

การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์
ที่มีวิธีการตรวจ จำนวนผู้ตรวจ และประสบการณ์ของผู้ตรวจ แตกต่างกัน

ปริญญาณิพนธ์
ของ
อุษณีย์ บัวศิริพันธุ์

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา
มีนาคม 2543
ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์
ที่มีวิธีการตรวจ จำนวนผู้ตรวจ และประสบการณ์ของผู้ตรวจ แตกต่างกัน

๑. ๕

บทคัดย่อ

ของ

อูษณีย์ บัวศิริพันธุ์

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา

มีนาคม 2543

อุษณีย์ บัวศิริพันธ์. (2543). การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจ จำนวนผู้ตรวจ และประสบการณ์ของผู้ตรวจ แตกต่างกัน. ปรินซ์นิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. คณะกรรมการควบคุม : รองศาสตราจารย์ อังคณา สายยศ , รองศาสตราจารย์ชูศรี วงศ์รัตน์

การศึกษาครั้งนี้มีจุดมุ่งหมายเพื่อเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนต่างกัน คือ การตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเองและการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีด และมีจำนวนผู้ตรวจต่างกัน คือ 3, 4, 5 และ 6 คน และผู้ตรวจมีประสบการณ์ต่างกัน คือ ผู้ตรวจที่มีประสบการณ์มากกว่า 5 ปีและผู้ตรวจที่มีประสบการณ์ไม่เกิน 5 ปี กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2540 ของโรงเรียนในสังกัดกรมสามัญศึกษา จังหวัดจันทบุรี จำนวน 60 คน ซึ่งได้มาโดยการสุ่มแบบหลายขั้นตอน

ผลการวิจัยพบว่า

1. ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อมีจำนวนผู้ตรวจเท่ากัน มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01
2. ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อผู้ตรวจมีประสบการณ์เหมือนกัน มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ
3. ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีจำนวนผู้ตรวจต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01
4. ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่ผู้ตรวจมีประสบการณ์ต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

A COMPARISON OF GENERALIZABILITY COEFFICIENTS OF
MATHEMATICS TESTS WITH DIFFERENT SCORING METHODS,
NUMBERS OF RATERS AND EXPERIENCE OF RATERS

AN ABSTRACT
BY
USANEE BOWSIRIPON

Presented in partial fulfillment of the requirements
For the Master of Education degree in Educational Measurement
at Srinakharinwirot University
March 2000

Usanee Bowsiripon. (2000). *A Comparison of Generalizability Coefficients of Mathematics Tests with Different Scoring Methods, Numbers of Raters and Experience of Raters*. Master thesis, M.Ed. (Educational Measurement). Bangkok: Graduate School, Srinakharinwirot University.
Advisor Committee: Assoc.Prof. Aungkana Saiyos, Assoc.Prof. Chusri Wongrattana.

The purpose of the study was to compare the generalizability coefficients of mathematics tests with different holistic scoring methods : scoring by judgement of rater and rubric scoring, different numbers of raters : three raters, four raters, five raters, and six raters and different experience of raters : more than five years and five years or less. The sample of this study were 60 Mathayomsuksa 1 students of academic year 1997 in the general education school in Chantaburi were selected by the Multi - stage random sampling technique.

The findings were summarized as follows :

1. The generalizability coefficient of mathematics tests with different scoring methods when the same numbers of raters were significantly different at .01 level.
2. The generalizability coefficient of mathematics tests with different scoring methods when the same experience of raters were not significantly different.
3. The generalizability coefficient of mathematics tests with different numbers of raters when the same scoring methods were significantly different at .01 level.
4. The generalizability coefficient of mathematics tests with different experience of raters when the same scoring methods were significantly different at .01 level.

ปริญญานิพนธ์
เรื่อง

การเปรียบเทียบค่าสัมประสิทธิ์การสรุปร่างของแบบทดสอบวิชาคณิตศาสตร์
ที่มีวิธีการตรวจ จำนวนผู้ตรวจ และประสบการณ์ของผู้ตรวจ แตกต่างกัน

ของ
อุษณีย์ บัวศิริพันธุ์

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตร
ปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา
ของมหาวิทยาลัยศรีนครินทรวิโรฒ

..... คณบดีบัณฑิตวิทยาลัย
(ศาสตราจารย์ ดร.เสริมศักดิ์ วิชาลาภรณ์)
วันที่ 1 เดือน สิงหาคม พ.ศ. 2543

คณะกรรมการสอบปริญญานิพนธ์

..... ประธาน
(รองศาสตราจารย์ อังคณา สายยศ)

..... กรรมการ
(รองศาสตราจารย์ ชูศรี วงศ์รัตนะ)

..... กรรมการที่แต่งตั้งเพิ่มเติม
(รองศาสตราจารย์ วัณญา วิชาลาภรณ์)

..... กรรมการที่แต่งตั้งเพิ่มเติม
(ผู้ช่วยศาสตราจารย์ ระวีวรรณ พันธุ์พานิช)

ปริญญานิพนธ์ฉบับนี้ได้รับทุนสนับสนุนการวิจัยจาก
กองทุนพัฒนามหาวิทยาลัย

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จลงได้ด้วยความเมตตาจากรองศาสตราจารย์ อังคณา สายยศ และรองศาสตราจารย์ ชูศรี วงศ์รัตนะ ที่กรุณาดูแลเอาใจใส่อย่างยิ่งในการให้คำปรึกษา แนะนำ แก้ไขข้อบกพร่องต่างๆ รวมทั้งท่านรองศาสตราจารย์ วิทยุญา วิชาลาภรณ์ และผู้ช่วยศาสตราจารย์ ระวีวรรณ พันธุ์พานิช กรรมการแต่งตั้งเพิ่มเติมที่ได้ให้ข้อแนะนำ ทำให้ปริญญานิพนธ์ฉบับนี้มีความสมบูรณ์ยิ่งขึ้น ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง

ในการตรวจสอบและพิจารณาเครื่องมือที่ใช้ในการวิจัย ได้รับความเอื้อเฟื้อจากรองศาสตราจารย์ ชูศรี วงศ์รัตนะ, อาจารย์ธนาอนุสรณ์ สรณะ, อาจารย์นริศ สวัสดิ์, อาจารย์พัชรา ทศนวิจิตรวงศ์, อาจารย์พรศิริ กองพล และอาจารย์อรอนงค์ ผลทวีวัฒน์ชัย ที่ให้คำแนะนำ แก้ไขข้อบกพร่องของแบบทดสอบ ซึ่งผู้วิจัยขอขอบพระคุณเป็นอย่างสูง

การเก็บรวบรวมข้อมูลได้รับความร่วมมือจากผู้บริหาร คณะครู-อาจารย์ และนักเรียนของโรงเรียนในสังกัดกรมสามัญศึกษา จังหวัดจันทบุรี ที่เป็นกลุ่มตัวอย่าง ที่กรุณาให้ความสะดวกในการเก็บรวบรวมข้อมูล และตั้งใจตอบแบบทดสอบเป็นอย่างดี สำหรับการตรวจแบบทดสอบได้รับความร่วมมือจากอาจารย์ผู้สอนวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 1 ในจังหวัดจันทบุรี ซึ่งผู้วิจัยรู้สึกซาบซึ้งในความกรุณาที่ทุกท่านสละเวลาและให้ความช่วยเหลือเป็นอย่างดี

และท้ายที่สุดนี้ผู้วิจัยขอกราบขอบพระคุณทุกท่านที่ไม่ได้กล่าวนามมา ณ ที่นี้ ที่สนับสนุน ให้กำลังใจ และให้ความช่วยเหลือในทุกๆ ด้าน ทำให้ปริญญานิพนธ์ฉบับนี้สำเร็จลุล่วงไปด้วยดี คุณค่าและประโยชน์ของปริญญานิพนธ์ฉบับนี้ ขอมอบเป็นเครื่องบูชาพระคุณของบิดา มารดา ครู อาจารย์ ตลอดจนผู้มีพระคุณของผู้วิจัยทุกท่าน ที่ช่วยให้ผู้วิจัยประสบความสำเร็จในการศึกษา

อุษณีย์ บัวศิริพันธุ์

สารบัญ

บทที่	หน้า
1 บทนำ.....	1
ภูมิหลัง.....	1
จุดมุ่งหมายของการศึกษาค้นคว้า.....	3
ความสำคัญของการศึกษาค้นคว้า.....	3
ขอบเขตของการศึกษาค้นคว้า.....	3
นิยามศัพท์เฉพาะ.....	4
2 เอกสารและงานวิจัยที่เกี่ยวข้อง.....	7
ความเป็นมาของทฤษฎีการสรุปอ้างอิง.....	7
โครงสร้างของทฤษฎีการสรุปอ้างอิง.....	8
หลักพื้นฐานของทฤษฎีการสรุปอ้างอิง.....	10
การประเมินผลจากผลงานของนักเรียน.....	17
คำถามแบบปลายเปิด (Open-ended Question).....	17
วิธีการตรวจให้คะแนนคำถามแบบปลายเปิด.....	20
เกณฑ์การให้คะแนนแบบรูบริค (Rubric scoring).....	24
งานวิจัยที่เกี่ยวข้อง.....	28
3 วิธีดำเนินการศึกษาค้นคว้า.....	32
ประชากร.....	32
กลุ่มตัวอย่าง.....	32
เครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	33
วิธีดำเนินการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	33
ลักษณะของเครื่องมือที่ใช้ในการวิจัย.....	37
วิธีดำเนินการเก็บรวบรวมข้อมูล.....	37
สถิติที่ใช้ในการวิเคราะห์ข้อมูล.....	39
4 ผลการวิเคราะห์ข้อมูล.....	43
สัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูล.....	43
การเสนอผลวิเคราะห์ข้อมูล.....	43
ผลการวิเคราะห์ข้อมูล.....	44
5 สรุป อภิปรายผล และข้อเสนอแนะ.....	51
จุดมุ่งหมายของการศึกษาค้นคว้า.....	51
กลุ่มตัวอย่าง.....	51
เครื่องมือที่ใช้ในการศึกษาค้นคว้า.....	51

สารบัญ (ต่อ)

บทที่	หน้า
5(ต่อ) วิธีดำเนินการเก็บรวบรวมข้อมูล.....	52
การเสนอผลการวิเคราะห์ข้อมูล.....	52
สรุปผลการวิเคราะห์ข้อมูล.....	53
อภิปรายผล.....	55
ข้อเสนอแนะ.....	56
บรรณานุกรม.....	58
ภาคผนวก.....	62
ประวัติย่อของผู้วิจัย.....	88

บัญชีตาราง

ตาราง		หน้า
1	การประมาณค่าความแปรปรวนใหม่ของแบบแผน $p \times I \times R$	15
2	จำนวนนักเรียนที่ใช้เป็นกลุ่มตัวอย่างในการเก็บรวบรวมข้อมูล.....	32
3	ค่าสถิติพื้นฐานของแบบทดสอบวิชาคณิตศาสตร์ จำแนกตามวิธีการตรวจให้คะแนน และ ประสบการณ์ของผู้ตรวจ.....	44
4	ผลการทดสอบความแตกต่างของค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อจำนวนผู้ตรวจเหมือนกัน.....	45
5	ผลการทดสอบความแตกต่างของค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อผู้ตรวจมีประสบการณ์เหมือนกัน.....	46
6	ผลการทดสอบความแตกต่างของค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีจำนวนผู้ตรวจต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน.....	47
7	ผลการทดสอบความแตกต่างของค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีจำนวนผู้ตรวจต่างกันเป็นรายคู่ เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน.....	48
8	ผลการทดสอบความแตกต่างของค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่ผู้ตรวจมีประสบการณ์ต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน.....	50
9	ดัชนีค่าความง่าย (P_E) และดัชนีค่าอำนาจจำแนก (D) ของแบบทดสอบวิชาคณิตศาสตร์เป็น รายข้อ และผลการวิเคราะห์ค่าสัมประสิทธิ์ความสอดคล้องภายในของผู้ตรวจ (W).....	64
10	การประมาณค่าความแปรปรวนขององค์ประกอบ 3 องค์ประกอบ ที่มีผลต่อค่าสัมประสิทธิ์การ สรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีการตรวจให้คะแนนแบบประเมินรวมโดย ใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง.....	83
11	การประมาณค่าความแปรปรวนขององค์ประกอบ 3 องค์ประกอบ ที่มีผลต่อค่าสัมประสิทธิ์การ สรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีการตรวจให้คะแนนแบบประเมินรวมโดย ใช้เกณฑ์การให้คะแนนแบบรูบริค.....	83
12	การประมาณค่าความแปรปรวนขององค์ประกอบ 3 องค์ประกอบ ที่มีผลต่อค่าสัมประสิทธิ์การ สรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีการตรวจให้คะแนนแบบประเมินรวมโดย ใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง และผู้ตรวจมีประสบการณ์มากกว่า 5 ปี.....	84
13	การประมาณค่าความแปรปรวนขององค์ประกอบ 3 องค์ประกอบ ที่มีผลต่อค่าสัมประสิทธิ์การ สรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีการตรวจให้คะแนนแบบประเมินรวมโดย ใช้เกณฑ์การให้คะแนนแบบรูบริค และผู้ตรวจมีประสบการณ์มากกว่า 5 ปี.....	84
14	การประมาณค่าความแปรปรวนขององค์ประกอบ 3 องค์ประกอบ ที่มีผลต่อค่าสัมประสิทธิ์การ สรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีการตรวจให้คะแนนแบบประเมินรวมโดย ใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง และผู้ตรวจมีประสบการณ์ไม่เกิน 5 ปี.....	85
15	การประมาณค่าความแปรปรวนขององค์ประกอบ 3 องค์ประกอบ ที่มีผลต่อค่าสัมประสิทธิ์การ สรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีการตรวจให้คะแนนแบบประเมินรวมโดย ใช้เกณฑ์การให้คะแนนแบบรูบริค และผู้ตรวจมีประสบการณ์ไม่เกิน 5 ปี.....	85

บัญชีภาพประกอบ

ภาพประกอบ	หน้า
1 แผนภาพ Venn แสดงองค์ประกอบความแปรปรวนของรูปแบบ $p \times i \times r$ เมื่อองค์ประกอบที่ศึกษาทั้งหมดเป็นองค์ประกอบสุ่ม.....	13
2 แผนภาพ Venn แสดงองค์ประกอบความแปรปรวนของรูปแบบ $p \times i \times r$ เมื่อ p และ i เป็นองค์ประกอบสุ่ม ส่วน r เป็นองค์ประกอบกำหนด.....	14
3 แผนภาพแสดงการเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีผู้ตรวจ 3 ถึง 6 คน และมีวิธีการตรวจให้คะแนน 2 วิธี.....	49

บทที่ 1

บทนำ

ภูมิหลัง

จุดหมายปลายทางที่สำคัญอย่างหนึ่งของการศึกษาคือ พยายามวัดว่าเด็กสามารถนำความรู้ที่เรียนแล้วไปปฏิบัติในชีวิตจริง หรือไปแก้ปัญหาจริง ๆ เหล่านั้นได้หรือไม่ (ชวาล แพรัตกุล. 2518 : 106-107) การวัดหรือการประเมินผลความสามารถของนักเรียนในลักษณะนี้สามารถทำได้หลายวิธี วิธีหนึ่งคือการประเมินผลงานของนักเรียน (Performance Assessment) โดยให้นักเรียนแสดงทักษะโดยทำงานหรือสร้างงาน เช่น การทดลองในห้องปฏิบัติการ การเขียนรายงาน การพูด ร้องเพลง พิมพ์ดีด ประดิษฐ์หรือซ่อมแซมชิ้นงาน รวมถึงการทำข้อสอบโดยการเขียนตอบ ถ้างานและระบบการให้คะแนนเหมือนการทดสอบในชีวิตจริง (Courtney. 1996 : 40-41) การประเมินผลงานของนักเรียนด้วยคำถามแบบปลายเปิด (Open-ended Questions) เป็นเทคนิคการประเมินแบบหนึ่งที่พัฒนาขึ้นเพื่อสนับสนุนและปรับปรุงการประเมินผลงานของนักเรียน ในลักษณะการทำข้อสอบที่นักเรียนต้องสร้างคำตอบจากคำถามหรือกิจกรรมที่ให้นักเรียนแสดงความคิดเห็น เหตุผลและกลยุทธ์กระบวนการต่างๆ ในการหาคำตอบซึ่งมีแนวทางปฏิบัติที่เป็นไปได้หลายทางและอาจได้คำตอบที่แตกต่างกันแต่เป็นคำตอบที่ถูกต้องเหมือนกัน จะเห็นว่าคำถามแบบปลายเปิดนี้ช่วยส่งเสริมความสามารถของนักเรียนในด้านความรู้ ความคิดสร้างสรรค์ การประยุกต์ใช้ความรู้ในสถานการณ์ที่ไม่คุ้นเคย ซึ่งนักเรียนจะแสดงออกมาในรูปของการอธิบายโดยการเขียน ดังนั้นการสื่อสารหรือการสื่อความหมายในด้านภาษาก็เป็นอีกส่วนหนึ่งที่สำคัญสำหรับการตอบคำถามแบบปลายเปิด

ในวิชาคณิตศาสตร์ คำถามแบบปลายเปิดถูกนำมาใช้กับสถานการณ์ปัญหาที่สร้างเป็นคำถามให้นักเรียนตอบด้วยการเขียน ลักษณะของปัญหาจะออกแบบเพื่อค้นหาความสามารถของนักเรียนในด้านการคิด การแก้ปัญหา และการสื่อสาร เกี่ยวกับสถานการณ์ที่ได้รับ (Mary Ann Heglie-King. 1995) ในงานวิจัยของซูซาน เลน และคณะ (Suzanne Lane, et al. 1996 : 73) ได้พัฒนาเครื่องมือที่ใช้ประเมินความรู้ QUASAR ให้ชื่อว่า QCAI (QUASAR Cognitive Assessment Instrument) ของซิลเวอร์ (Silver. 1991) มาใช้เป็นเครื่องมือที่ออกแบบมาเพื่อวัดผลสัมฤทธิ์และการพัฒนาในวิชาคณิตศาสตร์ ทั้งยังช่วยชี้แนะความสำเร็จในการเรียนการสอนของโรงเรียนระดับมัธยมซึ่งส่งเสริมให้นักเรียนมีทักษะการคิดและเหตุผลทางคณิตศาสตร์ เครื่องมือแบบ QCAI นี้ประเมินผลงานของนักเรียนด้วยข้อสอบแบบปลายเปิดที่ประกอบด้วย การแก้ปัญหาคณิตศาสตร์ คณิตศาสตร์เหตุผล และการสื่อสารทางคณิตศาสตร์ ในการศึกษาครั้งนี้ผู้วิจัยจึงสนใจที่จะศึกษาการประเมินผลงานของนักเรียนในวิชาคณิตศาสตร์ ด้วยแบบทดสอบวิชาคณิตศาสตร์ที่มีลักษณะคำถามแบบปลายเปิด

เนื่องจากแบบทดสอบวิชาคณิตศาสตร์ลักษณะนี้ เป็นแบบทดสอบที่นักเรียนต้องใช้ในการเขียนตอบ เช่นเดียวกับแบบทดสอบความเรียง (Essey Test) ดังนั้นจึงมีปัญหาในด้านการวัดผลที่สำคัญคือ แบบทดสอบมักจะมีคามเชื่อมั่นต่ำ ปัญหานี้แก้วัดผลการศึกษาหลายท่านได้ให้ข้อเสนอแนะวิธีการปรับปรุงเพื่อให้ผลการวัดมีความถูกต้องแม่นยำไว้ เช่น เมอเรนส์ และลีแมน (Mehrens and Lehmann. 1972 : 228) กล่าวว่า เพื่อให้การตรวจมีความเชื่อมั่นเพิ่มขึ้น ผู้ตรวจควรปฏิบัติดังนี้อย่างเคร่งครัด คือ 1) ใช้วิธีการตรวจที่เหมาะสม

สามารถจัดทำได้ 2) การตรวจต้องมุ่งสนใจอยู่เฉพาะประเด็นที่เกี่ยวข้องกับผลการเรียนที่มุ่งวัดเท่านั้น 3) ระมัดระวังอย่าให้มีส่วนตัวของผู้ตรวจมีอิทธิพลต่อคะแนน 4) ต้องใช้เกณฑ์การตรวจกับนักเรียนทุกคนอย่างคงเส้นคงวา เพื่อให้ผลการตรวจมีความสอดคล้องภายในผู้ตรวจสูงขึ้นจำเป็นต้องมีการออกแบบทั้งการให้คะแนนแบบรูบริก (Rubric) และการฝึกหรืออบรมผู้ตรวจ (Zuzanne Lane, et al. 1996 : 72 ; citing Linn. 1993) นอกจากนี้การตรวจควรตรวจที่ละข้อจนครบทุกคน ถ้าทำได้ควรตรวจหลายครั้งหรือตรวจหลายคน (Multiple Rating) คือ ให้ผู้ตรวจคนเดียวตรวจหลายครั้งหรือในการตรวจหนึ่งครั้งให้ผู้ตรวจหลายคน แล้วใช้คะแนนเฉลี่ยหรือคะแนนรวมแทนคะแนนความสามารถของผู้สอบ จะให้ผลการวัดที่มีความเชื่อมั่นสูงขึ้นได้ (Ebel. 1972 : 152; Mehrens and Lehmann. 1973 : 234-236; Gronlund. 1977 : 79; Bergman. 1981 : 130-131) และผู้ตรวจที่เชื่อถือได้ต้องมาจากการคัดเลือก การฝึกฝน การแนะนำการให้คะแนนที่ถูกต้อง และต้องมีการตรวจสอบการปฏิบัติงานผู้ประเมินเป็นระยะๆ (Zuzanne Lane, et al. 1996 : 72 ; citing Mehrens. 1992 : 8)

การประเมินผลคำถามแบบปลายเปิดสามารถทำได้ 2 วิธี (Stenmark. 1991 : 20) คือ 1) วิธีวิเคราะห์ (Analytical Method หรือ Point Method) เป็นการวิเคราะห์ด้วยประเด็นที่มีลักษณะแตกต่างกันของคำตอบ โดยแยกเป็นประเด็นตามที่กำหนด 2) วิธีการประเมินรวม (Holistic Assessment) ซึ่งผู้ประเมินหรือผู้ตรวจจะพิจารณาคำตอบโดยรวมมากกว่าตรวจสอบรายละเอียดประเด็นย่อยๆ โดยเฉพาะการให้คะแนนจะขึ้นอยู่กับคุณภาพของประเด็นหลัก โดยผู้ตรวจอ่านคำตอบแล้วใช้เกณฑ์เดียวกันในการตรวจให้คะแนน (Mehren and Lehmann. 1984 : 233) ซึ่งเกณฑ์ที่ใช้ในการตรวจให้คะแนนนี้ผู้ตรวจสามารถประยุกต์ใช้เกณฑ์การตรวจให้คะแนนแบบรูบริก (Rubric) สำหรับกำหนดคะแนนในการตอบของนักเรียนได้ เพื่อควบคุมความคลาดเคลื่อนจากผู้ตรวจ รวมทั้งควรมีการอบรมฝึกหัดผู้ตรวจอย่างเหมาะสมด้วย (Zuzanne Lane, et al. 1996 : 72 ; citing Dunbar, Koretz & Hoover. 1991; Messick. 1989)

สำหรับทฤษฎีที่ใช้ในการประมาณค่าความเชื่อมั่นของแบบทดสอบเป็นอีกปัญหาหนึ่งที่มีความสำคัญ ทฤษฎีการวัดแบบดั้งเดิม (Classical Test Theory) จะประมาณค่าความเชื่อมั่นโดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจริงกับคะแนนของผู้สอบ ภายใต้เงื่อนไขการวัดที่เท่าเทียมกันหรือการวัดนั้นต้องมีคุณสมบัติคู่ขนานกันกล่าวคือ เป็นการวัดสองครั้งที่อิสระต่อกันแต่ผลของการวัดจะมีคุณสมบัติความเท่าเทียมกัน คือ มีค่าเฉลี่ยเท่ากัน ความแปรปรวนเท่ากัน และค่าสหสัมพันธ์ระหว่างแบบทดสอบเท่ากัน (Gulliksen. 1950 : 173) ซึ่งในทางปฏิบัติ การวัดและประเมินผลโดยทั่วไปไม่สามารถจัดหาเครื่องมือที่มีคุณสมบัติความเท่าเทียมอย่างสมบูรณ์ได้ โดยเฉพาะถ้าเป็นการวัดผลที่ต้องคำนึงถึงผู้ตรวจด้วย การใช้ทฤษฎีการวัดแบบดั้งเดิมจึงไม่เหมาะสม จากข้อจำกัดดังกล่าว ครอนบัค และคณะ (Cronbach, et al. 1972) ได้พัฒนาทฤษฎีการสรุปอ้างอิง (Generalizability Theory) ขึ้น โดยทฤษฎีจะไม่ยึดข้อตกลงเกี่ยวกับคุณสมบัติเท่าเทียมเป็นทฤษฎีที่ปรับปรุงและพัฒนามาจากทฤษฎีการวัดแบบดั้งเดิม และใช้วิธีการวิเคราะห์ความแปรปรวนในการหาค่าความคลาดเคลื่อนจากหลายๆ แหล่ง (Brennan. 1983 : 1)

จากที่กล่าวมาข้างต้นพบว่า ตัวประกอบที่มีผลต่อค่าความเชื่อมั่นของแบบทดสอบวิชาคณิตศาสตร์ที่มีลักษณะคำถามแบบปลายเปิด ประกอบด้วย แบบทดสอบ ผู้ตรวจ วิธีการตรวจ และทฤษฎีที่ใช้ในการประมาณค่าความเชื่อมั่น ซึ่งการควบคุมแหล่งความคลาดเคลื่อนของตัวประกอบที่เกี่ยวข้อง จะทำให้ผลการวัดมีความถูกต้องแม่นยำเพิ่มขึ้น ดังนั้นผู้วิจัยจึงสนใจศึกษาเปรียบเทียบผลของการควบคุมแหล่งความคลาดเคลื่อนของตัวประกอบที่เกี่ยวข้อง และเนื่องจากทฤษฎีการสรุปอ้างอิงเป็นทฤษฎีที่เหมาะสมในการประมาณค่าความเชื่อมั่นของแบบทดสอบ ผู้วิจัยจึงเลือกใช้ทฤษฎีการสรุปอ้างอิงในการศึกษาดังนี้

จุดมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อมีจำนวนผู้ตรวจเท่ากัน
2. เพื่อเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อผู้ตรวจมีประสบการณ์เหมือนกัน
3. เพื่อเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีจำนวนผู้ตรวจต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน
4. เพื่อเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่ผู้ตรวจมีประสบการณ์ต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน

ความสำคัญของการศึกษาค้นคว้า

การศึกษาค้นคว้าครั้งนี้เป็นการศึกษาเปรียบเทียบการประเมินผลงานของนักเรียนในวิชาคณิตศาสตร์ด้วยแบบทดสอบวิชาคณิตศาสตร์ที่มีลักษณะคำถามแบบปลายเปิด ที่มีวิธีการตรวจให้คะแนน จำนวนผู้ตรวจ และประสบการณ์ของผู้ตรวจแตกต่างกัน โดยผลที่ได้จากการศึกษาค้นคว้าครั้งนี้จะเป็นแนวทางในการเลือกใช้รูปแบบการวัด วิธีการตรวจ จำนวนข้อสอบ และจำนวนผู้ตรวจ รวมถึงประสบการณ์ของผู้ตรวจที่เหมาะสมและให้ค่าความเชื่อมั่นสูงสุด อันจะเป็นประโยชน์ต่อการวัดผลการศึกษาและเป็นการประเมินผลงานของนักเรียนในวิชาคณิตศาสตร์ ซึ่งจะช่วยให้ครู อาจารย์ และผู้ที่เกี่ยวข้อง ได้สนใจและเห็นคุณค่าของการประเมินผลลักษณะนี้ กล่าวคือ การประเมินผลที่มีวิธีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเองกับการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูบริค ซึ่งมีจำนวนผู้ตรวจแตกต่างกันตั้งแต่ 3 – 6 คน และผู้ตรวจมีประสบการณ์ในการสอนวิชาคณิตศาสตร์แตกต่างกันคือ มีประสบการณ์มากกว่า 5 ปี กับมีประสบการณ์ไม่เกิน 5 ปี

ขอบเขตของการศึกษาค้นคว้า

1. ประชากร
ประชากรที่ใช้ในการค้นคว้า เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ของโรงเรียนกรมสามัญศึกษา จังหวัดจันทบุรี ในปีการศึกษา 2540 จำนวน 4,436 คน
2. กลุ่มตัวอย่าง
กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้า เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดจันทบุรี ในปีการศึกษา 2540 จำนวน 60 คน ที่ได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi - Stage Random Sampling)
3. ตัวแปรที่ใช้ในการวิจัย
ตัวแปรอิสระ ประกอบด้วย

- 3.1 วิธีการตรวจให้คะแนน แบ่งเป็น 2 วิธี คือ
 - 3.1.1 การตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง
 - 3.1.2 การตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูบริก(Rubric)
- 3.2 จำนวนผู้ตรวจ มี 6 คน แบ่งเป็น 4 กรณี คือ
 - 3.2.1 ผู้ตรวจ 3 คน
 - 3.2.2 ผู้ตรวจ 4 คน
 - 3.2.3 ผู้ตรวจ 5 คน
 - 3.2.4 ผู้ตรวจ 6 คน
- 3.3 ประสิทธิภาพของผู้ตรวจ แบ่งเป็น 2 กรณี คือ
 - 3.3.1 ผู้ตรวจที่มีประสิทธิภาพมากกว่า 5 ปี
 - 3.3.2 ผู้ตรวจที่มีประสิทธิภาพไม่เกิน 5 ปี
- 3.4 ตัวแปรตาม คือ ค่าสัมประสิทธิ์การสรุปอ้างอิง

นิยามศัพท์เฉพาะ

1. สัมประสิทธิ์การสรุปอ้างอิง (Generalizability Coefficient : ρ^2) หมายถึง คุณภาพด้านความเชื่อมั่นของแบบทดสอบโดยใช้ทฤษฎีการสรุปอ้างอิง (Generalizability Theory) ซึ่งคำนวณจากกระบวนการวิเคราะห์ความแปรปรวนด้วยโปรแกรมสำเร็จรูป GENOVA (Crick & Brennan. 1983)

2. ทฤษฎีการสรุปอ้างอิง (Generalizability Theory) ในการศึกษาครั้งนี้ใช้ทฤษฎีการสรุปอ้างอิงตามแนวคิดของเบรนนอน (Brennan. 1983) โดยใช้รูปแบบความสัมพันธ์ระหว่างองค์ประกอบแบบไขว้ (Crossed) คือ $p \times i \times r$ โดยมีองค์ประกอบ (Facets) ที่ศึกษา คือ ผู้ตอบ (p) ข้อสอบ (i) และผู้ตรวจ (r)

3. รูปแบบการวัด หมายถึง การกำหนดประเภทขององค์ประกอบแต่ละองค์ประกอบในการวัดว่าเป็นองค์ประกอบกลุ่มหรือองค์ประกอบกำหนด ในการศึกษาครั้งนี้กำหนดรูปแบบของการวัดเป็น 3 รูปแบบ คือ

รูปแบบที่ 1 ให้ผู้ตอบเป็นสิ่งที่ถูกวัดกำหนดให้เป็นองค์ประกอบกลุ่ม ให้ข้อสอบและผู้ตรวจเป็นเครื่องมือในการวัดกำหนดให้เป็นองค์ประกอบกลุ่มด้วย

รูปแบบที่ 2 ให้ผู้ตอบเป็นสิ่งที่ถูกวัดกำหนดให้เป็นองค์ประกอบกลุ่ม ให้ข้อสอบและผู้ตรวจเป็นเครื่องมือในการวัด โดยข้อสอบเป็นองค์ประกอบกำหนดแต่ผู้ตรวจเป็นองค์ประกอบกลุ่ม

รูปแบบที่ 3 ให้ผู้ตอบเป็นสิ่งที่ถูกวัดกำหนดให้เป็นองค์ประกอบกลุ่ม ให้ข้อสอบและผู้ตรวจเป็นเครื่องมือในการวัด โดยผู้ตรวจเป็นองค์ประกอบกำหนดแต่ข้อสอบเป็นองค์ประกอบกลุ่ม

4. องค์ประกอบ (Facets) หมายถึง ตัวประกอบ (Factor) หรือ ลักษณะร่วมกันของกลุ่มเงื่อนไขการวัดชนิดเดียวกัน ในการศึกษาวิจัยครั้งนี้ 3 ตัวประกอบ ได้แก่ ผู้ตอบ ข้อสอบ และผู้ตรวจ

5. เอกภพ (Universe) หมายถึง เงื่อนไขการวัดทั้งหมดของแต่ละองค์ประกอบ มีความหมายคล้ายกับคำว่า ประชากร (Population) แต่คำว่า เอกภพ ใช้กับองค์ประกอบที่เกี่ยวกับการวัด ส่วนคำว่า ประชากร ใช้กับองค์ประกอบที่เป็นสิ่งที่ถูกวัดเท่านั้น เอกภพในการศึกษาครั้งนี้ได้แก่ นักเรียนชั้นมัธยม

ศึกษาปีที่ 1 ของโรงเรียนในสังกัดกรมสามัญศึกษาจังหวัดจันทบุรี ปีการศึกษา 2540 ทุกคน และอาจารย์ที่สอนวิชาคณิตศาสตร์ระดับมัธยมศึกษาปีที่ 1 และข้อสอบปลายเปิดวิชาคณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 ทุกข้อ

เอกภพของการศึกษาค้นคว้าครั้งนี้ แบ่งเป็น 2 ชนิด ตามแนวทฤษฎีการสรุปอ้างอิง ได้แก่

5.1 เอกภพของการสังเกตที่ยอมรับได้ (Universe of Admissible Observation) หมายถึง กลุ่มเงื่อนไขของการวัดที่สามารถวัดหรือสังเกตได้ในแต่ละองค์ประกอบ ได้แก่ ผู้สอบทุกคน ข้อสอบทุกข้อ และผู้ตรวจทุกคน

5.2 เอกภพของการสรุปอ้างอิง (Universe of Generalizability) หมายถึง กลุ่มเงื่อนไขการวัดทั้งหมดขององค์ประกอบที่ผู้วิจัยต้องการสรุปอ้างอิงผลของการวัด จากกลุ่มตัวอย่างของเงื่อนไขการวัดเหล่านี้ไปยังกลุ่มเงื่อนไขการวัดทั้งหมดขององค์ประกอบนั้นๆ ในการศึกษาครั้งนี้ต้องการอ้างอิงไปยังเอกภพที่มีขอบเขตต่างกัน 3 ระดับ ได้แก่ 1) เอกภพของข้อสอบและเอกภพของผู้ตรวจพร้อมๆ กัน 2) เอกภพของข้อสอบเพียงอย่างเดียว 3) เอกภพของผู้ตรวจอย่างเดียว

6. เงื่อนไขการวัด (Condition) หมายถึง ขนาดของกลุ่มตัวอย่างขององค์ประกอบที่ใช้ในการวัด เช่น จำนวนข้อสอบ หรือจำนวนผู้ตรวจที่สามารถสุ่มมาใช้ในการประเมินผลแบบทดสอบแบบปลายเปิด

7. แบบทดสอบวิชาคณิตศาสตร์ หมายถึง แบบทดสอบที่ใช้ประเมินผลงานของนักเรียนวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 1 ที่ได้จากการทำแบบทดสอบที่มีลักษณะคำถามแบบปลายเปิด ซึ่งแบบทดสอบจะประกอบด้วยข้อคำถามที่วัดความสามารถของนักเรียน 3 ด้าน ดังนี้

7.1 การแก้ปัญหาในวิชาคณิตศาสตร์ หมายถึง ความสามารถในการนำความรู้และความเข้าใจในวิชาคณิตศาสตร์ที่มีอยู่ มาประยุกต์ใช้ในการคิดและหาวิธีการแก้ปัญหาหรือสถานการณ์ต่างๆ ที่กำหนดให้ได้สำเร็จ

7.2 คณิตศาสตร์เหตุผล หมายถึง ความสามารถในการนำความรู้และความเข้าใจในวิชาคณิตศาสตร์ที่มีอยู่ มาใช้ประยุกต์ในการคิดอย่างเป็นเหตุเป็นผล หรือการคิดอย่างมีเหตุผลว่า สถานการณ์ที่กำหนดให้เกิดขึ้นได้หรือไม่ อย่างไร และเพราะเหตุใด

7.3 การสื่อสารทางคณิตศาสตร์ หมายถึง ความสามารถในการใช้ทักษะการเขียน การใช้ภาษา และสัญลักษณ์ทางคณิตศาสตร์ แสดงความคิดออกมาอย่างมีระบบ และสามารถสื่อสารความคิดของตนกับผู้อื่นได้อย่างมีประสิทธิภาพ

8. คำถามแบบปลายเปิด (Open-ended Questions) หมายถึง คำถามที่ให้นักเรียนแสดงความคิดเห็น เหตุผล และกลยุทธ์กระบวนการต่างๆ ในการหาคำตอบที่เป็นไปได้หลายทาง ซึ่งนักเรียนแต่ละคนอาจเลือกแนวทางปฏิบัติที่แตกต่างกันและได้คำตอบที่แตกต่างกัน แต่เป็นคำตอบที่ถูกต้องเหมือนกัน

9. การประเมินรวม (Holistic) หมายถึง การประเมินที่ผู้ประเมินพิจารณาข้อสอบโดยรวมมากกว่าการวิเคราะห์รายละเอียดปลีกย่อย วิธีการตรวจจะพิจารณาข้อสอบทีละข้อแล้วจัดแบ่งข้อสอบออกเป็นกองๆ ตามระดับคุณภาพของข้อสอบ ในการประเมินครั้งนี้ใช้การตรวจด้วยวิธีประเมินรวมโดยวิธีการให้คะแนน 2 วิธี คือ

9.1 การตรวจโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง เป็นการตรวจให้คะแนนข้อสอบโดยการประเมินรวม ที่ใช้เกณฑ์การให้คะแนนที่ผู้ตรวจแต่ละคนกำหนดขึ้นเอง แต่มีการกำหนดคะแนนไว้ 5 ระดับ คือ 0-4 คะแนน กล่าวคือ ผู้ตรวจจะต้องตรวจข้อสอบของนักเรียนทุกคนทีละข้อแล้วแบ่งข้อสอบออก

เป็น 5 กอง ตามระดับคุณภาพที่ผู้ตรวจเห็นว่าเหมาะสม โดยผู้วิจัยเพียงแต่ชี้แจงวิธีการประเมินรวมให้ผู้ตรวจทุกคนเข้าใจตรงกันเท่านั้น

9.2 การตรวจโดยใช้เกณฑ์การให้คะแนนแบบรูบริก (Rubric) เป็นการตรวจให้คะแนนข้อสอบโดยการประเมินรวม ที่มีระบบการให้คะแนนพื้นฐานของเกณฑ์ที่กำหนดขึ้น มีคำแนะนำการให้คะแนนและคำตอบตัวแบบของข้อสอบแต่ละข้อกำหนดไว้ให้ซึ่งการตั้งเกณฑ์กำหนดไว้โดยผู้เชี่ยวชาญแล้วว่า คำตอบของนักเรียนควรอยู่ในระดับคะแนนเท่าใด โดยมีการกำหนดระดับคะแนนไว้ 5 ระดับคือ 0-4 คะแนน ผู้ตรวจจะต้องตรวจให้คะแนนตามเกณฑ์การให้คะแนนแบบรูบริกที่กำหนดให้

10. จำนวนผู้ตรวจ หมายถึง จำนวนผู้ตรวจแบบทดสอบวิชาคณิตศาสตร์ มี 6 คน แบ่งออกเป็น 4 กรณี คือ

10.1 ผู้ตรวจ 3 คน

10.2 ผู้ตรวจ 4 คน

10.3 ผู้ตรวจ 5 คน

10.4 ผู้ตรวจ 6 คน

11. ประสบการณ์ของผู้ตรวจ หมายถึง ประสบการณ์ในการสอนวิชาคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 1 ของอาจารย์ผู้ตรวจแบบทดสอบวิชาคณิตศาสตร์ ซึ่งในการศึกษาค้างนี้แบ่งออกเป็น 2 กรณี คือ

11.1 ผู้ตรวจที่มีประสบการณ์มากกว่า 5 ปี

11.2 ผู้ตรวจที่มีประสบการณ์ไม่เกิน 5 ปี

12. ผู้เชี่ยวชาญ หมายถึง ผู้ที่มีวุฒิการศึกษาย่างน้อยปริญญาโททางการวัดผลทางการศึกษาและมีความชำนาญในสาขาวิชาต่อไปนี้

12.1 สาขาวิชาคณิตศาสตร์ จำนวน 3 ท่าน

12.2 สาขาวิชาภาษาไทย จำนวน 2 ท่าน

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องกับการศึกษาค้นคว้าครั้งนี้ ผู้วิจัยได้เสนอรายละเอียดต่างๆ ไว้ดังนี้

1. ความเป็นมาของทฤษฎีการสรุปอ้างอิง
2. โครงสร้างของทฤษฎีการสรุปอ้างอิง
3. หลักพื้นฐานของทฤษฎีการสรุปอ้างอิง
4. การประเมินผลจากผลงานของนักเรียน
5. คำถามแบบปลายเปิด
6. วิธีการตรวจให้คะแนนแบบปลายเปิด
7. เกณฑ์การให้คะแนนแบบรูบริก (Rubric)
8. งานวิจัยที่เกี่ยวข้อง

1. ความเป็นมาของทฤษฎีการสรุปอ้างอิง

ในการวัดทางจิตวิทยาและการศึกษา โดยส่วนใหญ่การศึกษาความเชื่อมั่นและวิธีการประเมินค่าความเชื่อมั่นของแบบทดสอบใช้หลักของทฤษฎีการวัดแบบดั้งเดิม (Classical Test Theory) ที่ว่า คะแนนที่ได้จากการวัดที่สังเกตได้เกิดจากคะแนนจริงรวมกับคะแนนจากแหล่งความคลาดเคลื่อน ซึ่งเขียนเป็นสมการได้ดังนี้ (Gulliksen. 1950)

$$X_i = T_i + E_i$$

เมื่อ	X_i	แทน คะแนนที่สังเกตได้ของผู้สอบคนที่ i
	T_i	แทน คะแนนจริงของผู้สอบคนที่ i
	E_i	แทน คะแนนความคลาดเคลื่อนของผู้สอบคนที่ i

ตามทฤษฎีการวัดแบบดั้งเดิมนี้อธิบายว่าค่า E_i เป็นค่าเดียวกล่าวคือเป็นคะแนนรวมความคลาดเคลื่อนทุกแหล่งเข้าด้วยกัน ดังนั้นการใช้ทฤษฎีการวัดแบบดั้งเดิมจึงไม่สามารถแยกแยะความแตกต่างระหว่างแหล่งความคลาดเคลื่อนหลายๆ แหล่งได้ชัดเจน ดังนั้นฟิชเชอร์ (Fisher. 1925) จึงได้เสนอวิธีการวิเคราะห์ความแปรปรวน โดยใช้วิธีการทางสถิติแบบทดลองแฟกทอเรียล (Factorial Experiment) ซึ่งสามารถประมาณค่าความคลาดเคลื่อนจากแหล่งต่างๆ และหาค่าความเชื่อมั่นของแบบทดสอบ การใช้การวิเคราะห์ความแปรปรวนเพื่อหาค่าความเชื่อมั่นเริ่มเผยแพร่โดย เบิร์ต (Burt. 1936), ฮอยท์ (Hoyt. 1941) และเฟอร์กูสัน (Ferguson. 1941) ต่อมาได้มีผู้ศึกษาวิธีการนี้อีก เช่น อเล็กซานเดอร์ (Alexander. 1947), อีเบล (Ebel. 1951), ฟินเลย์สัน (Finlayson. 1951), เลิฟแลนด์ (Loveland. 1952), เบิร์ต (Burt. 1955)

รวมทั้งลินด์ควิสต์ (Lindquist. 1953) ในงานของเบิร์ตและลินด์ควิสต์นั้นมีความคาดหมายว่า มีการพัฒนา ทฤษฎีอ้างอิง หลังจากนั้นยังมีงานของเวบสเตอร์ (Webster. 1960) และงานของเมดเลย์และมิทเซล (Medley and Mitzel. 1963) ได้เสนอเกี่ยวกับทฤษฎีการสรุปอ้างอิงให้เห็นเด่นชัดมากขึ้น ซึ่งต่อมา ครอนบัค และคณะ (Cronbach, Rajaratnam and Gleser. 1963) ได้เสนอพื้นฐานการสร้างทฤษฎีการสรุปอ้างอิงไว้ และในปี 1972 ได้มีอธิบายอย่างกว้างขวางเกี่ยวกับทฤษฎีการสรุปอ้างอิง ในหนังสือของครอนบัคและคณะ (Cronbach, et al. 1972) ให้ชื่อหนังสือว่า ความเชื่อถือได้ของการวัดทางพฤติกรรม (The Dependability of Behavioral Measurements) หนังสือเล่มนี้มีวิธีการของทฤษฎีการสรุปอ้างอิงอย่างละเอียด ต่อมาเบรนนอน และเคน (Brennan and Kane. 1979) ได้ตรวจสอบเกี่ยวกับทฤษฎีนี้อีกครั้ง และมีการขยายให้กว้างขวาง มากขึ้นโดย เซฟเวลสัน และเวบบ์ (Brennan. 1983 : 1 ; citing Shavelson and Webb. 1981)

หลักการของทฤษฎีการสรุปอ้างอิง อธิบายคะแนนที่ได้จากการวัดที่สังเกตได้ด้วยสมการต่อไปนี้

$$X_i = T_i + E_{i1} + E_{i2} + \dots + E_{in}$$

เมื่อ X_i แทน คะแนนที่สังเกตได้ของผู้สอบคนที่ i

T_i แทน คะแนนจริงของผู้สอบคนที่ i

$E_{i1}, E_{i2}, \dots, E_{in}$ แทน คะแนนความคลาดเคลื่อนจากแหล่งต่างๆ ของผู้ที่สอบคนที่ i

2. โครงสร้างของทฤษฎีการสรุปอ้างอิง

การที่จะนำทฤษฎีการสรุปอ้างอิงไปใช้ในการหาค่าสัมประสิทธิ์การสรุปอ้างอิงหรือหาค่าความเชื่อมั่นของแบบทดสอบนั้น จะต้องคำนึงถึงความสามารถในการกำหนดเงื่อนไขต่างๆ ของการวัด ดังนั้นในการศึกษาทฤษฎีการสรุปอ้างอิงจะต้องมีความรู้เกี่ยวกับโครงสร้างของทฤษฎีการสรุปอ้างอิงในเรื่องต่างๆ ดังนี้ (Brennan. 1983 : 2-6)

1. องค์ประกอบ (Facets) หมายถึง กลุ่มเงื่อนไขในการวัดชนิดเดียวกัน เช่น องค์ประกอบของข้อสอบ องค์ประกอบของผู้ตรวจ เป็นต้น

2. เงื่อนไขของการวัด (Condition) หมายถึง ระดับขององค์ประกอบที่ทำให้ได้ค่าสังเกตแต่ละค่าในการวัดครั้งหนึ่งๆ อาจกล่าวได้ว่าเป็นขนาดของกลุ่มตัวอย่างขององค์ประกอบนั้น เช่น ข้อสอบแต่ละข้อถือว่าเป็นเงื่อนไขของการวัดครั้งหนึ่งๆ ขององค์ประกอบ

3. รูปแบบความสัมพันธ์ระหว่างองค์ประกอบ

ความสัมพันธ์ระหว่างองค์ประกอบเป็นสิ่งที่ชี้ถึงแหล่งความแปรปรวนของค่าที่วัดได้ในแบบจำลองการวิเคราะห์ความแปรปรวน ความสัมพันธ์ระหว่างองค์ประกอบมี 3 ลักษณะ คือ

3.1 ความสัมพันธ์แบบไขว้ (Crossed) เป็นความสัมพันธ์ในลักษณะที่แต่ละระดับของสิ่งที่ถูกวัดจะถูกวัดภายใต้เงื่อนไขเดียวกันทั้งหมด สัญลักษณ์ที่ใช้คือ "x" อ่านว่า crossed with เช่น รูปแบบความสัมพันธ์ระหว่างองค์ประกอบเป็น $p \times i \times r$ หมายถึง ผู้สอบ (p) ทำข้อสอบทุกข้อ (i) และผู้ตรวจ (r) ตรวจข้อสอบทุกข้อ

3.2 ความสัมพันธ์แบบแฝง (Nested) เป็นความสัมพันธ์ในลักษณะที่แต่ละระดับของสิ่งที่ถูกวัดจะถูกวัดภายใต้เงื่อนไขของการวัดที่แตกต่างกัน สัญลักษณ์ที่ใช้คือ ":" อ่านว่า nested within เช่น รูปแบบความสัมพันธ์ระหว่างองค์ประกอบเป็น $i:r$ หมายถึง ผู้ตรวจ (r) จะตรวจข้อสอบ (i) ต่างข้อกัน

3.3 ความสัมพันธ์แบบผสม (Confounded) เป็นความสัมพันธ์ที่มีทั้งความสัมพันธ์แบบไขว้และความสัมพันธ์แบบแฝงรวมกัน เช่น $p \times (i:r)$ หมายถึง ผู้สอบ (p) ทำข้อสอบ (i) ทุกข้อที่แฝงอยู่ในผู้ตรวจ (r) ต่างกัน

4. เอกภพ (Universe) หมายถึง เงื่อนไขของการวัดทั้งหมดของแต่ละองค์ประกอบ เช่น จำนวนข้อสอบทุกข้อ จำนวนผู้ตรวจทั้งหมด คำว่าเอกภพมีความหมายคล้ายกับคำว่าประชากร (Population) แต่คำว่าเอกภพใช้กับองค์ประกอบที่เกี่ยวข้องกับการวัด ส่วนคำว่าประชากรใช้กับองค์ประกอบที่ถูกวัด

คะแนนเอกภพ (Universe Scores) หมายถึง คะแนนของเงื่อนไขของการวัดทั้งหมดของแต่ละองค์ประกอบ ซึ่งคล้ายกับคะแนนจริง (Ideal Score) ของสิ่งที่จะวัดในทฤษฎีการวัดแบบมาตรฐานเดิม

5. เอกภพของการสังเกตที่ยอมรับได้ (Universe of Admissible Observation) หมายถึง กลุ่มเงื่อนไขของการวัดที่สามารถวัดหรือสังเกตได้ในแต่ละองค์ประกอบ ดังตัวอย่าง (Brennan. 1983 : 2) สมมติต้องการวัดผลการเรียนของนักเรียนด้วยข้อสอบความเรียงที่มีประสิทธิภาพสามารถนำไปใช้ได้ และผู้ตรวจที่มีศักยภาพ จากจุดนี้เขาจึงได้กำหนดเงื่อนไขของการวัดเป็นองค์ประกอบข้อสอบและองค์ประกอบผู้ตรวจ ดังนั้นเราจึงเรียกว่า เอกภพของการสังเกตที่ยอมรับได้ของสมมติประกอบด้วยองค์ประกอบข้อสอบและองค์ประกอบของผู้ตรวจ

6. เอกภพของการสรุปอ้างอิง (Univer of Generalization) หมายถึง เงื่อนไขของการวัดทั้งหมดขององค์ประกอบที่ผู้วิจัยต้องการสรุปอ้างอิง ผลการวัดจากกลุ่มตัวอย่างของเงื่อนไขของการวัดเหล่านี้ไปยังกลุ่มเงื่อนไขของการวัดทั้งหมดขององค์ประกอบนั้นๆ ดังนั้นเงื่อนไขขององค์ประกอบในเอกภพของการสรุปอ้างอิงจึงเป็นจึงเป็นสับเซตในเอกภพของการสังเกตที่ยอมรับได้ เอกภพของการสรุปอ้างอิงนี้ ผู้วิจัยจะต้องระบุประเด็นต่างๆ ให้ชัดเจน ซึ่งจะนำไปใช้ในการออกแบบการวัดของการศึกษาเพื่อการตัดสินใจ

7. การศึกษาเพื่อการสรุปอ้างอิง (Generalizability Study : G Study) เป็นขั้นแรกของการวิเคราะห์สรุปอ้างอิง เพื่อประมาณค่าความแปรปรวนจากแหล่งต่างๆ ของรูปแบบการวัด โดยเริ่มจากการกำหนดองค์ประกอบต่างๆ กำหนดความสัมพันธ์ขององค์ประกอบและกำหนดเอกภพของการสังเกตที่ยอมรับได้และมีการวิเคราะห์ความแปรปรวน การประมาณค่าความแปรปรวนจากแหล่งต่างๆ

8. การศึกษาเพื่อการตัดสินใจ (Decision Study : D Study) เป็นการศึกษาเพื่อประเมินวิธีการวัดที่เหมาะสมเพื่อการตัดสินใจให้หรือลงข้อสรุป เป็นขั้นสุดท้ายของการวิเคราะห์การสรุปอ้างอิง โดยเริ่มจากผู้วิจัยจะต้องระบุสิ่งที่ต้องการวัด เอกภพของการสรุปอ้างอิง จำนวนเงื่อนไขของการวัดของแต่ละองค์ประกอบ และรูปแบบของการวัด แล้วหาค่าความแปรปรวนจากการศึกษา ประมาณค่าความแปรปรวน หาค่าสัมประสิทธิ์ของการสรุปอ้างอิง และปรับปรุงแบบการวัดเพื่อหาวิธีที่เหมาะสมเพื่อให้เกิดประโยชน์สูงสุด

ขนาดของกลุ่มตัวอย่างใน D Study ไม่จำเป็นต้องมีขนาดเหมือนกับขนาดของกลุ่มตัวอย่างที่ใช้ใน G Study แต่ขึ้นอยู่กับนำไปใช้ในการตัดสินใจเกี่ยวกับสิ่งที่จะวัดจริงๆ ซึ่งจะถูกกำหนดขึ้นโดยมีขนาดของกลุ่มตัวอย่างใน G Study เป็นพื้นฐาน

9. ความแปรปรวนของความคลาดเคลื่อนแบบสมบูรณ์ (Absolute Error Variance : $\sigma^2(\Delta)$) หมายถึง ความแปรปรวนของผลต่างระหว่างคะแนนที่สังเกตได้และคะแนนเอกภพ ซึ่งมีค่าเท่ากับผลรวมของความแปรปรวนที่ประมาณได้ทั้งหมด ยกเว้นความแปรปรวนของแหล่งบุคคล

10. ความแปรปรวนของความคลาดเคลื่อนแบบสัมพัทธ์ (Relation Error Variance : $\sigma^2(\delta)$) หมายถึง ความแปรปรวนของผลต่างระหว่างส่วนเบี่ยงเบนของคะแนนที่สังเกตได้ (ความแตกต่างของคะแนนที่สังเกตได้จากคะแนนเฉลี่ยของประชากรของคะแนนที่สังเกตได้) กับส่วนเบี่ยงเบนของคะแนนเอกภพ (ความแตกต่างของคะแนนเอกภพจากคะแนนเฉลี่ยของประชากรของคะแนนเอกภพ)

11. สัมประสิทธิ์การสรุปอ้างอิง (Generalizability Coefficient : ρ^2) หมายถึง อัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพกับความแปรปรวนของคะแนนสังเกตที่คาดหวัง ซึ่งค่าสัมประสิทธิ์ของการสรุปอ้างอิงสามารถประมาณค่าได้จากกำลังสองของค่าสหสัมพันธ์ระหว่างคะแนนเอกภพและคะแนนสังเกต ดังนี้

11.1 สัมประสิทธิ์การสรุปอ้างอิง สำหรับการวัดที่มีการตัดสินใจเชิงเปรียบเทียบหรือการวัดแบบอิงกลุ่ม ความคลาดเคลื่อนในการวัดจะเป็นความคลาดเคลื่อนแบบสัมพัทธ์ ดังนั้นสัมประสิทธิ์การสรุปอ้างอิงจะประมาณค่าได้จาก อัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพและผลบวกของความแปรปรวนของสิ่งที่ถูกวัดกับความแปรปรวนของความคลาดเคลื่อนแบบสัมพัทธ์

11.2 สัมประสิทธิ์การสรุปอ้างอิง สำหรับการวัดที่มีการตัดสินใจแบบสมบูรณ์หรือการวัดแบบอิงเกณฑ์ ความคลาดเคลื่อนในการวัดจะเป็นความคลาดเคลื่อนแบบสมบูรณ์ ดังนั้นสัมประสิทธิ์การสรุปอ้างอิงจะประมาณค่าได้จาก อัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพและผลบวกของความแปรปรวนของสิ่งที่ถูกวัดกับความแปรปรวนของความคลาดเคลื่อนแบบสมบูรณ์

11.3 สัมประสิทธิ์การสรุปอ้างอิง สำหรับการวัดที่มีการตัดสินใจแบบสมบูรณ์หรือการวัดแบบอิงเกณฑ์ที่คำนึงถึงคะแนนจุดตัด สัมประสิทธิ์การสรุปอ้างอิงจะประมาณค่าได้จาก อัตราส่วนที่เหมือนข้อ 11.2 แต่บวกทั้งเศษและส่วนด้วยผลต่างของคะแนนเฉลี่ยกับสัดส่วนของคะแนนจุดตัดยกกำลังสอง

3. หลักพื้นฐานของทฤษฎีการสรุปอ้างอิง

ทฤษฎีการสรุปอ้างอิงมีหลักพื้นฐานสำคัญดังนี้

1. ข้อตกลงของทฤษฎีการสรุปอ้างอิง อาศัยข้อตกลงแบบแผนการวิเคราะห์ความแปรปรวนที่สัมพันธ์กับรูปแบบการวัดที่ต้องการศึกษา โดยทั่วไปมีข้อตกลงดังนี้

1.1 ต้องระบุเอกภพที่ต้องการอ้างอิงให้ชัดเจน สามารถบอกได้ว่ามีเงื่อนไขใดบ้างที่เป็นสมาชิกของเอกภพนั้น

1.2 เงื่อนไขของการวัดที่เป็นอิสระต่อกัน กล่าวคือ คะแนนของผู้สอบที่ตอบข้อสอบข้อ i ถูกหรือผิด ไม่ขึ้นการตอบข้อสอบข้ออื่น

1.3 คะแนนสังเกต (X_{pi}) เป็นค่าการวัดในมาตราอันตรภาค (Interval Scale)

2. ความหมายของค่าสัมประสิทธิ์การสรุปอ้างอิง

2.1 ค่าสัมประสิทธิ์ของการสรุปอ้างอิง เป็นดัชนีที่อธิบายความแม่นยำของการวัดเช่นเดียวกับค่าสัมประสิทธิ์ความเชื่อมั่นแบบดั้งเดิม สามารถใช้คำนวณช่วงความเชื่อมั่นของคะแนนเอกภพ หรือใช้ในการสมการถดถอยในการประมาณค่าคะแนนเอกภพ และใช้ในการปรับแก้ค่าสหสัมพันธ์ที่ลดลงอันเนื่องมาจากความคลาดเคลื่อน (Correcting Correlations For Attenuation) (Brennan. 1983 : 1)

2.2 ค่าสัมประสิทธิ์ของการสุปร้องอิง เป็นค่าประมาณของค่าเฉลี่ยของสหสัมพันธ์ระหว่างค่าการวัดที่สุ่มมาจากเอกภพรายคู่ (Cronbach, et al. 1972 : 157) เช่น ค่าสัมประสิทธิ์การสุปร้องอิงเมื่ออ้างไปยังชุดข้อสอบ (แบบทดสอบ) ซึ่งประกอบด้วยข้อสอบ 20 ข้อ มีค่าเป็น .83 หมายความว่า ถ้าเราสุ่มนักเรียนจากประชากรหนึ่งสมมุติเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ของจังหวัดหนึ่งมาทำการทดสอบ สุ่มแบบทดสอบมาทีละฉบับๆ ฉบับละ 20 ข้อที่ไม่ซ้ำกัน ค่าเฉลี่ยของค่าสัมประสิทธิ์ระหว่างแบบทดสอบที่สุ่มมาจะมีค่าเป็น .83

2.3 ค่าสัมประสิทธิ์ของการสุปร้องอิง เป็นค่ากำลังสองของค่าสหสัมพันธ์ระหว่างคะแนนเอกภพกับคะแนนสังเกต (Feldt and Brennan. 1989 : 140) เช่น ค่าสัมประสิทธิ์การสุปร้องอิงของแบบทดสอบวิชาคณิตศาสตร์ของชั้นมัธยมศึกษาปีที่ 1 จำนวน 20 ข้อ มีค่าเป็น .90 แสดงว่า ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนเอกภพของแบบทดสอบวิชาคณิตศาสตร์กับคะแนนสังเกตของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ยกกำลังสอง มีค่าเป็น .90 ถ้าถอยรอกที่ 2 จะได้ค่าสหสัมพันธ์ระหว่างคะแนนเอกภพของแบบทดสอบวิชาคณิตศาสตร์กับคะแนนสังเกต

2.4 ค่าสัมประสิทธิ์ของการสุปร้องอิง สามารถอธิบายในรูปอัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพกับคะแนนสังเกต (Brennan. 1983 : 4) เช่น ค่าสัมประสิทธิ์ของการสุปร้องอิงเป็น .90 แสดงว่า ความแตกต่างที่วัดได้ร้อยละ 90 เป็นความแตกต่างเนื่องมาจากคะแนนเอกภพ อีกเพียงร้อยละ 10 เป็นความแตกต่างเนื่องมาจากความคลาดเคลื่อน

3. ข้อแตกต่างระหว่างค่าความเชื่อมั่นกับค่าสัมประสิทธิ์การสุปร้องอิง ถึงแม้ว่าค่าสัมประสิทธิ์การสุปร้องอิงจะมีความหมายเช่นเดียวกับความเชื่อมั่นของทฤษฎีการวัดแบบดั้งเดิม แต่ก็มีประเด็นที่แตกต่างกันดังนี้

3.1 การวัดแต่ละครั้งมีค่าสัมประสิทธิ์การสุปร้องอิงได้มากกว่า 1 ค่า

3.2 การอ้างอิงไปยังเอกภพใด จะต้องระบุและอธิบายเอกภพนั้นให้ชัดเจนและต้องสุ่มเงื่อนไขนั้นมาศึกษาด้วย

3.3 ค่าสัมประสิทธิ์ของการสุปร้องอิง สามารถบอกถึงความเป็นเอกพันธ์ของเอกภพได้ด้วย (Cronbach, et al. 1963 : 159) ถ้าข้อสอบที่นำมาศึกษาเป็นตัวอย่างสุ่มจากเอกภพข้อสอบที่มีความเป็นเอกพันธ์ เราจะสามารถใช้คะแนนสังเกตแทนคะแนนเอกภพได้อย่างมั่นใจ

4. การวิเคราะห์สัมประสิทธิ์การสุปร้องอิง

แนวคิดในการประมาณค่าพารามิเตอร์ของทฤษฎีการสุปร้องอิง (Brennan. 1983:105-146) เสนอวิธีการประมาณค่าพารามิเตอร์เป็น 2 ขั้นตอนคือ ขั้นที่ 1 การศึกษาเพื่อสุปร้องอิง (Generalizability Study : G Study) และขั้นที่ 2 การศึกษาเพื่อการตัดสินใจ (Decision Study : D Study) แต่ละขั้นมีรายละเอียดดังนี้

ขั้นที่ 1 การศึกษาเพื่อสุปร้องอิง (G Study) เป็นขั้นแรกของการวิเคราะห์ข้อมูลเพื่อการประมาณค่าความแปรปรวนจากแหล่งต่างๆ ภายใต้เอกภพของการสังเกตที่ยอมรับได้ โดยมีกระบวนการดังนี้

1.1 กำหนดสิ่งที่วัด (Object of Measurement) เช่น นักเรียนชั้นมัธยมศึกษาปีที่ 1

1.2 กำหนดองค์ประกอบ (Facets) ของการวัด เช่น องค์ประกอบข้อสอบ องค์ประกอบผู้ตรวจ เป็นต้น

1.3 กำหนดเอกภพของการสังเกตที่ยอมรับได้ เช่น ในการวิจัยครั้งนี้เอกภพของการสังเกตที่ยอมรับได้ คือ องค์ประกอบของข้อสอบ และองค์ประกอบของผู้ตรวจ เป็นต้น

1.4 กำหนดความสัมพันธ์ระหว่างองค์ประกอบว่าเป็นแบบ Crossed หรือแบบ Nested เช่น ในการวิจัยครั้งนี้ความสัมพันธ์ระหว่างองค์ประกอบเป็นแบบ Crossed คือ $p \times i \times r$ กล่าวคือ ผู้สอบทุกคนทำข้อสอบทุกข้อและผู้ตรวจแต่ละคนตรวจข้อสอบทุกข้อ (เมื่อ p คือผู้สอบ, i คือข้อสอบ และ r คือผู้ตรวจ)

1.5 กำหนดชนิดขององค์ประกอบว่าเป็นชนิดใด ดังต่อไปนี้

1.5.1 องค์ประกอบสุ่มแท้จริง (Purely Random Facet) ได้แก่ องค์ประกอบทุกเงื่อนไขของการวัด ได้มาจากการสุ่มอย่างง่ายจากเอกภพที่มีขนาดไม่จำกัด (α)

1.5.2 องค์ประกอบสุ่มจำกัด (Finite Random Facet) ได้แก่ องค์ประกอบทุกเงื่อนไขของการวัด ได้มาจากการสุ่มอย่างง่ายจากเอกภพที่มีขนาดจำกัด

1.5.3 องค์ประกอบกำหนด (Fixed Facet) ได้แก่ องค์ประกอบทุกเงื่อนไขของการวัดมีจำนวนจำกัด อยู่ในเอกภพและเงื่อนไขของการวัดที่ถูกเลือกมาศึกษาทั้งหมด

1.6 กำหนดรูปแบบ (Model) การวัดว่าเป็นรูปแบบใด ดังต่อไปนี้

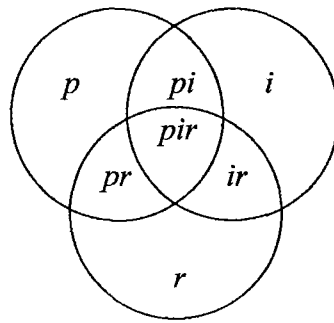
1.6.1 รูปแบบกำหนด (Fixed Model) เป็นรูปแบบที่ประกอบด้วยองค์ประกอบกำหนดทั้งหมด

1.6.2 รูปแบบสุ่ม (Random Model) เป็นรูปแบบที่ประกอบด้วยองค์ประกอบสุ่มแท้จริง และ / หรือองค์ประกอบสุ่มจำกัด

1.6.3 รูปแบบผสม (Mixed Model) เป็นรูปแบบที่ประกอบด้วยองค์ประกอบกำหนด และองค์ประกอบสุ่มแท้จริง และ / หรือองค์ประกอบสุ่มจำกัด ผสมกัน

จากการกำหนดสิ่งที่จะวัด องค์ประกอบของการวัด เอกภพของการสังเกตที่ยอมรับได้ และความสัมพันธ์ระหว่างองค์ประกอบแล้ว ใช้ ANOVA ทำการวิเคราะห์ให้สอดคล้องกับแบบแผนของความสัมพันธ์ขององค์ประกอบ เพื่อประมาณค่า Mean Square แล้วประมาณค่าความแปรปรวนของแหล่งต่างๆ เรียกค่าประมาณความแปรปรวนของสิ่งที่ถูกวัดว่าค่าประมาณความแปรปรวนของคะแนนเอกภพ (Estimated Universe Score Variance) ซึ่งสอดคล้องกับค่าประมาณความแปรปรวนของคะแนนจริง (True Score Variance) ของทฤษฎีการวัดแบบดั้งเดิม เรียกค่าประมาณความแปรปรวนขององค์ประกอบอื่นๆ ว่า ค่าประมาณความแปรปรวนของความคลาดเคลื่อน (Estimated Error Score Variance) การวิเคราะห์ในชั้น G Study นั้นใช้ค่าขนาดของกลุ่มตัวอย่างศึกษาจริงๆ แต่ในการวิเคราะห์ชั้น D Study สามารถกำหนดขนาดของกลุ่มตัวอย่างแตกต่างไปจากที่วิเคราะห์ใน G Study ได้

1.7 การเก็บรวบรวมและการวิเคราะห์ข้อมูล เช่น ในรูปแบบสุ่มโดยการสุ่มกลุ่มตัวอย่างมาศึกษา คือ ผู้สอบ n_p คน ข้อสอบ n_i ข้อ และผู้ตรวจ n_r คน ซึ่งเป็นการศึกษา G Study แบบ $p \times i \times r$ คือ ผู้สอบทุกคนทำข้อสอบทุกข้อและผู้ตรวจแต่ละคนจะตรวจข้อสอบทุกข้อ แล้วคำนวณค่าความแปรปรวนขององค์ประกอบต่างๆ วิธีการดังกล่าวนี้ จะได้แหล่งความแปรปรวน 7 แหล่ง (Brennan. 1983 : 27-28) เขียนแทนด้วยแผนภาพ Venn ได้ดังภาพประกอบ 1 ดังนี้



ภาพประกอบ 1 แผนภาพ Venn แสดงองค์ประกอบความแปรปรวนของรูปแบบ $p \times i \times r$ เมื่อองค์ประกอบที่ศึกษาทั้งหมดเป็นองค์ประกอบสุ่ม

จากแผนภาพจะพบว่า มีผลหลัก (Main Effects) 3 แหล่ง คือ ผลของผู้สอบ (p) ผลของข้อสอบ (i) และผลของผู้ตรวจ (r) ส่วนผลของปฏิสัมพันธ์ (Interaction Effects) 4 แหล่งคือ ปฏิสัมพันธ์ระหว่างผู้สอบกับข้อสอบ (pi) ปฏิสัมพันธ์ระหว่างผู้สอบกับผู้ตรวจ (pr) ปฏิสัมพันธ์ระหว่างข้อสอบกับผู้ตรวจ (ir) และปฏิสัมพันธ์ระหว่างผู้สอบ ข้อสอบ และผู้ตรวจ (pir)

ค่าความแปรปรวนของแหล่งต่างๆ เหล่านี้ คำนวณจาก Mean Square โดยใช้ ANOVA แบบ Factorial Design $p \times i \times r$ ในการประมาณค่าความแปรปรวนจาก Mean Square นั้น ใช้สูตรเฉพาะอย่าง ที่สอดคล้องกับรูปแบบ เช่น สูตรในการประมาณค่าความแปรปรวนของแหล่งต่างๆ เมื่อองค์ประกอบที่ศึกษาเป็นองค์ประกอบสุ่มทั้งหมด และกำหนดรูปแบบความสัมพันธ์ระหว่างองค์ประกอบเป็นแบบ Crossed ใช้สูตรของเบรนนอน (Brennan. 1983 : 130) ดังนี้

$$EMS_p = \sigma^2(pir) + n_i \sigma^2(pr) + n_r \sigma^2(pi) + n_i n_r \sigma^2(p)$$

$$EMS_i = \sigma^2(pir) + n_p \sigma^2(ir) + n_r \sigma^2(pi) + n_p n_r \sigma^2(i)$$

$$EMS_r = \sigma^2(pir) + n_i \sigma^2(pr) + n_p \sigma^2(ir) + n_p n_i \sigma^2(r)$$

$$EMS_{pi} = \sigma^2(pir) + n_r \sigma^2(pi)$$

$$EMS_{pr} = \sigma^2(pir) + n_i \sigma^2(pr)$$

$$EMS_{ir} = \sigma^2(pir) + n_p \sigma^2(ir)$$

$$EMS_{pir} = \sigma^2(pir)$$

โครงสร้างของสมการทั้งหมดนี้สัมพันธ์กับแผนภาพ Venn ข้างต้น ตัวอย่างเช่น ในวงกลม p ประกอบด้วยพจน์ (Term) ต่างๆ 4 พจน์ คือ p, pi, pr และ pir เป็นเทอมทางขวามือของสูตรในการหาค่า EMS_p หรือในพื้นที่ pi ซึ่งเป็นส่วนร่วมระหว่างวงกลม p และวงกลม i จะประกอบด้วย 2 พจน์ คือ pi และ pir เป็นเทอมทางขวามือของสูตรในการหาค่า EMS_{pi} เป็นต้น ดังนั้นจึงสามารถสร้างสูตรเพื่อหาค่า EMS ได้โดยอาศัยภาพ Venn โดยใช้ n_p เป็นตัวคูณควบถ่วงองค์ประกอบนั้นไม่รวม p ใช้ n_i เป็นตัวคูณควบถ่วงองค์ประกอบนั้นไม่มี i และใช้ n_r เป็นตัวคูณควบถ่วงองค์ประกอบนั้นไม่มี r

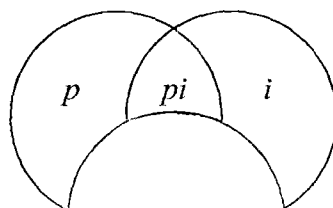
สูตรในการคำนวณเมื่อองค์ประกอบที่ศึกษาเป็นแบบผสมคือ มีทั้งองค์ประกอบกำหนดและองค์ประกอบสุ่ม และรูปแบบความสัมพันธ์ระหว่างองค์ประกอบเป็นแบบ Crossed ถ้าให้ผู้ตรวจ (r) เป็นองค์ประกอบกำหนด สูตรในการคำนวณรวมทั้งแหล่งความแปรปรวนที่ต้องการประมาณค่าจะเป็นดังนี้ (Brennan. 1983 : 71-72)

$$EMS_p = n_r \sigma^2(pi) + n_i n_r \sigma^2(p)$$

$$EMS_i = n_r \sigma^2(pi) + n_p n_r \sigma^2(i)$$

$$EMS_{pi} = n_r \sigma^2(pi)$$

เนื่องจากองค์ประกอบผู้ตรวจ (r) เป็นองค์ประกอบกำหนด จึงหายไปจากแผนภาพเดิม เหลือเพียงวงกลม p และ i เท่านั้น ดังแสดงในแผนภาพ Venn ในภาพประกอบ 2 ดังนี้



ภาพประกอบ 2 แผนภาพ Venn แสดงองค์ประกอบความแปรปรวนของรูปแบบ $p \times i \times r$ เมื่อ p และ i เป็นองค์ประกอบสุ่ม ส่วน r เป็นองค์ประกอบกำหนด

ค่าความแปรปรวนที่ได้จาก G Study เป็นค่าประมาณความแปรปรวนที่แท้จริงของเอกภพของเงื่อนไขการวัดที่มีขนาด 1 หน่วย กล่าวคือ เป็นความแปรปรวนของข้อสอบ 1 ข้อ ผู้ตรวจ 1 คน ของนักเรียน 1 คน เช่น $\sigma^2(p)$ เป็นค่าประมาณความแปรปรวนของ $\sigma^2(p)$ หมายความว่าจากประชากรที่ทำการวัดนั้นจะได้คะแนนของนักเรียน (p) แต่ละคนจากการทำข้อสอบทุกข้อ (N_i) และได้รับการตรวจจากผู้ตรวจทุกคน (N_r) ในเอกภพของการสังเกต คะแนนเฉลี่ยของนักเรียนแต่ละคน คือ คะแนนเอกภพ (μ_p) ดังนั้นค่าความแปรปรวนของ μ_p จากการสอบวัดทุกๆ คนในประชากร คือ $\sigma^2(\mu_p)$ หรือเขียนให้ง่ายว่า $\sigma^2(p)$ เป็นความแปรปรวนของคะแนนเอกภพ ซึ่งตรงกับค่าความแปรปรวนของคะแนนจริงตามทฤษฎีการวัดแบบดั้งเดิม

ขั้นที่ 2 การศึกษาเพื่อการตัดสินใจ (D Study) ในขั้นนี้เน้นการใช้ค่าประมาณและการตีความองค์ประกอบความแปรปรวน สำหรับการตัดสินใจภายใต้วิธีการวัดที่เหมาะสม มีประเด็นสำคัญดังนี้ (Brennan. 1983 : 3-5)

2.1 กำหนดเอกภพของการสรุปอ้างอิง (Universe of Generalization) เป็นเป้าหมายสำคัญของ D Study ได้แก่การกำหนดลักษณะเฉพาะของเอกภพของการสรุปอ้างอิงที่ผู้ตัดสินใจต้องการอ้างอิงถึง อาจจะประกอบด้วยเงื่อนไขทั้งหมดในเอกภพของการสังเกตที่ยอมรับได้ หรืออาจเป็นเซตย่อย (Subset) ของเอกภพของการสังเกตที่ยอมรับได้

2.2 ขนาดของกลุ่มตัวอย่างของ D Study (D Study Sample Sizes) หมายถึง จำนวนเงื่อนไขขององค์ประกอบ D Study สามารถกำหนดแตกต่างจากจำนวนเงื่อนไขใน G Study ใช้สัญลักษณ์ (') แทนขนาดของกลุ่มตัวอย่างใน D Study เช่น n'_i และ n'_r แทนจำนวนข้อสอบและจำนวนผู้ตรวจ

2.3 กำหนดโครงสร้างแบบแผนของ D Study (D Study Design Structure) หมายถึง การกำหนดโครงสร้างของแบบแผนหรือความสัมพันธ์ขององค์ประกอบที่ศึกษา อาจจะเหมือนหรือแตกต่างจาก G Study ก็ได้เช่น ในการวิจัยครั้งนี้แบบแผนของ D Study เป็น $p \times I \times R$ ซึ่งตรงกับแบบแผนของ G Study (สังเกตว่าตัวอักษรตัวใหญ่แทนองค์ประกอบใน D Study ซึ่งในแบบแผน G Study ใช้ตัวอักษรเล็ก)

2.4 กำหนดรูปแบบ (Model) ที่จะใช้รูปแบบใด เช่น รูปแบบสุ่ม (Random Model) หรือรูปแบบกำหนด (Fixed Model) หรือรูปแบบผสม (Mixed Model) ซึ่งโดยหลักการแล้วควรใช้รูปแบบสุ่มดีที่สุด เพราะจะทำให้การสรุปอ้างอิงได้กว้างขวางกว่ารูปแบบผสมและรูปแบบกำหนดตามลำดับ

2.5 การประมาณค่าความแปรปรวนในชั้น D Study (Estimate D Study Variance Component) หมายถึง การประมาณค่าความแปรปรวนขึ้นมาใหม่ ขึ้นอยู่กับการตัดสินใจของนักวิจัย โดยอาศัยผลจากการประมาณค่าในชั้น G Study เป็นฐาน และให้สอดคล้องกับแบบแผนและขนาดของกลุ่มตัวอย่างที่ต้องการตัดสินใจ เช่น ถ้าแบบแผนเป็น $p \times I \times R$ ขนาดตัวอย่างของข้อสอบ = n'_i และขนาดของผู้ตรวจ = n'_r สามารถประมาณค่าความแปรปรวนใหม่ได้ดังตาราง 1 (Brennan. 1983 : 58)

ตาราง 1 แสดงการประมาณค่าความแปรปรวนใหม่ของแบบแผน $p \times I \times R$

ความแปรปรวนใหม่	ความแปรปรวนเดิม/ขนาดของกลุ่มตัวอย่าง
$\sigma^2(p)$	$\sigma^2(p)$
$\sigma^2(I)$	$\sigma^2(i) / n'_i$
$\sigma^2(R)$	$\sigma^2(r) / n'_r$
$\sigma^2(pI)$	$\sigma^2(pi) / n'_i$
$\sigma^2(pR)$	$\sigma^2(pr) / n'_r$
$\sigma^2(IR)$	$\sigma^2(ir) / n'_i n'_r$
$\sigma^2(pIR)$	$\sigma^2(pir) / n'_i n'_r$

2.6 การประมาณค่าความแปรปรวนของความคลาดเคลื่อน นอกจากการประมาณค่าองค์ประกอบความแปรปรวน (Variance Components) ตามรูปแบบและขนาดของกลุ่มตัวอย่างที่ต้องการแล้ว จะต้องประมาณค่าความแปรปรวนของความคลาดเคลื่อนซึ่งแบ่งออกเป็น 2 ชนิด คือ

2.6.1 ความแปรปรวนคลาดเคลื่อนแบบสัมบูรณ์ (Absolute Error Variance) ใช้สัญลักษณ์ $\sigma^2(\Delta)$ หมายถึง ความแปรปรวนของความแตกต่างระหว่างคะแนนที่ได้จากการสังเกตกับคะแนนเอกภพของผู้เข้าสอบ คำนวณจากผลบวกของค่าองค์ประกอบความแปรปรวนทั้งหมด ยกเว้นความแปรปรวนของคะแนนเอกภพ จากแบบแผน $p \times I \times R$ จะได้

$$\sigma^2(\Delta) = \sigma^2(I) + \sigma^2(R) + \sigma^2(pI) + \sigma^2(pR) + \sigma^2(IR) + \sigma^2(pIR)$$

2.6.2 ความแปรปรวนคลาดเคลื่อนแบบสัมพัทธ์ (Relative Error Variance) ใช้สัญลักษณ์ $\sigma^2(\delta)$ หมายถึง ความแปรปรวนของผลต่างระหว่างส่วนเบี่ยงเบนของคะแนนที่สังเกตได้กับส่วนเบี่ยงเบนของคะแนนเอกภพ คำนวณจากผลบวกของค่าองค์ประกอบความแปรปรวนระหว่างที่ถูกวัดกับองค์ประกอบอื่นๆ จากแบบแผน $p \times I \times R$ จะได้

$$\sigma^2(\delta) = \sigma^2(pI) + \sigma^2(pR) + \sigma^2(pIR)$$

2.7 การประมาณค่าสัมประสิทธิ์การสรุปอ้างอิง (Generalizability Coefficient) เป็นขั้นสุดท้ายของการศึกษาเพื่อการตัดสินใจ (D Study) ซึ่งเป็นดัชนีชี้ถึงความเชื่อถือได้ของการวัด (Dependability of Measurement) (Feldt and Brennan. 1989 : 141) โดยคำนวณได้จากอัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพกับความแปรปรวนของคะแนนที่สังเกต ที่คาดหวัง

สูตรในการคำนวณค่าสัมประสิทธิ์การสรุปอ้างอิงมีดังนี้ (Brennan. 1983 : 17)

$$\rho^2 = \frac{\sigma^2(p)}{\sigma^2(p) + \sigma^2(\Delta)}$$

หรือ

$$\rho^2 = \frac{\sigma^2(p)}{\sigma^2(p) + \sigma^2(\delta)}$$

เมื่อ ρ^2 คือ ค่าสัมประสิทธิ์การสรุปอ้างอิงซึ่งมีความหมายเช่นเดียวกับความเชื่อมั่น

5. โปรแกรมสำเร็จรูปสำหรับวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิง

ในปี ค.ศ.1982 คริกและเบรนนอน (Brennan. 1983 : 141 ; citing Crick and Brennan. 1982) ได้คิดค้นโปรแกรมสำเร็จรูป GENOVA (Generalized Analysis of Variance System) ขึ้นเพื่อใช้ในการวิเคราะห์ในทฤษฎีการสรุปอ้างอิง

โปรแกรม GENOVA สามารถอธิบายถึงการวิเคราะห์การสรุปอ้างอิงของแหล่งความแปรปรวนของความสมบูรณ์ ความสมดุลของรูปแบบ ซึ่งจะเป็นประโยชน์สำหรับผู้ที่น่าทฤษฎีการสรุปอ้างอิงไปใช้ จากความสัมพันธ์ของทฤษฎีการสรุปอ้างอิง โปรแกรม GENOVA จะสามารถคำนวณหาค่า Parameter ซึ่งแบ่งการวิเคราะห์ออกเป็น 2 ขั้นตอน ได้แก่ ขั้นการศึกษาเพื่อการสรุปอ้างอิง (G Study) เป็นขั้นที่ประมาณค่าความแปรปรวนจากแหล่งต่างๆ และขั้นการศึกษาเพื่อการตัดสินใจ (D Study) เป็นขั้นที่นำผลจากการประมาณค่าความแปรปรวนมาเลือกวิธีการวัดที่เหมาะสมที่สุดภายใต้การตัดสินใจ เพื่อลดความคลาดเคลื่อนและเพิ่มความเชื่อมั่นของการสรุปอ้างอิง ดังนั้นผู้วิจัยจึงใช้โปรแกรมสำเร็จรูป GENOVA ในการวิเคราะห์ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์

4. การประเมินผลจากผลงานของนักเรียน

การประเมินผลจากผลงานของนักเรียน (Performance Assessment) เป็นการวัดว่านักเรียนสามารถทำอะไรได้บ้าง โดยนักเรียนจะแสดงทักษะการทำงานบางอย่าง หรือสร้างอะไรบางอย่าง ตัวอย่าง เช่น การทดลองในห้องปฏิบัติการ การเขียนรายงาน การพูด ร้องเพลง เล่นดนตรี พิมพ์ดีด ใช้โปรแกรมคอมพิวเตอร์ ประดิษฐ์หรือซ่อมแซมชิ้นงาน หรือแสดง สาธิต แข่งขัน ในการเล่นยืดหยุ่น รวมถึงการทำข้อสอบโดยการเขียนตอบ ถ้างานและระบบการให้คะแนนเหมือนกับการทดสอบในชีวิตจริง

คุณลักษณะของการประเมินผลจากผลงานของนักเรียน

1. มีพื้นฐานจากบริบทของโลกภายนอกที่แท้จริง
2. มักจะเป็นการประเมินทั้งงานที่ทำในชั้นเรียนและนอกชั้นเรียน
3. เป็นการบูรณาการหลาย ๆ วิชาเข้าด้วยกัน
4. เป็นทั้งการทำงานเป็นรายบุคคลและเป็นกลุ่ม
5. ใช้เกณฑ์การให้คะแนนเป็นหลัก
6. ให้ความสำคัญกับการประเมินตนเองของนักเรียน
7. รวมเอากลยุทธ์ในการสื่อสารเข้าไปในการประเมินด้วย

ในการประเมินผลจากผลงานของนักเรียนนั้น มีเทคนิคการประเมินผลอยู่หลายแบบ และเทคนิคการประเมินผลทางเลือกใหม่แบบหนึ่ง คือ การใช้คำถามแบบปลายเปิด (Open-ended Questions) (Courtney. 1996)

5. คำถามแบบปลายเปิด (Open-ended Questions)

คำถามแบบปลายเปิดเป็นคำถามที่มีแนวทางให้นักเรียนตอบได้หลายทาง

คำถามแบบปลายเปิด (Open-ended Questions) แตกต่างจากคำถามแบบมีการตอบอย่างเสรี (Open Response) ตรงที่ คำถามแบบมีการตอบอย่างเสรี เป็นคำถามที่ผู้ตอบจะตอบคำถามในรูปของการแสดงรายละเอียด ขั้นตอนต่างๆ เกี่ยวกับวิธีการ ข้อเสนอแนะ และความคิดเห็นต่าง ๆ อย่างเสรี เพื่อนำไปสู่คำตอบที่เฉพาะเจาะจง ส่วนคำถามแบบปลายเปิด ผู้ตอบอาจจะให้คำตอบที่ถูกต้องได้หลายแบบ ผู้ตอบอาจจะเลือกประเมินในแนวทางปฏิบัติต่าง ๆ เพื่อให้ได้คำตอบเหล่านั้น ซึ่งเราจะเน้นที่ความคิดของผู้ตอบมากกว่าการเขียน

การสร้างคำถามแบบปลายเปิด

หลักทั่วๆ ไปยังคงพัฒนาเพื่อสนับสนุนการปรับปรุงการประเมินแบบใหม่ในชั้นเรียน แต่เปลี่ยนบางอย่างให้ถูกต้องเหมาะสม ดังนี้

1. เมื่อนักเรียนใช้ขั้นตอนทางคณิตศาสตร์ในการหาคำตอบ ถามให้นักเรียนอธิบายความหมายของขั้นตอนที่ใช้ และอธิบายว่าขั้นตอนนั้นทำอย่างไร ด้วยการเขียนหรืออธิบายด้วยแผนภาพ

2. จากปัญหาที่มีในการเรียนการสอน เพิ่มคำถามว่า "จงอธิบายว่าทำอย่างไรจึงจะได้คำตอบ"
3. นอกจากต้องการคำตอบที่ถูกต้องแล้ว คำถามยังต้องกระตุ้นให้นักเรียนเกิดความคิดด้วย
4. ควบคุมคำแนะนำและการตอบคำถามกับนักเรียน เพื่อให้นักเรียนได้อธิบายมากกว่าอาจารย์เป็นผู้อธิบายให้นักเรียนฟัง
5. หาสถานการณ์ที่ถามให้นักเรียนได้ใช้สูตร เขียนแนวทาง ทำตามกระบวนการและอื่นๆ (Stenmark. 1991 : 20)

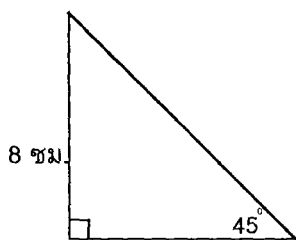
สำหรับครูผู้สอน คำถามแบบปลายเปิดทำให้ทราบถึงความเข้าใจของนักเรียนได้ดีกว่าคำถามซึ่งต้องการคำตอบเพียงคำตอบที่ถูกเท่านั้น เราพบว่า คำตอบเพียงอย่างเดียวไม่ได้ชี้ให้เห็นความเข้าใจของนักเรียนได้อย่างมั่นใจ ในทางตรงข้าม เมื่อเราถามนักเรียนให้อธิบายหรือแสดงเหตุผลในคำตอบเหล่านั้น เราสามารถประเมินความเข้าใจในแต่ละขั้นตอนที่นักเรียนใช้ว่าอยู่ในระดับใด คำถามแบบปลายเปิดมีประโยชน์สำหรับนักเรียนเป็นอย่างดี เมื่อนักเรียนถูกถามด้วยสถานการณ์ปัญหาแล้วให้อธิบาย โดยประยุกต์ใช้ความรู้ในสภาพแวดล้อมที่ไม่คุ้นเคย นักเรียนต้องสร้างคำตอบด้วยตัวเอง ด้วยการบรรยายสิ่งที่เขารู้เกี่ยวกับปัญหาที่ได้รับ คำถามแบบนี้ช่วยส่งเสริมนักเรียนเข้าใจเกี่ยวกับคณิตศาสตร์ ว่ามีความสำคัญในกิจกรรมที่เหมาะสม สร้างความสัมพันธ์ที่เห็นได้ชัดเจนในชีวิตประจำวัน และเข้าถึงได้ง่ายกับนักเรียนทุกคน ไม่ว่าจะอายุเท่าใดหรือมีความสามารถในระดับใด (Stenmark. 1991 : 21 ; citing Massachusetts Department of Education. 1989 : 41)

ตัวอย่างคำถามและปัญหา (Stenmark. 1991 : 21-22)

ต่อไปเป็นตัวอย่างของคำถามแบบปลายเปิด ซึ่งสามารถใช้ได้กับหลาย ๆ ระดับชั้น ปัญหาเหล่านี้ส่วนมากไม่แตกต่างจากปัญหาแบบเดิมนัก แต่คำถามเหล่านี้ถามเพื่อให้นักเรียนอธิบายเหตุผลที่เขาใช้ในการตอบคำถาม

- ข้อ 1 ถ้าท่านวัดวัตถุชิ้นหนึ่งด้วยไม้บรรทัดที่แตกต่างกัน 5 อัน และได้ผลลัพธ์ที่แตกต่างกัน 5 คำตอบ ท่านจะตัดสินใจอย่างไรว่าคำตอบไหนถูกต้อง จงอธิบาย
- ข้อ 2 นักเรียนต้องการสอนญาติพี่น้องของท่านเกี่ยวกับการลบเลขสองหลัก จงเขียนว่าท่านจะพูดอะไรและให้ตัวอย่างประกอบด้วย
- ข้อ 3 แจกมีดพกที่ขาวไธต์ 3 ชิ้น จีมีดพกที่ช็อคโกแลตชิพ 3 ชิ้น และมูฟฟิมีดพกที่เนยถั่ว 3 ชิ้น พวกเขาแบ่งให้คนอื่น ๆ เท่า ๆ กัน ให้นักเรียนวาดภาพเพื่อแสดงดุกที่พวกเขามีครั้งแรก และดุกที่พวกเขามีหลังจากแบ่งกับคนอื่น ๆ ใช้ตารางประกอบได้
- ข้อ 4 ยอร์ชสะสมรถจักรยานและรถสามล้อ เขามีล้อรวมกัน 6 ล้อ หลังจากนั้นเขาหามาเพิ่มได้อีก 3 ล้อ ให้นักเรียนวาดภาพว่า ยอร์ชมีรถจักรยานและรถสามล้ออย่างละกี่คัน ที่สามารถเป็นไปได้ พร้อมทั้งอธิบายว่าทำไมจึงตอบได้มากกว่า 1 คำตอบ

ข้อ 5 บอกทุกสิ่งที่ท่านรู้หรือสามารถหาได้ จากรูปสามเหลี่ยมข้างล่างนี้



ข้อ 6 ด้านล่างนี้เป็นรายการราคาสินค้า เครื่องคิดเลขของท่านเสีย จงอธิบายวิธีการอย่างน้อย 2 วิธี ที่จะหาผลรวมของราคาสินค้าที่ท่านสามารถทำได้

65 \$
88 \$
23 \$
42 \$
77 \$
35 \$
99 \$
<u>2 \$</u>
<u><u>?</u></u>

นอกจากนี้ ในงานวิจัยของซูซาน เลน และคณะ (Suzanne Lane, et al. 1996:89) มีตัวอย่างคำถามแบบปลายเปิดในวิชาคณิตศาสตร์ สำหรับชั้นประถมศึกษาปีที่ 6 และชั้นมัธยมศึกษาปีที่ 1 ที่วัดความสามารถของนักเรียนในด้านการแก้ปัญหาในวิชาคณิตศาสตร์ คณิตศาสตร์เหตุผล และการสื่อสารทางคณิตศาสตร์ ดังนี้

1. กิจกรรมทฤษฎีจำนวน

โยลันดากำลังบอกกับพี่ชายของเธอที่ชื่อดาเมียน เกี่ยวกับปัญหาคณิตศาสตร์ในชั้นเรียน

โยลันดาพูดว่า "ดาเมียน ฉันประยุกต์ใช้บล็อกในวิชาคณิตศาสตร์วันนี้ เมื่อฉันจับกลุ่มบล็อกในกลุ่ม 2, 1 แล้วยังมีบล็อกเหลือ 1 บล็อก เมื่อฉันจับกลุ่มบล็อกในกลุ่ม 3, 1 แล้วมีบล็อกเหลือ 1 บล็อก และเมื่อฉันจับกลุ่มบล็อกในกลุ่ม 4, 1 ยังคงมีบล็อกเหลือ 1 บล็อก"

ดาเมียนถามว่า "เธอมีบล็อกอยู่ที่บล็อก"

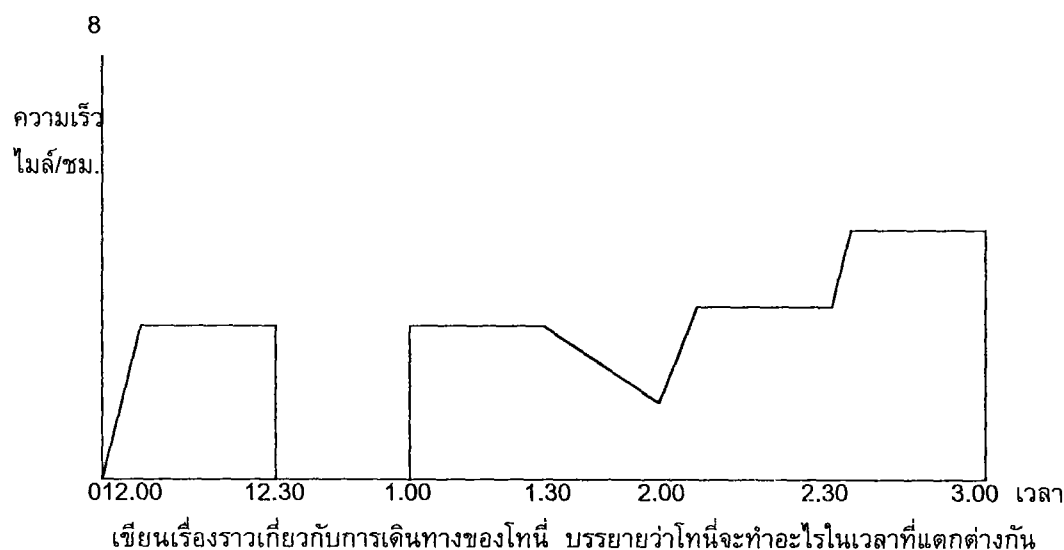
โยลันดาจะตอบคำถามพี่ชายอย่างไร แสดงวิธีการทำของท่าน

2. กิจกรรมการตีความหมาย(อธิบาย) กราฟ

ใช้ข้อมูลต่อไปนี้และกราฟเขียนเรื่องราวเกี่ยวกับการเดินทางของโทนี่

ตอนเที่ยงโทนี่เริ่มเดินทางเพื่อไปที่บ้านของยาย เขาไปถึงบ้านยายเวลา 3.00 น.

กราฟข้างล่างนี้แสดงความเร็วเป็นไมล์ต่อชั่วโมงตลอดการเดินทางของเขา



6. วิธีการตรวจให้คะแนนคำถามแบบปลายเปิด

การตรวจให้คะแนนคำถามแบบปลายเปิด ทำได้ 2 วิธี (Stenmark. 1991 : 20)

1. วิธีวิเคราะห์ (Analytical Method หรือ Point Method) ใช้การวิเคราะห์ด้วยประเด็นที่มีลักษณะแตกต่างกันของคำตอบ

2. วิธีประเมินรวม (Holistic Method) เป็นวิธีที่ผู้ประเมินพิจารณาคำตอบโดยรวมมากกว่าตรวจสอบรายละเอียดปลีกย่อยเฉพาะ

นักวัดผลหลายท่าน ได้ให้ความหมายและวิธีการตรวจให้คะแนนแบบทดสอบทั้ง 2 วิธีนี้ไว้ เช่น

อีเบล (Ebel. 1972 : 149-152) ได้ให้ข้อเสนอแนะสำหรับการตรวจให้คะแนนไว้ ดังนี้

1. การให้คะแนนด้วยวิธีวิเคราะห์ หรือวิธีการประเมินคุณภาพโดยรวม (Analytic Scoring or Global-quality Scaling)

1.1 วิธีวิเคราะห์ วิธีนี้ส่วนประกอบที่สำคัญต่างๆ ของคำตอบจะเป็นตัวกำหนดการให้คะแนนมาก-น้อย อย่างอิสระ การพิจารณาจะพิจารณาทั้งส่วนประกอบที่สำคัญของคำตอบ และความสัมพันธ์ระหว่างส่วนประกอบเหล่านั้นด้วยว่าเป็นองค์ประกอบเดียวกันของคำตอบหรือไม่ แต่ถ้าความสัมพันธ์นั้นซับซ้อนและเข้าใจยาก วิธีวิเคราะห์ก็อาจจะทำได้ไม่สะดวกและใช้เวลานาน

1.2 วิธีประเมินคุณภาพโดยรวม วิธีนี้ผู้ตรวจจะอ่านคำตอบโดยรวม และพิจารณาจากความประทับใจ แล้วจึงเปลี่ยนความประทับใจนั้นเป็นระดับคะแนนบันทึกไว้ ทำแบบนี้ทุก ๆ ข้อ วิธีการที่ดีกว่าเพื่อควบคุมความสม่ำเสมอหรือความมีมาตรฐานในการตรวจกระดาษคำตอบที่ต่างกัน คือ การแบ่งกลุ่มคำ

ตอบเป็นหลายๆ กองตามระดับคุณภาพที่ต่างกัน โดยในแต่ละกองจะมีลักษณะคำตอบที่คล้ายๆ กัน เมื่อแบ่งกลุ่มแล้วผู้ตรวจจึงมาพิจารณาตัดสินใหม่อีกครั้งว่าแต่ละกองเหมาะสมหรือยัง

ผู้ให้คะแนนจะกำหนดขนาดของแต่ละกองไว้ล่วงหน้า และพยายามแบ่งให้มีการแจกแจงใกล้เคียงกับที่กำหนดไว้ ในการแบ่งกลุ่มนั้นอาจแบ่งได้ดังนี้ :

1. แบ่งเป็น 3 กอง

ระดับต่ำ	ระดับปานกลาง	ระดับสูง
25%	50%	25%

หรือแบ่งเป็น 5 กอง

ต่ำที่สุด	ต่ำกว่า	ปานกลาง	สูงกว่า	สูงที่สุด
5%	25%	40%	25%	5%

2. ตรวจให้คะแนนคำตอบแบบคำถามต่อคำถาม มากกว่า คนต่อคน
3. ถ้าเป็นไปได้ ควรปิดบังชื่อของนักเรียนที่ทำข้อสอบไม่ให้ผู้ตรวจทราบ
4. ถ้าเป็นไปได้ ควรจัดให้มีการตรวจให้คะแนนอย่างอิสระ

เมอร์เรน และลีแมน (Mehrens and Lehmann. 1973 : 229-238) ได้อธิบายถึงการตรวจให้คะแนนด้วยวิธีวิเคราะห์และวิธีประเมินรวมไว้ ดังนี้

1. วิธีวิเคราะห์ (Analytical Method หรือ Point-score Method) รูปแบบคำตอบจะประกอบด้วยประเด็นเฉพาะที่กำหนดไว้ก่อนแล้ว คะแนนของนักเรียนที่ได้จะขึ้นอยู่กับจำนวนประเด็นที่เขาตอบ รวมไปถึงถึงส่วนอื่นๆ เช่น แสดงความคิดเห็นได้ชัดเจน การให้เหตุผลและการยกตัวอย่างสนับสนุนในประเด็นคำตอบ และการกำหนดคะแนนในแต่ละประเด็นจะขึ้นอยู่กับ เวลาที่ใช้ในการตอบ ความซับซ้อนของคำถาม และเนื้อหาที่ครูสอน

ข้อดี ของการตรวจให้คะแนนด้วยวิธีวิเคราะห์ มีดังนี้

1. ให้ผลคะแนนที่เชื่อถือได้ เมื่อผู้ตรวจตรวจอย่างยุติธรรม
2. เมื่อมีการเตรียมรายละเอียดของคำตอบไว้แล้ว ครูสามารถพิจารณาความผิดพลาด จากการใช้คำ ความยาก หรือความซับซ้อนของคำถาม รวมทั้งเวลาที่ไม่เหมาะสม และสามารถเปลี่ยนแปลงคำถามหรือขยายเวลาให้เหมาะสมได้
3. มีการแบ่งคำตอบเป็นกอง ๆ ทำให้สามารถตรวจให้คะแนน จัดระดับคุณภาพข้อสอบของนักเรียนได้ง่ายขึ้น

ข้อจำกัด

1. ตรวจยาก และเสียเวลา
2. การกำหนดประเด็นหรือส่วนประกอบมากเกินไป อาจทำให้ได้รับคำตอบด้านต่างๆ ไม่ลึกซึ้ง

ถึงแม้การตรวจด้วยวิธีวิเคราะห์จะใช้ได้ทั้งคำถามชนิดจำกัดคำตอบและไม่จำกัดคำตอบ แต่จะใช้ได้ดีกับชนิดจำกัดคำตอบและไม่มากเกินไป จึงจะเชื่อมั่นได้มากกว่า

2. วิธีประเมินรวม (Global Scoring หรือ Holistic หรือ Rating Method) วิธีนี้คำตอบจะไม่ถูกแบ่งเป็นส่วนๆ เป็นประเด็นเฉพาะ แต่ผู้ตรวจจะอ่านคำตอบอย่างรวดเร็ว แล้วใช้ความประทับใจและใช้มาตรฐานบางอย่างกำหนดระดับของคำตอบ การตรวจคำตอบจะขึ้นอยู่กับของการแบ่ง อาจแบ่งข้อสอบเป็น 2 กลุ่ม

คือ "กลุ่มที่ยอมรับได้ - กลุ่มที่ยอมรับไม่ได้" หรือ 5 กลุ่ม คือ "ดีมาก-ต่ำกว่ามาตรฐาน" โดยมากจะแบ่งประมาณ 4 หรือ 5 กลุ่ม

สำหรับกรณีนี้แบ่งเป็น 5 กลุ่ม ประกอบด้วย

1. คุณภาพดีมาก
2. คุณภาพดีกว่าปานกลาง
3. คุณภาพปานกลาง
4. คุณภาพต่ำกว่าปานกลาง
5. คุณภาพต่ำกว่ามาตรฐาน

ในการอ่านคำตอบอย่างรวดเร็วนั้น ผู้ตรวจจะกำหนดคุณภาพของคำตอบให้อยู่ 1 ใน 5 กลุ่มนี้ แต่ละคำตอบจะต้องอ่านอย่างน้อยที่สุด 2 ครั้ง เพื่อจัดอันดับคุณภาพวิธีนี้จะไม่ลำบากและไม่เสียเวลามากเกินไป จึงทำได้เร็วกว่าวิธีวิเคราะห์

วิธีประเมินรวมจะมีประสิทธิภาพมาก เมื่อแบบทดสอบมีจำนวนมาก

ข้อเสนอแนะสำหรับการตรวจให้คะแนน

เป็นข้อเสนอแนะเพื่อให้การตรวจให้คะแนนมีความเชื่อมั่นเพิ่มขึ้น

1. ตรวจคะแนนที่ให้ไปแล้วอีกครั้งกับระดับคะแนนที่กำหนดไว้
2. พยายามให้การตรวจให้คะแนนของท่าน คงที่ สม่าเสมอ
3. ตรวจให้คะแนนที่ละคำถามให้ครบทุกคนก่อน
4. ตรวจให้คะแนนคำตอบโดยไม่บอกชื่อผู้ทำ
5. การพิจารณาต้องแยกกันระหว่างเนื้อหาความสามารถในการเขียนของนักเรียน
6. พยายามให้คะแนนคำตอบทั้งหมดในแต่ละคำถาม โดยไม่มีการหยุดชะงัก
7. ถ้าเป็นไปได้ ควรมีผู้ตรวจ 2 คนที่เป็นอิสระจากกัน ในการตรวจแบบทดสอบ 1 ฉบับ

และใช้ค่าเฉลี่ยเป็นคะแนนสรุป

8. จัดให้มีข้อคิดเห็นและการแก้ไขข้อผิดพลาด
9. จัดชุดคำตอบมาตรฐานที่เป็นไปได้จริง

โกลแลนด์ (Groulund. 1977 : 79) ได้อธิบายถึงวิธีการตรวจให้คะแนนไว้ดังนี้

1. การให้คะแนนคำถามแบบจำกัดคำตอบ (Restricted-response) ด้วยวิธีแยกประเด็น (Point-method) โดยใช้รูปแบบคำตอบเป็นแนวทาง

การให้คะแนนแบบนี้จะต้องมีการเตรียมกำหนดคะแนนไว้ก่อน ขั้นตอน คือ ต้องกำหนดจำนวนประเด็นที่จะให้คะแนนแต่ละคำตอบ ซึ่งประเด็นเหล่านี้จะต้องสามารถวัดจุดประสงค์การเรียนรู้ที่ต้องการได้

2. การให้คะแนนคำถามแบบไม่จำกัดคำตอบ (Extended-response) ด้วยวิธีการจัดระดับ (Rating method) โดยใช้การกำหนดเกณฑ์เป็นแนวทาง

ข้อสอบแบบไม่จำกัดคำตอบจะให้อิสระในการตอบ ผู้ตรวจแบบทดสอบจะต้องตรวจให้คะแนนแต่ละคำตอบด้วยการพิจารณาคุณภาพ จากเกณฑ์ที่กำหนดไว้แล้ว เกณฑ์ที่ใช้ประเมินคำตอบกำหนดคะแนนเป็น 5 ระดับ ดังนั้น เมื่ออ่านคำตอบแล้วจะต้องกำหนดระดับคุณภาพของคำตอบจาก 1-5 คะแนน ในแต่ละระดับคะแนนอาจกำหนดให้เป็นคุณภาพโดยรวมของทั้งคำตอบ หรือแยกการพิจารณาตามเกณฑ์ก็ได้

โดยปกติ วิธีนี้ผู้ตรวจจะอ่านคำตอบแต่ละข้อ 2 ครั้ง ครั้งแรกจะจัดระดับคร่าวๆ เข้ากอง 1 ใน 5 จากคุณภาพสูงไปต่ำ การอ่านครั้งที่สองจะช่วยตรวจสอบคำตอบแต่ละกอง และสามารถเปลี่ยนการจัดระดับได้ถ้าต้องการ

นอกจากนี้ คอทนี (Courtney. 1996 : 34) ยังได้เสนอลำดับขั้นตอนการให้คะแนนแบบประเมินรวม (Holistic) ไว้ดังนี้

จัดแบ่งงานของนักเรียนออกเป็น 3 กอง

กองที่ 1 งานที่มีคุณภาพพิเศษ

กองที่ 2 งานที่ได้รับการยอมรับ

กองที่ 3 งานที่ไม่ได้รับการยอมรับ

แบ่งงานแต่ละกองออกเป็น 2 ระดับ

กองที่ 1 ให้คะแนน 6 หรือ 5 คะแนน

กองที่ 2 ให้คะแนน 4 หรือ 3 คะแนน

กองที่ 3 ให้คะแนน 2 หรือ 1 คะแนน

งานที่แสดงว่าไม่ได้ใช้ความพยายามเลยให้ 0 คะแนน

และ ไพศาล หวังพานิช (2528 : 96-97) กล่าวว่า ไม่ว่าจะใช้วิธีใดตรวจให้คะแนน การตรวจที่จะช่วยเพิ่มความเชื่อถือได้ของคะแนน ควรกระทำดังต่อไปนี้

1. ไม่ควรให้เด็กเขียนชื่อในกระดาษคำตอบ เพื่อป้องกันการให้คะแนนจากความรู้สึกประทับใจในเรื่องอื่นๆ ของเด็ก ซึ่งเรียกว่า halo effect เช่น ให้คะแนนจากความคุ้นเคย ความรู้สึกว่าเด็กมีความตั้งใจและขยันขันแข็ง เป็นต้น

2. ตรวจคำตอบทีละข้อ ไม่ควรตรวจทุกข้อของแต่ละคน เพราะจะก่อให้เกิด halo effect เช่นกัน เช่น เห็นว่าข้อแรกๆ ของเด็กได้คะแนนมาก ข้อต่อไปจึงให้คะแนนน้อย (ทั้งๆ ที่คำตอบดี) หรือในทางตรงกันข้าม การตรวจคำถามข้อเดียวกันของทุกๆ คนให้เสร็จ จะช่วยในแง่การเปรียบเทียบคุณภาพการตอบของเด็กทั้งกลุ่มได้ด้วย อีกทั้งช่วยให้การตรวจแต่ละข้อนั้นๆ ของแต่ละคนยึดเกณฑ์ที่เหมือนกัน

3. ไม่ควรย้อนไปดูคะแนนของเด็กจากข้อที่ตรวจแล้วในการตรวจข้อต่อไปทั้งนี้ เพื่อป้องกันไม่ให้คะแนนจากข้ออื่นๆ มีผลกระทบต่อการให้คะแนนในข้ออื่นๆ

4. ไม่ควรให้คะแนนโดยยึดความถูกต้องทางภาษาเป็นหลัก ถ้าหากมิได้มุ่งวัดความถูกต้องในการเขียนและการใช้ภาษา ความถูกต้องสละสลวยในการใช้ถ้อยคำสำนวนในการตอบ ไม่ควรมีอิทธิพลต่อการให้คะแนนมากหรือน้อย ผู้ตรวจควรมีพิจารณาเฉพาะเป้าหมายการตอบในแง่ของความสมบูรณ์ครบถ้วนของเนื้อหาหรือความสมเหตุสมผลของความคิดเป็นหลักในการให้คะแนน

5. ถ้าเป็นไปได้ควรให้คนอื่นช่วยตรวจสอบผลการตรวจให้คะแนน ทั้งนี้เพื่อให้คนอื่นได้ประเมินความเหมาะสมในการให้คะแนนของเรา ตามหลักการที่ถูกต้องนั้น คะแนนที่เด็กได้ควรเป็นคะแนนเฉลี่ยที่เกิดจากผู้ตรวจให้คะแนนหลายๆ คน ซึ่งคงเป็นเรื่องยากในเชิงปฏิบัติ แต่อย่างไรก็ตามถ้าไม่อาจปฏิบัติได้อย่างน้อยผู้ตรวจควรจะได้ทบทวนความเหมาะสมในการตรวจให้คะแนนของตนอีกครั้งหนึ่ง ก่อนที่จะนำคะแนนเหล่านั้นไปใช้

6. ควรเขียนวิจารณ์ ข้อท้วงติง (Comments) ลงบนคำตอบ เพื่อประโยชน์ในการเรียนรู้ของผู้เรียน หรืออย่างน้อยให้เป็นหลักฐานว่า ทำไมจึงให้คะแนนเท่านี้ สำหรับคำตอบของเด็กคนนี้

7. การตรวจให้คะแนนต้องกระทำอย่างตั้งใจ โดยเฉพาะอย่างยิ่งต้องมีเกณฑ์การให้คะแนน และต้องอ่านคำตอบของเด็กอย่างถี่ถ้วน เมื่อใดที่ผู้ตรวจขาดหลักเกณฑ์การให้คะแนนหรือไม่ได้อ่านคำตอบอย่างตั้งใจ คะแนนที่ให้กับเด็กมักจะออกมาในรูปกลางๆ ซึ่งเป็นไปตามหลักของ Central Tendency Error ซึ่งเป็นไปในลักษณะที่ว่า "เมื่อไม่แน่ใจก็ให้คะแนนกลางๆ ไว้ก่อน" ซึ่งแน่นอนคะแนนที่ได้จากการสอบวัดนั้นจะมีความเชื่อมั่นต่ำลง

จากแนวคิดเกี่ยวกับการตรวจให้คะแนนที่กล่าวมาข้างต้น พอสรุปถึงการตรวจให้คะแนนด้วยวิธีวิเคราะห์ และวิธีประเมินรวม ได้ดังนี้

1. การตรวจวิธีวิเคราะห์ (Analytical Method หรือ Point Method) เป็นวิธีการตรวจที่เหมาะสมกับการตรวจประเภทจำกัดคำตอบ ในสถานการณ์การประเมินผลการเรียนการสอนที่มีผู้สอบไม่มากนัก การตรวจแบบนี้ผู้ตรวจต้องมีคำตอบไว้ล่วงหน้าซึ่งได้แยกเป็นประเด็นย่อยๆ ตามองค์ประกอบของการตอบ ตามความสำคัญหรือตามประเด็นที่ถาม เช่น การจัดเรียงความคิด การยกตัวอย่าง การตอบได้ตรงประเด็น เป็นต้น โดยกำหนดคะแนนของแต่ละประเด็นย่อยๆ ไว้เป็นเกณฑ์ ตรวจเทียบกับคำตอบของผู้ตอบแต่ละตอนย่อยๆ จากนั้นจึงเอาคะแนนของแต่ละตอนย่อยๆ มารวมกันเป็นคะแนนที่ได้รับทั้งข้อ จุดอ่อนของการตรวจแบบวิธีวิเคราะห์ คือ เสียเวลาตรวจมากเพราะต้องตรวจทีละประเด็นและการแยกประเด็นย่อยมากๆ อาจวัดคุณลักษณะที่ไม่ตรงหรือไม่ลึกซึ้งในด้านจุดมุ่งหมายการเรียนการสอนที่ต้องการได้

2. การตรวจวิธีประเมินรวม หรือการตรวจคะแนนแบบจัดระดับคุณภาพ (Holistic Method หรือ Global Method หรือ Rating Method) เป็นวิธีการตรวจให้คะแนนรวมๆ ทั้งข้อ การให้คะแนนจะขึ้นอยู่กับคุณภาพของประเด็นหลัก สามารถตรวจได้รวดเร็ว การตรวจแบบนี้ผู้ตรวจจะอ่านคำตอบทุกฉบับอย่างรวดเร็วในขั้นต้น แล้วแยกกระดาษคำตอบเป็นกลุ่มๆ ตามระดับคุณภาพของการตอบเช่น แยกเป็นคุณภาพ 3 ระดับ คือ ดี ปานกลาง ไม่ดี หรือแยกเป็น 5 ระดับ คือ ดีมาก ดี ก่อนข้างดี พอใช้ อ่อน เมื่อจัดเรียงคำตอบเป็นกลุ่มเสร็จแล้วให้อ่านอย่างละเอียดทีละกลุ่มคุณภาพเพื่อให้คะแนนตามลำดับคุณภาพอีกทีหนึ่ง ในการจัดแยกกลุ่มกระดาษคำตอบอาจเปรียบเทียบคุณภาพระหว่างกันเองภายในกลุ่ม หรืออาจมีคำตอบตัวแบบ หรือกำหนดเกณฑ์ในการเปรียบเทียบคุณภาพคำตอบก็ได้

สำหรับการศึกษาค้นคว้าในครั้งนี้ ผู้วิจัยเลือกใช้วิธีการตรวจให้คะแนนแบบประเมินรวม (Holistic) และใช้เกณฑ์การให้คะแนนแบบรูบริก (Rubric Scoring) ซึ่งจะกล่าวถึงในหัวข้อต่อไป

7. เกณฑ์การให้คะแนนแบบรูบริก (Rubric scoring)

คอร์ทนี่ (Courtney. 1996 : 77-79) ได้อธิบายว่า

รูบริก (Rubric) คือ คำแนะนำเกี่ยวกับการให้คะแนน

รูบริก (Rubric) คือ ระบบการให้คะแนนบนพื้นฐานของเกณฑ์ ถ้ามีเกณฑ์การประเมิน 5 ระดับ นักเรียนจะได้ 5 คะแนนถ้าสามารถทำทุกอย่างตามที่ระบุไว้ในเกณฑ์

เกณฑ์การประเมินแบบรูบริก (Rubric Assessment) เน้นที่ความสามารถของนักเรียนในหัวข้อต่าง ๆ ดังนี้

1. ความเข้าใจในความคิดรวบยอด
2. ความรู้เกี่ยวกับกระบวนการหรือขั้นตอน
3. ทักษะการแก้ปัญหา / กลยุทธ์ในการแก้ปัญหา
4. ใช้เทคโนโลยีใหม่ๆ (เช่น คอมพิวเตอร์) ในการประมวลข้อมูล
5. การสื่อความหมายกับผู้อื่น

หลังจากตรวจผลงานของนักเรียนแล้ว ครูจะให้คะแนนตามเกณฑ์ 5 ระดับ ที่กล่าวข้างต้น

จากเอกสารการประเมินผลทางคณิตศาสตร์ (Mathematics Assessment)(Stenmark. 1991 : 24) ได้สรุปเกี่ยวกับเกณฑ์การให้คะแนนแบบรูบริกทั่วไป (General Rubric) ที่ใช้สำหรับคำถามแบบปลายเปิด (Open-ended Questions) ซึ่งปรับปรุงมาจาก California Assessment Program Grade 12 Generalized Rubric

เกณฑ์รูบริกเฉพาะ (Specific Rubric) เป็นเกณฑ์เฉพาะของคำถามแบบปลายเปิดแต่ละข้อ สำหรับเกณฑ์รูบริกทั่วไป (General Rubric) จะเป็นตัวแบบของเกณฑ์การให้คะแนน เพื่อใช้ในการพิจารณาเกณฑ์เฉพาะ รายละเอียดจะเปลี่ยนแปลงสำหรับระดับคะแนนที่แตกต่างกัน แต่แนวคิดพื้นฐานประยุกต์ได้ในทุกระดับ

ในการกำหนดเกณฑ์การให้คะแนนเพื่อใช้ในการตรวจให้คะแนนแบบประเมินรวม (Holistic) ขั้นแรกแบ่งข้อสอบออกเป็นกองต่างๆ ดังนี้ ข้อสอบที่มีคะแนนระดับสูงสุด ระดับกลาง และระดับต่ำ ถ้าต้องการเราสามารถเพิ่มกองในการแบ่งระดับคะแนนได้ ทุกระดับคะแนนจะอธิบายถึงการประเมินคำตอบของการแก้ปัญหาหนึ่งๆ โดยเฉพาะ ไม่ใช่การจัดหมวดหมู่ของนักเรียน ระดับคะแนนในแต่ละข้อไม่จำเป็นต้องเหมือนกันในการสร้างเกณฑ์รูบริกเฉพาะ (Specific Rubric)

เกณฑ์การให้คะแนนแบบรูบริกทั่วไป (General Rubric)

ระดับสูงสุด

- ประกอบด้วยคำตอบที่สมบูรณ์ ชัดเจน ปะติดปะต่อ ไม่คลุมเครือและอธิบายได้ดี
- รวมถึงแผนภาพที่ชัดเจนและไม่ซับซ้อน
- สื่อสารอย่างมีประสิทธิภาพเพื่อแสดงต่อผู้อ่าน
- แสดงความเข้าใจในแนวคิดและกระบวนการของคำตอบเกี่ยวกับคณิตศาสตร์
- ระบุส่วนประกอบที่สำคัญของปัญหาทั้งหมด
- รวมถึงตัวอย่างประกอบ
- ให้เหตุผลสนับสนุนที่มีน้ำหนัก
- ความต้องการที่นอกเหนือจากนี้ของปัญหา

ระดับที่สอง

- ประกอบด้วยคำตอบที่ดี แต่อาจมีบางส่วนไม่สมบูรณ์

- อธิบายได้ดีน้อยกว่า สมบูรณ์น้อยกว่า
- ไม่แสดงความต้องการของปัญหามากเกินไปจากนี้

ระดับที่สาม

- ประกอบด้วยคำตอบที่สมบูรณ์ แต่การอธิบายสับสน
- แสดงเหตุผลแต่ไม่สมบูรณ์
- รวมถึงแผนภาพไม่เหมาะสม หรือไม่ชัดเจน
- แสดงความเข้าใจในแนวคิดเกี่ยวกับคณิตศาสตร์ แต่ไม่ชัดเจนแน่นอน

ระดับที่สี่

- ยกเว้นส่วนที่น่าสนใจ มีความหมาย หรือทั้งหมดของคำถามและคำตอบ
- มีส่วนใหญ่ผิดพลาด
- ใช้กลยุทธ์ที่ไม่เหมาะสม

จากงานวิจัยของ ซูซาน เลน และคณะ (Zuzanne Lane, et al. 1996 : 89-91) ได้เสนอเกณฑ์การให้คะแนนแบบรูปรีดทั่วๆ ไป (General Rubric) ซึ่งพัฒนาโดยโปรแกรมการประเมินผลของแคลิฟอร์เนีย (California State Department of Education. 1989) เพื่อเป็นแนวทางในการสร้างเกณฑ์รูปรีดเฉพาะ (Specific Rubric) สำหรับการตรวจให้คะแนนด้วยวิธีประเมินรวม (Holistic) ไว้ 5 ระดับ คือ 0-4 คะแนน

เกณฑ์การให้คะแนนแบบรูปรีด (General Rubric)

ระดับคะแนน 4	
ความรู้ทางคณิตศาสตร์ :	แสดงความเข้าใจในแนวคิดและหลักการด้านคณิตศาสตร์ปัญหา; ใช้คำศัพท์เฉพาะ และสัญลักษณ์ (เครื่องหมาย) ทางคณิตศาสตร์อย่างเหมาะสม; ปฏิบัติตามขั้นตอนการคำนวณให้สมบูรณ์ถูกต้อง
ความรู้เกี่ยวกับการกลยุทธ์ :	ใช้ข้อมูลภายในให้ตรงประเด็น ตามคุณสมบัติที่เป็นแบบแผนและไม่เป็นแบบแผน; ระบุส่วนประกอบที่สำคัญทั้งหมดของปัญหาและแสดงความเข้าใจในความสัมพันธ์ระหว่างส่วนประกอบนั้น; พิจารณาความเหมาะสมและวิธีที่เป็นระบบสำหรับการแก้ปัญหา; แสดงหลักฐานอธิบายกระบวนการแก้ไขให้ชัดเจน และอธิบายกระบวนการให้สมบูรณ์และเป็นระบบ
การสื่อสารทางคณิตศาสตร์ :	อธิบายคำตอบให้สมบูรณ์ ชัดเจน ไม่คลุมเครือ; อาจจะมีแผนภาพประกอบที่สมบูรณ์ สื่อสารอย่างมีประสิทธิภาพ เพื่อชี้แจงผู้อ่าน (ผู้ตรวจ); แสดงความเชี่ยวชาญในการให้เหตุผลอย่างสมบูรณ์ อาจมีการยกตัวอย่างประกอบการให้เหตุผล
ระดับคะแนน 3	
ความรู้ทางคณิตศาสตร์ :	แสดงความเข้าใจในแนวคิดและหลักการด้านคณิตศาสตร์และปัญหาเกือบสมบูรณ์; ใช้คำศัพท์เฉพาะ และสัญลักษณ์ (เครื่องหมาย) ทางคณิตศาสตร์ถูกต้องเกือบหมด; ปฏิบัติตามขั้นตอนการคำนวณส่วนมากถูกต้อง แต่อาจมีความผิดพลาดอยู่เล็กน้อย

ระดับคะแนน 3 (ต่อ)	
ความรู้เกี่ยวกับกลยุทธ์ :	ใช้ข้อมูลภายนอกให้ตรงประเด็น ตามคุณสมบัติที่เป็นแบบแผนและไม่เป็นแบบแผน; ระบุส่วนประกอบที่สำคัญที่สุดของปัญหาและแสดงความเข้าใจทั่วไปของความสัมพันธ์ระหว่างส่วนประกอบนั้น; แสดงหลักฐานอธิบายกระบวนการแก้ไขได้ชัดเจน และอธิบายกระบวนการได้สมบูรณ์หรือเกือบสมบูรณ์และเป็นระบบ
การสื่อสารทางคณิตศาสตร์ :	อธิบายคำตอบให้สมบูรณ์ ชัดเจน ไม่คลุมเครือ; อาจจะมีแผนภาพประกอบที่สมบูรณ์หรือเกือบสมบูรณ์; การสื่อสารส่วนใหญ่มีประสิทธิภาพ เพื่อชี้แจงผู้อ่าน (ผู้ตรวจ); แสดงการสนับสนุนการให้เหตุผลอย่างเหมาะสม แต่อาจจะมีช่องว่างเล็กน้อย
ระดับคะแนน 2	
ความรู้ทางคณิตศาสตร์ :	แสดงความเข้าใจในแนวคิดและหลักการบางส่วนในคณิตศาสตร์ปัญหา; ใช้คำศัพท์เฉพาะ และสัญลักษณ์ (เครื่องหมาย) ทางคณิตศาสตร์ส่วนมากผิด; การคำนวณอาจผิดพลาด
ความรู้เกี่ยวกับกลยุทธ์ :	ระบุส่วนประกอบที่สำคัญของปัญหาได้บ้างแต่แสดงความเข้าใจในความสัมพันธ์ระหว่างส่วนประกอบนั้นได้เพียงเล็กน้อย; แสดงหลักฐานอธิบายกระบวนการแก้ไขได้บ้าง แต่การอธิบายกระบวนการแก้ไขอาจจะไม่สมบูรณ์หรือบางที่ไม่เป็นระบบ
การสื่อสารทางคณิตศาสตร์ :	อธิบายคำตอบไม่ชัดเจนหรือมีสองนัย; แผนภาพประกอบบกพร่องหรือไม่ชัดเจน; การสื่อสารคลุมเครือหรือตีความได้ยาก; การให้เหตุผลอาจไม่สมบูรณ์หรือไม่มีความสนับสนุน
ระดับคะแนน 1	
ความรู้ทางคณิตศาสตร์ :	แสดงความเข้าใจในแนวคิดและหลักการในคณิตศาสตร์ปัญหาได้น้อยมาก; ใช้คำศัพท์เฉพาะ และสัญลักษณ์ (เครื่องหมาย) ทางคณิตศาสตร์ผิด; การคำนวณผิดพลาด
ความรู้เกี่ยวกับกลยุทธ์ :	พยายามใช้ข้อมูลภายนอกที่ไม่ตรงประเด็น; ระบุส่วนประกอบที่สำคัญของปัญหาผิดหรือเน้นส่วนประกอบที่ไม่สำคัญมากเกินไป; แสดงหลักฐานอธิบายกระบวนการแก้ไขไม่สมบูรณ์หรือไม่เหมาะสม การอธิบายกระบวนการแก้ไขผิดพลาด หรือไม่เป็นระบบ
การสื่อสารทางคณิตศาสตร์ :	อธิบายคำตอบอาจจะผิดหรือเข้าใจยาก; แผนภาพประกอบไม่ถูกต้องตามสถานการณ์ปัญหา หรือแผนภาพไม่ชัดเจนตีความหมายยาก
ระดับคะแนน 0	
ความรู้ทางคณิตศาสตร์ :	แสดงความไม่เข้าใจในแนวคิดและหลักการในคณิตศาสตร์ปัญหา
ความรู้เกี่ยวกับกลยุทธ์ :	พยายามใช้ข้อมูลภายนอกที่ไม่ตรงประเด็น (ไม่เกี่ยวข้อง); ระบุส่วนประกอบของปัญหาผิด; ลอกส่วนปัญหาของโจทย์มา แต่ไม่พยายามแก้ปัญห
การสื่อสารทางคณิตศาสตร์ :	การสื่อสารไม่มีประสิทธิภาพ; คำที่ใช้ไม่เกี่ยวกับปัญหา; แผนภาพประกอบผิดหมด

ในการสร้างแบบทดสอบวิชาคณิตศาสตร์ที่มีลักษณะคำถามแบบปลายเปิด ที่ผู้วิจัยทำการศึกษาค้นคว้านั้น ได้ยึดตามแนวการสร้างแบบทดสอบประเมินผลงานของนักเรียนด้วยคำถามแบบปลายเปิดของ ชูชาน เลน และคณะ (Zuzanne Lane, et al. 1996 : 71-91) และในเอกสารประเมินผลทางคณิตศาสตร์ (Mathematics Assessment) (Stenmark. 1991 : 24) และเกณฑ์การให้คะแนนแบบรูบริก (Rubric) ปรับปรุงมาจากการกำหนดเกณฑ์การให้คะแนนแบบรูบริกทั่วไป (General Rubric) ของชูชาน เลน และคณะ และในเอกสารการประเมินผลทางคณิตศาสตร์ เช่นกัน

8. งานวิจัยที่เกี่ยวข้อง

งานวิจัยที่เกี่ยวกับทฤษฎีการสรุปอ้างอิง

สมิท (Smith. 1978 : 319-346) ได้ศึกษาถึงการประมาณค่าความคลาดเคลื่อนในการสุ่มกลุ่มตัวอย่างของการศึกษาการสรุปอ้างอิงชนิดหลายตัวประกอบ (Multifacet Study) ที่ใช้กลุ่มตัวอย่างจำนวนน้อยๆ พบว่าความคลาดเคลื่อนพบมากที่สุดเป็นหลัก (Main Effects) และได้เสนอวิธีลดความคลาดเคลื่อนเชิงสุ่มไว้ 3 วิธีคือ 1) สุ่มระดับขององค์ประกอบแต่ละตัวที่น่าสนใจในการศึกษา G Study ให้ได้จำนวนมากที่สุดเท่าที่จะทำได้ 2) สุ่มระดับของผลลัพธ์อื่นๆ ให้เป็นสัดส่วนของเอกภพขององค์ประกอบ และ 3) เปลี่ยนรูปแบบความสัมพันธ์ระหว่างองค์ประกอบ เพื่อลดจำนวนเทอมให้เท่ากับข้อ (2) และควรใช้เมื่อไม่สามารถสุ่มกลุ่มตัวอย่างให้ได้จำนวนมากๆ นอกจากนี้ สมิทได้เสนอทางเลือกหนึ่งเพื่อศึกษา G Study ซึ่งประมาณค่าความคลาดเคลื่อนขององค์ประกอบทั้งหมดที่สนใจในเวลาเดียวกัน ได้จัดให้เพิ่มข้อมูลเพื่อควบคุมผลของปฏิสัมพันธ์ขององค์ประกอบว่า ถ้าเป็นไปได้ ขนาดของกลุ่มตัวอย่างที่สุ่มมาสำหรับใช้ในการศึกษา 3 องค์ประกอบ ควรมีขนาด $n_p \times n_i \times n_j$ อย่างน้อย 1,000 ค่าสังเกต

สมิท (Smith. 1979 : 77-87) ได้ใช้ทฤษฎีการสรุปอ้างอิงตรวจสอบความเชื่อถือได้ของการจัดลำดับนักเรียน สำหรับการประเมินเกี่ยวกับหลักสูตรและการสอน งานวิจัยนี้สรุปผลตรงกันข้ามกับการศึกษาที่ผ่านมา คือ ความเชื่อถือได้ของการประเมินหลักสูตรสามารถทำได้บนพื้นฐานของการจัดลำดับนักเรียน เขาเสนอแนะว่าความขัดแย้งระหว่างผลที่เสนอกับงานวิจัยที่ผ่านมา โดยส่วนใหญ่เนื่องมาจากความแตกต่างในเอกภพของการสังเกตที่ยอมรับได้ของการศึกษาทั้ง 2 ครั้ง ดังนั้น เราควรให้ความสำคัญกับการพิจารณาเอกภพของการสังเกตที่ยอมรับได้ในทฤษฎีการสรุปอ้างอิง

มาเคลดี (Macready. 1983 : 149-157) ได้ประยุกต์ทฤษฎีการสรุปอ้างอิงในการประเมินความยากและความเป็นเอกพันธ์ (Homogeneity) ของข้อสอบอินโดเมนที่ใช้ในการวินิจฉัย โดยมีองค์ประกอบที่ศึกษาประกอบด้วย โดเมนการคูณจำนวนเต็ม (d) ห้องเรียน (c) จำนวนหลักของตัวคูณ (n) ข้อสอบซึ่งแบ่งอยู่ในโดเมนและจำนวนหลัก $i:(d \times n)$ และนักเรียนซึ่งแบ่งอยู่ในชั้นเรียน ($s:c$) จากรูปแบบการวิเคราะห์ G Study แบบ $(s:c) \times (i:(d \times n))$ โดยใช้ห้องเรียน นักเรียน และข้อสอบเป็นองค์ประกอบสุ่มให้โดเมนและจำนวนหลักของตัวคูณเป็นองค์ประกอบกำหนด พบว่า $(s:c) \times (i:(d \times n))$ เป็นแหล่งความแปรปรวนที่มีค่ามากที่สุด คือ 47% อีก 4 แหล่งที่มีค่ารองลงไป ได้แก่ ห้องเรียน นักเรียนซึ่งแบ่งอยู่ในห้องเรียน โดเมน และผลรวมระหว่างนักเรียนกับโดเมน $(s:c) \times d$ โดยทั้ง 4 แหล่งมีค่ารวมกันถึง 51% เฉพาะแหล่งความแปรปรวนสุดท้าย หมายถึง ความยากของข้อสอบในแต่ละโดเมนสำหรับนักเรียนแต่ละคน

มีค่าไม่เท่ากัน แหล่งความแปรปรวนอื่นๆ ที่เหลือมีค่าน้อยมาก โดยเฉพาะจำนวนหลักของตัวคูณ มาเคลดีสรุปว่าไม่ควรจะใช้ตัวคูณ 4 ตำแหน่ง ใช้เพียง 3 ตำแหน่งก็พอเพราะให้ค่าความยากไม่แตกต่างกัน อีกแหล่งหนึ่งที่มีค่าน้อย คือ $(i:(d \times n))$ แสดงว่าความยากของข้อสอบในทุกโดเมนที่มีตำแหน่งตัวคูณเท่ากันมีค่าพอกๆ กัน เมื่อตรวจดูความยากรายโดเมน พบว่าเกือบทุกโดเมนมีข้อสอบที่มีความยากเท่าเทียมกัน มีเพียงโดเมนที่ 15 ที่ค่อนข้างแตกต่างกัน มาเคลดีเสนอว่าควรจะแยกให้เป็นโดเมนย่อยหรือนำไปรวมกับโดเมนอื่นจากการประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงของข้อสอบแต่ละโดเมน โดยใช้รูปแบบการวิเคราะห์ 2 รูปแบบ คือ $s \times i$ และ $s \times r$ เมื่อ r หมายถึง การสอบซ้ำ พบว่าค่าสัมประสิทธิ์การสรุปอ้างอิงของข้อสอบหนึ่งข้อมีค่าอยู่ระหว่าง 0.338 ถึง 0.606

โอบราฮิม (Ibrahim, 1984 : 499-A) ได้ประยุกต์ทฤษฎีการสรุปอ้างอิง เพื่อประมาณค่าความแปรปรวนที่มีต่อการประเมินวัตถุประสงค์ทางการศึกษา (The Rating of Evaluational Goals) โดยสุ่มกลุ่มตัวอย่างครู 80 คน และนักเรียน 80 คน ในประเทศชูดาน ประเมินวัตถุประสงค์ทางการศึกษา 2 ชนิด คือ วัตถุประสงค์ที่สำคัญจริงๆ และวัตถุประสงค์ตามที่คาดหวัง องค์ประกอบในการศึกษามีผู้ประเมิน กลุ่มผู้ประเมิน จำนวนครั้งของการประเมิน ถิ่นที่อยู่ของผู้ประเมิน ชนิดของวัตถุประสงค์ สถานที่ทำงานของผู้ประเมิน และเพศของผู้ประเมิน พบว่า องค์ประกอบที่มีผลต่อการประเมินมากที่สุด ได้แก่ ผู้ประเมินและกลุ่มผู้ประเมิน ส่วนสถานที่ทำงานของผู้ประเมินมีผลเล็กน้อยเท่านั้นส่วนองค์ประกอบอื่นที่เหลือไม่มีผลต่อการประเมินเลย

คุนนอน (Kunnan, 1992 : 30-47) ได้ทำการศึกษาวิธีการวิเคราะห์ 3 ชนิดโดยแบ่งข้อสอบเป็น 3 ฉบับ คือ การฟัง การอ่าน และศัพท์ ไวยากรณ์ ฉบับละ 30 ข้อเท่าๆ กันทุกฉบับ เป็นแบบทดสอบชนิดเลือกตอบ และแบ่งกลุ่มตัวอย่างออกเป็น 4 กลุ่มตามความสามารถในการสอบและการคำนวณหาค่าความเชื่อมั่นโดยทฤษฎีการสรุปอ้างอิง โดยใช้โปรแกรมสำเร็จรูป GENOVA ซึ่งออกแบบแผนเป็น $p \times (i:t)$ คือ ข้อสอบที่แบ่งอยู่ในแบบทดสอบซึ่งเป็นโดเมนผสม และการคำนวณความเที่ยงตรงโดยใช้วิเคราะห์องค์ประกอบและการวิเคราะห์กลุ่มใช้โปรแกรมสำเร็จรูป SPSS-X โดยแบ่งแบบทดสอบแต่ละฉบับออกเป็น 2 ฉบับ ผลของการวิจัยพบว่า ความเชื่อมั่นของแบบทดสอบแตกต่างกันทุกๆ กลุ่ม และค่าจะสูงในกลุ่มที่มีความสามารถต่ำได้ดีกว่ากลุ่มที่มีความสามารถสูง

เบเกอร์ และคณะ (Baker, et al. 1995 : 197-205) ได้ทำการศึกษาเพื่อประเมินระดับซึ่งเป็นการควบคุมหัวเรื่อง (Topic) และความไม่คงที่ของผู้ประเมิน, มุ่งประเด็นไปที่การสรุปอ้างอิง, ความเชื่อมั่นของผู้ประเมิน และมิติของเกณฑ์การให้คะแนนแบบรูปรีด (Rubric Scoring) นักเรียนที่เรียนวิชาประวัติศาสตร์ 2 ชั้นเรียน ถูกจัดให้ทำกิจกรรมภาคปฏิบัติ 3 กิจกรรมๆ ละ 1 สัปดาห์ ในแต่ละหัวเรื่อง นักเรียนทั้งหมดต้องทำแบบทดสอบเกี่ยวกับความรู้เดิม, อ่านเนื้อเรื่องที่กำหนดให้ และเขียนข้อเขียนสั้นๆ ใช้เกณฑ์การให้คะแนนแบบรูปรีด อบรมเกี่ยวกับการให้คะแนนข้อเขียนกับผู้ประเมิน 4 คนศึกษาค่าความเชื่อมั่นทั้งภายในและภายนอกผู้ประเมิน และศึกษาผลของ G Study จากผลที่ได้แสดงความสัมพันธ์อย่างมีประสิทธิภาพของการประเมิน การวิเคราะห์มิติสนับสนุน 2 องค์ประกอบ คือ ความเข้าใจที่ลึกซึ้ง และความเข้าใจแบบผิวเผิน ในทั้ง 3 หัวเรื่อง คะแนนความรู้เดิมและ GPA ในวิชาประวัติศาสตร์ มีความสัมพันธ์กับส่วนประกอบของความเข้าใจที่ลึกซึ้งในเกณฑ์การให้คะแนนแบบรูปรีด

ซูซาน เลน และคณะ (Zuzanne Lane, et al. 1996 : 71-91) ได้ศึกษาค่าสัมประสิทธิ์การสรุปอ้างอิงและความเที่ยงตรงของการประเมินผลงานของนักเรียนในวิชาคณิตศาสตร์ที่พัฒนาจากเครื่องมือประเมินความรู้แบบ QUASAR ชื่อว่า QCAI เป็นชุดของข้อสอบแบบปลายเปิด (Open-ended) โดยเน้นที่วิธีการตรวจ

ให้คะแนนแบบการประเมินรวม (Holistic) และมีเกณฑ์การให้คะแนนแบบรูบริค (Rubric Scoring) แล้วจึงใช้ทฤษฎีการอ้างอิงตรวจสอบความสอดคล้องภายในข้อสอบและภายในผู้ตรวจ โดยมีแบบแผน 3 รูปแบบคือ 1) $p \times t$ ผู้สอบทุกคนทำข้อสอบทุกข้อ 2) $p \times t \times r$ ผู้สอบทุกคนทำข้อสอบทุกข้อและผู้ตรวจแต่ละคนตรวจข้อสอบทุกข้อ 3) $(p:s) \times t$ ผู้สอบที่แฝงอยู่ในโรงเรียนทำข้อสอบทุกข้อ โดยใช้โปรแกรม GENOVA ในการคำนวณ ผลจากการศึกษาพบว่า ความคลาดเคลื่อนเนื่องจากผู้ตรวจมีน้อยมาก เนื่องมาจากการให้รายละเอียดเกี่ยวกับเกณฑ์การให้คะแนนแบบรูบริคและตัวอย่างการตอบของนักเรียน รวมทั้งการฝึกฝนวิธีการตรวจให้คะแนนกับผู้ตรวจถูกต้องเพียงพอ มีความแตกต่างที่ต้องคำนึงถึงในการทำข้อสอบของนักเรียนในแต่ละข้อ ค่าความเชื่อมั่นที่ได้จากการศึกษา D Study แบบสมบูรณ์พบว่าเมื่อจำนวนนักเรียนเท่ากับ 350 คน ค่าสัมประสิทธิ์อยู่ระหว่าง 0.80 - 0.97 ขึ้นอยู่กับชุดของข้อสอบ และระดับชั้น

แดง กลางทำไค (2531) ได้ประยุกต์ทฤษฎีการสรุปอ้างอิงในการหาค่าความเชื่อมั่นของการประเมินความเที่ยงตรงเชิงเนื้อหา รูปแบบการวัดในการศึกษานั้น ผู้เชี่ยวชาญทุกคนประเมินความเที่ยงตรง (ความเหมาะสม) ของข้อสอบทุกข้อ ว่าวัดตรงจุดประสงค์เพียงใด รูปแบบการวัดในการศึกษาค้างนี้ คือ $i \times (r:s)$ โดยผู้เชี่ยวชาญทุกคนประเมินข้อสอบทุกข้อ และผู้เชี่ยวชาญถูกสุ่มมาจากโรงเรียนแต่ละโรงเรียนจำนวนเท่าๆ กัน เครื่องมือที่ใช้ในการวิจัยประกอบด้วย แบบทดสอบวัดผลสัมฤทธิ์ทางเรียนวิชาคณิตศาสตร์ เรื่องเซต จำนวน 30 ข้อ และแบบประเมินความเที่ยงตรงเชิงเนื้อหาชนิด 5 มาตรา ผลการศึกษาพบว่า แหล่งความแปรปรวนที่มีอิทธิพลต่อความเชื่อมั่น ได้แก่ ความแปรปรวนของข้อสอบ ความแปรปรวนของผู้เชี่ยวชาญซึ่งแฝงอยู่ในโรงเรียน และผลร่วมระหว่างข้อสอบกับโรงเรียน ส่วนรูปแบบการวัดที่ให้ค่าความเชื่อมั่นในการประเมินความเที่ยงตรงเชิงเนื้อหาโดยผู้เชี่ยวชาญที่สูงกว่ารูปแบบการวัดอื่นๆ ภายใต้ความสัมพันธ์ขององค์ประกอบแบบ $i \times (r:s)$ คือ $M(I-I/S/R)$ ส่วนขนาดของกลุ่มตัวอย่างที่จะเป็นตัวแทนของสมาชิกทั้งหมดในเอกภพขององค์ประกอบพบว่า ข้อสอบอย่างน้อย 9 ข้อ และผู้เชี่ยวชาญไม่เกิน 9 คนต่อโรงเรียน สุ่มจากโรงเรียนอย่างน้อย 7 โรงเรียน จึงจะให้ค่าสัมประสิทธิ์การสรุปอ้างอิงอย่างน้อย 0.80

ไพรัตน์ วงษ์นาม (2533) ได้ศึกษาและเปรียบเทียบผลของวิธีตรวจ การชี้แจงและการรู้ผลการเรียนของผู้ตอบ ที่มีต่อค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบความเรียง วัดความสามารถในการแสดงความคิดเห็นเกี่ยวกับข่าวและเหตุการณ์ ของนักเรียนชั้นประถมศึกษาปีที่ 6 พร้อมทั้งเลือกวิธีตรวจ จำนวนผู้ตรวจ และจำนวนข้อสอบที่ให้ค่าสัมประสิทธิ์การสรุปอ้างอิง 0.50 เป็นอย่างต่ำ ผลการวิจัยพบว่า เมื่อสรุปอ้างอิงไปยังเอกภพของข้อสอบและผู้ตรวจพร้อมกัน (M1) ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบที่ตรวจโดยวิธีประเมินรวมมีค่าอยู่ระหว่าง 0.3328 - 0.4782 และตรวจด้วยวิธีวิเคราะห์มีค่าอยู่ระหว่าง 0.3348 - 0.5895 เมื่อสรุปอ้างอิงไปยังเอกภพของผู้ตรวจเพียงอย่างเดียว (M2) ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบที่ตรวจโดยวิธีประเมินรวมมีค่าอยู่ระหว่าง 0.4743 - 0.6865 และตรวจด้วยวิธีวิเคราะห์มีค่าอยู่ระหว่าง 0.5985 - 0.7761 วิธีตรวจการชี้แจงและการรู้ผลการเรียนของผู้ตอบต่างกันไม่มีผลต่อความแตกต่างของค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับ M1 แต่ใน M2 พบว่าค่าสัมประสิทธิ์การสรุปอ้างอิงของการตรวจตามที่ได้รับการชี้แจงมีค่าสูงกว่าการตรวจโดยใช้ประสบการณ์เดิม เฉพาะกรณีที่ผู้ตรวจที่ได้รับการฝึกอบรมไม่รู้ผลการเรียนของผู้ตอบเท่านั้น ถ้ารู้ผลการเรียนของผู้ตอบค่าสัมประสิทธิ์จะไม่แตกต่างกัน ค่าสัมประสิทธิ์ของการตรวจที่ผู้ตรวจไม่รู้ผลการเรียนของผู้ตอบ มีค่าสูงกว่าการตรวจที่รู้ผลการเรียนของผู้ตอบ วิธีตรวจวิเคราะห์ตามที่ได้รับการชี้แจงและผู้ตรวจไม่รู้ผลการเรียนของผู้ตอบ ใช้ข้อสอบ 6 ข้อ ผู้ตรวจอย่างน้อย 5 คน สำหรับ M1 และผู้ตรวจ 5 คน ใช้ข้อสอบ 5 ข้อ สำหรับ M2

ชาญ วชิรเดช (2538) ได้ศึกษาหารูปแบบของแบบทดสอบการแก้ปัญหาที่สอบด้วยไมโครคอมพิวเตอร์ กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 4 มีแบบแผนการวิเคราะห์แบบผสม คือ $p \times (i:c)$ นักเรียนทุกคนทำข้อสอบทุกข้อที่แฝงอยู่ในเงื่อนไขสถานการณ์ของโจทย์ปัญหาที่กำหนด เครื่องมือที่ใช้วิจัยเป็นแบบทดสอบความสามารถในการแก้ปัญหาที่สอบด้วยไมโครคอมพิวเตอร์จำนวน 3 รูปแบบ ผลการวิจัยพบว่าความแปรปรวนของผลร่วมระหว่างนักเรียนทุกคนทำข้อสอบทุกข้อที่แฝงอยู่ในเงื่อนไขสถานการณ์ของโจทย์ปัญหาที่กำหนด มีผลต่อค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบการแก้ปัญหาทั้งสามรูปแบบมากที่สุด ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบแก้ปัญหาที่สอบด้วยไมโครคอมพิวเตอร์ทั้งสามรูปแบบ มีค่าสัมประสิทธิ์การสรุปอ้างอิงประมาณที่ใกล้เคียงกับพารามิเตอร์ของความเชื่อมั่นที่แท้จริง

จากงานวิจัยที่กล่าวมาข้างต้น กล่าวได้ว่า ทฤษฎีการสรุปอ้างอิงสามารถนำมาใช้กับการวัดและประเมินผลที่ได้เป็นอย่างดีในหลายกรณี ทั้งการศึกษาคุณภาพของแบบทดสอบ ศึกษาแหล่งความแปรปรวนที่มีผลต่อการประเมิน รวมทั้งการตีความหมายคะแนนด้วย

สมมติฐานการศึกษาครั้งนี้ว่า

จากการศึกษาทฤษฎีและงานวิจัยที่เกี่ยวข้องกับสัมประสิทธิ์การสรุปอ้างอิงที่ได้เสนอไว้ สามารถกำหนดสมมติฐานตามจุดมุ่งหมายของการศึกษาค้นคว้าได้ดังนี้

1. ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อมีจำนวนผู้ตรวจเท่ากัน มีค่าแตกต่างกัน
2. ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อผู้ตรวจมีประสบการณ์เหมือนกัน มีค่าแตกต่างกัน
3. ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีจำนวนผู้ตรวจต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน มีค่าแตกต่างกัน
4. ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่ผู้ตรวจมีประสบการณ์แตกต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน มีค่าแตกต่างกัน

บทที่ 3

วิธีดำเนินการศึกษาค้นคว้า

ประชากร

ประชากรที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2540 ของโรงเรียนในสังกัดกรมสามัญศึกษา จังหวัดจันทบุรี จำนวน 20 โรงเรียน จำนวนห้องเรียน 112 ห้องเรียน รวมทั้งหมด 4,436 คน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าครั้งนี้ เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2540 ของโรงเรียนในสังกัดกรมสามัญศึกษา จังหวัดจันทบุรี จำนวน 4 โรงเรียน จำนวนห้องเรียน 60 คน ซึ่งได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi-Stage Random Sampling) มีขั้นตอนการสุ่มดังนี้

ขั้นที่ 1 สุ่มโรงเรียน แบ่งตามขนาดโรงเรียน มีทั้งหมด 4 ขนาด โดยวิธีการสุ่มอย่างง่ายขนาดละ 1 โรงเรียน ได้โรงเรียนทั้งหมด 4 โรงเรียน

ขั้นที่ 2 สุ่มห้องเรียนจากโรงเรียนที่สุ่มได้ โรงเรียนละ 1 ห้องเรียน

ขั้นที่ 3 สุ่มนักเรียนจากแต่ละห้องตามสัดส่วนจำนวนนักเรียน ได้นักเรียน 60 คน ดังแสดงในตารางต่อไปนี้

ตาราง 2 จำนวนนักเรียนที่ใช้เป็นกลุ่มตัวอย่างในการเก็บรวบรวมข้อมูล

ขนาดโรงเรียน	โรงเรียน	จำนวนนักเรียนในห้องเรียนที่สุ่มได้	จำนวนนักเรียนที่สุ่มมาทำการวิจัย
ใหญ่พิเศษ	ศรียานุสรณ์	45	15
ใหญ่	เบญจมานุสรณ์	45	15
กลาง	นายายอามพิทยาคม	45	15
เล็ก	ศิษย์คุณวิทย์	45	15
รวม		180	60

หมายเหตุ เนื่องจากการศึกษาค้นคว้าครั้งนี้ใช้ผู้ตรวจและวิธีการตรวจต่างกัน ดังนั้นการกำหนดขนาดของกลุ่มตัวอย่างจึงคำนึงถึงความเหมาะสม ซึ่งเกี่ยวข้องกับผู้ตรวจและสามารถนำไปใช้ประมาณค่าสัมประสิทธิ์การสุ่มอย่างอิสระได้อย่างมั่นใจ สมิท (Smith, 1978 : 319 - 346) ได้เสนอว่าขนาดของกลุ่มตัวอย่างขั้นต่ำที่ใช้ในการศึกษา 3 องค์ประกอบควรมีขนาด $n_p \times n_i \times n_r$ อย่างน้อย 1,000 ค่า ดังนั้นผู้วิจัยจึงกำหนดจำนวนนักเรียน (n_p) 60 คน จำนวนข้อสอบ (n_i) 6 ข้อ และจำนวนผู้ตรวจ (n_r) 6 คน จะได้ขนาดของกลุ่มตัวอย่างเท่ากับ $60 \times 6 \times 6 = 2,160$ ค่า ซึ่งเป็นไปตามเกณฑ์ของสมิท

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้า เป็นแบบทดสอบวิชาคณิตศาสตร์ที่ลักษณะคำถามแบบปลายเปิด ซึ่งลักษณะของแบบทดสอบเป็นสถานการณ์ปัญหาทางคณิตศาสตร์ที่กำหนดให้ เพื่อวัดความสามารถของนักเรียน 3 ด้าน คือ

- | | |
|--|-------------|
| 1. การแก้ปัญหาในวิชาคณิตศาสตร์ | จำนวน 2 ข้อ |
| 2. คณิตศาสตร์เหตุผล | จำนวน 2 ข้อ |
| 3. การสื่อสารทางคณิตศาสตร์ | จำนวน 2 ข้อ |
| รวมเป็นข้อสอบ 1 ฉบับ จำนวน 6 ข้อ ใช้เวลา 60 นาที | |

วิธีดำเนินการสร้างเครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้าครั้งนี้เป็นแบบทดสอบวิชาคณิตศาสตร์ที่มีลักษณะคำถามแบบปลายเปิด ที่ผู้วิจัยสร้างขึ้น มีขั้นตอนการสร้างดังรายละเอียดต่อไปนี้

1. กำหนดจุดมุ่งหมายในการสร้างแบบทดสอบ
 - 1.1 เพื่อสร้างแบบทดสอบวิชาคณิตศาสตร์ที่มีลักษณะคำถามแบบปลายเปิด ซึ่งวัดความสามารถในการแก้ปัญหาในวิชาคณิตศาสตร์ คณิตศาสตร์เหตุผล และการสื่อสารทางคณิตศาสตร์
 - 1.2 เพื่อนำคะแนนจากแบบทดสอบไปใช้ในการศึกษาเปรียบเทียบค่าสัมประสิทธิ์การสุ่มอย่างอิสระ
2. ศึกษาทฤษฎีและเอกสารงานวิจัยที่เกี่ยวข้อง
 - 2.1 ศึกษาทฤษฎีและเอกสารงานวิจัยที่เกี่ยวข้องกับคำถามแบบปลายเปิด (Open-ended Questions) เพื่อเป็นแนวทางในการสร้างแบบทดสอบวิชาคณิตศาสตร์
 - 2.2 ศึกษาวิธีการสร้างและเกณฑ์การตรวจให้คะแนนคำถามแบบปลายเปิด เพื่อเป็นแนวทางในการสร้างแบบทดสอบ
3. เขียนนิยามปฏิบัติการของแบบทดสอบวิชาคณิตศาสตร์ เพื่อเป็นแนวทางในการสร้างแบบทดสอบ
4. เขียนข้อสอบตามนิยามเชิงปฏิบัติการ ซึ่งมีรายละเอียดดังนี้
 - 4.1 ลักษณะของแบบทดสอบที่สร้างขึ้น มีลักษณะคำถามแบบปลายเปิด โดยผู้สอบพิจารณา

สถานการณ์ปัญหาที่กำหนดให้ แล้วเขียนคำตอบแสดงแนวคิดและหลักการ วิธีดำเนินการ และขั้นตอนการคำนวณ ที่ถูกต้องเหมาะสม จากสถานการณ์ปัญหาที่กำหนดให้

4.2 สร้างข้อคำถามให้ตรงตามนิยามเชิงปฏิบัติการของแบบทดสอบวิชาคณิตศาสตร์ จำนวน 12 ข้อ โดยแบ่งเป็น

4.2.1 การแก้ปัญหาในวิชาคณิตศาสตร์ จำนวน 4 ข้อ

4.2.2 คณิตศาสตร์เหตุผล จำนวน 4 ข้อ

4.2.3 การสื่อสารทางคณิตศาสตร์ จำนวน 4 ข้อ

4.3 แบ่งข้อคำถามที่สร้างขึ้นออกเป็น 2 ชุดๆ ละ 6 ข้อ

4.4 สร้างเกณฑ์ในการตรวจให้คะแนนแบบรูบริก (General rubric) ซึ่งปรับปรุงมาจากการกำหนดเกณฑ์การให้คะแนนแบบรูบริกของ ซูซาน เลน และคณะ (Zuzanne Lane, et al. 1996 : 71-91) และจากเอกสารการประเมินผลทางคณิตศาสตร์ (Mathematics Assessment) (Stenmark, J.K. 1991 : 24) ดังนี้

ระดับคะแนน 4

- อธิบายความเข้าใจในแนวคิดและกระบวนการในการหาคำตอบได้สมบูรณ์
- ปฏิบัติตามขั้นตอนการคำนวณได้สมบูรณ์ถูกต้อง
- อธิบายคำตอบได้สมบูรณ์ ชัดเจน ไม่คลุมเครือ
- ให้เหตุผลสนับสนุนการอธิบาย และยกตัวอย่างประกอบได้สมบูรณ์ถูกต้อง
- แผนภาพประกอบสมบูรณ์
- ใช้คำศัพท์เฉพาะ และสัญลักษณ์ (เครื่องหมาย) ทางคณิตศาสตร์ ได้อย่างถูกต้อง เหมาะสม
- สื่อสารกับผู้อ่านได้อย่างดี เมื่ออ่านแล้วตีความ / ทำความเข้าใจได้ตรงกัน

ระดับคะแนน 3

- อธิบายความเข้าใจในแนวคิดและกระบวนการในการหาคำตอบได้เกือบสมบูรณ์
- ปฏิบัติตามขั้นตอนการคำนวณได้ถูกต้อง แต่อาจมีความผิดพลาดในการคำนวณเล็กน้อย
- อธิบายคำตอบได้ชัดเจน ไม่คลุมเครือ แต่อาจมีบางส่วนไม่สมบูรณ์หรือไม่ได้อธิบาย
- ให้เหตุผลสนับสนุนการอธิบาย และยกตัวอย่างประกอบได้เกือบสมบูรณ์ อาจมีช่องว่าง
- แผนภาพประกอบเกือบสมบูรณ์ อาจมีข้อบกพร่องบ้างเล็กน้อย
- ใช้คำศัพท์เฉพาะ และสัญลักษณ์ (เครื่องหมาย) ทางคณิตศาสตร์ ถูกต้องเกือบหมด
- สื่อสารกับผู้อ่านได้พอสมควร อาจมีบางส่วนที่ไม่ชัดเจนอยู่บ้าง

ระดับคะแนน 2

- อธิบายความเข้าใจในแนวคิดและกระบวนการในการหาคำตอบได้บางส่วน
- ปฏิบัติตามขั้นตอนการคำนวณได้ถูกต้องเพียงครั้งหนึ่งของขั้นตอนการคำนวณทั้งหมด
- อธิบายคำตอบไม่ชัดเจน หรือมีสองนัย
- ให้เหตุผลสนับสนุนการอธิบายไม่สมบูรณ์ หรือไม่มีการยกตัวอย่างประกอบการอธิบาย
- แผนภาพประกอบไม่ชัดเจน หรือมีข้อบกพร่องมาก
- ใช้คำศัพท์เฉพาะ และสัญลักษณ์ (เครื่องหมาย) ทางคณิตศาสตร์ ได้เพียงบางส่วน
- การสื่อสารกับผู้อ่าน คลุมเครือ อ่านแล้วตีความ / ทำความเข้าใจได้ยาก

ระดับคะแนน 1
- อธิบายความเข้าใจในแนวคิดและกระบวนการในการหาคำตอบได้น้อยมาก - ขั้นตอนการคำนวณมีที่ผิดพลาดมาก - อธิบายคำตอบผิด หรือเข้าใจยาก - ให้เหตุผลสนับสนุนการอธิบายได้น้อยมาก หรือยกตัวอย่างประกอบการอธิบาย ไม่เหมาะสม - แผนภาพประกอบส่วนใหญ่ไม่ถูกต้อง หรือตีความได้น้อย - ใช้คำศัพท์เฉพาะ และสัญลักษณ์ (เครื่องหมาย) ทางคณิตศาสตร์ส่วนมากผิด - การสื่อสารกับผู้อ่านไม่ดี ผู้อ่านๆ แล้วไม่ค่อยเข้าใจ
ระดับคะแนน 0
- ไม่สามารถอธิบายความเข้าใจในแนวคิดและกระบวนการในการหาคำตอบได้ - ไม่มีการคำนวณ - อธิบายคำตอบไม่ตรงประเด็น - ไม่ให้เหตุผลสนับสนุนการอธิบาย หรือไม่ยกตัวอย่างประกอบการอธิบาย - แผนภาพประกอบผิดหมด - ใช้คำศัพท์เฉพาะ และสัญลักษณ์ (เครื่องหมาย) ทางคณิตศาสตร์ที่ไม่เกี่ยวกับปัญหา - ไม่สามารถสื่อสารกับผู้อ่านให้เข้าใจตรงกันได้

5. ตรวจสอบคุณภาพขั้นต้น โดยตรวจสอบความเที่ยงตรงเชิงพินิจ (Face Validity) ของแบบทดสอบ รวมทั้งเกณฑ์ในการตรวจให้คะแนนแบบรูบริก (Rubric scoring) โดยผู้เชี่ยวชาญทางการวัดผล ในสาขาวิชาคณิตศาสตร์ จำนวน 3 ท่าน และสาขาวิชาภาษาไทย จำนวน 2 ท่าน โดยพิจารณาจากค่า IOC ที่มีค่าตั้งแต่ 0.5 ขึ้นไป ถือว่าเป็นข้อสอบที่มีความเที่ยงตรงเชิงพินิจ จากผลการตรวจสอบปรากฏว่า ข้อสอบจำนวน 12 ข้อ ได้รับการพิจารณาว่ามีความสอดคล้องกับเนื้อหา และความสามารถทางคณิตศาสตร์ทั้ง 3 ด้านที่ต้องการวัด ทุกข้อ โดยมีค่า IOC อยู่ระหว่าง 0.6 – 1 และเกณฑ์ในการตรวจให้คะแนนแบบรูบริกก็ได้รับการพิจารณา และปรับปรุงแก้ไขข้อความบางส่วนให้ถูกต้องเหมาะสมยิ่งขึ้น

6. นำข้อสอบที่ผ่านการพิจารณาและได้รับการปรับปรุงแก้ไขแล้ว แบ่งออกเป็น 2 ฉบับๆ ละ 6 ข้อ โดยแต่ละฉบับจะแบ่งข้อสอบเป็น 3 ด้านๆ ละ 2 ข้อ ดังนี้

- การแก้ปัญหาในวิชาคณิตศาสตร์ จำนวน 2 ข้อ
- คณิตศาสตร์เหตุผล จำนวน 2 ข้อ
- การสื่อสารทางคณิตศาสตร์ จำนวน 2 ข้อ

จากนั้นจึงนำแบบทดสอบทั้ง 2 ฉบับ ไปทดลองครั้งที่ 1

6. ทดลองครั้งที่ 1 นำแบบทดสอบวิชาคณิตศาสตร์ทั้ง 2 ฉบับไปทดลองสอบกับกลุ่มตัวอย่างซึ่งเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ที่ได้จากการสุ่มอย่างง่ายจากประชากรได้โรงเรียนเบญจมราชูทิศ จังหวัดจันทบุรี ซึ่งมีนักเรียนชั้นมัธยมศึกษาปีที่ 1 อยู่ 10 ห้องเรียน สุ่มห้องเรียนมา 2 ห้องเรียน และให้นักเรียนห้องละ 30 คน ให้นักเรียนทำแบบทดสอบห้องละ 1 ฉบับ โดยในแต่ละห้องให้นักเรียนทำข้อสอบทีละข้อเพื่อตรวจสอบเวลาที่เหมาะสมสำหรับใช้ในการทำข้อสอบแต่ละข้อ โดยพิจารณาจากจำนวนนักเรียนที่ทำเสร็จ

ประมาณ 80% ของนักเรียนทั้งห้อง ปรากฏว่าในแต่ละข้อนักเรียนใช้เวลาในการทำแบบทดสอบประมาณข้อละ 10 นาที

8. นำผลการทดลองครั้งที่ 1 มาตรวจให้คะแนนตามเกณฑ์การให้คะแนนที่กำหนดในข้อ 4.4 โดยใช้ผู้ตรวจเป็นอาจารย์ผู้สอนวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 1 จำนวน 3 คน โดยมีวิธีการตรวจดังนี้

8.1 ใช้วิธีการตรวจแบบประเมินรวม (Holistic) โดยผู้วิจัยต้องชี้แจงวิธีการตรวจแบบประเมินรวมให้ผู้ตรวจมีความเข้าใจตรงกัน

8.2 ให้ผู้ตรวจตรวจแบบทดสอบทีละฉบับ แต่ละฉบับให้ตรวจข้อสอบทีละข้อจากแบบทดสอบทั้ง 30 ข้อที่นักเรียนทำ โดยผู้ตรวจให้คะแนนอย่างอิสระต่อกัน พิจารณาจากเกณฑ์การให้คะแนนแบบรูปริก (General Rubric) ที่กำหนดไว้ในข้อ 4.4 แล้วทำการแบ่งคำตอบของนักเรียนในแต่ละข้อออกเป็น 5 ระดับคะแนน คือ 0 - 4 คะแนน ทำเช่นเดียวกันนี้ทีละข้อให้ครบทั้ง 6 ข้อ ในแบบทดสอบฉบับที่ 1

8.3 ให้ผู้ตรวจทำการตรวจแบบทดสอบฉบับที่ 2 โดยใช้วิธีการเดียวกับการตรวจแบบทดสอบฉบับที่ 1

8.4 นำผลการให้คะแนนของผู้ตรวจทั้ง 3 คน จากแบบทดสอบทั้ง 2 ฉบับ มาพิจารณาดัชนีค่าความง่าย และดัชนีค่าอำนาจจำแนกเป็นรายข้อ พร้อมทั้งหาค่าความสอดคล้องของการให้คะแนน โดยการหาค่าความเชื่อมั่นของผู้ตรวจ 3 คน

ผลการวิเคราะห์ปรากฏว่า ข้อสอบทั้ง 12 ข้อ มีดัชนีค่าความง่ายอยู่ระหว่าง 0.50 - 0.82 และมีดัชนีค่าอำนาจจำแนกอยู่ระหว่าง -0.14 - 0.64 ส่วนค่าความเชื่อมั่นของผู้ตรวจ (W) มีค่าอยู่ระหว่าง 0.74 - 0.93 และเมื่อเปลี่ยนค่า W เป็นค่า χ^2 แล้ว ค่า χ^2 ที่คำนวณได้ทุกข้อมีค่ามากกว่าค่าวิกฤตของ χ^2 ที่เปิดจากตารางอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 หลังจากนั้นได้พิจารณาเลือกข้อสอบที่มีดัชนีค่าความง่ายตั้งแต่ 0.20 ถึง 0.80 มีดัชนีค่าอำนาจจำแนกตั้งแต่ 0.20 ขึ้นไป และค่าความเชื่อมั่นของผู้ตรวจต้องมียุทธศาสตร์ทางสถิติที่ระดับ .01 ไว้จำนวน 6 ข้อ ซึ่งมีดัชนีค่าความง่ายอยู่ระหว่าง 0.50 - 0.64 มีดัชนีค่าอำนาจจำแนกอยู่ระหว่าง 0.24 - 0.64 และมีค่าความเชื่อมั่นของผู้ตรวจอยู่ระหว่าง 0.84 - 0.92 ซึ่งแสดงไว้ในภาคผนวก ก

8.5 ผู้วิจัยสร้างเกณฑ์รูปริกเฉพาะ (Specific Rubric) ของข้อสอบที่คัดเลือกไว้ทั้ง 6 ข้อ โดยสร้างเกณฑ์รูปริกเฉพาะสำหรับข้อสอบแต่ละข้อโดยใช้คำตอบของนักเรียนเป็นคำตอบตัวแบบ (Essey Model) โดยคัดเลือกจากคำตอบของนักเรียนจากการทดลองสอบและได้รับการแบ่งระดับคะแนนแล้ว ในข้อ 8.2 - 8.3 ซึ่งแสดงไว้ในภาคผนวก ก

8.6 เมื่อได้เกณฑ์เฉพาะสำหรับข้อสอบทุกข้อ แต่ละข้อมีครบทั้ง 5 ระดับคะแนนแล้ว จัดเป็นชุดเฉลยคำตอบและเกณฑ์การให้คะแนนเฉพาะในแต่ละข้อไว้สำหรับใช้ในการวิจัย พร้อมทั้งแนบเกณฑ์การให้คะแนนแบบรูปริก (General Rubric) ด้วย เพื่อให้ผู้ตรวจใช้ในการกรณีที่คำตอบของนักเรียนคนใดแตกต่างไปจากชุดเฉลยคำตอบที่เตรียมไว้แล้ว ก็ให้ผู้ตรวจพิจารณาให้คะแนนคำตอบนั้นตามเกณฑ์แบบรูปริกได้ เอกสารชุดนี้จัดไว้สำหรับผู้ตรวจที่ตรวจโดยใช้เกณฑ์การให้คะแนนแบบรูปริก ในขั้นตอนของการเก็บรวบรวมข้อมูลเพื่อการวิจัยต่อไป

9. นำข้อสอบที่คัดเลือกไว้ทั้ง 6 ข้อ มาจัดพิมพ์เป็นแบบทดสอบวิชาคณิตศาสตร์ ที่วัดความสามารถของนักเรียน 3 ด้าน ดังนี้

1. การแก้ปัญหาในวิชาคณิตศาสตร์ จำนวน 2 ข้อ
2. คณิตศาสตร์เหตุผล จำนวน 2 ข้อ

3. การสื่อสารทางคณิตศาสตร์ จำนวน 2 ข้อ

รวมเป็นแบบทดสอบ 1 ฉบับ จำนวน 6 ข้อ ใช้เวลา 60 นาที แล้วนำไปทดสอบกับกลุ่มตัวอย่างเพื่อใช้ในการวิจัยต่อไป

ลักษณะของเครื่องมือที่ใช้ในการวิจัย

1. แบบทดสอบการแก้ปัญหาในวิชาคณิตศาสตร์

ตัวอย่างข้อคำถาม

ข้อ (0) ปรียาและอาภรณ์ไปรับประทานอาหารที่ภัตตาคารแห่งหนึ่ง ค่าอาหารคิดเป็นเงิน 1,350 บาท ทั้งสองต้องการให้ทิปแก่บริกร 15% ของค่าอาหาร นักเรียนมีวิธีการคิด อย่างไรว่า ทั้งสองต้องให้ทิปแก่บริกรเป็นเงินเท่าไร และเมื่อรวมแล้วทั้งสองต้องจ่ายเงินทั้งหมดเท่าไร

2. แบบทดสอบคณิตศาสตร์เหตุผล

ตัวอย่างข้อคำถาม

ข้อ (00) สมควรทราบว่าผลการสอบเข้ามหาวิทยาลัยในปีนี้ นักเรียนจากโรงเรียนของเขา สอบเข้ามหาวิทยาลัยของรัฐบาลได้ครึ่งหนึ่งของนักเรียนทั้งหมดที่สอบ และสอบเข้ามหาวิทยาลัยเอกชนได้ครึ่งหนึ่งของนักเรียนทั้งหมดที่สอบเช่นกัน สมควรคิดว่าทั้งสองส่วนรวมกันเป็น 100% หรือไม่ ดังนั้นเขาจึงมั่นใจว่าเขาต้องเป็นหนึ่งในผู้สอบได้ใหม่มหาวิทยาลัยแห่งใดแห่งหนึ่ง จึงอธิบายว่าสมควรอาจจะคิดผิดเพราะอะไร และเขียนแผนภาพประกอบการอธิบายถ้าเป็นไปได้

3. แบบทดสอบการสื่อสารทางคณิตศาสตร์

ตัวอย่างข้อคำถาม

ข้อ (000) ถ้านักเรียนต้องอธิบายให้เพื่อนในห้องเรียนเข้าใจในเรื่องการหา ค.ร.น. ของจำนวน 3 จำนวน จงเขียนอธิบายว่านักเรียนต้องพูดอะไรบ้าง มีลำดับขั้นตอนอย่างไร พร้อมทั้งยกตัวอย่างประกอบการอธิบายด้วย

วิธีดำเนินการเก็บรวบรวมข้อมูล

ในการเก็บรวบรวมข้อมูลที่ใช้ในการวิจัยครั้งนี้ ผู้วิจัยได้ดำเนินการเป็นขั้นตอนดังนี้

1. ทำหนังสือจากบัณฑิตวิทยาลัยมหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อขอความร่วมมือในการเก็บรวบรวมข้อมูลจากโรงเรียนที่เป็นกลุ่มตัวอย่าง และนัดหมายวันและเวลาในการสอบ
2. จัดเตรียมแบบทดสอบวิชาคณิตศาสตร์ ที่ผ่านการตรวจสอบและปรับปรุงแก้ไขแล้ว ให้เพียงพอกับจำนวนนักเรียนที่เป็นกลุ่มตัวอย่างในแต่ละโรงเรียน

3. ศึกษาคู่มือการดำเนินการสอบ โดยผู้วิจัยเป็นผู้ดำเนินการสอบเอง
4. เมื่อนักเรียนได้รับแบบทดสอบ และอ่านคำชี้แจงในการทำแบบทดสอบแล้ว หากมีข้อสงสัยให้สอบถามได้ เมื่อเข้าใจตรงกันแล้วให้นักเรียนเริ่มลงมือทำแบบทดสอบได้
5. นำกระดาษคำตอบของนักเรียนที่ได้มาทำการตรวจ โดยมีวิธีการตรวจดังนี้
 - 5.1 ขอความร่วมมือจากโรงเรียน และเรียนเชิญอาจารย์ผู้สอนวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 1 มาเป็นผู้เชี่ยวชาญในการตรวจแบบทดสอบ และนัดหมายวันและเวลาในการตรวจแบบทดสอบ
 - 5.2 ถ่ายเอกสารกระดาษคำตอบของนักเรียนจำนวน 12 ชุด สำหรับผู้ตรวจ 6 คน (แต่ละคนจะต้องตรวจให้คะแนนแบบทดสอบฉบับเดียวกัน 2 ครั้ง) โดยผู้ตรวจแบ่งเป็น
 - 5.2.1 ผู้ตรวจที่มีประสบการณ์มากกว่า 5 ปี จำนวน 3 คน
 - 5.2.2 ผู้ตรวจที่มีประสบการณ์ไม่เกิน 5 ปี จำนวน 3 คน
 - 5.3 ในการตรวจครั้งที่ 1 เป็นการตรวจโดยวิธีประเมินรวม (Holistic) โดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง เมื่อถึงวันเวลาที่นัดหมาย ผู้วิจัยชี้แจงวิธีการตรวจแบบประเมินรวมให้ผู้ตรวจทุกคนมีความเข้าใจตรงกัน และกำหนดระดับคะแนนออกเป็น 5 ระดับ คือ 0 - 4 คะแนน
 - 5.4 เมื่อผู้ตรวจมีความเข้าใจตรงกันแล้ว ให้ผู้ตรวจแต่ละคนตรวจข้อสอบทีละข้อให้ครบทุกคน โดยให้คะแนนแบบประเมินรวม และใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง ตรวจให้ครบทุกข้อ
 - 5.5 การตรวจครั้งที่ 2 จัดขึ้นหลังจากการตรวจครั้งที่ 1 เป็นเวลา 1 สัปดาห์ เป็นการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบบริบทที่จัดเตรียมไว้ให้กับผู้ตรวจทุกคน พร้อมทั้งชี้แจงเกี่ยวกับเกณฑ์การให้คะแนนแบบบริบทให้ผู้ตรวจทุกคนมีความเข้าใจตรงกัน
 - 5.6 เมื่อพร้อมแล้ว ให้ผู้ตรวจแต่ละคนทำการตรวจข้อสอบชุดเดิมอีกครั้ง แต่ครั้งนี้ให้ตรวจตามคำตอบตัวแบบ และเกณฑ์การให้คะแนนแบบบริบทในเอกสารที่มอบให้
6. นำผลที่ได้จากการตรวจมาทำการวิเคราะห์ข้อมูล ดังนี้
 - 6.1 คำนวณค่าสถิติพื้นฐาน ($\bar{X}, S.D.$)
 - 6.2 คำนวณหาค่าสัมประสิทธิ์การสรุปอ้างอิง โดยใช้โปรแกรมสำเร็จรูป GENOVA ซึ่งมีขั้นตอนดังนี้
 - 6.2.1 ประเมินค่าความแปรปรวนขององค์ประกอบ 3 องค์ประกอบที่ศึกษา ในชั้นการศึกษาเพื่อการสรุปอ้างอิง (G Study) โดยใช้รูปแบบการวัดแบบผสม (Mixed model) ให้สอดคล้องกับรูปแบบการวัดที่กำหนดไว้ 3 รูปแบบในนิยามศัพท์เฉพาะ
 - 6.2.2 ประเมินค่าความแปรปรวนในชั้นการศึกษาเพื่อการตัดสินใจ (D Study)
 - 6.2.3 ประเมินค่าสัมประสิทธิ์การสรุปอ้างอิงของรูปแบบการวัดแต่ละแบบ ภายใต้เงื่อนไขที่ทำการศึกษา
 - 6.3 ทดสอบสมมติฐานของการวิจัย

ทดสอบความแตกต่างของค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์แต่ละแบบ ตามสมมติฐานของการวิจัย โดยใช้สูตร UX_1 ของวูดรอฟฟ์ และเฟลด์ท์ (Woodruff and Feldt. 1986 : 393-413) ถ้ามีความแตกต่างอย่างมีนัยสำคัญทางสถิติจะต้องทำการทดสอบเพื่อเปรียบเทียบความแตกต่างเป็นรายคู่ ด้วยสูตรของพิทแมน (Feldt, et al. 1978 : 99; Citing Pitman. 1939)

สถิติที่ใช้ในการวิเคราะห์ข้อมูล

1. วิเคราะห์ข้อสอบเป็นรายข้อ เพื่อหาดัชนีค่าความง่าย และดัชนีค่าอำนาจจำแนก โดยใช้เทคนิค 25% ของนักเรียนที่เข้าสอบทั้งหมด โดยคำนวณจากสูตรที่ D.R Whitney and D.L Sabers, 1970 ได้เสนอไว้ (ล้วน สายยศ และอังคณา สายยศ. 2539 : 199 - 201)

1.1 ดัชนีค่าความง่าย (P_E) คำนวณจากสูตร ดังนี้

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

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

1.2 ดัชนีค่าอำนาจจำแนก คำนวณจากสูตร ดังนี้

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

2. ค่าสถิติพื้นฐาน ได้แก่ ค่าเฉลี่ย ($\bar{X}$) และส่วนเบี่ยงเบนมาตรฐาน ($S.D.$)

2.1 ค่าเฉลี่ย (Ferguson. 1981 : 49)

$$\bar{X} = \frac{\sum X}{N}$$

เมื่อ $\bar{X}$	แทน ค่าเฉลี่ย
$\sum X$	แทน ผลรวมของคะแนนทั้งหมด
N	แทน จำนวนคนในกลุ่มตัวอย่าง

2.2 ส่วนเบี่ยงเบนมาตรฐาน (Ferguson. 1981 : 68)

$$S.D. = \sqrt{\frac{N \sum X^2 - (\sum X)^2}{N(N-1)}}$$

เมื่อ $S.D.$ แทน ส่วนเบี่ยงเบนมาตรฐาน
 $\sum X$ แทน ผลรวมของคะแนนทั้งหมด
 $\sum X^2$ แทน ผลรวมของคะแนนแต่ละตัวยกกำลังสอง

3. ค่าความเชื่อมั่นของผู้ตรวจ 3 คน โดยใช้สูตรของเคนดอลล์ (Kendall Coefficient of Concordance : W) (Siegel. 1956 : 229-239)

$$W = \frac{S}{\frac{1}{12} k^2 (N^3 - N) - k \sum T}$$

$$\text{เมื่อ } S = \sum \left(R_j - \frac{\sum R_j}{N} \right)^2$$

$$T = \frac{\sum (t^3 - t)}{12}$$

R_j แทน ผลรวมของตำแหน่งที่ของนักเรียนคนที่ j

N แทน จำนวนนักเรียนทั้งหมด

$\frac{R_j}{N}$ แทน ค่าเฉลี่ยของตำแหน่งที่ของนักเรียน

k แทน จำนวนผู้ตรวจ

t แทน จำนวนนักเรียนที่อยู่ในตำแหน่งที่เดียวกันในแต่ละข้อ

การทดสอบนัยสำคัญทางสถิติของค่า W ในกรณีที่ $N > 7$ ให้เปลี่ยนค่า W เป็น χ^2

โดยใช้สูตร

$$\chi^2 = k(N-1)W$$

$$\text{ที่ } df = N - 1$$

4. ประมาณค่าสัมประสิทธิ์การสรุปอ้างอิง (ρ^2) โดยใช้สูตรของเบรนนอน และเคน ที่วิเคราะห์โดยใช้โปรแกรมสำเร็จรูป GENOVA (Generalized Analysis of Variance System) (Brennan. 1983 : 17)

$$\rho^2 = \frac{\sigma^2(p)}{\sigma^2(p) + \sigma^2(\delta)}$$

เมื่อ ρ^2 แทน ค่าสัมประสิทธิ์การสรุปอ้างอิง
 $\sigma^2(p)$ แทน ความแปรปรวนของคะแนนเอกภพ
 $\sigma^2(\delta)$ แทน ความแปรปรวนของคะแนนความคลาดเคลื่อน

5. ทดสอบความแตกต่างของค่าสัมประสิทธิ์การสรุปอ้างอิงที่มากกว่า 2 ค่า ที่ได้จากการทดสอบกลุ่มตัวอย่างกลุ่มเดียวกัน โดยใช้สูตร UX_1 (Woodruff and Feldt. 1986 : 393-413)

$$UX_1 = \frac{\sum_{i=1}^m (u_i - \bar{u})^2}{S_u^2 - C_u}$$

$$\bar{u} = \sum_{i=1}^m \frac{u_i}{m}$$

$$u_i = \left(\frac{1}{1 - a_i} \right)^{\frac{1}{3}}$$

$$S_u^2 = \frac{2}{9m(N_c - 1)} \sum_{i=1}^m u_i^2$$

$$C_u = \frac{4}{9m(m-1)(N_c - 1)} \sum_{i=2}^m \sum_{j=1}^{i-1} [r(x_i, x_j)]^2 u_i u_j$$

$$N_c = N \left(\frac{n-1}{n+1} \right)$$

เมื่อ UX_1 แทน ค่าสถิติทดสอบที่มีการแจกแจงคล้ายกับ χ^2 ที่ $df = m - 1$
 N แทน จำนวนนักเรียนที่ทดสอบ

n	แทน จำนวนค่าสังเกต
m	แทน จำนวนค่าสัมประสิทธิ์การสรุปอ้างอิง
a_i	แทน ค่าสัมประสิทธิ์การสรุปอ้างอิง
S_u^2	แทน ความแปรปรวนของคะแนน u_i
$\bar{u}$	แทน ค่าเฉลี่ยของคะแนน u_i
N_c	แทน การปรับแก้ตัวสมการของความเป็นอิสระ
C_u	แทน ค่าที่ปรับแก้ความแปรปรวนของ u_i ในกรณีที่ค่าสัมประสิทธิ์การสรุปอ้างอิงคำนวณจากกลุ่มตัวอย่างกลุ่มเดียวกัน
$r(x_i, x_j)$	แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างแบบทดสอบวิชาคณิตศาสตร์แต่ละรูปแบบที่จับคู่กัน

6. การทดสอบภายหลังเมื่อผลการทดสอบ UX_1 มีนัยสำคัญทางสถิติ หรือเมื่อจำนวน ρ^2 มี 2 ค่า ใช้สูตรของพิทแมน (Pitman. 1936) ในกรณีที่ค่าสัมประสิทธิ์การสรุปอ้างอิงที่คำนวณจากกลุ่มตัวอย่างที่เหมือนกัน และใช้ตาราง Dunn Multiple Comparison Test ซึ่งเรียกว่า Dunn - Sidak Table ซึ่งเป็นตารางแบบ Two - tailed t - test

$$t = \frac{(W-1)(N-2)^{\frac{1}{2}}}{[4W(1-r_{x_1x_2}^2)]^{\frac{1}{2}}}$$

$$W = \frac{1-r_{x_2}}{1-r_{x_1}}$$

$$df = N - 2$$

เมื่อ	N	แทน จำนวนนักเรียน
	r	แทน สัมประสิทธิ์การสรุปอ้างอิง
	$r_{x_1x_2}$	แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างแบบทดสอบวิชาคณิตศาสตร์แต่ละรูปแบบ โดยวิธี Phi Correlation

บทที่ 4

ผลการวิเคราะห์ข้อมูล

ในการศึกษารุ่นนี้ เพื่อให้เกิดความเข้าใจตรงกันในการนำเสนอ และแปลผลการวิเคราะห์ข้อมูล ผู้วิจัยจึงได้กำหนดสัญลักษณ์ที่ใช้ในการวิเคราะห์ข้อมูลดังนี้

สัญลักษณ์ที่ใช้ในการวิเคราะห์ข้อมูล

$\bar{X}$	แทน	ค่าเฉลี่ย
$S.D.$	แทน	ส่วนเบี่ยงเบนมาตรฐาน
H_1	แทน	การตรวจให้คะแนนแบบประเมินรวม โดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง
H_2	แทน	การตรวจให้คะแนนแบบประเมินรวม โดยใช้เกณฑ์การให้คะแนนแบบรูบริค (Rubric scoring)
R_1	แทน	ผู้ตรวจที่มีประสบการณ์มากกว่า 5 ปี
R_2	แทน	ผู้ตรวจที่มีประสบการณ์ไม่เกิน 5 ปี
ρ^2	แทน	ค่าสัมประสิทธิ์การสรุปอ้างอิง
UX_1	แทน	สถิติทดสอบที่มีการแจกแจงคล้ายกับ Chi-Square

การเสนอผลการวิเคราะห์ข้อมูล

- ในการศึกษารุ่นนี้ ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลตามลำดับดังนี้ คือ
1. ค่าสถิติพื้นฐานของแบบทดสอบวิชาคณิตศาสตร์
 2. เปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อมีจำนวนผู้ตรวจเท่ากัน
 3. เปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อผู้ตรวจมีประสบการณ์เหมือนกัน
 4. เปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีจำนวนผู้ตรวจต่างกัน เมื่อมีวิธีการตรวจให้เหมือนกัน
 5. เปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่ผู้ตรวจมีประสบการณ์ต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน

ผลการวิเคราะห์ข้อมูล

1. ค่าสถิติพื้นฐานของแบบทดสอบวิชาคณิตศาสตร์

ผู้วิจัยนำแบบทดสอบวิชาคณิตศาสตร์ที่มีลักษณะคำถามแบบปลายเปิด ที่วัดความสามารถของนักเรียน 3 ด้าน 1 ฉบับ จำนวน 6 ข้อ ไปทดสอบกับกลุ่มตัวอย่างจำนวน 60 คน และนำกระดาษคำตอบของนักเรียนมาตรวจให้คะแนน 2 วิธี คือ การตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง และการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูบริค ซึ่งมีผู้ตรวจ 6 คน แบ่งเป็นผู้ตรวจที่มีประสบการณ์มากกว่า 5 ปี จำนวน 3 คน และผู้ตรวจที่มีประสบการณ์ไม่เกิน 5 ปี จำนวน 3 คน แล้วนำคะแนนที่ได้มาคำนวณค่าสถิติพื้นฐาน ปรากฏผลดังแสดงในตาราง 3

ตาราง 3 ค่าสถิติพื้นฐานของแบบทดสอบวิชาคณิตศาสตร์ จำแนกตามวิธีการตรวจให้คะแนน ประสบการณ์ของผู้ตรวจ และจำนวนผู้ตรวจ

วิธีการตรวจให้คะแนน	ประสบการณ์ของผู้ตรวจ / จำนวนผู้ตรวจ (คน)	คะแนนเต็ม (คะแนน)	$\bar{X}$	$S.D.$
H_1	R_1	24	8.47	5.22
	R_2	24	9.12	4.67
H_2	R_1	24	10.36	6.13
	R_2	24	9.28	5.47
H_1	3	24	8.80	4.95
	4	24	9.05	5.12
	5	24	9.03	5.03
	6	24	8.79	4.91
H_2	3	24	9.82	5.80
	4	24	10.05	5.90
	5	24	9.65	5.76
	6	24	9.82	5.78

จากตาราง 3 พบว่า แบบทดสอบวิชาคณิตศาสตร์ที่มีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง สำหรับผู้ตรวจที่มีประสบการณ์มากกว่า 5 ปี กับผู้ตรวจที่มีประสบการณ์ไม่เกิน 5 ปี มีคะแนนเฉลี่ยเท่ากับ 8.47 และ 9.12 ตามลำดับ โดยการตรวจให้คะแนนโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเองนั้น ผู้ตรวจที่มีประสบการณ์ไม่เกิน 5 ปี จะตรวจให้คะแนนมีแนวโน้มสูงกว่าผู้ตรวจที่มีประสบการณ์มากกว่า 5 ปี ส่วนการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูบริค สำหรับผู้ตรวจที่มีประสบการณ์มากกว่า 5 ปี กับผู้ตรวจที่มีประสบการณ์ไม่เกิน 5 ปี มีคะแนนเฉลี่ยเท่ากับ

10.36 และ 9.28 ตามลำดับ โดยการตรวจให้คะแนนโดยใช้เกณฑ์การให้คะแนนแบบรูปรีคนั้น ผู้ตรวจที่มีประสบการณ์มากกว่า 5 ปี จะตรวจให้คะแนนมีแนวโน้มสูงกว่าผู้ตรวจที่มีประสบการณ์ไม่เกิน 5 ปี

ในกรณีที่แบบทดสอบวิชาคณิตศาสตร์มีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง เมื่อมีจำนวนผู้ตรวจ 3, 4, 5 และ 6 คน มีคะแนนเฉลี่ยอยู่ระหว่าง 8.79 - 9.05 โดยการตรวจให้คะแนนเมื่อมีจำนวนผู้ตรวจ 4 คน คะแนนเฉลี่ยมีแนวโน้มสูงที่สุด รองลงมาคือมีจำนวนผู้ตรวจ 5 คน และเมื่อมีจำนวนผู้ตรวจ 6 คน คะแนนเฉลี่ยมีแนวโน้มต่ำที่สุด ส่วนการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีคนั้น เมื่อมีจำนวนผู้ตรวจ 3, 4, 5 และ 6 คน มีคะแนนเฉลี่ยอยู่ระหว่าง 9.82 - 10.05 โดยการตรวจให้คะแนนเมื่อมีจำนวนผู้ตรวจ 4 คน คะแนนเฉลี่ยมีแนวโน้มสูงที่สุด รองลงมาคือมีจำนวนผู้ตรวจ 3 และ 6 คน และเมื่อมีจำนวนผู้ตรวจ 5 คน คะแนนเฉลี่ยมีแนวโน้มต่ำที่สุด

สำหรับคะแนนเบี่ยงเบนมาตรฐานของแบบทดสอบนั้น เมื่อจำแนกตามวิธีการตรวจให้คะแนน ประสบการณ์ของผู้ตรวจ และจำนวนผู้ตรวจ ปรากฏว่ามีค่าอยู่ระหว่าง 4.67 – 6.13 แสดงว่าแบบทดสอบมีการกระจายของคะแนนใกล้เคียงกัน

2. เปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อมีจำนวนผู้ตรวจเหมือนกัน

ในการเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อมีจำนวนผู้ตรวจเหมือนกันนั้น ผู้วิจัยได้คำนวณค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบ จากนั้นนำค่าที่ได้มาทดสอบความแตกต่างด้วยสูตร t -test ของพิทแมน (Pitman, 1936) ปรากฏผลดังแสดงในตาราง 4

ตาราง 4 ผลการทดสอบความแตกต่างของค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อมีจำนวนผู้ตรวจเท่ากัน

จำนวนผู้ตรวจ (คน)	วิธีการตรวจให้คะแนน	ρ^2	t
3	H_1	0.69542	2.91434**
	H_2	0.76047	
4	H_1	0.75273	3.89190**
	H_2	0.80891	
5	H_1	0.79190	4.53375**
	H_2	0.84105	
6	H_1	0.82035	4.55845**
	H_2	0.86394	

** มีนัยสำคัญทางสถิติที่ระดับ .01

จากตาราง 4 พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง กับใช้เกณฑ์การให้คะแนนแบบรูบริค เมื่อมีจำนวนผู้ตรวจ 3 คน มีค่าเท่ากับ 0.69542 และ 0.76047 ตามลำดับ เมื่อมีจำนวนผู้ตรวจ 4 คน มีค่าเท่ากับ 0.75273 และ 0.80891 ตามลำดับ เมื่อมีจำนวนผู้ตรวจ 5 คน มีค่าเท่ากับ 0.79190 และ 0.84105 ตามลำดับ และเมื่อมีจำนวนผู้ตรวจ 6 คน มีค่าเท่ากับ 0.82035 และ 0.86394 ตามลำดับ โดยทั้ง 4 กรณีพบว่า การตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูบริค มีค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบสูงกว่าการตรวจให้คะแนนโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง และเมื่อนำค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบมาเปรียบเทียบกับระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน 2 วิธี เมื่อจำนวนผู้ตรวจเป็น 3, 4, 5 และ 6 คน พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกกรณี

3. เปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อผู้ตรวจมีประสบการณ์เหมือนกัน

ในการเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนต่างกันเมื่อผู้ตรวจมีประสบการณ์เหมือนกันนั้น ผู้วิจัยได้คำนวณค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบ จากนั้นนำค่าที่ได้มาทดสอบความแตกต่างด้วยสูตร t -test ของพิทแมน (Pitman, 1936) ปรากฏผลดังแสดงในตาราง 5

ตาราง 5 ผลการทดสอบความแตกต่างของค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อผู้ตรวจมีประสบการณ์เหมือนกัน

ประสบการณ์ของผู้ตรวจ	วิธีการตรวจให้คะแนน	ρ^2	t
R_1	H_1	0.83238	0.709969
	H_2	0.84193	
R_2	H_1	0.77003	1.263188
	H_2	0.79035	

จากตาราง 5 พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง กับใช้เกณฑ์การให้คะแนนแบบรูบริค ในกรณีที่ผู้ตรวจมีประสบการณ์มากกว่า 5 ปี มีค่าเท่ากับ 0.83238 และ 0.84193 ตามลำดับ ส่วนกรณีที่ผู้ตรวจมีประสบการณ์ไม่เกิน 5 ปี มีค่าเท่ากับ 0.77003 และ 0.79035 ตามลำดับ โดยทั้ง 2 กรณีพบว่า การตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูบริค มีค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบสูงกว่าการตรวจให้คะแนนโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง และเมื่อนำค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบมาเปรียบเทียบกับระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน 2 วิธี เมื่อ

ผู้ตรวจมีประสบการณ์มากกว่า 5 ปี และผู้ตรวจมีประสบการณ์ไม่เกิน 5 ปี พบว่า แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติทั้ง 2 กรณี

4. เปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีจำนวนผู้ตรวจต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน

ในการเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีจำนวนผู้ตรวจต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกันนั้น ผู้วิจัยได้คำนวณค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบ จากนั้นนำค่าที่ได้มาทดสอบความแตกต่างด้วยสูตร UX_1 ของวูดรUFF และเฟลด์ (Woodruff and Feldt. 1986 : 398 - 413) ปรากฏผลดังแสดงในตาราง 6

ตาราง 6 ผลการทดสอบความแตกต่างของค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีจำนวนผู้ตรวจต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน

วิธีการตรวจให้คะแนน	จำนวนผู้ตรวจ (คน)	ρ^2	ทดสอบความแตกต่าง UX_1
H_1	3	0.69542	557.5813**
	4	0.75273	
	5	0.79190	
	6	0.82035	
H_2	3	0.76047	561.3359**
	4	0.80891	
	5	0.84105	
	6	0.86394	

** มีนัยสำคัญทางสถิติที่ระดับ .01

จากตาราง 6 พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง เมื่อมีจำนวนผู้ตรวจ 3, 4, 5 และ 6 คน มีค่าเท่ากับ 0.69542, 0.75273, 0.79190 และ 0.82035 ตามลำดับ ส่วนค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบบุคลิก เมื่อมีจำนวนผู้ตรวจ 3, 4, 5 และ 6 คน มีค่าเท่ากับ 0.76047, 0.80891, 0.84105 และ 0.86394 ตามลำดับ โดยทั้ง 2 กรณีพบว่า เมื่อมีจำนวนผู้ตรวจ 6 คน ค่าสัมประสิทธิ์การสรุปอ้างอิงสูงที่สุด รองลงมาคือเมื่อมีจำนวนผู้ตรวจ 5 คน และเมื่อมีจำนวนผู้ตรวจ 3 คน ค่าสัมประสิทธิ์การสรุปอ้างอิงมีค่าต่ำที่สุด

เมื่อทำการทดสอบความแตกต่างของค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบพบว่า จำนวนผู้ตรวจที่แตกต่างกันมีผลทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบอย่างน้อยหนึ่งคู่แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ดังนั้นเพื่อให้ทราบว่า จำนวนผู้ตรวจที่แตกต่างกันคู่ใดทำให้ค่า

สัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบแตกต่างกัน ผู้วิจัยจึงทดสอบความแตกต่างเป็นรายคู่ด้วยสูตร t -test ของพิทแมน (Pitman, 1936) ปรากฏผลดังแสดงในตาราง 7

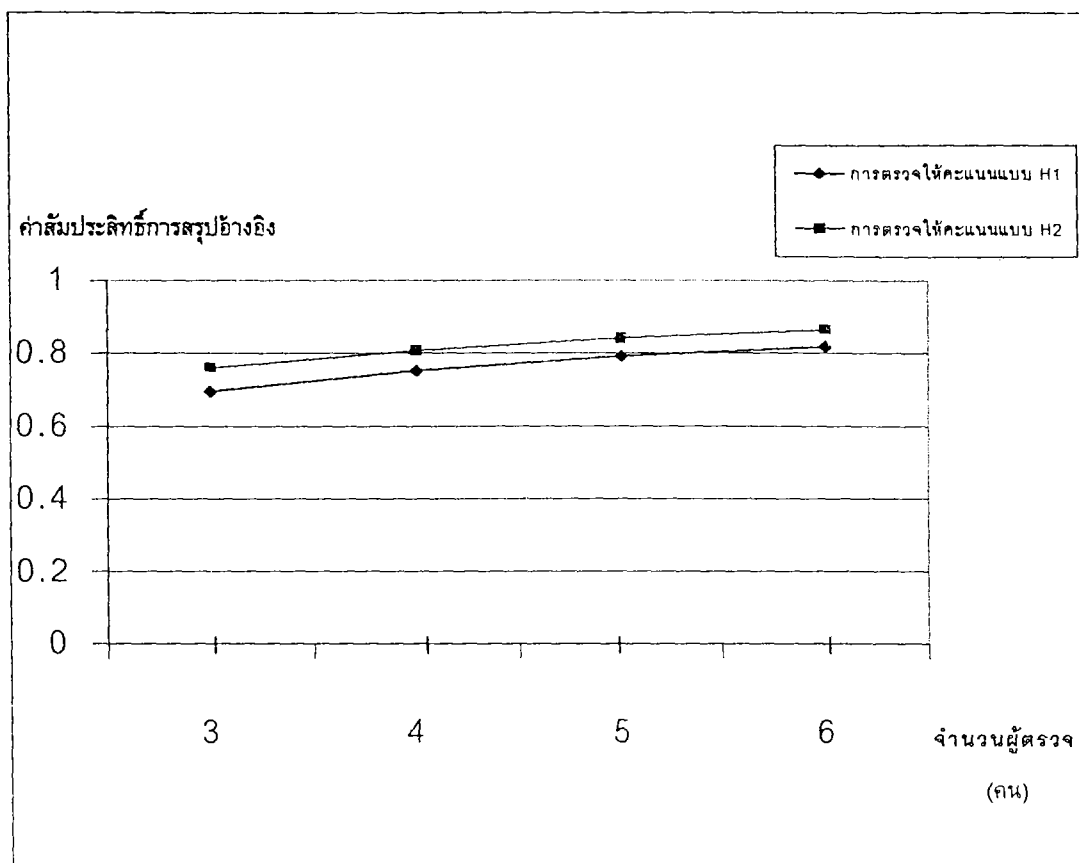
ตาราง 7 ผลการทดสอบความแตกต่างของค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีจำนวนผู้ตรวจต่างกันเป็นรายคู่ เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน

วิธีการตรวจให้คะแนน	จำนวนผู้ตรวจ (คน)	3	4	5	6
H_1	3	-	7.7339**	13.8081**	16.7479**
	4		-	12.0093**	16.9528**
	5			-	11.4404**
	6				-
H_2	3	-	11.7405**	16.5947**	26.4892**
	4		-	13.7764**	25.4949**
	5			-	15.8433**
	6				-

** มีนัยสำคัญทางสถิติที่ระดับ .01

จากตาราง 7 พบว่า เมื่อใช้วิธีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การตรวจให้คะแนนของผู้ตรวจเอง แบบทดสอบที่มีผู้ตรวจ 3 คน กับ 4 คน แบบทดสอบที่มีผู้ตรวจ 3 คน กับ 5 คน แบบทดสอบที่มีผู้ตรวจ 3 คน กับ 6 คน แบบทดสอบที่มีผู้ตรวจ 4 คน กับ 5 คน แบบทดสอบที่มีผู้ตรวจ 4 คน กับ 6 คน และแบบทดสอบที่มีผู้ตรวจ 5 คน กับ 6 คน มีค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกคู่ ในทำนองเดียวกัน เมื่อใช้วิธีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีค แบบทดสอบที่มีผู้ตรวจ 3 คน กับ 4 คน แบบทดสอบที่มีผู้ตรวจ 3 คน กับ 5 คน แบบทดสอบที่มีผู้ตรวจ 3 คน กับ 6 คน แบบทดสอบที่มีผู้ตรวจ 4 คน กับ 5 คน แบบทดสอบที่มีผู้ตรวจ 4 คน กับ 6 คน และแบบทดสอบที่มีผู้ตรวจ 5 คน กับ 6 คน มีค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกคู่

และเพื่อให้เห็นการเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีการตรวจให้คะแนน 2 วิธี และมีจำนวนผู้ตรวจต่างกันตั้งแต่ 3 ถึง 6 คน ตามรายละเอียดที่กล่าวมาข้างต้นอย่างชัดเจน ผู้วิจัยได้นำเสนอการเปรียบเทียบในรูปของกราฟเส้น ดังแสดงในภาพประกอบ 3



ภาพประกอบ 3 แสดงการเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีผู้ตรวจ 3 ถึง 6 คน และมีวิธีการตรวจให้คะแนน 2 วิธี

จากภาพประกอบ 3 แสดงให้เห็นว่าเมื่อมีจำนวนผู้ตรวจเพิ่มขึ้นจะทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบมีค่าสูงขึ้นด้วย ทั้งที่ใช้วิธีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเองและใช้เกณฑ์การให้คะแนนแบบรูปรีด

5. เปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่ผู้ตรวจมีประสบการณ์ต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน

ในการเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่ผู้ตรวจมีประสบการณ์ต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกันนั้น ผู้วิจัยได้คำนวณค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบ จากนั้นนำค่าที่ได้มาทดสอบความแตกต่างด้วยสูตร t -test ของพิทแมน (Pitman, 1939) ปรากฏผลดังแสดงในตาราง 8

ตาราง 8 ผลการทดสอบความแตกต่างของค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่ผู้ตรวจมีประสบการณ์ต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน

วิธีการตรวจให้คะแนน	ประสบการณ์ของผู้ตรวจ	ρ^2	t
H_1	R_1	0.83238	4.73933**
	R_2	0.77003	
H_2	R_1	0.84193	5.17212**
	R_2	0.79035	

** มีนัยสำคัญทางสถิติที่ระดับ .01

จากตาราง 8 พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่ผู้ตรวจมีประสบการณ์มากกว่า 5 ปี กับผู้ตรวจมีประสบการณ์ไม่เกิน 5 ปี เมื่อมีการตรวจให้คะแนนแบบประเมินรวม โดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง มีค่าเท่ากับ 0.83238 และ 0.77003 ตามลำดับ และเมื่อมีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีด มีค่าเท่ากับ 0.84193 และ 0.79035 ตามลำดับ โดยทั้ง 2 กรณี เมื่อผู้ตรวจมีประสบการณ์มากกว่า 5 ปี ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบมีค่าสูงกว่าเมื่อผู้ตรวจมีประสบการณ์ไม่เกิน 5 ปี และเมื่อนำค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบมาเปรียบเทียบกันระหว่างผู้ตรวจที่มีประสบการณ์ต่างกัน 2 กรณี เมื่อใช้วิธีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเองและใช้เกณฑ์การให้คะแนนแบบรูปรีด พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทั้ง 2 กรณี

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

จุดมุ่งหมายของการศึกษาค้นคว้า

1. เพื่อเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อมีจำนวนผู้ตรวจเท่ากัน
2. เพื่อเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อผู้ตรวจมีประสบการณ์เหมือนกัน
3. เพื่อเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีจำนวนผู้ตรวจต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน
4. เพื่อเปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่ผู้ตรวจมีประสบการณ์ต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน

กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการศึกษาค้นคว้าเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ของโรงเรียนสังกัดกรมสามัญศึกษา จังหวัดจันทบุรี ภาคเรียนที่ 2 ปีการศึกษา 2540 จำนวน 60 คน ที่ได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi - Stage Random Sampling)

เครื่องมือที่ใช้ในการศึกษาค้นคว้า

เครื่องมือที่ใช้ในการศึกษาค้นคว้า เป็นแบบทดสอบวิชาคณิตศาสตร์ที่มีลักษณะคำถามแบบปลายเปิด ซึ่งลักษณะของแบบทดสอบจะเป็นสถานการณ์ปัญหาทางคณิตศาสตร์ที่กำหนดให้ เพื่อวัดความสามารถของนักเรียน 3 ด้าน คือ

- | | |
|--|-------------|
| 1. การแก้ปัญหาในวิชาคณิตศาสตร์ | จำนวน 2 ข้อ |
| 2. คณิตศาสตร์เหตุผล | จำนวน 2 ข้อ |
| 3. การสื่อสารทางคณิตศาสตร์ | จำนวน 2 ข้อ |
| รวมเป็นข้อสอบ 1 ฉบับ จำนวน 6 ข้อ ใช้เวลา 60 นาที | |

วิธีดำเนินการเก็บรวบรวมข้อมูล

ในการเก็บรวบรวมข้อมูลที่ใช้ในการวิจัยครั้งนี้ ผู้วิจัยดำเนินการเป็นขั้นตอนดังนี้

1. ติดต่อโรงเรียนที่เป็นกลุ่มตัวอย่าง เพื่อขอความร่วมมือในการเก็บรวบรวมข้อมูล และนัดหมายวันและเวลาในการสอบ
2. จัดเตรียมแบบทดสอบให้เพียงพอกับจำนวนนักเรียนที่เป็นกลุ่มตัวอย่างในแต่ละโรงเรียน
3. ทำการทดสอบโดยผู้วิจัยเป็นผู้ดำเนินการสอบ เมื่อนักเรียนได้รับแบบทดสอบและอ่านคำสั่งแจ้งในการทำแบบทดสอบแล้ว หากมีข้อสงสัยให้สอบถามได้ เมื่อเข้าใจตรงกันแล้วให้นักเรียนเริ่มทำแบบทดสอบได้
4. หลังจากดำเนินการทดสอบครบทุกโรงเรียนแล้ว นำกระดาษคำตอบของนักเรียนที่เป็นกลุ่มตัวอย่างมาทำการตรวจ โดยมีวิธีการตรวจ ดังนี้
 - 4.1 นำกระดาษคำตอบมาถ่ายเอกสารจำนวน 12 ชุด สำหรับผู้ตรวจซึ่งเป็นอาจารย์ผู้สอนวิชาคณิตศาสตร์ ระดับชั้นมัธยมศึกษาปีที่ 1 จำนวน 6 คน (แต่ละคนจะต้องตรวจให้คะแนนแบบทดสอบฉบับเดียวกัน 2 ครั้ง) ผู้ตรวจแบ่งเป็น
 - 4.1.1 ผู้ตรวจที่มีประสบการณ์มากกว่า 5 ปี จำนวน 3 คน
 - 4.1.2 ผู้ตรวจที่มีประสบการณ์ไม่เกิน 5 ปี จำนวน 3 คน
 - 4.2 ในการตรวจครั้งที่ 1 เป็นการตรวจแบบประเมินรวมที่ใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง โดยกำหนดระดับคะแนนออกเป็น 5 ระดับ คือ 0 - 4 คะแนน เมื่อถึงวันและเวลาที่นัดหมาย ผู้วิจัยชี้แจงวิธีการตรวจแบบประเมินรวมให้ผู้ตรวจทุกคนมีความเข้าใจตรงกัน เมื่อพร้อมแล้วจึงให้ผู้ตรวจทำการตรวจข้อสอบทีละข้อให้ครบทุกคน และตรวจให้ครบทุกข้อ
 - 4.3 ในการตรวจครั้งที่ 2 เป็นการตรวจแบบประเมินรวมที่ใช้เกณฑ์การให้คะแนนแบบรูปรีค ผู้วิจัยแจกเอกสารชุดเฉลยคำตอบและเกณฑ์การให้คะแนนเฉพาะในแต่ละข้อ พร้อมทั้งเกณฑ์การให้คะแนนแบบรูปรีคที่จัดเตรียมไว้สำหรับผู้ตรวจทุกคน พร้อมทั้งชี้แจงวิธีการตรวจให้คะแนนโดยใช้เกณฑ์การให้คะแนนแบบรูปรีคให้ผู้ตรวจทุกคนมีความเข้าใจตรงกัน เมื่อพร้อมแล้วให้ผู้ตรวจทำการตรวจข้อสอบชุดเดิมอีกครั้ง แต่ครั้งนี้ให้ตรวจตามเอกสารชุดเฉลยคำตอบที่มอบให้
5. นำผลการตรวจทั้งหมดมาทำการวิเคราะห์ทางสถิติและทดสอบสมมติฐาน

การเสนอผลการวิเคราะห์ข้อมูล

ในการศึกษาครั้งนี้ ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลตามลำดับดังนี้ คือ

1. ค่าสถิติพื้นฐานของแบบทดสอบวิชาคณิตศาสตร์
2. เปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อมีจำนวนผู้ตรวจเท่ากัน
3. เปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีวิธีการตรวจให้คะแนนต่างกัน เมื่อผู้ตรวจมีประสบการณ์เหมือนกัน

4. เปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีจำนวนผู้ตรวจต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน

5. เปรียบเทียบสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่ผู้ตรวจมีประสบการณ์ต่างกัน เมื่อมีวิธีการตรวจให้คะแนนเหมือนกัน

สรุปผลการวิเคราะห์ข้อมูล

ผลจากการวิเคราะห์ข้อมูล และการศึกษาค้นคว้าสรุปผลได้ดังนี้

1. คะแนนเฉลี่ยแบบทดสอบวิชาคณิตศาสตร์ที่มีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง สำหรับผู้ตรวจมีประสบการณ์มากกว่า 5 ปี กับผู้ตรวจมีประสบการณ์ไม่เกิน 5 ปี มีค่าเท่ากับ 8.47 และ 9.12 ตามลำดับ โดยการตรวจให้คะแนนโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเองนั้น ผู้ตรวจที่มีประสบการณ์ไม่เกิน 5 ปี จะตรวจให้คะแนนมีแนวโน้มสูงกว่าผู้ตรวจที่มีประสบการณ์มากกว่า 5 ปี ส่วนการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีคสำหรับผู้ตรวจมีประสบการณ์มากกว่า 5 ปี กับผู้ตรวจมีประสบการณ์ไม่เกิน 5 ปี มีคะแนนเฉลี่ยเท่ากับ 10.36 และ 9.28 ตามลำดับ โดยการตรวจให้คะแนนโดยใช้เกณฑ์การให้คะแนนแบบรูปรีค นั้น ผู้ตรวจที่มีประสบการณ์มากกว่า 5 ปี จะตรวจให้คะแนนมีแนวโน้มสูงกว่าผู้ตรวจที่มีประสบการณ์ไม่เกิน 5 ปี

ในกรณีที่แบบทดสอบวิชาคณิตศาสตร์มีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง เมื่อมีจำนวนผู้ตรวจ 3, 4, 5 และ 6 คน มีคะแนนเฉลี่ยอยู่ระหว่าง 8.79 - 9.05 โดยการตรวจให้คะแนนเมื่อมีจำนวนผู้ตรวจ 4 คน คะแนนเฉลี่ยมีแนวโน้มสูงที่สุด รองลงมาคือมีจำนวนผู้ตรวจ 5 คน และเมื่อมีจำนวนผู้ตรวจ 6 คน คะแนนเฉลี่ยมีแนวโน้มต่ำที่สุด ส่วนการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีค เมื่อมีจำนวนผู้ตรวจ 3, 4, 5 และ 6 คน มีคะแนนเฉลี่ยอยู่ระหว่าง 9.82 - 10.05 โดยการตรวจให้คะแนนเมื่อมีจำนวนผู้ตรวจ 4 คน คะแนนเฉลี่ยมีแนวโน้มสูงที่สุด รองลงมาคือมีจำนวนผู้ตรวจ 3 และ 6 คน และเมื่อมีจำนวนผู้ตรวจ 5 คน คะแนนเฉลี่ยมีแนวโน้มต่ำที่สุด

สำหรับคะแนนเบี่ยงเบนมาตรฐานของแบบทดสอบนั้น เมื่อจำแนกตามวิธีการตรวจให้คะแนน ประสบการณ์ของผู้ตรวจ และจำนวนผู้ตรวจ ปรากฏว่ามีค่าอยู่ระหว่าง 4.67 – 6.13 แสดงว่าแบบทดสอบมีการกระจายของคะแนนใกล้เคียงกัน

2. ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง กับใช้เกณฑ์การให้คะแนนแบบรูปรีค เมื่อมีจำนวนผู้ตรวจ 3 คน มีค่าเท่ากับ 0.69542 และ 0.76047 ตามลำดับ เมื่อมีจำนวนผู้ตรวจ 4 คน มีค่าเท่ากับ 0.75273 และ 0.80891 ตามลำดับ เมื่อมีจำนวนผู้ตรวจ 5 คน มีค่าเท่ากับ 0.79190 และ 0.84105 ตามลำดับ และเมื่อมีจำนวนผู้ตรวจ 6 คน มีค่าเท่ากับ 0.82035 และ 0.86394 ตามลำดับ โดยทั้ง 4 กรณีพบว่า การตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีค มีค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบสูงกว่าการตรวจให้คะแนนโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง และเมื่อนำค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบมาเปรียบเทียบกันระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน 2 วิธี เมื่อจำนวนผู้ตรวจเป็น 3, 4, 5 และ 6 คน พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทุกกรณี

5. ค่าสัมประสิทธิ์การสุป้อ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่ผู้ตรวจมีประสบการณ์มากกว่า 5 ปี กับผู้ตรวจมีประสบการณ์ไม่เกิน 5 ปี เมื่อมีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง มีค่าเท่ากับ 0.83238 และ 0.77003 ตามลำดับ และเมื่อมีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีด มีค่าเท่ากับ 0.84193 และ 0.79035 ตามลำดับ โดยทั้ง 2 กรณี เมื่อผู้ตรวจมีประสบการณ์มากกว่า 5 ปี ค่าสัมประสิทธิ์การสุป้อ้างอิงของแบบทดสอบมีค่าสูงกว่าเมื่อผู้ตรวจมีประสบการณ์ไม่เกิน 5 ปี และเมื่อนำค่าสัมประสิทธิ์การสุป้อ้างอิงของแบบทดสอบมาเปรียบเทียบกันระหว่างผู้ตรวจที่มีประสบการณ์ต่างกัน 2 กรณี เมื่อใช้วิธีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง และใช้เกณฑ์การให้คะแนนแบบรูปรีด พบว่า แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ทั้ง 2 กรณี

อภิปรายผล

การศึกษาวิจัยครั้งนี้สามารถอภิปรายผลได้ดังนี้

1. ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีจำนวนผู้ตรวจเท่ากัน เมื่อมีวิธีการตรวจให้คะแนนต่างกัน 2 วิธี คือ การตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง กับ การตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีด มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งสอดคล้องกับสมมติฐานการวิจัยที่ตั้งไว้ และยังพบว่าถ้าจำนวนผู้ตรวจเท่ากัน แบบทดสอบที่มีวิธีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีด มีค่าสัมประสิทธิ์การสรุปอ้างอิงสูงกว่าแบบทดสอบที่มีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง ทั้งนี้อาจเป็นเพราะการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีดทำให้ผลการตรวจของผู้ตรวจแต่ละคนมีความสอดคล้องกันมากขึ้น จึงทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบมีค่าสูงกว่าการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์ของผู้ตรวจเอง ซึ่งสอดคล้องกับงานวิจัยของซูซาน เลน และคณะ (Zuzanne Lane, et al. 1996 : 71 - 91) ที่ว่าความคลาดเคลื่อนที่เนื่องมาจากผู้ตรวจจะมีน้อยมากเมื่อมีการตรวจให้คะแนนแบบประเมินรวม และมีเกณฑ์การให้คะแนนแบบรูปรีด ดังนั้นเพื่อให้ผลการตรวจให้คะแนนมีความสอดคล้องภายในผู้ตรวจสูงขึ้น จึงควรมีการออกแบบทั้งการให้คะแนนแบบรูปรีด และการฝึกหรืออบรมผู้ตรวจ

2. ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบที่ผู้ตรวจมีประสบการณ์มากกว่า 5 ปี เมื่อมีวิธีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง กับ ตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีด มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ในทำนองเดียวกันค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบที่ผู้ตรวจมีประสบการณ์ไม่เกิน 5 ปี เมื่อมีวิธีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง กับ ตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีด มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติเช่นกัน ซึ่งขัดแย้งกับสมมติฐานการวิจัยที่ตั้งไว้ แสดงว่าถ้าผู้ตรวจมีประสบการณ์เหมือนกัน วิธีการตรวจให้คะแนนที่ต่างกัน 2 วิธี คือ การให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเองกับการให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีด ไม่มีผลต่อค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ทั้งนี้อาจเนื่องจากกลุ่มผู้ตรวจที่มีประสบการณ์ในการสอนวิชาคณิตศาสตร์เหมือนกันจะมีความชำนาญและแนวทางในการตรวจให้คะแนนแบบทดสอบที่คล้ายคลึงกัน ส่งผลให้ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบมีค่าไม่แตกต่างกัน

3. ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง เมื่อมีจำนวนผู้ตรวจต่างกัน คือ 3, 4, 5 และ 6 คน มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทำนองเดียวกันค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีด เมื่อมีจำนวนผู้ตรวจต่างกัน คือ 3, 4, 5 และ 6 คน มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เช่นกัน ซึ่งสอดคล้องกับสมมติฐานการวิจัยที่ตั้งไว้ และยังพบว่าวิธีการตรวจให้คะแนนทั้ง 2 วิธี เมื่อจำนวนผู้ตรวจแตกต่างกัน ยิ่งจำนวนผู้ตรวจมากค่าสัมประสิทธิ์ของแบบทดสอบจะมีค่าสูงขึ้น ซึ่งสอดคล้องกับคำกล่าวของอีเบล, เมอร์เรน และลีแมน, โกลแลนด์ และเบอร์กแมน (Ebel. 1972 : 152; Mehrens and Lehmann. 1973 : 234-236; Gronlund. 1977 : 79; Bergman. 1981 : 130-131) และไพศาล หวังพานิช

(2528 : 96-97) ที่ว่าไม่ว่าจะใช้วิธีใดในการตรวจให้คะแนน ถ้าทำได้ควรตรวจให้คะแนนโดยผู้ตรวจให้คะแนนหลาย ๆ คน แล้วใช้คะแนนเฉลี่ยหรือคะแนนรวมแทนคะแนนความสามารถของผู้สอบ ซึ่งจะให้ผลการวัดที่มีความเชื่อมั่นสูงขึ้นได้

4. คำสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ ที่มีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง เมื่อผู้ตรวจมีประสบการณ์มากกว่า 5 ปี กับผู้ตรวจที่มีประสบการณ์ไม่เกิน 5 ปี มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ในทำนองเดียวกันคำสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีด เมื่อผู้ตรวจมีประสบการณ์มากกว่า 5 ปี กับผู้ตรวจที่มีประสบการณ์ไม่เกิน 5 ปี มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 เช่นกัน ซึ่งสอดคล้องกับสมมุติฐานการวิจัยที่ตั้งไว้ และยังพบว่าเมื่อมีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเองและการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีด ผู้ตรวจที่มีประสบการณ์การมากกว่า 5 ปี จะส่งผลให้คำสัมประสิทธิ์การสรุปอ้างอิงมีค่าสูงกว่าผู้ตรวจที่มีประสบการณ์การไม่เกิน 5 ปี แสดงว่าประสบการณ์ที่แตกต่างกันมีผลต่อคำสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบที่มีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง ทั้งนี้อาจเป็นเพราะว่าผู้ตรวจที่มีประสบการณ์มากกว่า 5 ปี มีความเชี่ยวชาญในเนื้อหาวิชาและผ่านการตรวจให้คะแนนแบบทดสอบมามาก มีเกณฑ์การให้คะแนนที่เป็นมาตรฐานมากกว่าผู้ตรวจที่มีประสบการณ์ไม่เกิน 5 ปี ส่งผลให้คำสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบมีค่าสูงกว่า ส่วนในกรณีที่มีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีด ผู้ตรวจที่มีประสบการณ์มากกว่า 5 ปี อาจมีเกณฑ์การให้คะแนนของตนเองสอดคล้องกับเกณฑ์การให้คะแนนแบบรูปรีดที่ผู้วิจัยสร้างขึ้น หรือมีความเข้าใจในการใช้เกณฑ์การให้คะแนนแบบรูปรีดมากกว่า ดังนั้นผลการตรวจให้คะแนนจึงส่งผลให้คำสัมประสิทธิ์การสรุปอ้างอิงสูงกว่าผู้ตรวจที่มีประสบการณ์ไม่เกิน 5 ปี

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1) แบบทดสอบวิชาคณิตศาสตร์ที่มีลักษณะคำถามแบบปลายเปิด ควรมีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนแบบรูปรีด (Rubric scoring) เพราะคะแนนที่ได้จากการตรวจให้คะแนนแบบนี้ส่งผลให้คำสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบมีค่าสูงกว่าการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง

2) ในการตรวจให้คะแนนแบบทดสอบวิชาคณิตศาสตร์ที่มีลักษณะคำถามแบบปลายเปิดควรมีผู้ตรวจหลายคน ซึ่งในทางปฏิบัติอาจเป็นไปได้ยาก แต่เมื่อการตรวจให้คะแนนเป็นการตัดสินใจที่สำคัญและเป็นข้อสอบอัตนัยก็ควรมีผู้ตรวจหลายคน เพื่อให้ได้คำสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบที่เชื่อถือได้

3) ในกรณีที่จำนวนผู้ตรวจเท่ากัน การตรวจให้คะแนนโดยใช้เกณฑ์การให้คะแนนแบบรูปรีดจะดีกว่าการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจเอง

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

1) ควรจะมีการศึกษาลักษณะเดียวกันนี้โดยเพิ่มจำนวนผู้ตรวจ และศึกษาดูว่าจำนวนผู้ตรวจเท่าใด จึงจะให้ค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบที่ดีและเหมาะสมที่สุด

2) ควรจะมีการศึกษาลักษณะเดียวกันนี้กับรายวิชาอื่นๆ เพื่อดูว่ารายวิชาใดมีความเหมาะสมกับวิธีการตรวจให้คะแนนแบบใด จำนวนผู้ตรวจเท่าใด และประสบการณ์ของผู้ตรวจมีผลต่อการตรวจให้คะแนนหรือไม่ เพื่อเป็นประโยชน์ในทางการศึกษาต่อไป

บรรณานุกรม

บรรณานุกรม

- ชวาล แพรัตกุล. (2518). เทคนิคการวัดผล. กรุงเทพฯ: สำนักพิมพ์ วัฒนาพานิช.
- ชาญ วชิรเดช. (2538). *สัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบการแก้ปัญหาที่สอบด้วยไมโครคอมพิวเตอร์*. ปริญญาโท กศ.ม. กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร. ถ่ายเอกสาร.
- แดง กลางท่าไค้. (2531). *การประยุกต์ทฤษฎีการสรุปอ้างอิงในการหาความเชื่อมั่นการประเมินความตรงเชิงเนื้อหา*. วิทยานิพนธ์ ค.ม. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- นิภา ศรีไพโรจน์. (2528). *สถิตินอนพาราเมตริก*. มหาสารคาม: ภาควิชาพื้นฐานการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.
- บุญธรรม กิจปรีดาบริสุทธิ์. (2530). *สูตรการคำนวณทางสถิติและตารางสถิติ*. ภาควิชาศึกษาศาสตร์ คณะสังคมศาสตร์และมนุษยศาสตร์ มหาวิทยาลัยมหิดล.
- ผจงจิต อินทสุวรรณ. (2536). *แบบแผนเชิงสถิติของการทดลอง*. สถาบันวิจัยพฤติกรรมศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.
- ไพรัตน์ วงษ์นาม. (2533). *สัมประสิทธิ์การอ้างอิงสรุปของแบบทดสอบแบบความเรียง*. วิทยานิพนธ์ ค.ด. กรุงเทพฯ: จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- ไพศาล หวังพานิช. (2528, มกราคม - เมษายน). "ข้อสอบแบบความเรียง (Essay Test)," *วารสารการวัดผลการศึกษา*. 6(18): 89-97.
- ล้วน สายยศ และอังคณา สายยศ. (2536). *เทคนิคการวิจัยทางการศึกษา*. พิมพ์ครั้งที่ 3. กรุงเทพฯ: สำนักพิมพ์ศูนย์ส่งเสริมวิชาการ.
- _____. (2539). *เทคนิคการวัดผลการเรียนรู้*. กรุงเทพฯ: ชมรมเด็ก.
- ศึกษาริการ, กระทรวง. (2524). *หนังสือเรียนวิชาคณิตศาสตร์ ค 101*. กรุงเทพฯ: โรงพิมพ์คุรุสภาลาดพร้าว.
- _____. (2524). *หนังสือเรียนวิชาคณิตศาสตร์ ค 102*. กรุงเทพฯ : โรงพิมพ์คุรุสภาลาดพร้าว.
- ส.วาสนา ประवालพฤษ. (2533, กันยายน - ธันวาคม). "การพิจารณาความเชื่อมั่นของแบบทดสอบการปฏิบัติ," *วารสารการวัดผลการศึกษา*. 12(35): 28 - 42.
- สวงศ์ ร่มเงิน. (2541). *การเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบวิชาคณิตศาสตร์ที่มีรูปแบบการจัดเรียงข้อคำถามตามค่าความยากทางสถิติและพฤติกรรมการเรียนรู้ต่างกัน*. ปริญญาโท กศ.ม. กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร. ถ่ายเอกสาร.
- อังคณา สายยศ. (2536, มกราคม - เมษายน). "สัมประสิทธิ์การสรุปอ้างอิง (Generalizability coefficient : ρ^2)," *วารสารการวัดผลการศึกษา*. 14(42): 1 - 7.
- อาวุธ วัฒนสิน. (2536). *วิธีการวิจัยทางพฤติกรรมศาสตร์และสังคมศาสตร์*. ภาควิชาการวัดผลและการวิจัยการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร.
- Baker, E.L., et al. (1995, March/April). "Dimensionality and Generalizability of Domain – Independent Performance Assessments," *The Journal of Educational Research*. 89 (4): 197 - 205.

- Bergman, J. (1983). *Understanding Educational Measurement and Evaluation*. Boston: Houghton Mifflin Company.
- Coffman, W.E. (1971). "Essay Examinations," *Educational Measurement*. p.271 - 280. 2nd ed. Edited by Thorndike, R.L. Washington D.C.: American Council on Education.
- Courtney, E.W. (1996). "Alternative Assessment Techniques at the Elementary and Secondary Levels," เอกสารประกอบการฝึกอบรมเชิงปฏิบัติการ. คณะศึกษาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์.
- Crick, J.E., and Brennan, R.L. (1983, August). "A GENERALized Analysis of Variance System," *The American College Testing Program*. Iowa: ACT Publications.
- Cronbach, L.J., et al. (1972). *The Dependability of Behavioral Measurement: Theory of Generalizability for Score and Profiled*. New York: John Wiley and Sons Inc.
- Ebel, R.L. (1972). *Essentials of Education Measurement*. Englewood Cliffs, N.J.: Prentice-Hall.
- Feldt, L.S., Woodruff, D.J. and Saiih, F.A. (1987, March). "Statistical Inference for Coefficient Alpha," *Applied Psychological Measurement*. 11(1): 93 - 103.
- Ferguson, G.A. (1981). *Statistical analysis in psychology and education*. 5th ed. Tokyo: McGraw - Hill International Book.
- Gronlund, N.E. (1977). *Constructing Achievement Tests*. 2nd ed. Englewood Cliffs, N.J.: Prentice - Hall.
- Ibrahim, A.M. (1984, August). "The Dependability of Need Assessment Data: An Application of Generalizability Theory to the Ratings of Education Goals," *Dissertation Abstracts International*. 45(02): 449 - A.
- Kunnan, A.J. (1992, November). "An Investigation of a criterion - referenced test using G-theory, and factor and cluster analysis," *Language testing*. 9(1): 30 - 49.
- Lane, S., et al. (1996, Spring). "Generalizability and Validity of a Mathematics Performance Assessment," *Journal of Educational Measurement*. 33(1): 71 - 91.
- Macready, G.B. (1983, Spring). "The Use of Generalizability Theory for Assessing Relations among Items within Domains in Diagnostic Testing," *Applied Psychological Measurement*. 7(2): 149 - 157.
- MaryAnn Heglie - King. (1995). "Oregon's open - ended math assessment: What it means for teachers & students," *Alternative Assessments in oregon*. Oregon State University.
- Mehrens, W.A. and Lehmann, I.J. (1973). *Measurement and Evaluation in Education and Psychology*. New York: Holt, Rinehart and Winston.
- Siegel, S. (1956). *Nonparametric Statistics For the Behavioral Sciences*. New York: McGraw - Hill Book Company.
- Smith, P.L. (1978). "Sampling Errors of Variance Components in Small Sample Multifacet Generalizability Studies," *Journal of Educational Statistics*. 3(4): 319 - 346.

- Smith, P.L. (1979, Summer). "The Generalizability of Student Ratings of Courses: Asking the Right Questions," *Journal of Educational Measurement*. 16(2): 77 - 87.
- Stenmark, J.K. (1991). *Mathematics Assessment : myths, models, good question, and practical suggestion*. University of California : The National Council of Teachers of Mathematics, INC.
- Woodruff, D.J. and Feldt, L.S. (1986, September). "Test for Equality of Several Alpha Coefficients when their Sample Estimates Dependent," *Psychometrika*. 51: 293 - 413.

ภาคผนวก

ภาคผนวก ก

1. ตารางแสดงดัชนีค่าความง่าย ดัชนีค่าอำนาจจำแนกของแบบทดสอบวิชาคณิตศาสตร์ และค่าสัมประสิทธิ์ของความสอดคล้องภายในของผู้ตรวจ
2. แบบทดสอบวิชาคณิตศาสตร์ คำตอบตัวแบบ เกณฑ์การให้คะแนนแบบรูบริค และ เกณฑ์การให้คะแนนเฉพาะในแต่ละข้อ

ตาราง 9 ดัชนีค่าความง่าย (P_E) และดัชนีค่าอำนาจจำแนก (D) ของแบบทดสอบวิชาคณิตศาสตร์
เป็นรายข้อ และผลการวิเคราะห์ค่าสัมประสิทธิ์ความสอดคล้องภายในของผู้ตรวจ (W)

ฉบับที่	ลักษณะของแบบทดสอบ	ข้อ ที่	P_E	D	W	χ^2	ผลการ พิจารณา
1	การแก้ปัญหาในวิชาคณิตศาสตร์	1	0.64	0.57	0.92	79.61**	คัดเลือกไว้
		2	0.82	0.36	0.78	67.86**	ตัดออก
	คณิตศาสตร์เหตุผล	1	0.71	0.50	0.78	67.86**	ตัดออก
		2	0.61	0.64	0.90	77.43**	คัดเลือกไว้
	การสื่อสารทางคณิตศาสตร์	1	0.76	0.29	0.85	73.95**	ตัดออก
		2	0.54	0.34	0.84	73.08**	คัดเลือกไว้
2	การแก้ปัญหาในวิชาคณิตศาสตร์	1	0.59	0.61	0.93	80.91**	ตัดออก
		2	0.54	0.57	0.91	79.17**	คัดเลือกไว้
	คณิตศาสตร์เหตุผล	1	0.61	0.50	0.85	73.52**	คัดเลือกไว้
		2	0.64	0.50	0.74	64.38**	ตัดออก
	การสื่อสารทางคณิตศาสตร์	1	0.50	0.24	0.87	75.69**	คัดเลือกไว้
		2	0.50	-0.14	0.86	74.82**	ตัดออก

** มีนัยสำคัญทางสถิติที่ระดับ .01

แบบทดสอบวิชาคณิตศาสตร์

คำชี้แจง

1. แบบทดสอบฉบับนี้ประกอบด้วยข้อคำถามทั้งหมด 6 ข้อ ให้เวลา 60 นาที โดยแบ่งข้อคำถามเป็น 3 ด้าน ดังนี้
 - 1.1 การแก้ปัญหาในวิชาคณิตศาสตร์ จำนวน 2 ข้อ
 - 1.2 คณิตศาสตร์เหตุผล จำนวน 2 ข้อ
 - 1.3 การสื่อสารทางคณิตศาสตร์ จำนวน 2 ข้อ
2. ในแต่ละข้อกำหนดสถานการณ์ปัญหาทางคณิตศาสตร์มาให้ ให้นักเรียนแสดงวิธีการคิดพร้อมทั้งแสดงเหตุผลในการคิด และแสดงขั้นตอนการคำนวณเพื่อให้ได้คำตอบที่ถูกต้องเหมาะสมกับสถานการณ์ปัญหาที่กำหนดให้
3. ในการตอบคำถามแต่ละข้อ ให้นักเรียนเขียนตอบทั้งวิธีการคิด เหตุผลในการคิด และขั้นตอนการคำนวณ ลงในกระดาษคำตอบที่จัดเตรียมไว้ให้ทั้งหมด เพราะทุกส่วนมีผลต่อการให้คะแนน
4. พยายามอธิบายคำตอบให้ชัดเจน ใช้สัญลักษณ์และเครื่องหมายทางคณิตศาสตร์ให้ถูกต้องเหมาะสม และสามารถใช้แผนภาพ, กราฟ หรือตาราง ประกอบการอธิบายได้ตามความเหมาะสม

แบบทดสอบ

1. แบบทดสอบการแก้ปัญหาในวิชาคณิตศาสตร์

ข้อ (1) กระดาษสี่แผ่นหนึ่ง ขนาดกว้าง 24 นิ้ว ยาว 60 นิ้ว ถ้าต้องการตัดกระดาษสี่ออกเป็นรูปสี่เหลี่ยมจัตุรัสที่มีขนาดเท่าๆ กัน โดยไม่ให้มีเศษกระดาษเหลือ นักเรียนคิดว่าจะตัดกระดาษได้กี่แผ่น และแต่ละแผ่นมีพื้นที่เท่าไร จงอธิบายวิธีการคิด พร้อมทั้งแสดงวิธีทำและวาดภาพประกอบด้วย

ข้อ (2) แม่ค้าซื้อเงาะมา 180 กิโลกรัม คิดเป็นเงิน 630 บาท แม่ค้าต้องจ่ายค่าขนส่งไป 20 บาท ค่าถุงพลาสติกใส่เงาะ 35 บาท และเมื่อจะนำมาขายพบว่า มีเงาะเสียประมาณ 5% นักเรียนคิดว่า แม่ค้าควรจะขายเงาะราคา กิโลกรัมละเท่าไรจึงจะได้กำไร จงอธิบายวิธีการคิดของนักเรียน

2. แบบทดสอบคณิตศาสตร์เหตุผล

ข้อ (1) สุเทพมีกิจการรถเช่า เขามีรถจักรยานยนต์และรถสามล้อเครื่องให้เช่า ถ้านับล้อรถทั้ง 2 ชนิดรวมกันจะได้ 24 ล้อ นักเรียนคิดว่าสุเทพมีรถชนิดใดมากกว่ากัน เพราะเหตุใด จงแสดงวิธีการคิด พร้อมทั้งเขียนแผนภาพประกอบการอธิบายด้วย

ข้อ (2) กำหนดให้รูปสามเหลี่ยม ABC มี $\overline{AB} = \overline{AC}$ และ $\angle ABC = 45^\circ$ จากสิ่งที่กำหนดให้นักเรียนสามารถใช้ความรู้เกี่ยวกับเรื่องสมบัติของรูปสามเหลี่ยมที่เรียนมาสรุปหรือหาค่าอะไรได้บ้าง ให้นักเรียนเขียนมาเป็นข้อๆ พร้อมทั้งบอกเหตุผลในการคิด รวมทั้งวาดภาพประกอบให้ถูกต้องด้วย

3. แบบทดสอบการสื่อสารทางคณิตศาสตร์

ข้อ (1) ครูให้นักเรียนทำรายงานสรุปเรื่อง ห.ร.ม. กับ ค.ร.น. ว่ามีส่วนที่เหมือนและแตกต่างกันอย่างไร ในแต่ละด้านต่อไปนี้

- ความหมาย
- วิธีการคำนวณ
- การนำไปใช้

ให้นักเรียนทำรายงานฉบับนี้ โดยเขียนตารางประกอบด้วย

เฉลยคำตอบ

แบบทดสอบการแก้ปัญหาในวิชาคณิตศาสตร์

ข้อ (1) กระดาษสี่แผ่นหนึ่งขนาดกว้าง 24 นิ้ว ยาว 60 นิ้ว ถ้าต้องการตัดกระดาษสี่ออกเป็นรูปสี่เหลี่ยมจัตุรัสที่มีขนาดเท่าๆ กัน โดยไม่ให้มีเศษกระดาษเหลือ นักเรียนคิดว่าจะตัดกระดาษได้กี่แผ่น และแต่ละแผ่นมีพื้นที่เท่าไร จงอธิบายวิธีการคิด พร้อมทั้งแสดงวิธีทำ และวาดภาพประกอบด้วย

คำตอบที่เป็นไปได้

วิธีที่ 1 ใช้หลักการหา ห.ร.ม. เพื่อหาความยาวของด้านของรูปสี่เหลี่ยมจัตุรัสที่จะตัด โดยวิธีใดวิธีหนึ่งต่อไปนี้

1. การตั้งหาร

$$\begin{array}{r} 2 \overline{) 24, 60} \\ 2 \overline{) 12, 30} \\ 3 \overline{) 6, 15} \\ \underline{2, 5} \end{array}$$

$$\text{ห.ร.ม.} = 2 \times 2 \times 3 = 12$$

2. การแยกตัวประกอบ

$$\begin{array}{l} 24 = 2 \times 2 \times 2 \times 3 \\ 60 = 2 \times 2 \times 3 \times 5 \end{array}$$

$$\begin{array}{l} \text{ห.ร.ม.} = 2 \times 2 \times 3 \\ = 12 \end{array}$$

ดังนั้น ต้องตัดกระดาษสี่เป็นรูปสี่เหลี่ยมจัตุรัส ที่มีด้านยาวด้านละ 12 นิ้ว

หาพื้นที่ของสี่เหลี่ยมจัตุรัสจากสูตร ด้าน $\times$ ด้าน

ดังนั้น รูปสี่เหลี่ยมจัตุรัส 1 แผ่น มีพื้นที่ $= 12 \times 12 = 144$ ตารางนิ้ว

หาพื้นที่ของกระดาษจากสูตร กว้าง $\times$ ยาว

ดังนั้น กระดาษสี่แผ่นใหญ่มีพื้นที่ $= 24 \times 60 = 1,440$ ตารางนิ้ว

ต้องการหว่า จะตัดกระดาษได้กี่แผ่น หาได้โดยนำพื้นที่ของกระดาษสี่หารด้วย

พื้นที่ของรูปสี่เหลี่ยมจัตุรัส 1 แผ่น จะได้ $1,440 \div 144 = 10$ แผ่น

สรุป จะตัดกระดาษได้ 10 แผ่น แต่ละแผ่นมีพื้นที่ 144 ตารางนิ้ว

วิธีที่ 2 จากวิธีที่ 1 จะเห็นว่ามีตัวประกอบร่วมของ 24 และ 60 อีก (จำนวนที่หาร 24 และ 60

ลงตัว คือ 2, 3, 4, 6 รวมทั้ง 1 ด้วย เพราะว่า 1 เป็นตัวประกอบของทุกจำนวน

ดังนั้น เราสามารถตัดกระดาษรูปสี่เหลี่ยมจัตุรัสให้มีด้านยาวด้านละ 1, 2, 3, 4 และ 6 นิ้ว ได้ด้วย

1. ถ้าให้รูปสี่เหลี่ยมจัตุรัสมีด้านยาวด้านละ 1 นิ้ว

รูปสี่เหลี่ยมจัตุรัสจะมีพื้นที่ $= 1 \times 1 = 1$ ตารางนิ้ว และ

สามารถตัดกระดาษได้ $= 1,440 \div 1 = 1,440$ แผ่น

2. ถ้าให้รูปสี่เหลี่ยมจัตุรัสมีด้านยาวด้านละ 2 นิ้ว
รูปสี่เหลี่ยมจัตุรัสจะมีพื้นที่ = $2 \times 2 = 4$ ตารางนิ้ว และ
สามารถตัดกระดาษได้ = $1,440 \div 4 = 360$ แผ่น
3. ถ้าให้รูปสี่เหลี่ยมจัตุรัสมีด้านยาวด้านละ 3 นิ้ว
รูปสี่เหลี่ยมจัตุรัสจะมีพื้นที่ = $3 \times 3 = 9$ ตารางนิ้ว และ
สามารถตัดกระดาษได้ = $1,440 \div 9 = 160$ แผ่น
4. ถ้าให้รูปสี่เหลี่ยมจัตุรัสมีด้านยาวด้านละ 4 นิ้ว
รูปสี่เหลี่ยมจัตุรัสจะมีพื้นที่ = $4 \times 4 = 16$ ตารางนิ้ว และ
สามารถตัดกระดาษได้ = $1,440 \div 16 = 90$ แผ่น
5. ถ้าให้รูปสี่เหลี่ยมจัตุรัสมีด้านยาวด้านละ 6 นิ้ว
รูปสี่เหลี่ยมจัตุรัสจะมีพื้นที่ = $6 \times 6 = 36$ ตารางนิ้ว และ
สามารถตัดกระดาษได้ = $1,440 \div 36 = 40$ แผ่น
(นักเรียนจะทำเพียงวิธีเดียว)

วิธีที่ 3 ใช้การหา ห.ร.ม. ของ 24 และ 60 ด้วยวิธีการตั้งหาร

$$\begin{array}{r}
 2 \overline{) 24, 60} \\
 2 \overline{) 12, 30} \\
 3 \overline{) 6, 15} \\
 \underline{2, 5}
 \end{array}$$

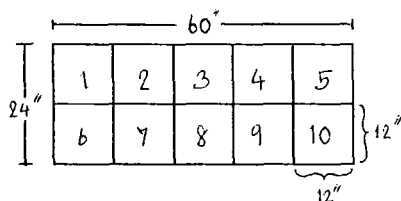
สรุป 1. ห.ร.ม. ของ 24 และ 60 คือ $2 \times 2 \times 3 = 12$

ดังนั้น รูปสี่เหลี่ยมจัตุรัสแต่ละแผ่นจะมีพื้นที่ = $12 \times 12 = 144$ ตารางนิ้ว

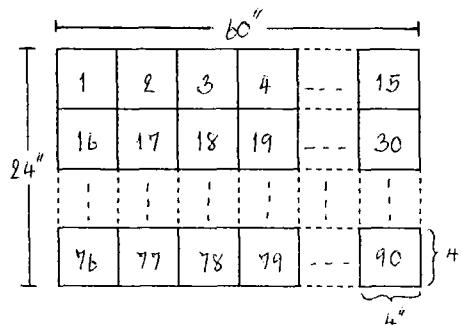
2. จากผลลัพธ์ของการตั้งหาร แสดงว่าจะตัดกระดาษได้ทั้งหมด = $2 \times 5 = 10$ แผ่น
(เมื่อ 2 คือ จำนวนแผ่นในด้านกว้าง และ 5 คือ จำนวนแผ่นในด้านยาว)

หมายเหตุ ทุกวิธีที่กล่าวมาข้างต้น นักเรียนต้องวาดรูปประกอบด้วย

เช่น ถ้าตัดกระดาษได้ 10 แผ่น แต่ละแผ่น
มีพื้นที่ 144 ตารางนิ้ว



ถ้าตัดกระดาษได้ 90 แผ่น แต่ละแผ่น
มีพื้นที่ 16 ตารางนิ้ว



ข้อ (2) แม่ค้าซื้อเงาะมา 180 กิโลกรัม คิดเป็นเงิน 630 บาท แม่ค้าต้องจ่ายค่าขนส่งไป 20 บาท
ค่าถุงพลาสติกใส่เงาะ 35 บาท และเมื่อนำมาขายพบว่า มีเงาะเสียประมาณ 5%
นักเรียนคิดว่า แม่ค้าควรขายเงาะราคา กิโลกรัมละเท่าไรจึงจะได้กำไร จงอธิบายวิธี
การคิดของนักเรียน

คำตอบที่เป็นไปได้

วิธีที่ 1 ขั้นที่ 1 คิดเงินค่าใช้จ่ายทั้งหมดจาก

$$\begin{aligned}\text{ราคาเงาะ} + \text{ค่าขนส่ง} + \text{ค่าถุงพลาสติก} &= 630 + 20 + 35 \\ &= 685 \text{ บาท}\end{aligned}$$

ขั้นที่ 2 หาจำนวนเงาะที่เสียไป 5%

เงาะทั้งหมด 180 กก. เสียไปประมาณ 5%

กล่าวคือ เงาะ 180 กก. เสียไปประมาณ 5 กก.

$$\text{เงาะ 180 กก. เสียไปประมาณ } \frac{180 \times 5}{100} = 9 \text{ กก.}$$

ดังนั้น จะเหลือเงาะที่ขายได้จริงประมาณ $180 - 9 = 171$ กก.

(ขั้นที่ 2 นี้ อาจทำอีกวิธีหนึ่ง ดังนี้)

หาจำนวนเงาะที่ขายได้จริง

เงาะทั้งหมด 180 กก. เสียไปประมาณ 5%

ฉะนั้น ขายได้จริงประมาณ $100 - 5 = 95\%$

กล่าวคือ เงาะ 180 กก. ขายได้จริง 95 กก.

$$\text{เงาะ 180 กก. ขายได้จริง } \frac{180 \times 95}{100} = 171 \text{ กก.}$$

ขั้นที่ 3 หาว่าควรขายเงาะ กก.ละเท่าไรจึงจะได้กำไร

จาก ค่าใช้จ่ายทั้งหมดที่แม่ค้าเสียไป 685 บาท

และ เงาะที่ขายได้จริง 171 กก.

ต้องขายเงาะ กก.ละ $685 \div 171 \approx 4.01$ บาท

ดังนั้น ถ้าจะขายให้ได้กำไรจะต้องขาย กก.ละ 5 บาทขึ้นไป (แล้วแต่นักเรียนจะตอบว่า
ควรขายราคาเท่าใด)

วิธีที่ 2 ขั้นที่ 1 หาราคาเงาะที่ซื้อเข้ามาว่า กก.ละเท่าไร

$$\text{โดยนำ } 630 \div 180 = 3.50$$

แสดงว่า ซื้อเงาะมาราคา กก.ละ 3.50 บาท

ขั้นที่ 2 หาจำนวนเงาะที่เสียไป 5% (คิดเช่นเดียวกับวิธีที่ 1 ขั้นที่ 2)

จะได้ว่า มีเงาะเสีย 9 กก.

และขายได้จริง 171 กก.

ขั้นที่ 3 คัดราคาเงาะที่เสียไป

จาก $9 \times 3.50 = 31.50$ บาท (เสีย 9 กก. ละ 3.50 บาท)

ขั้นที่ 4 คัดค่าใช้จ่ายไม่รวมค่าเงาะที่ซื้อมา

จาก ค่าขนส่ง + ค่าถุงพลาสติก + ค่าเงาะเสีย = $20 + 35 + 31.50$
= 86.50 บาท

ขั้นที่ 5 จากจำนวนเงาะที่ขายได้ 171 กก.

ราคาทุน กก.ละ 3.50 บาท

(1) ถ้าขาย กก. ละ 4 บาท = $171 \times 4 = 684$ บาท

จากเงินลงทุนครั้งแรก 630 บาท

แสดงว่า ถ้าขาย กก.ละ 4 บาท จะได้กำไร $684 - 630 = 54$ บาท

(ซึ่งเมื่อเทียบกับค่าใช้จ่าย 86.50 บาทแล้ว ยังขาดทุนอยู่)

(2) ถ้าขาย กก.ละ 5 บาท = $171 \times 5 = 855$ บาท

ขายราคานี้จะได้กำไร $855 - 630 = 225$ บาท

(ซึ่งเมื่อเทียบกับค่าใช้จ่าย 86.50 บาทแล้ว ได้กำไร)

ดังนั้น ต้องขายเงาะราคา กก.ละ 5 บาทขึ้นไป จึงจะได้กำไร

แบบทดสอบคณิตศาสตร์เหตุผล

ข้อ (1) สุเทพมีกิจการรถเช่า เขามีรถจักรยานยนต์และรถสามล้อเครื่องให้เช่า ถ้านับล้อรถทั้ง 2 ชนิดรวมกันจะได้ 24 ล้อ นักเรียนคิดว่าสุเทพมีรถชนิดใดมากกว่ากัน เพราะเหตุใด
จงแสดงวิธีการคิด พร้อมทั้งเขียนแผนภาพประกอบการอธิบายด้วย

คำตอบที่เป็นไปได้

วิธีการคิด - สุเทพมีรถ 2 ชนิด คือ รถจักรยานยนต์ ซึ่งมี 2 ล้อ และรถสามล้อเครื่อง ซึ่งมี 3 ล้อ
เมื่อนับล้อรถรวมกันแล้วได้ 24 ล้อ

- ต้องหาจำนวน 2 จำนวนที่รวมกันได้ 24

โดยคิดว่า จำนวน 2 จำนวนนั้น คือ จำนวนล้อของรถทั้ง 2 ชนิด

ถ้าให้จำนวนที่ 1 เป็นรถจักรยานยนต์ซึ่งมี 2 ล้อ จำนวนที่ 1 ต้องหารด้วย 2 ลงตัว
(หรือจำนวนคู่นั่นเอง)

ส่วนจำนวนที่ 2 เป็นรถสามล้อเครื่องซึ่งมี 3 ล้อ จำนวนที่ 2 ต้องหารด้วย 3 ลงตัว

- ดังนั้นมีวิธีที่เป็นไปได้ 3 วิธี คือ

วิธีที่ 1 $6 + 18 = 24$

คือ มีรถจักรยานยนต์ $6 \div 2 = 3$ คัน

มีรถสามล้อเครื่อง $18 \div 3 = 6$ คัน

ดังนั้น สุเทพจะมีรถสามล้อเครื่องมากกว่ารถจักรยานยนต์

วิธีที่ 2 $12 + 12 = 24$

คือ มีรถจักรยานยนต์ $12 \div 2 = 6$ คัน

มีรถสามล้อเครื่อง $12 \div 3 = 4$ คัน

ดังนั้น สุเทพจะมีรถจักรยานยนต์มากกว่ารถสามล้อเครื่อง

วิธีที่ 3 $18 + 6 = 24$

คือ มีรถจักรยานยนต์ $18 \div 2 = 9$ คัน

มีรถสามล้อเครื่อง $6 \div 3 = 2$ คัน

ดังนั้น สุเทพจะมีรถจักรยานยนต์มากกว่ารถสามล้อเครื่อง

หมายเหตุ 1. วิธีคิดของนักเรียนแต่ละคนอาจไม่เหมือนกัน ส่วนคำตอบที่ว่า สุเทพมีรถชนิดใดมากกว่ากัน

นักเรียนอาจจะตอบได้ 3 กรณี คือ

1.1 รถจักรยานยนต์มากกว่ารถสามล้อเครื่อง

1.2 รถสามล้อเครื่องมากกว่ารถจักรยานยนต์

1.3 ตอบทั้งข้อ 1 และ 2 เพราะคิดได้ทั้ง 2 แบบ

2. วิธีที่นักเรียนคิด ต้องมีการวาดภาพประกอบการอธิบายด้วย

ตัวอย่าง ถ้าตอบว่ามีรถจักรยานยนต์ 6 คัน

รถสามล้อเครื่อง 4 คัน

แผนภาพคือ

รถจักรยานยนต์	     
รถสามล้อเครื่อง	   

ข้อ (2) กำหนดให้รูปสามเหลี่ยม ABC มี $\overline{AB} = \overline{AC}$ และ $\hat{ABC} = 45^\circ$ จากสิ่งที่กำหนดให้นักเรียนสามารถใช้ความรู้เกี่ยวกับเรื่องสมบัติของรูปสามเหลี่ยมที่เรียนมาสรุปหรือหาค่าอะไรได้บ้าง ให้นักเรียนเขียนมาเป็นข้อๆ พร้อมทั้งบอกเหตุผลในการคิด รวมทั้งวาดภาพประกอบให้ถูกต้องด้วย

คำตอบที่เป็นไปได้

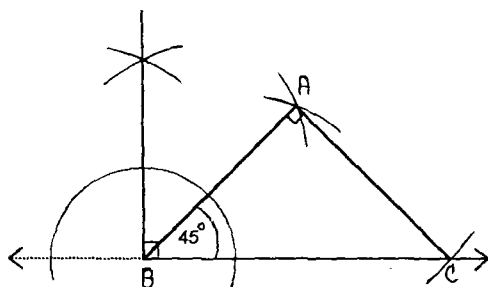
สรุปเป็นข้อๆ ดังนี้

1. สามเหลี่ยม ABC เป็นสามเหลี่ยมหน้าจั่ว เพราะมีแขนของมุมที่ฐานทั้งสองมุมยาวเท่ากัน ($\overline{AB} = \overline{AC}$)
2. $\hat{ABC} = \hat{ACB} = 45^\circ$ เพราะสามเหลี่ยมหน้าจั่วจะมีมุมที่ฐานกางเท่ากัน
3. $\hat{BAC} = 90^\circ$ เพราะมุมภายในของสามเหลี่ยมรวมกันได้ 180°

$$\begin{aligned}\hat{BAC} &= 180^\circ - \hat{ABC} - \hat{ACB} \\ &= 180^\circ - 45^\circ - 45^\circ = 90^\circ\end{aligned}$$

4. สามเหลี่ยม ABC เป็นสามเหลี่ยมมุมฉากด้วย เพราะมีมุมๆ หนึ่งเป็นมุมฉาก ($\hat{BAC} = 90^\circ$)
5. ถ้ารู้ความยาวของ $\overline{AB}, \overline{AC}$ หรือ $\overline{BC}$ สองด้านจะสามารถหาด้านที่เหลืออีกหนึ่งด้านได้จากสูตรพิทาโกรัส คือ $BC^2 = AC^2 + AB^2$
6. ถ้ารู้ความยาวของ $\overline{AB}, \overline{AC}$ หรือ $\overline{BC}$ สองด้านจะสามารถหาพื้นที่ของสามเหลี่ยม ABC ได้จากสูตรหาพื้นที่สามเหลี่ยม $= \frac{1}{2} \times \text{ฐาน} \times \text{สูง}$

หมายเหตุ นักเรียนต้องวาดภาพประกอบด้วย เช่น วาดภาพประกอบโดยใช้วงเวียน



(วิธีการของนักเรียนอาจแตกต่างกันไปจากนี้ แล้วแต่กรณี)

แบบทดสอบการสื่อสารทางคณิตศาสตร์

ข้อ (1) ครูให้นักเรียนทำรายงานสรุปเรื่อง ห.ร.ม. กับ ค.ร.น. ว่ามีส่วนที่เหมือนและแตกต่างกันอย่างไร ในแต่

ละด้านต่อไปนี้

- ความหมาย
- วิธีการคำนวณ
- การนำไปใช้

ให้นักเรียนทำรายงานฉบับนี้ โดยเขียนตารางประกอบด้วย

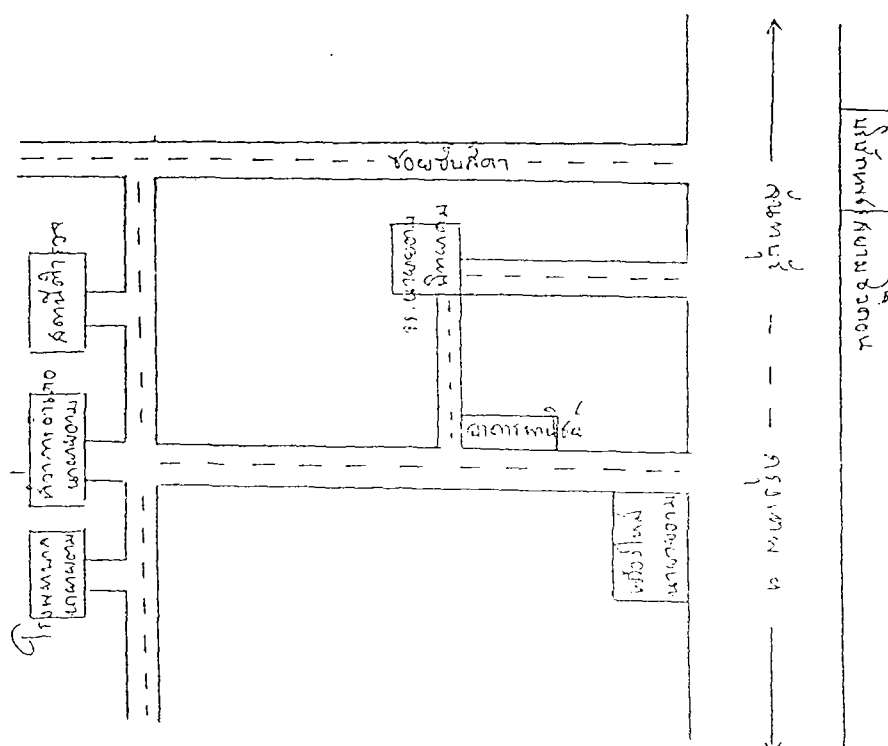
ต้องสร้างตารางประกอบ

ตารางสรุปเรื่อง ห.ร.ม. และ ค.ร.น.

ด้าน	เหมือน	ต่าง	ห.ร.ม.	ค.ร.น.
ความหมาย		/	ห.ร.ม. ตัวหารร่วมมาก คือ จำนวนนับที่เป็นตัวประกอบ ร่วมที่มากที่สุดของจำนวนที่ กำหนดให้ ตั้งแต่ 2 จำนวน ขึ้นไป	ค.ร.น. ตัวคูณร่วมน้อย คือ จำนวนนับที่น้อยที่สุดที่มี จำนวนที่กำหนดให้ตั้งแต่ 2 จำนวนขึ้นไปเป็นตัว ประกอบ
วิธีการคำนวณ	/	/	1. ใช้วิธีการตั้งหาร หรือ ใช้วิธีการแยกตัวประกอบในการคำนวณ	
			2. ถ้าเป็นการตั้งหาร ให้หา จำนวนเฉพาะที่หารจำนวนที่ กำหนดให้ทุกจำนวนลงตัว แล้วนำตัวหารทั้งหมดมาคูณ กันไม่รวมผลลัพธ์ จะได้ ห.ร.ม. ของจำนวนที่กำหนด (ยกตัวอย่างประกอบด้วย)	2. ถ้าเป็นการตั้งหาร ให้หา จำนวนเฉพาะที่หารจำนวน ที่กำหนดให้ลงตัว 1 จำนวน ขึ้นไป จนได้ผลลัพธ์เป็น 1 หรือจำนวนเฉพาะ แล้วนำ ทั้งตัวหารและผลลัพธ์มา คูณกัน จะได้ ค.ร.น. ของ จำนวนที่กำหนด (ยกตัวอย่างประกอบด้วย)
			3. ถ้าเป็นการแยกตัวประกอบ	
			3.1 ให้แยกตัวประกอบของทุกจำนวนที่จะหา ห.ร.ม. หรือ ค.ร.น.	
	/	/	3.2 หาตัวประกอบร่วมที่เหมือน กันทุกจำนวนที่จะหา ห.ร.ม. แล้วนำตัวประกอบร่วมนั้นมา คูณกัน จะได้ ห.ร.ม. ของ จำนวนที่ต้องการ (ยกตัวอย่างประกอบด้วย)	3.2 หาตัวประกอบร่วมที่เหมือน กันทุกจำนวนก่อน แล้วหา ตัวประกอบของจำนวนตั้ง แต่ 2 จำนวนขึ้นไป แล้วนำ ตัวประกอบร่วมนั้นมาคูณ ร่วมกับจำนวนที่เหลือทั้ง หมด จะได้ ค.ร.น. ของ

ด้าน	เหมือน	ต่าง	ห.ร.ม.	ค.ร.น.
วิธีการคำนวณ (ต่อ)		/		จำนวนที่กำหนด (ยกตัวอย่างประกอบ ด้วย)
	/		3.3 ผลการคำนวณ ถ้าหา ห.ร.ม. และ ค.ร.น. ของจำนวนชุดเดียวกัน ห.ร.ม. จะมีค่าน้อยกว่า ค.ร.น. เสมอ	
การนำไปใช้		/	1. ใช้เมื่อโจทย์ให้หาส่วนที่มากที่สุด, ใหญ่ที่สุด, ยาวที่สุด, กว้างที่สุด ฯลฯ ให้เท่าๆ กัน โดยไม่เหลือเศษ	1. ใช้เมื่อโจทย์ให้หาส่วนที่น้อยที่สุด, สั้นที่สุด ฯลฯ ส่วนมากจะเป็นเวลา ระยะทาง จำนวนหน้า ที่จะมาตรงกัน หรือทำพร้อมกันอีกครั้ง 2. ใช้ในการบวกลบเศษส่วน ที่มีส่วนไม่เท่ากัน ต้องทำส่วนให้เท่ากันก่อน โดยใช้การหา ค.ร.น. ของส่วน

ข้อ (2) จากแผนที่ด้านล่างนี้ ถ้ามีนักท่องเที่ยวมาถามนักเรียนว่าจะไปโรงพยาบาลนายายอามได้อย่างไร นักเรียนจะบอกทางไปกับเขาว่าอย่างไร โดยเริ่มออกเดินทางจากโรงเรียนนายายอามพิทยาคม



มาตราส่วน 1 เซนติเมตร : 100 เมตร

คำตอบที่เป็นไปได้ มีวิธีการเดินทางไปโรงพยาบาลนายายอาม 3 วิธี ดังนี้

วิธีที่ 1 จากโรงเรียนนายายอามพิทยาคม หันหน้าไปทางอาคารพาณิชย์ เดินตรงไปประมาณ 200 เมตร พบทางแยกให้เลี้ยวขวาเดินตรงไปอีกประมาณ 360 เมตร จะพบที่ว่าการอำเภอนายายอามอยู่ฝั่งตรงกันข้าม ให้เลี้ยวซ้ายเดินไปอีกประมาณ 100 เมตร จะพบโรงพยาบาลนายายอามอยู่ทางขวามือ

วิธีที่ 2 จากโรงเรียนนายายอามพิทยาคม เดินออกไปทางถนนใหญ่ที่จะไป จันทบุรี - กรุงเทพฯ ประมาณ 300 เมตร ถึงถนนใหญ่เลี้ยวซ้ายไปทางจันทบุรีเดินไปประมาณ 100 เมตร จะพบซอยทางซ้ายมือคือซอยขับสิดาซึ่งอยู่ตรงข้ามกับบริษัทเพชรสยามซีวิคอน เดินเข้าซอยไปประมาณ 700 เมตรจะพบทางแยกด้านซ้ายมือให้เลี้ยวซ้ายเข้าไปประมาณ

150 เมตร จะพบสถานีตำรวจอยู่ทางขวามือเดินไปอีกประมาณ 150 เมตร จะพบที่ว่าการอำเภอ นายายอามอยู่ทางขวามือให้เดินตรงไปอีกประมาณ 100 เมตร ก็จะพบโรงพยาบาลนายายอามอยู่ทางขวามือ

วิธีที่ 3 จากโรงเรียนนายายอามพิทยาคม เดินออกไปทางถนนใหญ่ที่จะไป จันทบุรี - กรุงเทพฯ ประมาณ 300 เมตร ถึงถนนใหญ่เลี้ยวขวาไปทางกรุงเทพฯ เดินไปประมาณ 200 เมตร จะพบทางแยกด้านขวามือให้เลี้ยวขวาไปประมาณ 300 เมตร จะมีทางแยกขวามือเข้าโรงเรียนนายายอามพิทยาคม เดินตรงไปอีกประมาณ 400 เมตร จะพบที่ว่าการอำเภอ นายายอามอยู่ฝั่งตรงกันข้ามให้เลี้ยวซ้ายเดินไปอีกประมาณ 100 เมตร จะพบโรงพยาบาลนายายอามอยู่ทางขวามือ

(นักเรียนแต่ละคนอาจมีวิธีการนำเสนอที่แตกต่างกัน)

เกณฑ์การให้คะแนนแบบรูบรีค (Rubric scoring)

ระดับคะแนน 4
- อธิบายความเข้าใจในแนวคิดและกระบวนการในการหาคำตอบได้สมบูรณ์ - ปฏิบัติตามขั้นตอนการคำนวณได้สมบูรณ์ถูกต้อง - อธิบายคำตอบได้สมบูรณ์ ชัดเจน ไม่คลุมเครือ - ให้เหตุผลสนับสนุนการอธิบาย และยกตัวอย่างประกอบได้สมบูรณ์ถูกต้อง - แผนภาพประกอบสมบูรณ์ - ใช้คำศัพท์เฉพาะ และสัญลักษณ์ (เครื่องหมาย) ทางคณิตศาสตร์ ได้อย่างถูกต้อง เหมาะสม - สื่อสารกับผู้อ่านได้อย่างดี เมื่ออ่านแล้วตีความ / ทำความเข้าใจได้ตรงกัน
ระดับคะแนน 3
- อธิบายความเข้าใจในแนวคิดและกระบวนการในการหาคำตอบได้เกือบสมบูรณ์ - ปฏิบัติตามขั้นตอนการคำนวณได้ถูกต้อง แต่อาจมีความผิดพลาดในการคำนวณเล็กน้อย - อธิบายคำตอบได้ชัดเจน ไม่คลุมเครือ แต่อาจมีบางส่วนไม่สมบูรณ์หรือไม่ได้อธิบาย - ให้เหตุผลสนับสนุนการอธิบาย และยกตัวอย่างประกอบได้เกือบสมบูรณ์ อาจมีช่องว่าง - แผนภาพประกอบเกือบสมบูรณ์ อาจมีข้อบกพร่องบ้างเล็กน้อย - ใช้คำศัพท์เฉพาะ และสัญลักษณ์ (เครื่องหมาย) ทางคณิตศาสตร์ ถูกต้องเกือบหมด - สื่อสารกับผู้อ่านได้ตีพอสมควร อาจมีบางส่วนที่ไม่ชัดเจนอยู่บ้าง
ระดับคะแนน 2
- อธิบายความเข้าใจในแนวคิดและกระบวนการในการหาคำตอบได้บางส่วน - ปฏิบัติตามขั้นตอนการคำนวณได้ถูกต้องเพียงครึ่งหนึ่งของขั้นตอนการคำนวณทั้งหมด - อธิบายคำตอบไม่ชัดเจน หรือมีสองนัย - ให้เหตุผลสนับสนุนการอธิบายไม่สมบูรณ์ หรือไม่มีการยกตัวอย่างประกอบการอธิบาย - แผนภาพประกอบไม่ชัดเจน หรือมีข้อบกพร่องมาก - ใช้คำศัพท์เฉพาะ และสัญลักษณ์ (เครื่องหมาย) ทางคณิตศาสตร์ ได้เพียงบางส่วน - การสื่อสารกับผู้อ่าน คลุมเครือ อ่านแล้วตีความ / ทำความเข้าใจได้ยาก
ระดับคะแนน 1
- อธิบายความเข้าใจในแนวคิดและกระบวนการในการหาคำตอบได้น้อยมาก - ขั้นตอนการคำนวณมีที่ผิดพลาดมาก - อธิบายคำตอบผิด หรือเข้าใจยาก - ให้เหตุผลสนับสนุนการอธิบายได้น้อยมาก หรือยกตัวอย่างประกอบการอธิบาย ไม่เหมาะสม - แผนภาพประกอบส่วนใหญ่ไม่ถูกต้อง หรือตีความได้น้อย - ใช้คำศัพท์เฉพาะ และสัญลักษณ์ (เครื่องหมาย) ทางคณิตศาสตร์ส่วนมากผิด - การสื่อสารกับผู้อ่านไม่ดี ผู้อ่านๆ แล้วไม่ค่อยเข้าใจ

ระดับคะแนน 0

- ไม่สามารถอธิบายความเข้าใจในแนวคิดและกระบวนการในการหาคำตอบได้
 - ไม่มีการคำนวณ
 - อธิบายคำตอบไม่ตรงประเด็น
 - ไม่ให้เหตุผลสนับสนุนการอธิบาย หรือไม่มียกตัวอย่างประกอบการอธิบาย
 - แผนภาพประกอบผิดพลาด
 - ใช้คำศัพท์เฉพาะ และสัญลักษณ์ (เครื่องหมาย) ทางคณิตศาสตร์ที่ไม่เกี่ยวกับปัญหา
 - ไม่สามารถสื่อสารกับผู้อ่านให้เข้าใจตรงกันได้
-

เกณฑ์การให้คะแนนเฉพาะในแต่ละข้อ

1. แบบทดสอบการแก้ปัญหาในวิชาคณิตศาสตร์

ข้อที่ 1

- 4 คะแนน - ทำได้เหมือนในคำตอบตัวแบบ พร้อมคำอธิบาย และมีภาพประกอบที่ถูกต้องสมบูรณ์
- 3 คะแนน - ทำได้เหมือนในคำตอบตัวแบบ พร้อมคำอธิบาย แต่ภาพประกอบไม่สมบูรณ์
หรือ ทำได้เหมือนในคำตอบตัวแบบ มีภาพประกอบสมบูรณ์ แต่คำอธิบายมีข้อผิดพลาด
- 2 คะแนน - ใช้หลักการหา ห.ร.ม. ในการคิด และตอบคำถามได้ 1 คำถามพร้อมคำอธิบาย ภาพประกอบไม่สมบูรณ์
- 1 คะแนน - ใช้หลักการหา ห.ร.ม. ในการคิด แต่ตอบคำถามผิด ภาพประกอบไม่ชัดเจน
- 0 คะแนน - ใช้วิธีการผิด คำนวณผิด และแผนภาพประกอบผิด

ข้อที่ 2

- 4 คะแนน - ทำได้เหมือนในคำตอบตัวแบบ พร้อมคำอธิบายถูกต้อง
- 3 คะแนน - ทำได้ทุกขั้นตอน แต่สรุปตอนสุดท้ายผิดนิดหน่อย หรือคำนวณผิดเพียงจุดใดจุดหนึ่ง
- 2 คะแนน - ทำได้ 2 ใน 3 ส่วนของคำถามทั้งหมด พร้อมคำอธิบาย (ได้ขั้นที่ 1 - 2 แต่ทำขั้นที่ 3 ไม่ได้)
- 1 คะแนน - ทำได้เพียงส่วนใดส่วนหนึ่ง (เช่น คิดค่าใช้จ่ายทั้งหมดได้, หาเงาที่เสียไปได้, หาเงาที่เหลือได้ ฯลฯ)
- 0 คะแนน - ทำไม่ได้เลย หรือทำผิด

2. แบบทดสอบคณิตศาสตร์เหตุผล

ข้อที่ 1

- 4 คะแนน - ทำได้เหมือนในคำตอบตัวแบบ พร้อมคำอธิบาย และวาดภาพประกอบถูกต้องสมบูรณ์
- 3 คะแนน - ทำได้เหมือนในคำตอบตัวแบบ พร้อมคำอธิบาย แต่วาดภาพประกอบไม่สมบูรณ์
- 2 คะแนน - วาดภาพประกอบได้ถูกต้อง แต่ไม่มีคำอธิบายหรืออธิบายผิด
- 1 คะแนน - ทำได้เล็กน้อย (เช่น รู้ว่าต้องแบ่งกล่องออกเป็น 2 ส่วน เพราะรถแบ่งเป็น 2 ชนิดที่จำนวนล้อไม่เท่ากัน แต่ทำต่อไม่ได้หรือทำผิด)
- 0 คะแนน - ไม่เข้าใจโจทย์ คิดผิด วาดภาพประกอบผิด

ข้อที่ 2

- 4 คะแนน - บอกได้ว่าเป็นสามเหลี่ยมมุมฉาก หรือสามเหลี่ยมหน้าจั่วอย่างใดอย่างหนึ่ง หรือทั้งสองอย่าง หามุมที่เหลือได้หมด และหาส่วนอื่นๆ เพิ่มเติมได้เหมือนในคำตอบตัวแบบ พร้อมทั้งให้เหตุผลประกอบได้ถูกต้อง และวาดภาพประกอบได้ถูกต้อง
- 3 คะแนน - บอกได้ว่าเป็นสามเหลี่ยมมุมฉาก หรือสามเหลี่ยมหน้าจั่วอย่างใดอย่างหนึ่ง หรือทั้งสองอย่าง หามุมที่เหลือได้หมด พร้อมทั้งให้เหตุผลประกอบได้ถูกต้อง ภาพประกอบเกือบสมบูรณ์

- 2 คะแนน - บอกได้เพียง 2 ข้อ พร้อมทั้งให้เหตุผล วาดภาพประกอบไม่สมบูรณ์
 1 คะแนน - บอกได้เพียง 1 ข้อ พร้อมทั้งให้เหตุผล วาดภาพประกอบผิด หรือไม่ได้วาดภาพประกอบ
 0 คะแนน - ตอบไม่ได้เลย หรือไม่ให้เหตุผล

3. แบบทดสอบการสื่อสารทางคณิตศาสตร์

ข้อที่ 1

- 4 คะแนน - ทำได้เหมือนในคำตอบตัวแบบ มีตารางประกอบถูกต้อง
 3 คะแนน - ทำได้ทุกส่วน แต่มีข้อบกพร่องหรือผิดพลาดเล็กน้อย
 2 คะแนน - ทำได้ 2 ใน 3 ส่วน หรือทำได้ทั้งหมดแต่ไม่ยกตัวอย่างประกอบ หรือไม่มีตาราง
 1 คะแนน - ทำได้เพียงเล็กน้อย
 0 คะแนน - ทำไม่ได้เลย

ข้อที่ 2

- 4 คะแนน - อธิบายได้ถูกต้องชัดเจนเหมือนในคำตอบตัวแบบ และใช้มาตราส่วนได้ถูกต้อง
 3 คะแนน - อธิบายได้ถูกต้องแต่ไม่สมบูรณ์ทุกส่วน ใช้มาตราส่วนได้ถูกต้องหรือผิดพลาดเล็กน้อย
 2 คะแนน - อธิบายได้ถูกต้อง แต่ไม่ใช้มาตราส่วน
 1 คะแนน - อธิบายเพียงเล็กน้อย ไม่ชัดเจน
 0 คะแนน - อธิบายผิด หรือไม่อธิบาย

ภาคผนวก ข

การประมาณค่าความแปรปรวนขององค์ประกอบ 3 องค์ประกอบ
ที่มีผลต่อค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์

ตาราง 10 การประมาณค่าความแปรปรวนขององค์ประกอบ 3 องค์ประกอบ ที่มีผลต่อค่าสัมประสิทธิ์การ
สรุปร่างของแบบทดสอบวิชาคณิตศาสตร์ ที่มีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้
คะแนนของผู้ตรวจเอง

แหล่งของ ความแปรปรวน	df	$SS(\alpha)$	$MS(\alpha)$	$\sigma^2(\alpha)$	คิดเป็น ร้อยละ
p	59	739.97917	12.54202	0.28580	11.97700
i	5	57.73194	11.54639	0.01306	0.54730
r	5	228.89306	45.77861	0.12716	5.32889
pi	295	840.18472	2.84808	0.27549	11.54494
pr	295	664.69028	2.25319	0.37553	15.73731
ir	25	171.08750	6.84350	0.11406	4.77991
pir	1475	1762.82917	1.19514	1.19514	50.08465
รวม	2159	4465.39583		2.38624	100

ตาราง 11 การประมาณค่าความแปรปรวนขององค์ประกอบ 3 องค์ประกอบ ที่มีผลต่อค่าสัมประสิทธิ์การ
สรุปร่างของแบบทดสอบวิชาคณิตศาสตร์ ที่มีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้
คะแนนแบบบุริค

แหล่งของ ความแปรปรวน	df	$SS(\alpha)$	$MS(\alpha)$	$\sigma^2(\alpha)$	คิดเป็น ร้อยละ
p	59	963.27037	16.32662	0.39181	14.60403
i	5	112.72037	22.54407	0.04471	1.66649
r	5	262.52593	52.50519	0.14585	5.43630
pi	295	1061.11296	3.59699	0.39483	14.71659
pr	295	655.30741	2.22138	0.37023	13.79967
ir	25	161.22407	6.44896	0.10748	4.00613
pir	1475	1811.27593	1.22798	1.22798	45.77079
รวม	2159	5027.43704		2.68289	100

ตาราง 12 การประมาณค่าความแปรปรวนขององค์ประกอบ 3 องค์ประกอบ ที่มีผลต่อค่าสัมประสิทธิ์การ
สรุปร่างของแบบทดสอบวิชาคณิตศาสตร์ ที่มีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้
คะแนนของผู้ตรวจเอง และผู้ตรวจมีประสบการณ์มากกว่า 5 ปี

แหล่งของ ความแปรปรวน	df	$SS(\alpha)$	$MS(\alpha)$	$\sigma^2(\alpha)$	คิดเป็น ร้อยละ
p	59	513.13333	8.69718	0.40218	15.67556
i	5	22.48889	4.49778	0.00839	0.32701
r	2	114.30556	57.15278	0.15876	6.18791
pi	295	621.84444	2.10795	0.20220	7.88104
pr	118	172.02778	1.45786	0.24298	9.47050
ir	10	29.87222	2.98722	0.04979	1.94064
pir	590	885.79444	1.50135	1.50135	58.51734
รวม	1079	2359.46667		2.56565	100

ตาราง 13 การประมาณค่าความแปรปรวนขององค์ประกอบ 3 องค์ประกอบ ที่มีผลต่อค่าสัมประสิทธิ์การ
สรุปร่างของแบบทดสอบวิชาคณิตศาสตร์ ที่มีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้
คะแนนแบบรูปรีค และผู้ตรวจมีประสบการณ์มากกว่า 5 ปี

แหล่งของ ความแปรปรวน	df	$SS(\alpha)$	$MS(\alpha)$	$\sigma^2(\alpha)$	คิดเป็น ร้อยละ
p	59	623.36574	10.56552	0.49419	17.65841
i	5	53.96019	10.79204	0.04460	1.59365
r	2	124.14630	62.07315	0.17243	6.16127
pi	295	737.09537	2.49863	0.36782	13.14295
pr	118	197.07593	1.67013	0.27836	9.94637
ir	10	27.63148	2.76315	0.04605	1.64546
pir	590	823.14630	1.39516	1.39516	49.85189
รวม	1079	2586.42130		2.79861	100

ตาราง 14 การประมาณค่าความแปรปรวนขององค์ประกอบ 3 องค์ประกอบ ที่มีผลต่อค่าสัมประสิทธิ์การ
สรุปร่างของแบบทดสอบวิชาคณิตศาสตร์ ที่มีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้
คะแนนของผู้ตรวจเอง และผู้ตรวจมีประสบการณ์ไม่เกิน 5 ปี

แหล่งของ ความแปรปรวน	df	$SS(\alpha)$	$MS(\alpha)$	$\sigma^2(\alpha)$	คิดเป็น ร้อยละ
p	59	434.31389	7.36125	0.31491	13.90485
i	5	112.27500	22.45500	0.10537	4.65261
r	2	56.01667	28.00833	0.07780	3.43526
pi	295	509.00278	1.72543	0.14953	6.60249
pr	118	199.76111	1.69289	0.28215	12.45833
ir	10	34.88333	3.48833	0.05814	2.56717
pir	590	753.33889	1.27685	1.27685	56.37929
รวม	1079	2099.59167		2.26475	100

ตาราง 15 การประมาณค่าความแปรปรวนขององค์ประกอบ 3 องค์ประกอบ ที่มีผลต่อค่าสัมประสิทธิ์การ
สรุปร่างของแบบทดสอบวิชาคณิตศาสตร์ ที่มีการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้
คะแนนแบบรูบริค และผู้ตรวจมีประสบการณ์ไม่เกิน 5 ปี

แหล่งของ ความแปรปรวน	df	$SS(\alpha)$	$MS(\alpha)$	$\sigma^2(\alpha)$	คิดเป็น ร้อยละ
p	59	527.64722	8.94317	0.39268	14.96859
i	5	44.13056	8.82611	0.02575	0.98157
r	2	106.20556	53.10278	0.14751	5.62294
pi	295	658.48056	2.23214	0.27852	10.61692
pr	118	221.23889	1.87491	0.31248	11.91144
ir	10	41.90556	4.19056	0.06984	2.66223
pir	590	823.98333	1.39658	1.39658	53.23601
รวม	1079	2423.59167		2.62336	100

ภาคผนวก ค

รายชื่อผู้ตรวจแบบทดสอบวิชาคณิตศาสตร์

รายชื่อผู้ตรวจแบบทดสอบวิชาคณิตศาสตร์

- | | | |
|-------------------------------|-------------------|--------------------------------------|
| 1. อาจารย์ลัดดา ชลายนวัฒน์ | อาจารย์ 2 ระดับ 7 | โรงเรียนศิษฏ์วิทยา |
| 2. อาจารย์ลิขิต สุขเมธานุสรณ์ | อาจารย์ 1 ระดับ 5 | โรงเรียนวัดสำโรง |
| 3. อาจารย์จิรภา แก้วงาม | อาจารย์ 1 ระดับ 5 | โรงเรียนบ้านโป่งน้ำร้อน |
| 4. อาจารย์เบญจวรรณ ภักดีพงษ์ | อาจารย์ 1 ระดับ 4 | โรงเรียนนายายอามพิทยาคม |
| 5. อาจารย์ละมัย รัตนสร้อย | อาจารย์ 1 ระดับ 3 | โรงเรียนบ้านจันทเขลม |
| 6. อาจารย์สุนันทา จิตรสมาน | อาจารย์ 1 ระดับ 3 | โรงเรียนเบญจมาชชุทิศ จังหวัดจันทบุรี |

ประวัติย่อของผู้วิจัย

ชื่อสกุล	อุษณีย์ บัวศิริพันธ์
วัน เดือน ปีเกิด	1 ตุลาคม 2513
สถานที่เกิด	อำเภอพลูตาหลวง จังหวัดชลบุรี
สถานที่อยู่ปัจจุบัน	2/52 หมู่ 1 ตำบลนายายอาม อำเภอนายายอาม จังหวัดจันทบุรี 22160
สถานที่ทำงาน	โรงเรียนนายายอามพิทยาคม อำเภอนายายอาม จังหวัดจันทบุรี 22160
ประวัติการศึกษา	
พ.ศ. 2527	มัธยมศึกษาปีที่ 3 โรงเรียนเซนต์โยเซฟระยอง จังหวัดระยอง
พ.ศ. 2530	มัธยมศึกษาปีที่ 6 โรงเรียนบ้านนา "นายกพิทยากร" จังหวัดนครนายก
พ.ศ. 2534	คบ. (คณิตศาสตร์) สถาบันราชภัฏรำไพพรรณี จังหวัดจันทบุรี
พ.ศ. 2543	กศ.ม. (การวัดผลการศึกษา) มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร