอนแก่น, 2527. อัดสำเนา.

ชวาล แพร่ตฤณ. เทคนิคการวัดผล. กรุงเทพฯ : วัฒนาพานิช, 2518.

ทัศนีย์ พิรมนตรี. การวิเคราะห์ความลำเอียงของแบบทดสอบวิชาคณิตศาสตร์ โครงการตรวจสอบคุณภาพการศึกษา ชั้นมัธยมศึกษาปีที่ 6 ปีการศึกษา 2526. วิทยานิพนธ์ ค.ศ. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2530. อัดสำเนา.

นิรมล ชัยขวลิต. การเปรียบเทียบผลการวิเคราะห์ความลำเอียงของแบบทดสอบความเข้าใจในการอ่านภาษาไทย ตามทฤษฎีคลาสสิกอลที่ใช้วิธีวิเคราะห์ต่างกัน. วิทยานิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร , 2537. อัดสำเนา.

บุญชม ศรีสะอาด. การวิจัยทางการวัดผลและประเมินผล. มหาสารคาม : มหาวิทยาลัยศรีนครินทรวิโรฒ มหาสารคาม, 2535.

พัชรี ปิยภักดิ์. การวิเคราะห์ความลำเอียงของข้อสอบจากแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน ชั้นประถมศึกษาปีที่ 6. วิทยานิพนธ์ กศ.ม. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2531. อัดสำเนา.

เรวดี อินทสระ. ผลการตรวจสอบความลำเอียงของข้อสอบต่อการศึกษาคำความเพียงตรงเชิงพยากรณ์ของแบบทดสอบคัดเลือกที่คิดคะแนนต่างกัน. วิทยานิพนธ์ กศ.ด. กรุงเทพฯ : มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2539. อัดสำเนา.

- ถ้วน สายยศ และ อังคณา สายยศ. เทคนิคการวิจัยทางการศึกษา. พิมพ์ครั้งที่ 3. กรุงเทพฯ : ศูนย์ส่งเสริมวิชาการ, 2536.
- สงบ ลักษณะ. "ความสามารถเกี่ยวกับการตอบข้อสอบ," วารสารวัดผลการศึกษา. 4 : 47 - 54; พฤษภาคม - สิงหาคม , 2535.
- สุพัฒน์ สุกมลสันต์. การวิเคราะห์ความลำเอียงของข้อสอบภาษาอังกฤษเข้ามหาวิทยาลัยปี 2531-2533. สถาบันภาษา กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2534.
- สุรศักดิ์ อมรรัตนศักดิ์. การศึกษาเปรียบเทียบผลของวิธีวิเคราะห์ความลำเอียงของข้อสอบที่แตกต่างกัน 4 วิธี. วิทยานิพนธ์ ค.ศ. กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย, 2531. อัดสำเนา.
- สำเร็จ บุญเรืองรัตน์. ทฤษฎีการวัดและการประเมินผลการศึกษา (ฉบับปรับปรุง). กรุงเทพฯ : สำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2529.
- อังคณา สายยศ. การศึกษามลภาวะวิเคราะห์ความลำเอียงของแบบทดสอบความเข้าใจในการอ่านตามลักษณะการใช้ภาษาพูดในครอบครัว. กรุงเทพฯ : ภาควิชาการวัดผลและวิจัยทางการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร, 2539.
- Allen, M.J. and W.M.Yen. Introduction to Measurement Theory. Monterey, Calif Brook / Cole, 1979.
- Angoff, W.H. "Use of Difficulty and Discrimination Indices of Detecting Item Bias," In Handbook of Methods for Detecting Test Bias. Baltimore and London : The Johns Hopkins University Press, 1982.
- Angoff, William H. and Susan F.Ford. "Item - Race Interaction on a Test of Scholastic Aptitude," Journal of Educational Measurement. 10(2) : 95 - 105; Summer, 1973.
- Anastasi, Anne. Differential Psychology : individual and group differences in behavior. 3rd ed. New York : Macmillan, 1976.

- Baghi, H. and S.F.Ferrara. Detecting Differential Item Functioning Using IRT and Mantel-Haenszel Techniques : Implementing Procedure and Comparing Results. 14 - 17; February, 1990.
- Berk, R. A. (Ed.). Handbook of Method for Detecting Test Bias. Baltimore and London MD : The Johns Hopkin University Press, 1982.
- Burrill, L. E. "Comparative Studies of Item Bias Method," in Handbook of Methods for Detecting Test Bias. Baltimore and London : The John Hopkins University Press, 1982.
- Clauser, Brian E. et al. Examination of Various Influence on the Mantel - Haenszel Statistic. Paper Presented at the Annual Meeting of the American Educational Research Association. Chicago, Illinois, April 3 - 7 ; 1991.
- Cronbach, L. J. Essentials of Psychological Testing. New York : Harper & Row Publishers. 1970.
- Doolittle, Allen E. and Cleary T. Anne. "Gender - Based Differential Item Performance in Mathematics Achivement Item," Journal of Educational Measurement. 24(2) : 157 - 166 ; Summer, 1987.
- Dorans, Neil J. and Samuel A. Livingston. "Male - Female Differences in SAT - Verbal Ability Among Students of High SAT - Mathematical Ability," Journal of Educational Measurement. 24(1) : 65 -71 ; Spring, 1987.
- Ferguson, G.A. Statistical Analysis in Psychology and Education. New York : Mc Graw-Hill,1981.
- Freedle,Roy. and Kostin Irene. "Item Difficulty of Four Verble Item Types and an Index of Differential Item Functioning for Black and White Examinees," Journal of Educational Measurement. 27(4) : 329-343 ; Winter,1990.

- Friedman, Lynn. "The Space Factor in Mathematics : Gender Difference," Review of Education Research. 56(1) : 22 - 50 ; Spring, 1995.
- Haebara, T. A. Method for Investigating Item Bias Using Birnbuan ' s Three - Parameter Logistic Model. Iowa Testing Programs Occasional Paper, The University of Iowa, 1979.
- Hambleton, Ronald K. and Roger H. Jane. " Detecting potentially biased test item : Comparison of IRT area and Mantel-Haenszel methods," Applied Measurement in Education. 2 : 313-334 ; 1989.
- Hambleton, Ronald K. and H.Swaminathan. Item Response Theory Principle and Application. Boston : Kluwer-Nijhoff Publishing, 1985.
- Hambleton, Ronald K.,H. Swaminathan and Roger H. Jame. Foundamentals of Item Response Theory. Newbury Park, California, Sage Publications, 1991.
- Hills, J. "Screening for Potentially Biased Item in Testing Program," Educational Measurement : Issues and Practice. 8 : 5 - 11 ; 1990.
- Hulin, Charles L. et al. Item Response Theory Application to Psychological Measurement. Illinois : Dow Jones Irwin, 1983.
- Intasuwan, P. A comparison of Three Approaches for Determining Item Bias in Gross - National Testing. Doctor' s Thesis The University of Pittsburgh, 1979.
- Laksana,S. Application of Analysis of Variance Approach and ITBS Form 7. The University of Iowa, 1979.
- Laksana, S. and W. E.Coffman. A Comparision of and ANOVA Approach and ICC Approach for Assessing Item Bias in an Achievement Test. Iowa Testing Programs Occasional Paper, The University of Iowa, 1980.

- Lord, F.M. and M.R. Novick. Statistical Theories of Mental Test Scores. Reading MA : Addison - Wesley. Publishing Company, 1968. ✓
- Lord, F.M. Application of Item Response Theory to Practical Testing Problems. New Jersey : Lawrence Erlbaum Associates Publishers, 1980. ✓
- Mazor, Kathleen M. et al. "The Effect of sample size on the functioning of The Mantel - Haenszel Statistic," Educational and Psychological Measurement. 52 : 443 - 451 ; 1991.
- Mc Canley, D. Cynthia and Jorge Mendoza . "A Simulation Study of Item Bias Using a Two - Parameter Item Response Model," Applied Psychological Measurement. 9(4) : 389 - 400 ; 1985.
- Mellenbergh, G.J."Item bias and item response theory," International Journal of Educational Research. 13 : 127-143 ;1989.
- Miller, M.D. and T.C.Oshima. "Effect of Sample Size, Number of Biased Item and Magnitude of Bias on a Two-Stage Item Bias Estimation Method," Applied Psychological of Measurement. 16(4) : 381-388; December,1992.
- Millsap, Roger E. and Howard T. Everson. "Methodology Review : Statistical Approaches for Assessing Measurement Bias," Applied Psychological Measurement. 17(4) : 297 - 334 ; 1993.
- ✓ Millsap,R.E. and W. Meredith. "Inferential conditions in the statistical detection of measurement bias," Applied Psychological Measurement. 16 : 389-402 ;1992.
- Narayanan, Pankaja and H.Swaminathan. "Performance of the Mantel - Haenszel and Simultaneous Item Bias Procedures for Detecting Differential Item Function," Applied Psychological Measurement. 18(4) : 315 - 328 ; December, 1994.
- Osterlind, S.I. Test Item Bias. Beverly Hill : Calieornis, Sara McCune, 1983.

Perlman, Carole L. et al. Investigating the Stability of Four Methods for Estimating Item Bias. Paper Presented at the Annual Meeting of the National Council for Measurement in Education. New Orleans, Los Angeles, April 5 - 9 ; 1988.

Popham, W.J. Modern Educational Measurement. Englewood Cliffs, New Jersey : Prentice - Hall, 1981.

Roger, H. Jane and H.Swaminathan. "A Comparison of Logistic Regression and Mantel - Haenszel Procedures for Detecting Differential Item Functioning," Applied Psychological Measurement, 17(2) : 105 - 116 ; 1993.

Roussos, Louis A. and William F. Stout. "Simulation Studies of the Effect of Small Sample Size and Studied Item Parameters on SIBTEST and Mantel-Haenszel Type I Error Performance," Journal of Educational Measurement . 33(2) : 215-230; Summer ,1996.

Ruder, L. M. , P. R.Geston and D. L.Knight. "A Monte Carlo Comparison of Seven Biased Item Detection Techniques," Journal of Educational Measurement. 17 : 1-10 ; 1980.

Shealy, Robin and William Stout. "A Model - Based Standardization Approach that Separates True Bias / DIF from Group Ability Difference and Detects Test Bias / DIF as Well as Item Bias / DIF," Psychometrika. 58 : 159-194 ;1993. ✓

Shepard, L. A. "Definition of Bias," in Handbook of Methods for Detecting Test Bias. Baltimore and London : The Johns Hopkins University Press, 1982.

Snedecor and Cochran. Statistical Method. Iowa :Iowa State University Press, 1967.

Stout, William F. and Louis A. Roussos. SIBTEST USER MANUAL. University of Illinois : December, 1992.

- Subkoviak, Micheal J. et al. "Empirical Comparison of Detected Item Bias Detection Procedure with Bias Manipulation," Journal of Education Measurement, 21(10) : 49 - 59 ; 1984.
- Sudweeks, Richard, R. and Richard R. Tolmen. The Use of Empirical Versus Subjective Procedures for Identifying Science Test Items which Function Differentially for Females and Males. Paper Presented at the Annual Meeting of National Association for Research in Science Teaching. Atlanta, Georgia . April, 8 - 11 ; 1980.
- Swaminathan, H. and H.J. Roger." Detecting differential item functioning using logistic regression procedures, " Journal of Educational Measurement, 27 : 361-370 ; 1990.
- Thissen David, Lynne Steinberg and Wainer Howard. "Use of Item Response Theory in the Study of Group Difference in Trace Lines." in Test Validity, p. 147 -170. Hillsdale, New Jersey : Lawrence Erlbaum Associates, 1988.
- Wise, L. L. Differential item difficulty indicators in small samples. Paper Presented at the meeting of the American Educational Research Association, Washington, April , 1987.

ภาคผนวก

ภาคผนวก ก
รายชื่อผู้เชี่ยวชาญ

ผู้เชี่ยวชาญ ในการพิจารณาความเที่ยงตรงเชิงเนื้อหาของแบบทดสอบคณิตศาสตร์โจทย์ปัญหา

- | | |
|--------------------------------|---------------------------------------|
| 1. ผศ.เขาวนา ชวลิตธำรง | มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร |
| 2. อาจารย์ระวีวรรณ พันธุ์พานิช | มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร |
| 3. อาจารย์ชวลิต รวยอาจิณ | มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร |
| 4. อาจารย์ดร.รชนี บุญทวีภ | โรงเรียนกุนนทีรุทธารามวิทยาคม |
| 5. อาจารย์แสงระวี มิ่งขวัญ | โรงเรียนลำลูกกา |

ภาคผนวก ข

ค่าความยาก (p) และค่าอำนาจจำแนก (r) รายข้อ
ของแบบทดสอบคณิตศาสตร์โจทย์ปัญหา

ตาราง 12 ค่าความยาก (p) ค่าอำนาจจำแนก (r) เฉพาะตัวถูกของแบบทดสอบคณิตศาสตร์ใจทย์
ปัญหา จากการทดสอบครั้งที่ 1

| ข้อ | p | r | ข้อ | p | r |
|-------|-------|-------|------|-------|--------|
| 1. | 0.303 | 0.470 | 25* | 0.371 | 0.457 |
| 2* | 0.448 | 0.455 | 26* | 0.248 | 0.398 |
| 3* | 0.326 | 0.391 | 27* | 0.432 | 0.388 |
| 4. | 0.829 | 0.395 | 28* | 0.210 | 0.378 |
| 5* | 0.487 | 0.407 | 29* | 0.294 | 0.434 |
| 6. | 0.387 | 0.459 | 30. | 0.165 | 0.030 |
| 7. | 0.352 | 0.022 | 31. | 0.239 | 0.176 |
| 8* | 0.216 | 0.374 | 32* | 0.284 | 0.418 |
| 9* | 0.610 | 0.467 | 33. | 0.148 | 0.113 |
| 10* | 0.471 | 0.410 | 34* | 0.258 | 0.436 |
| 11* | 0.352 | 0.398 | 35. | 0.190 | 0.088 |
| 12* | 0.423 | 0.421 | 36* | 0.303 | 0.411 |
| 13* | 0.365 | 0.380 | 37* | 0.261 | 0.431 |
| 14* | 0.342 | 0.451 | 38. | 0.184 | 0.069 |
| 15* | 0.358 | 0.345 | 39. | 0.135 | -0.026 |
| 16* | 0.390 | 0.383 | 40* | 0.281 | 0.447 |
| 17*.. | 0.445 | 0.383 | 41* | 0.265 | 0.440 |
| 18*.. | 0.332 | 0.360 | 42* | 0.265 | 0.414 |
| 19* | 0.242 | 0.362 | 43* | 0.232 | 0.482 |
| 20*.. | 0.300 | 0.394 | 44. | 0.161 | 0.074 |
| 21*.. | 0.397 | 0.400 | 45* | 0.230 | 0.482 |
| 22* | 0.371 | 0.419 | 46. | 0.097 | 0.074 |
| 23* | 0.255 | 0.442 | 47* | 0.245 | 0.499 |
| 24*.. | 0.374 | 0.425 | 48.. | 0.068 | 0.172 |

ตาราง 12 (ต่อ)

| ข้อ | p | r | ข้อ | p | r |
|------|-------|-------|------|-------|-------|
| 49* | 0.277 | 0.526 | 55. | 0.239 | 0.052 |
| 50.. | 0.094 | 0.069 | 56. | 0.200 | 0.039 |
| 51*. | 0.258 | 0.497 | 57*. | 0.277 | 0.439 |
| 52. | 0.245 | 0.043 | 58. | 0.216 | 0.157 |
| 53. | 0.181 | 0.083 | 59. | 0.097 | 0.023 |
| 54*. | 0.400 | 0.509 | 60*. | 0.313 | 0.495 |

* ข้อสอบที่คัดเลือกไว้ใช้ในการวิจัย

ภาคผนวก ค

ค่าเฉลี่ยของข้อสอบแต่ละข้อระหว่างผู้สอบเพศชายและ
ผู้สอบเพศหญิง จำแนกตามขนาดกลุ่มผู้สอบ

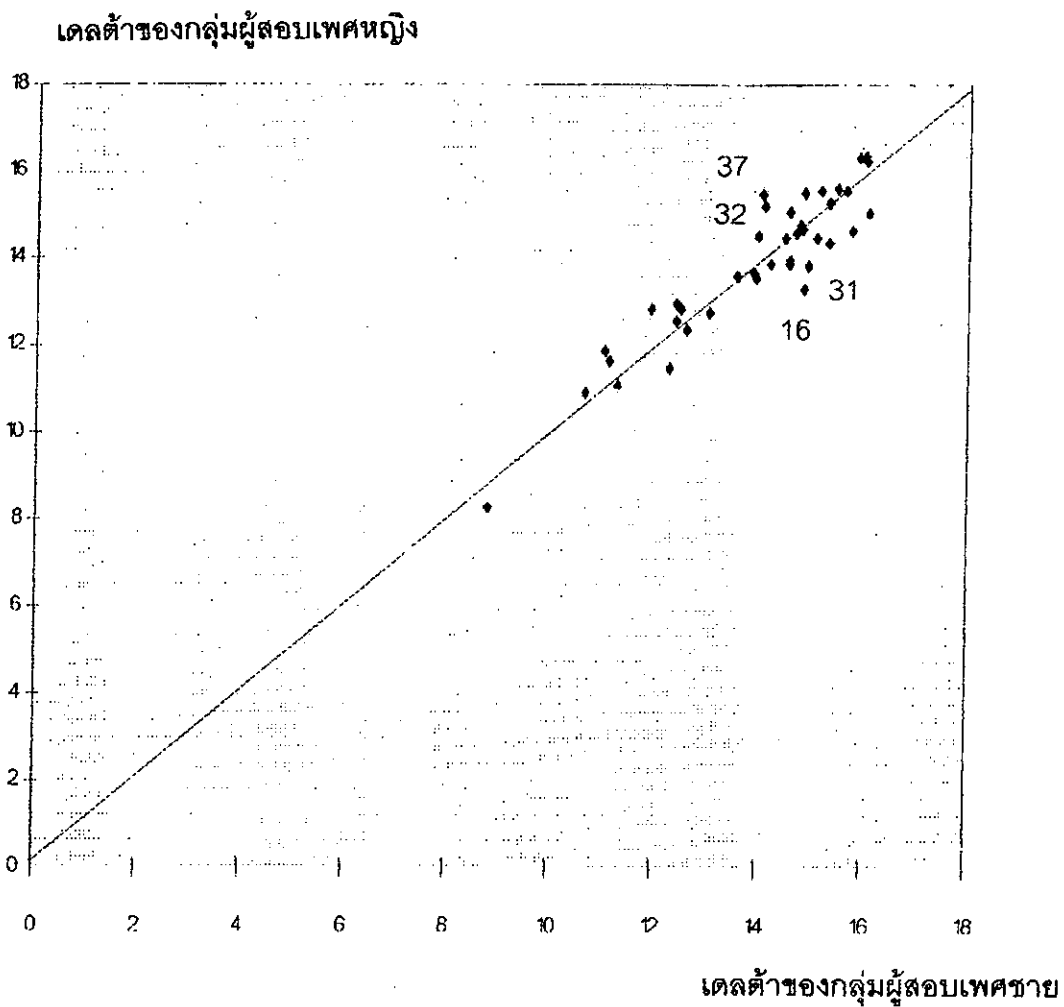
ภาพประกอบการกระจายของคู่อันดับค่าเฉลี่ย
ของข้อสอบแต่ละข้อระหว่างผู้สอบเพศชายและผู้สอบเพศหญิงจำแนกตาม
ขนาดกลุ่มผู้สอบ

ตาราง 13 ค่าเฉลี่ยของข้อสอบแต่ละข้อระหว่างกลุ่มผู้สอบเพศชายกับกลุ่มผู้สอบเพศหญิง
จำแนกตามขนาดกลุ่มผู้สอบ

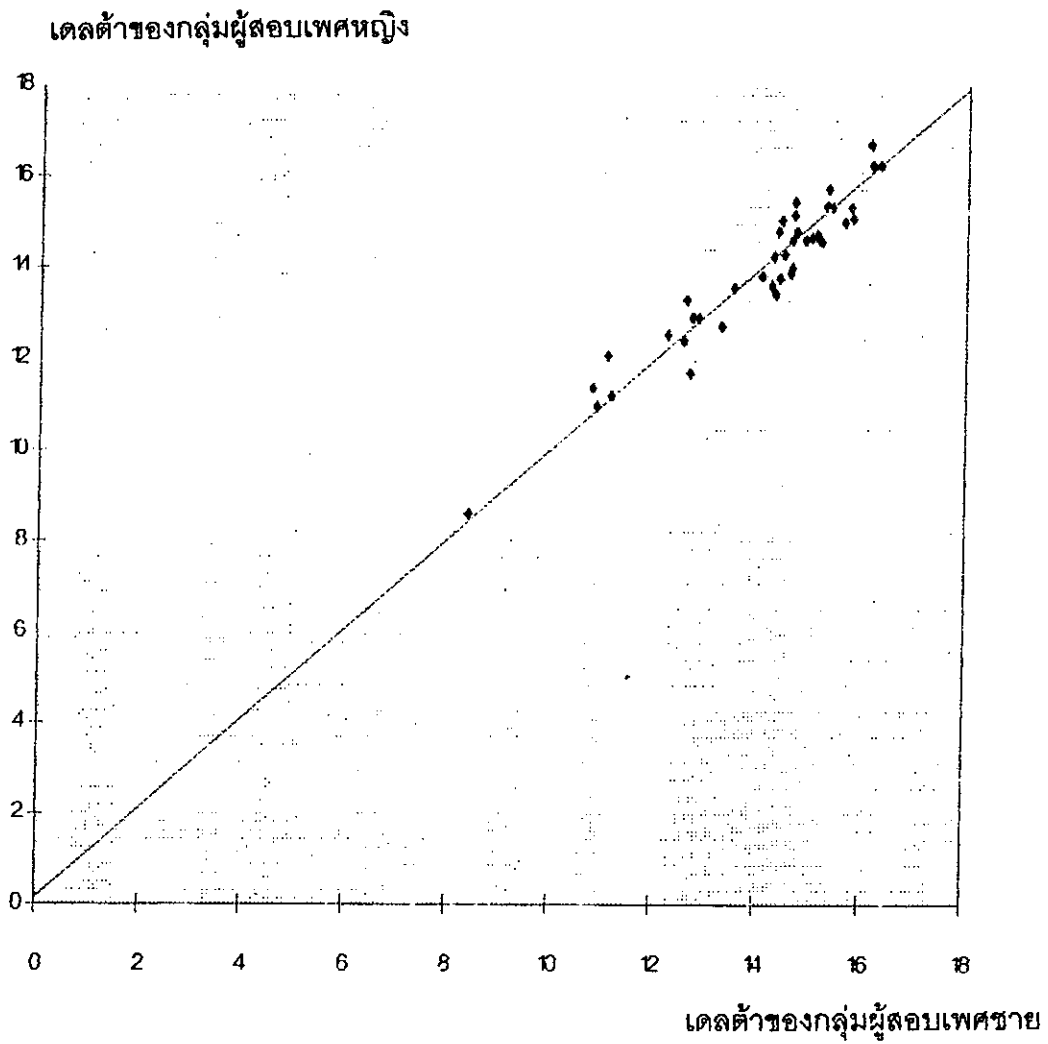
| ข้อ | ขนาดกลุ่มผู้สอบ 500 คน | | ขนาดกลุ่มผู้สอบ 1,000 คน | |
|-----|------------------------|---------|--------------------------|---------|
| | เพศชาย | เพศหญิง | เพศชาย | เพศหญิง |
| 1. | 8.8 | 8.28 | 8.40 | 8.60 |
| 2. | 11.12 | 11.64 | 10.80 | 11.36 |
| 3. | 10.68 | 10.92 | 10.88 | 10.96 |
| 4. | 11.04 | 11.88 | 11.08 | 12.04 |
| 5. | 15.88 | 16.32 | 16.12 | 16.68 |
| 6. | 12.60 | 12.36 | 12.56 | 12.40 |
| 7. | 12.40 | 12.96 | 12.60 | 13.28 |
| 8. | 11.92 | 12.84 | 12.24 | 12.52 |
| 9. | 12.40 | 12.56 | 12.72 | 12.92 |
| 10. | 13.88 | 13.68 | 14.04 | 13.80 |
| 11. | 11.28 | 11.08 | 11.16 | 11.20 |
| 12. | 14.56 | 13.88 | 14.48 | 14.28 |
| 13. | 13.56 | 13.60 | 13.52 | 13.56 |
| 14. | 12.28 | 11.48 | 12.68 | 11.68 |
| 15. | 12.48 | 12.84 | 12.84 | 12.88 |
| 16. | 14.84 | 13.32 | 14.32 | 13.40 |
| 17. | 15.16 | 15.56 | 15.28 | 15.36 |
| 18. | 15.48 | 15.60 | 15.32 | 15.72 |
| 19. | 13.04 | 12.76 | 13.28 | 12.72 |
| 20. | 13.96 | 14.52 | 14.28 | 14.24 |
| 21. | 14.84 | 15.52 | 14.68 | 15.16 |
| 22. | 13.92 | 13.56 | 14.24 | 13.60 |
| 23. | 14.76 | 14.76 | 14.64 | 14.60 |
| 24. | 14.56 | 13.96 | 14.64 | 14.00 |

ตาราง 13 (ต่อ)

| ข้อ | ขนาดกลุ่มผู้สอบ 500 คน | | ขนาดกลุ่มผู้สอบ 1,000 คน | |
|-----|------------------------|---------|--------------------------|---------|
| | เพศชาย | เพศหญิง | เพศชาย | เพศหญิง |
| 25. | 14.20 | 13.88 | 14.40 | 13.76 |
| 26. | 14.80 | 14.68 | 15.00 | 14.68 |
| 27. | 15.08 | 14.48 | 15.20 | 14.56 |
| 28. | 14.92 | 13.84 | 14.60 | 13.88 |
| 29. | 14.68 | 14.60 | 14.72 | 14.80 |
| 30. | 15.32 | 14.36 | 15.12 | 14.72 |
| 31. | 15.76 | 14.64 | 15.64 | 15.00 |
| 32. | 14.08 | 15.20 | 14.44 | 15.04 |
| 33. | 16.08 | 15.04 | 15.80 | 15.08 |
| 34. | 16.04 | 16.24 | 16.32 | 16.24 |
| 35. | 15.32 | 15.28 | 15.40 | 15.32 |
| 36. | 15.64 | 15.56 | 15.76 | 15.32 |
| 37. | 14.04 | 15.48 | 14.36 | 14.80 |
| 38. | 14.48 | 14.48 | 14.88 | 14.60 |
| 39. | 16.00 | 16.36 | 16.16 | 16.24 |
| 40. | 14.56 | 15.08 | 14.68 | 15.44 |



ภาพประกอบ 8 แสดงการกระจายของคู่อันดับค่าเดลต้าของข้อสอบแต่ละข้อระหว่างกลุ่มผู้สอบเพศชายกับกลุ่มผู้สอบเพศหญิง เมื่อวิเคราะห์จากขนาดกลุ่มผู้สอบ 500 คน



ภาพประกอบ 9 แสดงการกระจายของคู่อันดับค่าเดลด้าของข้อสอบแต่ละข้อระหว่างลุ่มผู้สอบเพศชายกับลุ่มผู้สอบเพศหญิง เมื่อวิเคราะห์จากขนาดลุ่มผู้สอบ 1,000 คน

ภาคผนวก ง

ค่าดัชนีเบต้า (β) ของข้อสอบแต่ละข้อ จำแนกตามขนาดกลุ่มผู้สอบ

ตาราง 14 ค่าดัชนีเบต้า(β)ของข้อสอบแต่ละข้อจำแนกตามขนาดกลุ่มผู้สอบ

| ข้อ | ค่าเบต้า (β) | |
|-----|------------------------|--------------------------|
| | ขนาดกลุ่มผู้สอบ 500 คน | ขนาดกลุ่มผู้สอบ 1,000 คน |
| 1. | -.041 | .011 |
| 2. | .047 | .054 |
| 3. | .029 | .024 |
| 4. | .097 | .111 |
| 5. | .035 | .037 |
| 6. | -.010 | -.008 |
| 7. | .053 | .078 |
| 8. | .108 | .040 |
| 9. | .017 | .046 |
| 10. | -.015 | -.002 |
| 11. | -.016 | .021 |
| 12. | -.083 | -.010 |
| 13. | .027 | .019 |
| 14. | -.075 | -.083 |
| 15. | .050 | .015 |
| 16. | -.143 | -.085 |
| 17. | .037 | .026 |
| 18. | .022 | .038 |
| 19. | -.014 | -.036 |
| 20. | .043 | -.020 |
| 21. | .077 | .050 |
| 22. | -.026 | -.051 |
| 23. | .023 | .019 |
| 24. | -.061 | -.048 |
| 25. | -.024 | -.047 |

ตาราง14(ต่อ)

| ข้อ | ค่าเบต้า (β) | |
|-----|------------------------|--------------------------|
| | ขนาดกลุ่มผู้สอบ 500 คน | ขนาดกลุ่มผู้สอบ 1,000 คน |
| 26. | .005 | -.004 |
| 27. | -.051 | -.043 |
| 28. | -.084 | -.050 |
| 29. | .004 | .026 |
| 30. | -.070 | -.034 |
| 31. | -.083 | -.044 |
| 32. | .109 | .071 |
| 33. | -.057 | -.052 |
| 34. | .023 | -.016 |
| 35. | -.002 | .011 |
| 36. | .005 | -.023 |
| 37. | .125 | .058 |
| 38. | .017 | -.021 |
| 39. | .033 | .037 |
| 40. | .054 | .067 |

ภาคผนวก จ
แบบทดสอบคณิตศาสตร์โจทย์ปัญหา

แบบทดสอบวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 1

คำชี้แจงวิธีทำแบบทดสอบ

1. แบบทดสอบฉบับนี้มี 40 ข้อ ให้เวลาทำ 60 นาที
2. คำถามทั้งหมดเป็นแบบปรนัยชนิดเลือกตอบ 5 ตัวเลือก จาก ก ข ค ง และ จ ซึ่งจะมีตัวถูกเพียงตัวเลือกเดียวเท่านั้น
3. การตอบให้นักเรียนเลือกตอบคำตอบที่ถูกที่สุดเพียงคำตอบเดียว เมื่อเลือกได้แล้วให้ทำเครื่องหมาย "x" ลงในกระดาษคำตอบให้ตรงกับข้อที่เลือก ดังตัวอย่างการตอบข้อ ก.

| | | | | |
|---|---|---|---|---|
| ก | ข | ค | ง | จ |
| x | | | | |

4. การเปลี่ยนคำตอบให้นักเรียนทำเครื่องหมาย "=" ทับคำตอบเดิมเสียก่อน แล้วจึงไปทำเครื่องหมาย "x" ที่คำตอบใหม่ ดังตัวอย่างการเปลี่ยนคำตอบจากข้อ ก เป็นข้อ ง ดังนี้

| | | | | |
|---|---|---|---|---|
| ก | ข | ค | ง | จ |
| = | | | x | |

5. ห้ามนักเรียนทำเครื่องหมายใด ๆ ลงในตัวข้อสอบ และต้องส่งข้อสอบคืนพร้อมกระดาษคำตอบด้วย
6. พยายามทำข้อสอบให้ครบทุกข้อ ถ้าพบข้อยากอย่าท้อใจ ควรข้ามไปทำข้ออื่นก่อนเมื่อมีเวลาเหลือ ค่อยย้อนกลับมาทำใหม่
7. การทศเลข เมื่อนักเรียนต้องการทศเลข หรือขีดเขียนสิ่งใด ให้ทศหรือเขียนลงในด้านหลังของกระดาษคำตอบ
8. ถ้าใครสงสัยให้ยกมือถามได้ในตอนนี เพราะในระหว่างที่ลงมือทำ นักเรียนจะถามอะไรอีกไม่ได้
9. ให้นักเรียนทุกคนเขียนชื่อ ชื่อสกุล เพศ โรงเรียน ให้เรียบร้อย และชัดเจนด้วยตัวบรรจง ลงในกระดาษคำตอบที่ครูแจกให้ แล้วลงมือทำได้

ขอให้นักเรียนพยายามคิดให้รอบคอบ และตัดสินใจตอบข้อสอบทุกข้อให้ถูกต้องมากที่สุดในเวลาที่กำหนดให้

1. เลขจำนวนหนึ่งน้อยกว่า 30 อยู่ 16 เลขอีกจำนวนหนึ่งมากกว่า 12 อยู่ 9 จงหาผลรวมของเลขสองจำนวนนี้ ?

- ก. 14
- ข. 21
- ค. 35
- ง. 46
- จ. 67

2. โรงเรียนแห่งหนึ่งมีนักเรียนชั้นมัธยมศึกษาปีที่หนึ่งจำนวน 660 คน เลือกแผนการเรียนคณิตศาสตร์ คิดเป็น $\frac{1}{3}$ ของจำนวนนักเรียนทั้งหมดและเลือกแผนการเรียนพาณิชยกรรมจำนวน 55 คน ดังนั้นนักเรียนที่ยังไม่ได้เลือกแผนการเรียนมีกี่คน ?

- ก. 165
- ข. 220
- ค. 275
- ง. 385
- จ. 440

3. ถ้านำ 4 ลบออกจากเลขจำนวนหนึ่ง แล้วคูณจำนวนที่เหลือด้วย 8 ผลคูณจะเป็น 144 เลขจำนวนนั้นมีค่าเท่าไร?

- ก. 36.0
- ข. 22.0
- ค. 18.5
- ง. 17.5
- จ. 14.0

4. ถังใบหนึ่งมีน้ำเต็มถัง เมื่อใช้ไปแล้ว $\frac{5}{6}$ ถัง จะเหลือน้ำ 13 ลิตร ถังใบนี้จะมีปริมาตรเท่าไร ?

- ก. 16
- ข. 43
- ค. 65
- ง. 78
- จ. 143

5. ปีติฝากเงิน 3,100 บาท กับธนาคารออมสิน ธนาคารคิดดอกเบี้ยเงินฝากทบต้นในอัตราร้อยละ 8.50 ต่อปี เมื่อครบ 3 ปี ปีติจะได้รับดอกเบี้ยเงินฝากทั้งหมดเท่าไร ?

- ก. 263.50
- ข. 790.50
- ค. 859.50
- ง. 3,363.50
- จ. 3,890.50

6. ก มีเงินเป็น 5 เท่าของ ข ถ้า ก แบ่งเงินให้ ข 30 บาท ก จะมีเงินเป็น 3 เท่าของ ข เดิม ก มีเงินเท่าไร ?

- ก. 15
- ข. 30
- ค. 45
- ง. 75
- จ. 120

7. ก และ ข นำเงินที่มีอยู่มาจับคู่ ปรากฏว่า ก มีเงินน้อยกว่า ข อยู่ 45 บาท แต่ ก มีเงินในกระปุกออมสินอีก 120 บาท ทำให้ ก มีเงินเป็นสองเท่าของ ข ข จะมีเงินเท่าไร ?

- ก. 30
- ข. 60
- ค. 75
- ง. 150
- จ. 165

8) สิ้นห้าของเงินจำนวนหนึ่ง เมื่อเพิ่มอีก 7 บาท
จะเป็น 92 บาท เงินจำนวนนี้เป็นเท่าไร?

- ก. 66.50
- ข. 80.50
- ค. 106.25
- ง. 108.00
- จ. 123.75

9) ปัจจุบัน อามีอายุ 54 ปี เมื่อ 6 ปีที่แล้ว หลาน
มีอายุเป็น $\frac{7}{12}$ ของอา ปัจจุบันหลานมีอายุ
เท่าไร?

- ก. 28.00
- ข. 29.00
- ค. 31.50
- ง. 34.00
- จ. 37.58

10) จำนวนเต็มสามจำนวนเรียงลำดับจาก
ค่าน้อยไปหาค่ามาก โดยที่จำนวนที่สองมี
ค่ามากเป็น 2 เท่าของจำนวนแรก และ
จำนวนที่สามมีค่ามากเป็น 6 เท่าของ
จำนวนที่สอง ผลรวมของเลขสามจำนวน
เป็น 180 จงหาผลต่างระหว่างจำนวนที่
สามกับผลรวมของจำนวนแรกกับจำนวนที่
สอง?

- ก. 60
- ข. 100
- ค. 108
- ง. 120
- จ. 135

11) สมชายหนัก 36 กิโลกรัม สุรชัยหนักกว่า
ครึ่งหนึ่งของน้ำหนักสมชายอยู่ 4 กิโลกรัม
ถ้าชั่งน้ำหนักพร้อมกันจะได้กี่กิโลกรัม?

- ก. 54
- ข. 56
- ค. 58
- ง. 74
- จ. 76

12) ก ขายวิทยุให้ ข ได้กำไร 4% ข ขายให้ ค ได้
กำไร 25% ค ขายให้ ง ขาดทุน 20% ถ้า
ง ซื้อวิทยุมาในราคา 312 บาท จงหาว่า ก
ซื้อวิทยุมาในราคาเท่าไร?

- ก. 284
- ข. 300
- ค. 334
- ง. 340
- จ. 390

13) ก และ ข มีเงินมากกว่า ค อยู่ 7 บาท และ 5
บาท ตามลำดับ ถ้าทั้งสามคนมีเงินรวมกัน
48 บาท อยากทราบว่า ข มีเงินเท่าไร?

- ก. 12
- ข. 16
- ค. 17
- ง. 19
- จ. 21

14) อีก 8 ปี มานีจะมีอายุเท่ากับสมศักดิ์ ถ้าในปี
พ.ศ. 2540 สมศักดิ์มีอายุครบวัยเบญจเพส
แล้วมานีจะมีอายุเท่าไร?

- ก. 12
- ข. 17
- ค. 28
- ง. 32
- จ. 33

15) หกเท่าของผลต่างของจำนวนหนึ่งกับ 5 เท่ากับ 12 สามเท่าของจำนวนนั้นมีค่าเท่าไร?

- ก. 7
- ข. 9
- ค. 11
- ง. 21
- จ. 77

16) ก และ ข มีเงินรวมกัน 30 บาท ถ้า ก จ่ายไป 6 บาท และ ข ได้รับแจก 9 บาท ทำให้ ก มีเงินเพียงครึ่งหนึ่งของ ข เดิม ก มีเงินเท่าไร?

- ก. 11
- ข. 13
- ค. 15
- ง. 17
- จ. 27

17) มีเงินอยู่ 252 บาท ซื้อสมุดราคาเล่มละ 6 บาทและหนังสือราคาเล่มละ 9 บาท เมื่อนับจำนวนสมุดและหนังสือรวมกันจะได้ 34 เล่ม สมุดมากกว่าหนังสือกี่เล่ม?

- ก. 2
- ข. 6
- ค. 12
- ง. 18
- จ. 24

18) ถุงใบหนึ่ง มีเหรียญบาทมากกว่าเหรียญห้าสิบลบาทอยู่ 7 เหรียญ เหรียญห้าบาทน้อยกว่าเหรียญบาทอยู่ 3 เหรียญ เมื่อนับจำนวนเหรียญทั้งหมดในถุงได้ 119 เหรียญ เหรียญที่มีจำนวนน้อยที่สุด คิดเป็นเงินกี่บาท?

- ก. 18
- ข. 21
- ค. 36
- ง. 40
- จ. 43

19) แบ่งลูกอม 3 โหล ให้เด็ก 3 คน คนแรกได้มากกว่าคนที่สองอยู่ 4 เม็ด คนที่สามได้น้อยกว่าคนแรก 8 เม็ด คนที่ได้มากที่สุดจะได้กี่เม็ด?

- ก. 13
- ข. 15
- ค. 16
- ง. 18
- จ. 20

20) มานีจ่ายค่าอาหารเดือนละ 6,000 บาท และฝากธนาคารเดือนละ 550 บาท ถ้ามานีต้องการฝากธนาคาร เดือนละ 1,250 บาท โดยการลดค่าอาหารลงแล้ว มานีจะต้องลดค่าอาหารลงร้อยละเท่าไรต่อสัปดาห์?

- ก. 2.9
- ข. 3.3
- ค. 11.7
- ง. 13.2
- จ. 22.1

21) ในการรับประทานอาหารมื้อหนึ่ง มานีจ่ายเงินทั้งสิ้น 540 บาท เป็นค่าเครื่องดื่ม 58 บาท ค่าผ้าเย็น 12 บาท นอกนั้นเป็นค่าอาหารซึ่งร้านค้าลดราคาให้แก่ลูกค้า 10% จงหาราคาค่าอาหารก่อนลดราคา ?

- ก. 423
- ข. 470
- ค. 486
- ง. 522
- จ. 600

22) ชายซื้อไปราคาตัวละ 168 บาท ขายขาดทุน 16% แต่ถ้าขายไปราคาตัวละ 286 บาท จะได้กำไรกี่เปอร์เซ็นต์ ?

- ก. 43
- ข. 60
- ค. 84
- ง. 86
- จ. 88

23) ถ้านี้สอบวิชาคณิตศาสตร์ได้ 47 คะแนน วิชาวิทยาศาสตร์ได้ 33 คะแนน ถ้าสี่เท่าของคะแนนวิชาภาษาไทย เท่ากับสี่ในห้าของคะแนนวิชาคณิตศาสตร์และวิชาวิทยาศาสตร์รวมกัน ถ้านี้สอบได้คะแนนทั้งหมดเท่าไร ?

- ก. 64
- ข. 80
- ค. 96
- ง. 100
- จ. 144

24) โรงเรียนแห่งหนึ่ง มีอัตราส่วนของจำนวนนักเรียนชายต่อจำนวนนักเรียนทั้งหมด เป็น 4: 9 ถ้ามีนักเรียนหญิง 360 คน จะมีนักเรียนทั้งหมดกี่คน ?

- ก. 160
- ข. 288
- ค. 450
- ง. 648
- จ. 810

25) แบ่งเชือกเส้นหนึ่งออกเป็นสามส่วน ให้ส่วนแรกมีความยาวเป็น $\frac{1}{8}$ ส่วนที่สองยาวเป็นครึ่งหนึ่งของส่วนแรก ส่วนที่สามยาว 39 เมตร เดิมเชือกยาวเท่าไร ?

- ก. 42
- ข. 45
- ค. 48
- ง. 55
- จ. 62

26) สี่เหลี่ยมผืนผ้ารูปหนึ่ง มีด้านยาว 12 นิ้ว และมีพื้นที่ 96 ตารางนิ้ว หากลดด้านยาวและด้านกว้างลงอีก 2 นิ้ว พื้นที่สี่เหลี่ยมผืนผ้าจะลดลงกี่ตารางนิ้ว ?

- ก. 80
- ข. 72
- ค. 60
- ง. 36
- จ. 24

27) นักเรียนชั้นมัธยมปีที่ 1 จำนวน 740 คน ลงทะเบียนซ่อมวิชาคณิตศาสตร์ 15% จงหาว่า มีนักเรียนสอบผ่านมากกว่านักเรียนสอบไม่ผ่านวิชาคณิตศาสตร์กี่คน ?

- ก. 111
- ข. 131
- ค. 222
- ง. 518
- จ. 629

28) ในการทำขนมเค้กมีส่วนผสมดังนี้ แป้งสาลี 3 ถ้วย น้ำตาลปน 1 ถ้วย เนยครึ่งถ้วย เมเปิ้ลหนึ่งในสี่ถ้วย และ ไข่ไก่ 2 ฟอง จงหาอัตราส่วนของน้ำตาลปนต่อเนยต่อเมเปิ้ลต่อแป้งสาลี เมื่อใช้ไข่ไก่ 8 ฟอง ?

- ก. 4 : 2 : 1 : 12
- ข. 4 : 4 : 2 : 12
- ค. 8 : 4 : 2 : 24
- ง. 12 : 4 : 2 : 1
- จ. 24 : 8 : 4 : 2

29) อัตราส่วนที่ใช้ในการแบ่งเงินให้เด็ก 2 คน เป็น 8 : 9 ถ้าให้เด็กคนแรกได้รับเงิน 104 บาท และเด็กคนที่สองได้รับเงิน 106 บาท จงหาว่า เด็กคนที่สองควรได้รับเงินเพิ่มอีกเท่าไร ?

- ก. 14
- ข. 13
- ค. 12
- ง. 11
- จ. 10

30) ดินสองราคาถุงละ 77 บาท ซื้อดินสอง 72 แห่ง คิดเป็นเงินกี่บาท ?

- ก. 36.00
- ข. 38.50
- ค. 46.25
- ง. 55.50
- จ. 93.50

31) ในการผสมปูนโบกตึก จะต้องใช้ปูน 3 ส่วนต่อน้ำ 5 ส่วน แต่ถ้าใช้น้ำ 3 ส่วน จะต้องใช้ทราย 4 ส่วน จงหาว่าจะต้องใช้ทรายมากกว่าน้ำกี่ถัง เมื่อใช้ปูน 27 ถังในการผสม ?

- ก. 5
- ข. 6
- ค. 11
- ง. 15
- จ. 18

32) ครึ่งหนึ่งของอายุมานี เมื่อ 7 ปีที่แล้ว เท่ากับ 15 ปี ถ้าอีก 7 ปีข้างหน้า มานีจะอายุเท่าไร ?

- ก. 23
- ข. 30
- ค. 37
- ง. 44
- จ. 51

33) สนามรูปวงกลมมีเส้นผ่านศูนย์กลางยาว 14 เมตร มานีวิ่งรอบสนาม 2 รอบใช้เวลา 4 นาที 30 วินาที มานีวิ่งรอบสนามด้วยอัตราเร็วที่เมตรต่อนาที ?

- ก. 19.6
- ข. 20.5
- ค. 39.1
- ง. 40.9
- จ. 68.4

34) สี่เหลี่ยมจัตุรัสมีด้านยาว 6 นิ้ว จงหาอัตราส่วนของเส้นรอบรูปต่อพื้นที่ของรูปสี่เหลี่ยมจัตุรัสสองรูปรวมกัน ?

- ก. 0.5 : 1.0
- ข. 1.0 : 1.5
- ค. 3.0 : 2.0
- ง. 24.0 : 36.0
- จ. 48.0 : 72.0

35) สี่เหลี่ยมจัตุรัส 2 รูป มีความยาวด้านเป็นอัตราส่วน 2 : 3 ถ้าสี่เหลี่ยมรูปใหญ่มีพื้นที่ 81 ตารางนิ้วแล้วสี่เหลี่ยมรูปเล็กจะมีพื้นที่เท่าไร ?

- ก. 4
- ข. 6
- ค. 9
- ง. 24
- จ. 36

36) มานีได้รับเงินค่าขนมวันละ 8,400 บาท แต่ใช้จ่ายหมด เมื่อถึงวันสิ้นปีมานียังมีหนี้สินอีก 3,600 บาท จงหาว่ามานีใช้เงินไปเท่าไรในเวลาสองเดือนครึ่ง ?

- ก. 750
- ข. 1,000
- ค. 1,750
- ง. 2,000
- จ. 2,500

37) เลขสามจำนวนเรียงกันและสามเท่าของจำนวนกลางมากกว่าผลบวกของเลขสองจำนวนที่เหลือ อยู่ 12 เลขจำนวนน้อยที่สุดมีค่าเท่าไร ?

- ก. 6
- ข. 7
- ค. 11
- ง. 13
- จ. 17

38) ซื้อกาแฟผงมา 2 ชนิด 20 กิโลกรัม ราคา 1,824 บาท ชนิดแรกราคากิโลกรัมละ 72 บาท ชนิดที่สองราคากิโลกรัมละ 120 บาท จงหาว่าซื้อกาแฟผงชนิดแรกมากกว่าชนิดที่สองกี่กิโลกรัม ?

- ก. 3
- ข. 4
- ค. 8
- ง. 12
- จ. 14

- 39) สี่เหลี่ยมมุมฉากรูปหนึ่ง มีเส้นรอบรูปยาว 28 นิ้ว ด้านยาวเป็น $\frac{4}{3}$ เท่าของด้านกว้าง จงหาพื้นที่สี่เหลี่ยมมุมฉากรูปนี้ ?
- ก. 12
ข. 14
ค. 20
ง. 36
จ. 48

- 40) ซื้อปากการาคาแท่งละ 12 บาท ดินสอราคาแท่งละ 3.50 บาท รวมเป็นเงิน 275 บาท โดยที่จำนวนปากกาเป็น 2 เท่าของจำนวนดินสอ จงหาว่าซื้อปากกามากี่แท่ง ?
- ก. 5
ข. 10
ค. 17
ง. 20
จ. 34

ประวัติย่อของผู้วิจัย

ชื่อ นางพรรณณี ชื่อสกุล จินตมาศ

เกิดวันที่ 30 เดือนสิงหาคม พุทธศักราช 2503

สถานที่เกิด อำเภอบางรัก กรุงเทพมหานคร

สถานที่อยู่ปัจจุบัน บ้านเลขที่ 7/263 หมู่ 10 เขตลาดพร้าว กรุงเทพมหานคร

ตำแหน่งหน้าที่การงาน อาจารย์ 1 ระดับ 5

สถานที่ทำงานปัจจุบัน โรงเรียนกุนนทีรุทธARAMวิทยาคม เขตดินแดง
กรุงเทพมหานคร

ประวัติการศึกษา

พ.ศ. 2520 มัธยมศึกษาตอนปลาย (วิทย์-คณิต) จากโรงเรียนศรีอยุธยา
กรุงเทพมหานคร

พ.ศ. 2523 ปกศ.สูง (วิชาเอกวิทยาศาสตร์ทั่วไป) จากวิทยาลัยครูบ้าน
สมเด็จพระเจ้าพระยา

พ.ศ. 2528 กศ.บ.(วิชาเอกการวัดผลการศึกษา) จากมหาวิทยาลัยศรีนคริน
ทรวิโรฒ ประสานมิตร

พ.ศ. 2540 กศ.ม. (การวัดผลการศึกษา) จากมหาวิทยาลัยศรีนครินทรวิโรฒ
ประสานมิตร