ผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง ที่มีจำนวนผู้ตรวจ และวิธีการตรวจต่างกัน

ปริญญานิพนธ์

ของ

นิศารัตน์ คงสวัสดิ์

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา

ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา

ตุลาคม 2544

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

ผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง ที่มีจำนวนผู้ตรวจ และวิธีการตรวจต่างกัน

บทคัดย่อ

ของ

นิศารัตน์ คงสวัสดิ์

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา

ตามหลักสูตรปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา

ตุลาคม 2544

นิศารัตน์ คงสวัสดิ์. (2544), *ผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง ที่มีจำนวนผู้ตรวจ และวิธีการตรวจต่างกัน.* ปริญญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. คณะกรรมการควบคุม: รองศาสตราจารย์ ดร.บุญเชิด ภิญโญอนันตพงษ์ , อาจารย์ชวลิต รวยอาจิณ.

การวิจัยครั้งนี้ มีจุดมุ่งหมายสำคัญเพื่อศึกษาและเปรียบเทียบผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียงที่มีผู้ตรวจจำนวนต่างกัน คือ สองคน สามคน และ สี่คน เมื่อใช้การตรวจวิธีเดียวกัน และ ใช้การตรวจต่างวิธีกัน คือ การตรวจวิธีวิเคราะห์ย่อย (Analytic Method) และ การตรวจวิธีประเมินรวม (Holistic Method) เมื่อมีผู้ตรวจจำนวนเท่ากัน กลุ่มตัวอย่างในการวิจัย เป็นนักเรียนชั้นประถมศึกษาปีที่ 5 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาเอกชน จำนวน 60 คน และ ผู้ตรวจแบบทดสอบความเรียงเป็นครูผู้สอนวิชาสร้างเสริมประสบการณ์ชีวิต ชั้นประถมศึกษาปีที่ 5 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาเอกชน จำนวนสี่คน ที่มีความสามารถตามเกณฑ์ที่กำหนด ตรวจแบบทดสอบความเรียงจำนวนห้าข้อ ด้วยการตรวจวิธีวิเคราะห์ย่อย และวิธีประเมินรวม ทั้งสี่คน หาค่าความเชื่อมั่นของผู้ตรวจโดยวิเคราะห์หาค่าดัชนีความสอดคล้องระหว่างผู้ตรวจ (Rater Agreement Indexes : RAI) ได้ผลการวิจัยดังนี้

1. การเปรียบเทียบความแตกต่างของค่าเฉลี่ยความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง ที่มีผู้ตรวจจำนวนต่างกัน คือ สองคน สามคน และ สี่คน โดยใช้การตรวจวิธีเดียวกัน พบว่า มีความแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ทั้งการตรวจวิธีวิเคราะห์ย่อย กับการตรวจวิธีประเมินรวม

2. การเปรียบเทียบความแตกต่างของค่าเฉลี่ยความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง ที่มีผู้ตรวจจำนวนเท่ากัน เมื่อใช้การตรวจต่างวิธีกัน คือ การตรวจวิธีวิเคราะห์ย่อย และ การตรวจวิธีประเมินรวม พบว่า มีความแตกต่างกันอย่างไม่มีนัยสำคัญ ทั้งสามกรณี คือ เมื่อมีผู้ตรวจจำนวนสองคน สามคน และ สี่คน

RESULTS OF ESSAY TEST RATER RELIABILITY BASED ON DIFFERENT NUMBER OF RATERS AND SCORING METHODS

AN ABSTRACT

BY

NISARATH KONGSAWAT

Presented in partial fulfillment of the requirements

for the Master of Education degree in Measurement Education

at Srinakharinwirot University

October 2001

Nisarath Kongsawat. (2001) *The Results of Essay Test Raters Reliability Based on Different Numbers of Raters and Scoring Methods.* Master Thesis, M.Ed. (Measurement Education). Bangkok : Graduate School, Srinakharinwirot University. Advisor Committee : Assoc. Prof. Dr. Boonchird Pinyoanuntapong and Mr. Chavalit Royarjin.

This research aimed to study and compare the results of essay test raters reliability based on two , three and four raters when they used the same scoring methods and compared the results of essay test raters reliability when an equal number of raters they used different scoring methods based on Analytic and Holistic scoring methods. The sample were sixty students and four essay test raters who taught the subject of Life Experience in Prathomsuksa V from three private schools in Bangkok. The raters checked five essays by using two scoring methods : the Analytic Method and the Holistic Method. The researcher of this study analyzed the research data by using the Rater Agreement Indexes (RAI).

The results of the research are as follows :

1. When comparing the different grand means of essay test raters reliability of two, three and four raters by using the same scoring methods, it was found that there was no significant difference in both Analytic and Holistic scoring methods.

2. When comparing the grand means of essay test raters reliability of the equal numbers of raters and by using different scoring methods including both Analytic Method and Holistic Method, it was found that there was no significant difference in three cases when we used two , three or four raters.

ปริญญานิพนธ์

เรื่อง

ผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง ที่มีจำนวนผู้ตรวจ และวิธีการตรวจต่างกัน

ของ

นางนิศารัตน์ คงสวัสดิ์

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาการศึกษามหาบัณฑิต วิชาเอกการวัดผลการศึกษา

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

.....คณบดีบัณฑิตวิทยาลัย

(ศาสตราจารย์ ดร. เสริมศักดิ์ วิศาลาภรณ์)

วันที่ ..... เดือน ..... พ.ศ. 2544

คณะกรรมการสอบปริญญานิพนธ์

.....ประธาน

(รองศาสตราจารย์ ดร. บุญเชิด ภิญโญอนันพงษ์)

.....กรรมการ

(อาจารย์ ชวลิต รวยอาจิณ)

.....กรรมการที่แต่งตั้งเพิ่มเติม

(รองศาสตราจารย์ ดร. ส.วาสนา ประวาลพฤกษ์)

.....กรรมการที่แต่งตั้งเพิ่มเติม

(รองศาสตราจารย์ นิภา ศรีไพโรจน์)

ประกาศคุณูปการ

ปริญญานิพนธ์ฉบับนี้สำเร็จสมบูรณ์ได้ เพราะได้รับความกรุณาจาก รองศาสตราจารย์ ดร. บุญเชิด ภิญโญอนันตพงษ์ ประธานคณะกรรมการควบคุม และ อาจารย์ชวลิต รวยอาจิณ กรรมการ ให้คำปรึกษา แนะนำ ชี้แนะ และ ตรวจสอบ งานวิจัยนี้จนสามารถบรรลุจุดมุ่งหมาย ข้าพเจ้าขอกราบขอบพระคุณท่านเป็นอย่างสูง

ขอกราบขอบพระคุณ รองศาสตราจารย์ ดร. ส.วาสนา ประวาลพฤกษ์ และ รองศาสตราจารย์นิภา ศรีไพโรจน์ กรรมการที่แต่งตั้งเพิ่มเติม ที่ได้ให้คำแนะนำ ชี้แนะ และปรับปรุงรูปเล่ม เนื้อหาสาระที่ถูกต้องชัดเจนยิ่งขึ้น เพื่อให้งานวิจัยนี้มีคุณภาพที่ดี ข้าพเจ้าขอกราบขอบพระคุณเป็นอย่างสูง

ขอกราบขอบพระคุณ คณะอาจารย์ ของภาควิชาวัดผลและวิจัยการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ ทุกท่านที่ได้ช่วยดูแล ให้คำปรึกษาแนะนำแนวทางในการทำวิจัยตลอดมา ผู้วิจัยขอกราบขอบพระคุณเป็นอย่างสูง

ขอกราบขอบพระคุณ รองศาสตราจารย์วัญญา วิศาลาภรณ์ ภาควิชาวัดผลและวิจัยการศึกษา อาจารย์ยุพาวดี รัชดาพรรณาธิกุล อาจารย์ปราณี วงศ์สมบุญ อาจารย์เพชรรัตน์ ชมจันทร์ และ อาจารย์สุกัญญา วิเชียรดิลก ที่ให้ความอนุเคราะห์ช่วยเหลือในการตรวจสอบเครื่องมือการวิจัย

ขอกราบขอบพระคุณ คณะภราดา ผู้บริหารโรงเรียนอัสสัมชัญแผนกประถม ในเครือมูลนิธิเซนต์คาเบรียลแห่งประเทศไทย ที่มีความเมตตาอนุมัติทุนการศึกษาตลอดหลักสูตรแก่ข้าพเจ้า รวมทั้งเพื่อน ครู และ นักเรียนโรงเรียนต่าง ๆ ที่ให้ความร่วมมือในการเก็บรวบรวมข้อมูลการวิจัย

ขอกราบขอบพระคุณ อาจารย์ปราณี วงศ์สมบุญ และ อาจารย์สุชิน บุญพิทักษ์กิจ อาจารย์ผู้สอนวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ชั้น ป. 5 โรงเรียนเปรมฤดีศึกษา รวมทั้ง อาจารย์เพชรรัตน์ ชมจันทร์ และ อาจารย์สุกัญญา วิเชียรดิลก อาจารย์ผู้สอนวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ชั้น ป. 5 โรงเรียนอัสสัมชัญแผนกประถม ที่ให้ความช่วยเหลือเป็นผู้ตรวจให้คะแนนแบบทดสอบความเรียง เพื่อนำผลมาวิเคราะห์ข้อมูลการวิจัย

ขอขอบคุณสมาชิกในครอบครัว และเพื่อน ๆ ผู้ร่วมงานทุกท่านที่ให้กำลังใจ และช่วยเหลืออย่างดีตลอดระยะเวลาที่ศึกษาและทำการวิจัย และขอบคุณเพื่อน ๆ นิสิตระดับปริญญาโท สาขาวิชาการวัดผลการศึกษา ภาคพิเศษ รุ่นที่ 11 ทุกคนที่ให้กำลังใจ ร่วมมือค้นคว้าข้อมูลด้วยกันตลอดมา

คุณค่าและประโยชน์ใด ๆ ที่ได้รับจากปริญญานิพนธ์ฉบับนี้ ผู้วิจัยขอมอบเป็นเครื่องบูชาพระคุณบิดา มารดา ครูอาจารย์ และผู้มีพระคุณทุกท่าน

นิศารัตน์ คงสวัสดิ์

สารบัญ

บทที่ หน้า

1 บทนำ 1

ภูมิหลัง 1

ความมุ่งหมายของการวิจัย 2

ความสำคัญของการวิจัย 3

ขอบเขตของการวิจัย 3

ประชากรที่ใช้ในการวิจัย 3

กลุ่มตัวอย่างที่ใช้ในการวิจัย 3

ตัวแปรที่ใช้ในการวิจัย 3

นิยามศัพท์เฉพาะ 4

2 เอกสารและงานวิจัยที่เกี่ยวข้อง 6

เอกสารที่เกี่ยวข้องกับการวิจัย 6

แบบทดสอบแบบอัตนัย หรือแบบทดสอบความเรียง 6

นิยามของแบบทดสอบความเรียง 6

ลักษณะและประเภทของแบบทดสอบความเรียง 8

การสร้างแบบทดสอบความเรียง 20

วิธีการตรวจให้คะแนนแบบทดสอบความเรียง 26

องค์ประกอบที่ส่งผลต่อคะแนนแบบทดสอบความเรียง 36

ความเชื่อมั่นของแบบทดสอบความเรียง 38

ความเชื่อมั่นของผู้ตรวจให้คะแนน 39

วิธีการหาความสอดคล้องระหว่างผู้ตรวจให้คะแนน แบบ RAI 43

งานวิจัยที่เกี่ยวข้อง

งานวิจัยในต่างประเทศ 44

งานวิจัยภายในประเทศ 47

สมมติฐานในการวิจัย 49

3 วิธีดำเนินการวิจัย 50

การกำหนดประชากรและเลือกกลุ่มตัวอย่าง 50

เครื่องมือที่ใช้ในการวิจัย 51

วิธีดำเนินการเก็บรวบรวมข้อมูล 57

การจัดกระทำและการวิเคราะห์ข้อมูล 58

สารบัญ (ต่อ)

บทที่ หน้า

4 ผลการวิเคราะห์ข้อมูล 62

5 สรุปผล อภิปราย และข้อเสนอแนะ 68

ความมุ่งหมายของการวิจัย 68
สมมุติฐานของการวิจัย 68
ประชากรที่ใช้ในการวิจัย 68
กลุ่มตัวอย่างที่ใช้ในการวิจัย 69
เครื่องมือที่ใช้ในการวิจัย 69
วิธีดำเนินการเก็บรวบรวมข้อมูล 69
สรุปผลการวิเคราะห์ข้อมูล 70
อภิปรายผล 71
ข้อเสนอแนะ 73

บรรณานุกรม 74

ภาคผนวก 80

ประวัติย่อผู้วิจัย 113

บัญชีตาราง

ตาราง หน้า

1 การเปรียบเทียบระหว่างลักษณะของข้อสอบปรนัยและข้อสอบแบบความเรียง 9
2 การเปรียบเทียบลักษณะของข้อสอบแบบต่าง ๆ 10
3 ผลรวมของการตรวจให้คะแนนแบบ 2 ขั้ว จากผู้ตรวจ 2 คน 40
4 ผลรวมของการให้คะแนนแบบหลายขั้ว โดยผู้ตรวจ 2 คน..... 41
5 จำนวนนักเรียนที่ใช้เป็นกลุ่มตัวอย่างในการเก็บรวบรวมข้อมูล 51
6 การวิเคราะห์หลักสูตรตามรายละเอียดของเนื้อหา..... 53
7 ค่าสถิติพื้นฐานของแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต(สังคม) จำแนกตามวิธีการตรวจให้คะแนน และจำนวนผู้ตรวจ..... 63
8 ค่าเฉลี่ยความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง จำแนกตามวิธีการตรวจให้คะแนน และจำนวนผู้ตรวจ..... 65
9 ผลการเปรียบเทียบค่าเฉลี่ยความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง ที่มีจำนวนผู้ตรวจสองคน, สามคน และ สี่คน เมื่อใช้การตรวจวิธีวิเคราะห์ย่อย กับ ตรวจวิธีประเมินรวม..... 66
10 แสดงการเปรียบเทียบค่าเฉลี่ยความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง เมื่อใช้การตรวจวิธีวิเคราะห์ย่อย กับ ตรวจวิธีประเมินรวม ที่มีผู้ตรวจจำนวนเท่ากัน..... 67
11 ดัชนีความสอดคล้องของผู้เชี่ยวชาญตรวจสอบคุณภาพเครื่องมือ..... 83
12 ค่าความยาก ค่าอำนาจจำแนก ของแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต (สังคม)..... 84
13 เปรียบเทียบสถิติพื้นฐาน จากการตรวจวิธีวิเคราะห์ย่อย และวิธีประเมินรวม จำแนกเป็นรายบุคคล 103
14 ผลการตรวจแบบทดสอบความเรียงวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) เมื่อใช้การตรวจวิธีวิเคราะห์ย่อย และค่าสถิติพื้นฐาน 104
15 ผลการตรวจแบบทดสอบความเรียงวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) เมื่อใช้การตรวจวิธีประเมินรวม และค่าสถิติพื้นฐาน 106
16 ค่าความเชื่อมั่นของผู้ตรวจ (RAI) แบบทดสอบความเรียง เมื่อใช้การตรวจวิธีวิเคราะห์ย่อย 108
17 ค่าความเชื่อมั่นของผู้ตรวจ (RAI) แบบทดสอบความเรียง เมื่อใช้การตรวจวิธีประเมินรวม 110

บัญชีภาพประกอบ

ภาพประกอบ หน้า

ลำดับขั้นการสร้างแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต (สังคม)..... 52

บทที่ 1

บทนำ

ภูมิหลัง

พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ. 2542 หมวด 4 การจัดการศึกษา มาตรา 26 ระบุไว้ว่า ให้สถานศึกษาจัดการประเมินผู้เรียน โดยพิจารณาจากพัฒนาการของผู้เรียน ความประพฤติ การสังเกตพฤติกรรมการเรียน การร่วมกิจกรรม และ การทดสอบควบคู่ไปในกระบวนการเรียนการสอน ให้มีวิธีการประเมินที่หลากหลาย (พ.ร.บ.การศึกษาแห่งชาติ พ.ศ. 2542. 2542 : 12-14) ให้มีความเหมาะสมกับวัย และลักษณะความสามารถของเด็ก ดังนั้นในการประเมินนักเรียนจึงจำเป็นต้องใช้เครื่องมือชนิดต่าง ๆ ทั้งการวัดแบบไม่เป็นทางการ เช่น การสังเกต การบันทึก การสนทนาบันทึก และ การวัดแบบทางการ ได้แก่ แบบทดสอบชนิดต่างๆ คือ แบบทดสอบถูกผิด จับคู่ เติมคำ เลือกตอบ และ แบบทดสอบอัตนัย หรือ แบบทดสอบความเรียง (บุญเชิด ภิญโญอนันตพงษ์. 2544) สำหรับแบบทดสอบความเรียงเหมาะสำหรับใช้วัดความสามารถด้านความเข้าใจ ทักษะในการคิด ผลการเรียนที่ซับซ้อน ความสามารถในการเลือกสรร การประมวลความคิด ความสามารถในการเขียน และทักษะในการแก้ปัญหาได้เป็นอย่างดี และยังเปิดโอกาสให้นักเรียนแสดงความสามารถในการเปรียบเทียบ แสดงความคิด และเขียนอธิบายอย่างอิสระ (Hopkins and Antes. 1990 : 151) และแบบทดสอบความเรียงก็มีจุดแข็งที่สามารถสร้างได้ในระยะสั้น จึงทำให้ครูสามารถใช้ในระหว่างการเรียนการสอนได้เป็นอย่างดี ดีกว่าที่จะเสียเวลาในการสร้างข้อสอบปรนัยแบบเลือกตอบ (Hopkins and Antes. 1990 : 222)

ข้อสอบอัตนัยหรือแบบทดสอบความเรียงเป็นข้อสอบที่มีคุณค่ามากชนิดหนึ่ง สามารถวัดพฤติกรรมด้านความรู้ความคิดระดับสูงได้ดี (บุญเชิด ภิญโญอนันตพงษ์. 2525 : 46) และมีความเป็นอัตนัย คือ ใช้การตัดสินตามความรู้สึกส่วนตน ซึ่งการตัดสินของแต่ละคนอาจไม่ตรงกัน หรืออาจจะตรงกันข้ามก็ได้ ทั้งนี้เพราะบุคคลย่อมแตกต่างกันในด้านต่าง ๆ การนำไปใช้ต้องพยายามวัดพฤติกรรมที่เหมาะสมแล้วยังต้องทำให้การตรวจให้คะแนนมีความเชื่อถือได้มากที่สุด ปัญหาใหญ่ของคะแนนแบบทดสอบความเรียงที่จะให้ความเชื่อมั่นสูงขึ้นอยู่กับผู้ตรวจ ดังนั้น หัวใจสำคัญของการให้คะแนนจึงต้องมีพัฒนาการให้คะแนนด้วยวิธีที่จะให้คะแนนได้อย่างมีความเชื่อถือได้จากการวัดนั้น ๆ ปัญหาในการให้คะแนนไม่สอดคล้องกัน มีสาเหตุสำคัญมาจาก 3 กรณี คือ 1) กรณีที่มีผู้ตรวจต่างกัน ตรวจให้คะแนนข้อเดียวกัน 2) กรณีที่ผู้ตรวจคนเดียวกัน ให้คะแนนข้อเดียวกันในเวลาต่างกัน และ 3) การตรวจคำตอบที่มีอิสระในการตอบ (Hopkins and Antes. 1990 : 236 ; citing Coffman. 1971 : 276-277) การลดความคลาดเคลื่อนดังกล่าวจำเป็นต้องมีวิธีการตรวจที่รัดกุมซึ่งมีวิธีการตรวจให้คะแนน 2 วิธีที่เหมาะสมกับการตรวจแบบทดสอบความเรียงที่สุด คือ การตรวจด้วยวิธีวิเคราะห์ย่อย (Analytic Method) หรือ วิธีที่ให้ตามประเด็น (Point Method) การตรวจแบบนี้ขึ้นอยู่กับการเตรียมรายละเอียดคำแนะนำในการตรวจด้วยประเด็นย่อย ๆ อย่างละเอียด และวิธีประเมินภาพรวม (Global Method) หรือเรียกว่า วิธีประเมินรวม (Holistic Method) ที่จะให้คะแนนในภาพรวมของคำตอบซึ่งจะขึ้นอยู่กับมาตรฐานของการตัดสินใจของผู้ตรวจเอง (Hopkins and Antes. 1990 : 237) และ วิธีการตรวจก็เป็นปัจจัยหนึ่งที่มีอิทธิพลต่อความเชื่อมั่นของผู้ตรวจ (Rater Reliability) ที่จะมีความเห็นสอดคล้องกันได้สูง นักวัดผลการศึกษา

และนักวิจัยเสนอว่า การชี้แจงให้ผู้ตรวจมีความรู้ความเข้าใจในเรื่องที่จะตรวจ รู้จักใช้แบบแผนหรือคู่มือการตรวจที่เหมาะสม สามารถทำให้การตรวจมีความเชื่อมั่นสูงได้ โดยทั่วไปการตรวจโดยวิธีวิเคราะห์ย่อย (Analytic Method) จะให้ค่าความเชื่อมั่นสูงกว่าวิธีประเมินรวม (Holistic Method) นอกจากนี้การตรวจทีละข้อจะดีกว่าการตรวจทีละคน (วัญญา วิศาลาภรณ์. 2533 : 79) เพราะเป็นการลดความแปรปรวนในการตรวจจากการเปลี่ยนข้อบ่อย ๆ นอกจากนี้สิ่งสำคัญของการตรวจทั้งสองวิธี คือ คำตอบเหล่านั้นต้องไม่คำนึงว่าเป็นของใคร เพื่อหลีกเลี่ยงความลำเอียงที่เกี่ยวกับความประทับใจของครูที่มีต่อนักเรียน และการตรวจทีละมาก ๆ ก็อาจทำให้เกิดความเหน็ดเหนื่อยขึ้นได้ สิ่งนี้เป็นเหตุให้การให้คะแนนไม่แน่นอนขึ้นได้ ดังนั้นการให้คะแนนจะเป็นมาตรฐานได้นั้น ต้องยึดถึงมาตรฐานคะแนนและเกณฑ์การตรวจเดียวกันทุกแผ่นคำตอบเสมอ (บุญเชิด ภิญโญอนันตพงษ์. 2544)

ดังนั้นในการสร้างแบบทดสอบความเรียง จึงต้องพัฒนาวิธีการตรวจให้คะแนนควบคู่กันไปด้วย และต้องตรวจสอบความเชื่อมั่นของการตรวจให้คะแนน สำหรับวิธีการคำนวณความเชื่อมั่นของผู้ตรวจมีหลายวิธีด้วยกัน แต่ละวิธีมีจุดเด่นจุดด้อยในตัวทั้งสิ้น การพิจารณาเลือกใช้วิธีใดจึงขึ้นอยู่กับความง่าย และใช้ได้สะดวก เบอร์รี่-สต๊อก และคนอื่น ๆ (Burry-Stock and others.1996) ได้นำเสนอสูตรวิธีการหาความสอดคล้องของผู้ตรวจแนวใหม่ที่คำนวณได้ทันทีเมื่อมีการวัดผลเสร็จสิ้น และดัชนีนี้ยังง่ายต่อการแปลผล คือ หากค่าที่ได้เท่ากับ 1.00 แสดงถึงผู้ตรวจมีความคิดเห็นสอดคล้องกันสูงที่สุด และ หากคำนวณได้ค่าใกล้เคียงหรือเท่ากับ 0.00 แสดงถึงผู้ตรวจมีความคิดเห็นแตกต่างกันมากที่สุดหรือแตกต่างกันอย่างสิ้นเชิง และ ยังคำนวณกับพฤติกรรมการสังเกตที่มีหลาย ๆ พฤติกรรม มีหลายสูตรให้เลือกใช้กับสถานการณ์ต่างๆ ดังนั้นผู้วิจัยจึงสนใจนำสูตรนี้มาศึกษาผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจ ที่มีจำนวนผู้ตรวจต่างกัน โดยใช้วิธีการตรวจต่างกัน เพื่อเป็นแนวทางในการหาคุณภาพของเครื่องมือที่ใช้ในการวัดผลต่อไป โดยตั้งเป็นคำถามของการวิจัยว่า 1) ควรใช้ผู้ตรวจกี่คน จึงจะให้ค่าความเชื่อมั่นที่เหมาะสม และ 2) ควรใช้วิธีการตรวจวิธีใด จึงจะให้ค่าความเชื่อมั่นสูงกว่า

ความมุ่งหมายของการวิจัย

การวิจัยครั้งนี้ มีจุดมุ่งหมายสำคัญ เพื่อศึกษาและเปรียบเทียบค่าความเชื่อมั่นของผู้ตรวจให้คะแนนแบบทดสอบความเรียง ที่มีจำนวนผู้ตรวจต่างกัน และตรวจให้คะแนนวิธีวิเคราะห์ย่อย และวิธีประเมินรวม โดยมีจุดมุ่งหมายเฉพาะของการวิจัย ดังนี้

1. เพื่อเปรียบเทียบผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียงที่มีจำนวนผู้ตรวจสองคน สามคน และ สี่คน โดยแยกเป็น
   1.1 เมื่อตรวจให้คะแนนวิธีวิเคราะห์ย่อย
   1.2 เมื่อตรวจให้คะแนนวิธีประเมินรวม
2. เพื่อเปรียบเทียบผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง ที่ใช้การตรวจให้คะแนนวิธีวิเคราะห์ย่อย และให้คะแนนวิธีประเมินรวม โดยแยกเป็น

2.1 เมื่อใช้ผู้ตรวจสองคน

2.2 เมื่อใช้ผู้ตรวจสามคน

2.3 เมื่อใช้ผู้ตรวจสี่คน

ความสำคัญของการวิจัย

ผลการวิจัยนี้มีประโยชน์ต่อวงการศึกษาในด้านการจัดการเรียนการสอนและการวัดผลระหว่างเรียนอย่างแท้จริง ครูผู้สอนสามารถนำวิธีการทดสอบนักเรียนด้วยแบบทดสอบความเรียงไปใช้อย่างแพร่หลายมากขึ้น นำวิธีการตรวจที่มีรูปแบบชัดเจน และมีความเชื่อถือได้ไปใช้ในการจัดกิจกรรมการเรียนและการประเมินไปพร้อม ๆ กัน รวมทั้งได้นำวิธีการตรวจสอบความสอดคล้องของผู้ตรวจ หรือผู้ประเมินด้วยวิธีง่าย ๆ ไปใช้โดยครูไม่ต้องมีความรู้ในเชิงสถิติขั้นสูง ฉะนั้นจึงสะดวกต่อการใช้งานได้ทั่วไปเป็นอย่างดี

ขอบเขตของการวิจัย

ประชากรที่ใช้ในการวิจัย

ประชากรที่ใช้ในการวิจัย เป็นนักเรียนชั้นประถมศึกษาปีที่ 5 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาเอกชน สาขากลุ่มกรุงเทพฯ ตะวันออก เขตสาทร เขตบางรัก และ เขตบางคอแหลม กรุงเทพมหานคร ในปีการศึกษา 2543 จำนวน 3,325 คน

กลุ่มตัวอย่างที่ใช้ในการวิจัย

กลุ่มตัวอย่างที่ใช้ในการวิจัย เป็นนักเรียนชั้นประถมศึกษาปีที่ 5 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาเอกชน สาขากลุ่มกรุงเทพฯ ตะวันออก เขตสาทร เขตบางรัก และเขตบางคอแหลม กรุงเทพมหานคร ในปีการศึกษา 2543 จำนวน 60 คน ที่ได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi – Stage Random Sampling)

ตัวแปรที่ใช้ในการวิจัย

1. ตัวแปรอิสระ ได้แก่

   1.1 จำนวนผู้ตรวจ แบ่งเป็น 3 กลุ่ม ได้แก่

      1.1.2 ผู้ตรวจจำนวนสองคน

      1.1.3 ผู้ตรวจจำนวนสามคน

      1.1.4 ผู้ตรวจจำนวนสี่คน

   1.2 วิธีการตรวจ แบ่งเป็น 2 วิธี ได้แก่

      1.2.1 วิธีวิเคราะห์ย่อย (Analytic Method)

      1.2.2 วิธีประเมินรวม (Holistic Method)

2. ตัวแปรตาม ได้แก่ ค่าความเชื่อมั่นของผู้ตรวจ

นิยามศัพท์เฉพาะ

1. แบบทดสอบความเรียง หมายถึง ชุดข้อคำถามที่เกี่ยวกับการวัดความสามารถทางสมอง ในกลุ่มวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ชั้นประถมศึกษาปีที่ 5 ตามหลักสูตรประถมศึกษา พุทธศักราช 2521 (ฉบับปรับปรุง พ.ศ. 2533) หน่วยที่ 5 หัตถกรรมและอุตสาหกรรม หน่วยที่ 8 ประเทศเพื่อนบ้าน และหน่วยที่ 12 ข่าวและเหตุการณ์ จำนวน 5 ข้อ ที่ผู้วิจัยสร้างขึ้น เพื่อวัดระดับการเรียนรู้ขั้นสูงคือ ความเข้าใจ การนำไปใช้ การวิเคราะห์ การสังเคราะห์ และการประเมินค่า ให้ผู้ตอบมีอิสระในการเขียนคำตอบแสดงหลักวิชา ความรู้ ข้อเท็จจริงต่าง ๆ ที่สะสมความรู้จากการเรียนมาใช้ในการตอบ

2. ผู้ตรวจ หมายถึง ผู้ตรวจให้คะแนนแบบทดสอบซึ่งเป็นครูผู้สอนวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ชั้นประถมศึกษาปีที่ 5 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาเอกชน สาขากลุ่มกรุงเทพฯ ตะวันออก เขตสาทร เขตบางรัก และเขตบางคอแหลม กรุงเทพมหานคร ในปีการศึกษา 2543 จำนวน 4 คน ซึ่งสำเร็จการศึกษาไม่ต่ำกว่าระดับปริญญาตรี เอกสังคมศึกษา และมีประสบการณ์การสอนวิชาสังคมศึกษาตั้งแต่ 5 ปีขึ้นไป

3. จำนวนผู้ตรวจ หมายถึง ผู้ตรวจแบบทดสอบความเรียงวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) จำนวนสี่คน แบ่งออกเป็น 3 กลุ่ม คือ

3.1 ผู้ตรวจสองคน มี 6 กรณี ดังนี้

3.1.1 ผู้ตรวจคนที่ 1 กับ 2

3.1.2 ผู้ตรวจคนที่ 1 กับ 3

3.1.3 ผู้ตรวจคนที่ 1 กับ 4

3.1.4 ผู้ตรวจคนที่ 2 กับ 3

3.1.5 ผู้ตรวจคนที่ 2 กับ 4

3.1.6 ผู้ตรวจคนที่ 3 กับ 4

3.2 ผู้ตรวจสามคน มี 4 กรณี ดังนี้

3.2.1 ผู้ตรวจคนที่ 1 , 2 กับ 3

3.2.2 ผู้ตรวจคนที่ 1 , 2 กับ 4

3.2.3 ผู้ตรวจคนที่ 1 , 3 กับ 4

3.2.3 ผู้ตรวจคนที่ 2 , 3 กับ 4

3.3 ผู้ตรวจสี่คน ได้แก่ ผู้ตรวจคนที่ 1 , 2 , 3 กับ 4

4. การตรวจโดยวิธีวิเคราะห์ย่อย (Analytic Method) หมายถึง การตรวจคำตอบ ตามเกณฑ์ที่เฉลยคำตอบไว้ว่าตอบอย่างไรได้กี่คะแนน โดยแบ่งออกเป็น 3 เกณฑ์ คือ ความถูกต้องของเนื้อหา การนำเสนอหลักการที่ถูกต้อง และการจัดเรียบเรียงความคิด (Ward. 1999 : 133) และมีการกำหนดคะแนนเป็นส่วนย่อย ๆ ในแต่ละประเด็นหลักเหล่านั้น

5. การตรวจให้คะแนนโดยวิธีประเมินรวม (Holistic Method) หมายถึง การตรวจให้คะแนนคำตอบแบบทดสอบความเรียง โดยพิจารณาคุณภาพรวมตามความประทับใจของผู้ตรวจเมื่อเทียบกับคำตอบ

ตัวแบบ โดยแยกกระดาษคำตอบออกเป็นกอง ๆ ตามระดับคุณภาพของคำตอบ สำหรับงานวิจัยนี้ การตรวจโดยวิธีประเมินรวมจะตรวจโดยใช้เกณฑ์การให้คะแนนตามวิธีของวอร์ด (Ward. 1999 : 133) ที่กำหนดเกณฑ์การให้คะแนนแบบทดสอบความเรียงไว้ 6 ระดับ

6. ความเชื่อมั่นของผู้ตรวจ หมายถึง ระดับความสอดคล้องของการให้คะแนนคำตอบของผู้ตรวจด้วยตัวเลขที่ใกล้เคียงกัน ในแต่ละวิธีการตรวจของผู้ตรวจแต่ละคน ซึ่งตรวจสอบโดยใช้ดัชนีความความสอดคล้องกัน หรือ RAI (Rater Agreement Indexes) ซึ่งมีค่าแปรเปลี่ยนตั้งแต่ 0.0 ถึง 1.0 การวิจัยครั้งนี้ต้องการค่าความเชื่อมั่นสูงตั้งแต่ 0.5 ขึ้นไป

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

เอกสารและงานวิจัยที่เกี่ยวข้องกับการวิจัยครั้งนี้ จำแนกตามหัวข้อสำคัญดังต่อไปนี้

ก. เอกสารที่เกี่ยวข้องกับการวิจัย

1. แบบทดสอบแบบอัตนัย หรือ แบบทดสอบความเรียง (Essay Type Test)
   - 1.1 นิยามของแบบทดสอบความเรียง
   - 1.2 ลักษณะและประเภทของแบบทดสอบความเรียง
   - 1.3 ขั้นตอนและแนวทางการสร้างแบบทดสอบความเรียง
   - 1.4 วิธีการตรวจให้คะแนนแบบทดสอบความเรียง
   - 1.5 องค์ประกอบที่ส่งผลต่อคะแนนแบบทดสอบความเรียง
   - 1.6 การหาความเชื่อมั่นของแบบทดสอบความเรียง
2. ความเชื่อมั่นของผู้ตรวจให้คะแนน (Rater Reliability)
3. วิธีการหาความสอดคล้องระหว่างผู้ตรวจให้คะแนนแบบ RAI

ข. งานวิจัยที่เกี่ยวข้อง

1. งานวิจัยในต่างประเทศ
2. งานวิจัยภายในประเทศ

ก. เอกสารที่เกี่ยวข้องกับการวิจัย

1. แบบทดสอบแบบอัตนัย หรือ แบบทดสอบความเรียง (Essay Type Test)

1.1 นิยามของแบบทดสอบความเรียง

คอฟแมน (Coffman.1971:271) ได้นิยามการทดสอบแบบความเรียง (Essay Examination) ว่า หมายถึง การสอบที่ใช้ข้อสอบแบบความเรียง ตั้งแต่หนึ่งข้อขึ้นไปกับกลุ่มนักเรียนภายใต้สถานการณ์อย่างหนึ่ง เพื่อเก็บข้อมูลในการประเมินผล คำถามแบบความเรียงแตกต่างกับคำถามแบบตอบสั้น ๆ ตรงที่ต้องใช้ผู้เชี่ยวชาญตัดสิน ไม่ใช่การตรวจอย่างง่าย โดยใช้ตัวเฉลยแบบเดียวกับแบบทดสอบปรนัย แตกต่างจากรายงานที่ทำส่งครูในการเรียนวิชาต่าง ๆ ตรงที่การดำเนินการสอบจะต้องดำเนินให้เป็นมาตรฐานเดียวกัน และต้องการข้อสอบที่เป็นตัวแทนผลการสอบที่ต้องการ

กลอนลันด์ (Gronlund. 1977 : 75) สรุปความหมายของแบบทดสอบความเรียงไว้ว่า แม้ว่าจะกล่าวว่าแบบทดสอบแบบความเรียงจะไม่มีประสิทธิภาพในการวัดผลการเรียนรู้ แต่ว่าข้อสอบแบบนี้จะให้อิสระในการตอบซึ่งเป็นสิ่งที่จำเป็นในการวัดผลที่ซับซ้อน รวมไปถึงการวัดความสามารถในการสร้าง การจัดระเบียบ

การรวมเป็นหนึ่งเดียว การแสดงออกและพฤติกรรมที่คล้ายคลึงกันซึ่งก่อให้เกิดการสร้างและการสังเคราะห์ความคิด

เวียสม่า และคนอื่น (Wiersma and others. 1985:165) ให้ความหมายของแบบทดสอบความเรียงว่า เป็นแบบทดสอบที่ให้นักเรียนแสดงความต้องการที่จะตอบอย่างมีโครงสร้าง ใช้ข้อมูลที่ละเอียดและยาวมากขึ้น จะประกอบด้วยการบรรยายหลาย ๆ ส่วน

ฮอบกินส์ และ แอนเตส (Hopkins and Antes. 1990 : 231) ได้ให้ความหมายของแบบทดสอบความเรียง (Essay Items) ไว้ว่า เป็นข้อสอบประเภทข้อสอบขยายความ (Constructed – response Items) ที่สมบูรณ์ที่สุด เป็นข้อสอบที่ต้องการเห็นพัฒนาการในการตอบคำถามของนักเรียน

เกย์ (Gay.1991 : 224) กล่าวว่าแบบทดสอบความเรียง เป็นแบบทดสอบหนึ่งที่ประกอบด้วยหัวข้อที่ต้องการให้นักเรียนเขียนตอบ ซึ่งโดยปกติแล้วจะเป็นคำตอบที่ยาว จริง ๆ แล้ว แบบทดสอบความเรียงไม่เหมาะสมสำหรับวัดผลระดับความรู้ความจำ แต่เหมาะสำหรับวัดผลในระดับสูง เพราะแบบทดสอบแบบนี้จะต้องมีการสังเคราะห์ และการประเมินค่า

คูบิสซ์ และ บอร์ริช (Kubiszyn and Borich. 1996 : 104) กล่าวว่า แบบทดสอบความเรียงเป็นแบบทดสอบสำหรับให้นักเรียนนำเสนอ มากกว่าเลือกตัวเลือกที่ถูกต้อง แบบทดสอบชนิดนี้ต้องให้นักเรียนเรียบเรียงคำตอบที่กว้างกว่าเดิมเพื่อตอบคำถามที่ไม่ใช่คำถามเดี่ยว

และ วอร์ด (Ward. 1999 : 129) ให้ความหมายของแบบทดสอบความเรียงว่า เป็นแบบทดสอบที่วัดทักษะขั้นสูงที่ยากที่จะสร้างด้วยแบบทดสอบแบบปรนัยได้ แบบทดสอบประเภทนี้ต้องการให้นักเรียนแสดงความรู้ กระบวนการของความรู้ การรวบรวมข้อมูล และ นำเสนอความรู้นั้นในรูปของเหตุผลที่ประติดประต่อกันได้

ส่วนนักวัดผลการศึกษาของไทยหลายท่านก็ได้นิยามแบบทดสอบความเรียงไว้ ดังนี้

รุจิร์ ภู่สาระ (2520:31) ได้กล่าวว่า การสอบแบบบรรยาย (Essay Examination) เป็นการตอบคำถามซึ่งมีจำนวนข้อจำกัด โดยปกติมีประมาณ 5 – 10 ข้อ ครูเป็นผู้ตั้งคำถามให้นักเรียนเป็นผู้ตอบ ครูจะถามในแต่ละวิชาที่นักเรียนได้เรียนมา เพื่อให้นักเรียนตอบตามความสามารถของแต่ละคน และมักตั้งคำถามว่า ใคร เมื่อไร อะไร หรือ ทำไม เพื่อถามนักเรียนถึงชื่อ สถานที่ ให้อภิปราย ประเมินค่า หาเหตุผล หาความหมายของสถานการณ์ต่าง ๆ เป็นต้น

โกวิท ประวาลพฤกษ์ และ สมศักดิ์ สินธุระเวชญ์ (2525 : 101) ให้นิยามแบบทดสอบความเรียงว่า เป็นข้อสอบที่ให้เด็กได้มีอิสระในการตอบเต็มที่ ให้โอกาสเด็กได้แสดงว่าเขารู้อะไร สามารถวัดสมรรถภาพด้านบูรณาการความคิด ความสามารถในการจัดรวบรวมความคิดที่สมบูรณ์ (Organization Ability) การสังเคราะห์และการประเมินค่าอย่างมีประสิทธิภาพ เช่น ความรู้ในการแก้ปัญหาใหม่ ๆ และเป็นต้นคิดหรือค้นหาวิธีการใหม่ ๆ เพื่อแก้ปัญหาต่าง ๆ ได้

พวงรัตน์ ทวีรัตน์ (2530 : 110) ให้ความหมายของแบบทดสอบอัตนัย หรือ แบบทดสอบความเรียง (Essay Test) ว่าเป็นแบบทดสอบที่มีคุณค่ามากชนิดหนึ่ง สามารถใช้วัดพฤติกรรมด้านพุทธิพิสัยหรือด้านความรู้และความคิดระดับสูงได้ดี เช่น วัดพฤติกรรมขั้นการสังเคราะห์ และการประเมินค่า เป็นต้น

วัญญา วิศาลาภรณ์ (2533 : 60) ให้ความหมายของแบบทดสอบความเรียง (Essay Test) หรือ (Subjective Test) เป็นข้อสอบที่ให้เสรีภาพเป็นอย่างมากแก่นักเรียนในการตอบ นักเรียนมีอิสระในการที่จะเสนอความคิดด้วยภาษาของตนเอง นักเรียนต้องใช้ความสามารถอธิบายปัญหาต่าง ๆ ได้ ดังนั้น ข้อสอบแบบความเรียงหรือข้อสอบอัตนัยจึงสามารถใช้วัดความรู้ความคิดขั้นสูงที่ซับซ้อนได้ ใช้วัดทัศนคติได้ นักเรียนต้องมีความสามารถในการจัดระเบียบความรู้ความคิด แสดงความริเริ่มและรู้จักสังเคราะห์

บุญชม ศรีสะอาด (2534 : 80) กล่าวถึงแบบทดสอบความเรียง(Essay Test) ว่าเป็นข้อสอบที่ให้ผู้สอบเขียนตอบตามความคิดของตนเอง แม้ว่าโดยทั่วไปจะได้รับการนำไปใช้น้อย แต่ก็เป็นเครื่องมือที่จำเป็นในการวัดผลโดยเฉพาะในการวัดเกี่ยวกับความคิดเห็น เจตคติ ความสามารถในการคิดสังเคราะห์ (Synthesis) ความสามารถในการอธิบายให้คนอื่นเข้าใจ ความสามารถในการบูรณาการความรู้ การวัดในลักษณะเหล่านี้ข้อสอบแบบอัตนัยจะวัดได้ดีมาก

ล้วน สายยศ และอังคณา สายยศ (2539 : 86) ได้กล่าวถึงแบบทดสอบความเรียง ว่ามีจุดประสงค์วัดความสามารถในการบรรยาย อธิบายและแสดงเหตุผลตามความคิดเห็นของตน อาจจำกัดความยาวหรือให้เขียนได้ตามสบายก็ได้

จากการศึกษานิยามของแบบทดสอบความเรียงจากนักการศึกษาหลาย ๆ ท่านที่กล่าวมานี้ ผู้วิจัยสรุปได้ว่า แบบทดสอบความเรียงหรือเรียกโดยทั่วไปว่าแบบทดสอบอัตนัย เป็นแบบทดสอบที่ใช้เพื่อการวัดผลสัมฤทธิ์ทางการเรียนในพฤติกรรมการเรียนรู้ขั้นสูง ที่ต้องการให้นักเรียนแสดงความคิดของตนเองออกมาด้วยภาษาและถ้อยคำที่ต้องอาศัยการเรียบเรียง และรวบรวมความรู้มาตอบด้วยตนเอง

1.2 ลักษณะและประเภทของแบบทดสอบความเรียง

1.2.1 การเปรียบเทียบลักษณะของแบบทดสอบความเรียงกับแบบทดสอบอื่น ๆ

มีนักวัดผลการศึกษาหลายท่านได้สรุปเปรียบเทียบลักษณะของแบบทดสอบความเรียง กับ แบบทดสอบแบบอื่น ไว้หลายท่าน เช่น

กลอนลันด์ (Gronlund. 1977 : 77-78) ได้เปรียบเทียบสรุปลักษณะระหว่างข้อสอบแบบปรนัย และข้อสอบแบบความเรียง ไว้ดังต่อไปนี้

ตาราง 1 การเปรียบเทียบระหว่างลักษณะข้อสอบปรนัย และข้อสอบแบบความเรียง

	ข้อสอบแบบปรนัย	ข้อสอบแบบความเรียง
การวัดผลการเรียนรู้	ดีสำหรับวัดผลการเรียนรู้ขั้นความรู้ความเข้าใจ การนำไปใช้ และการวิเคราะห์ แต่ไม่เพียงพอต่อการวัดการสังเคราะห์ และการประเมินค่า	ไม่มีประสิทธิภาพในการวัดผลการเรียนรู้ แต่ดีในการวัดความเข้าใจ การนำไปใช้ และการวิเคราะห์ และดีที่สุดสำหรับการวัดผลการสังเคราะห์และประเมินค่า
การสุ่มตัวอย่างเนื้อหา	การใช้ข้อสอบจำนวนหลายข้อจะมีความครอบคลุมกว้าง ซึ่งทำให้การสุ่มตัวอย่างเนื้อหาได้ครอบคลุม	การใช้ข้อสอบจำนวนไม่มากข้อ ส่งผลให้มีขอบเขตจำกัด ทำให้การสุ่มตัวอย่างเนื้อหาไม่ครอบคลุม
การเตรียมการสร้าง	การเตรียมคำถามที่ดีทำได้ยาก และใช้เวลานาน	การเตรียมคำถามที่ดีทำได้ยาก แต่ง่ายกว่าคำถามแบบปรนัย
การให้คะแนน	มีความเป็นปรนัย ง่าย และ มีความเชื่อมั่นสูง	มีความเป็นอัตนัย ยาก และให้ความเชื่อมั่นต่ำ
ปัจจัยที่ส่งผลต่อคะแนนที่ได้	ใช้ความสามารถในการอ่านและการเดา	ใช้ความสามารถในการเขียนและการลวง
ผลกระทบต่อการเรียนที่เป็นไปได้	กระตุ้นให้นักเรียนจำ ตีความ และวิเคราะห์ความคิดของผู้อื่น	กระตุ้นให้นักเรียนจัดระเบียบรวบรวมและแสดงความคิดเห็นของตนเอง

เวียสม่า และคนอื่น ๆ (Wiersma and others. 1985 : 168) ก็ได้เปรียบเทียบลักษณะที่แตกต่างระหว่างแบบทดสอบความเรียงกับแบบทดสอบปรนัย ไว้ดังนี้

1. แม้ว่าการให้คะแนนแบบทดสอบความเรียงจะมีความเป็นปรนัยน้อยกว่าแบบทดสอบปรนัย แต่ยังมีจุดสำคัญต่าง ๆ ที่ดีกว่าแบบทดสอบปรนัย

2. ข้อสอบปรนัยถูกฟอร์มด้วยส่วนคำถาม และ ส่วนคำตอบตามรูปแบบของแบบทดสอบที่ดี แต่แบบทดสอบความเรียงจะมีส่วนคำถามที่มีตั้งแต่ขนาดสั้นบรรทัดเดียว จนถึงยาวเป็นหน้า ๆ

3. การตอบแบบทดสอบความเรียงต่อข้อ ใช้เวลานานกว่าการตอบแบบทดสอบปรนัย

4. คำตอบของแบบทดสอบความเรียง ต้องอาศัยทักษะในการเขียนมากกว่าการตอบคำถามแบบทดสอบปรนัยที่ไม่ต้องอาศัยทักษะกระบวนการเขียนเลย

เช่นเดียวกัน ต่อมา ฮอบกินส์ และแอนเตส (Hopkins and Antes. 1990 : 150) ได้เปรียบเทียบลักษณะของข้อสอบแบบต่าง ๆ โดยสรุปได้ดังนี้

ตาราง 2 เปรียบเทียบลักษณะของข้อสอบแบบต่าง ๆ

ลักษณะ	ข้อสอบแบบปรนัย	ข้อสอบอัตนัย		ข้อสอบโจทย์ปัญหา
	ถูก-ผิด,เลือกตอบ, จับคู่,จัดให้เข้าพวก	เติมคำและตอบคำถามสั้น ๆ	ความเรียง	โจทย์ปัญหาคณิตฯ และเทคนิค
1. การเตรียมการ	เหนื่อย,ยาก	ค่อนข้างง่าย	ยาก	ยาก
2. จำนวนข้อ	มากข้อ	มาก	น้อย	น้อย
3. หัวข้อ	เฉพาะเจาะจง	เฉพาะเจาะจง	ทั่ว ๆ ไป	ทั่ว ๆ ไป
4. การใช้เวลาในการตอบ	อ่านและคิด	อ่าน และคิด	คิด และเขียน	แก้ปัญหา
5. การตอบ	สัญลักษณ์ (ใช้ตัวอักษรหรืออื่น ๆ)	คำเดียวหรือวลี	บรรยายด้วยถ้อยคำของตนเอง	เขียนออกมาเป็นตัวเลข และด้วยถ้อยคำของตนเอง
6. การตอบ	จากตัวเลือก	สั้น ๆ	ขยายความ	สั้น ๆ ถึงขยายความ
7. การให้คะแนน	มีความเป็นปรนัย	มีความเป็นปรนัยหรือเกือบจะเป็นปรนัย	อาศัยการตัดสินของผู้ตรวจ	มีความเป็นปรนัย
8. การตรวจให้คะแนน	ง่าย	ค่อนข้างง่าย	ใช้เวลานาน	ใช้เวลานาน
9. คะแนนที่ได้	กำหนดค่าโดยผู้สร้างข้อสอบ	กำหนดค่าโดยผู้สร้างข้อสอบ	ยืนยันคุณค่าโดยผู้ตรวจคำตอบ	ยืนยันคุณค่าโดยผู้สร้างข้อสอบ
10. จุดมุ่งหมายการเรียนรู้	ทุกพฤติกรรมยกเว้น สังเคราะห์	ความรู้	ทุกพฤติกรรม	ความเข้าใจและพฤติกรรมขั้นสูง
11. อิสระในการตอบ	จำกัดอยู่ที่ตัวเลือกที่ให้มา	น้อยมาก	มากที่สุด	ค่อนข้างมากถึงมากที่สุด
12. การเดา	สามารถทำได้ ,อาจมีแนวทางช่วยในการเดา	น้อย	น้อยที่สุด	น้อยที่สุด
13. การลวง	ไม่มี	มีบ้าง	มีโอกาส	ไม่มี
14. นัยของการเรียนรู้	เรียนรู้ข้อเท็จจริงและการใช้ความรู้ระดับสูงขึ้น	เรียนรู้ข้อเท็จจริง	เรียนรู้ข้อเท็จจริงและพฤติกรรมขั้นสูง	เรียนรู้ข้อเท็จจริง การคำนวณ และพฤติกรรมขั้นสูง

1.2.2 ข้อดี ข้อเสียของแบบทดสอบความเรียง

ฮอบกินส์ และแอนเตส (Hopkins and Antes. 1990 : 232-233) ได้สรุปจุดแข็งจุดอ่อนของแบบทดสอบความเรียง ไว้ว่า การใช้แบบทดสอบความเรียงสอบนักเรียนนั้นทำให้นักเรียนมีอิสระในการตอบเต็มที่ ความมีอิสระมากหรือน้อยขึ้นอยู่กับลักษณะข้อสอบที่สร้างมาว่าต้องการการตอบแบบใด ทำให้สามารถทดสอบพฤติกรรมการเรียนรู้ของนักเรียนในระดับที่สูงขึ้น เช่น การวิเคราะห์ การสังเคราะห์ และการประเมินค่า โดยการให้นักเรียนนำความรู้และหลักการต่าง ๆ ไปประยุกต์ใช้ , เปรียบเทียบ , อธิบาย และ อธิบายความเหมือนความต่างของความรู้และหลักการ ซึ่งทำให้นักเรียนได้แสดงความคิดเห็นโดยไม่มีการตัดสินว่าความคิดเห็นนั้นถูกหรือผิด ไม่เหมือนข้อสอบปรนัยซึ่งให้คะแนนตามความถูกผิด ปัจจัยสำคัญในการตอบข้อสอบแบบความเรียงคือ ความสามารถของนักเรียนในการสนับสนุนความคิดเห็นของตน ความจำเป็นในการจัดระเบียบและการแสดงความคิดเห็น ช่วยเปิดโอกาสให้นักเรียนมีพัฒนาการทางทักษะในการเขียน นอกจากนี้การให้อิสระในการตอบเต็มที่ ทำให้สามารถวิเคราะห์หรือประเมินการเขียน และจัดองค์ประกอบต่าง ๆ ได้ และ ข้อสอบแบบความเรียงที่ดี จะใช้เวลาในการออกข้อสอบน้อยกว่าข้อสอบแบบปรนัย เพราะมีหัวข้อให้เขียนเพียงไม่กี่หัวข้อ ไม่เหมือนข้อสอบแบบปรนัยที่ต้องคิดและเขียนตัวเลือกมากมาย

ส่วนจุดอ่อนของแบบทดสอบความเรียงจะมีผลต่อการตรวจคำตอบของนักเรียนที่ใช้เวลานานมาก ยากที่จะให้คะแนน ถ้าไม่มีเกณฑ์ในการให้คะแนน ผลจากการวัดก็จะไม่มีประโยชน์ คำตอบที่ได้มักเป็นส่วนหนึ่งของเนื้อหาวิชาเท่านั้น ซึ่งอาจน้อยกว่าสิ่งที่ต้องการวัด และถ้านักเรียนไม่สามารถเขียนแสดงความคิดเห็นออกมาได้ ก็อาจจะแสดงผลสำเร็จที่ไม่ดี นักเรียนคนที่มีความสำเร็จในการเรียนรู้เท่ากัน แต่มีระดับของทักษะการแสดงออกทางความคิดต่างกัน ก็อาจจะได้คะแนนที่ต่างกันมากในการสอบด้วยแบบทดสอบความเรียง

คูบิสซ์ และ บอร์ริช (Kubiszyn and Borich. 1996 : 109) ก็ได้รวบรวมข้อดีและข้อเสียของข้อสอบแบบความเรียง ไว้ 4 อย่าง ดังนี้

ข้อดี

1. มีประสิทธิผลมากที่สุดในการประเมินผลการเรียนรู้ที่ซับซ้อน
2. การสร้างข้อสอบง่าย
3. เน้นย้ำในด้านทักษะการสื่อสาร หมายความว่าครูต้องใช้เวลาในการสอนทักษะการสื่อสารที่ตรงกับหลักสูตร รวมถึงคำศัพท์พิเศษและรูปแบบในการเขียน
4. ไม่มีการเดา

ข้อเสีย

1. ยากต่อการให้คะแนน
2. คะแนนที่ได้ไม่มีความแน่นอนเชื่อถือได้
3. จำกัดเนื้อหาจากการสอนทั้งหมด
4. การลวง เนื่องจากว่ายิ่งคำตอบของนักเรียนยาวเท่าไร ก็มีแนวโน้มที่จะได้คะแนนมากกว่าคำตอบที่สั้นมากเท่านั้น โดยไม่ได้คำนึงถึงเนื้อหา

1.2.3 ประเภทของแบบทดสอบความเรียง

กลอนลันด์ (Gronlund. 1977 : 78-80) ได้แบ่งรูปแบบคำถามของแบบทดสอบความเรียงไว้ 2 ประเภท คือ

1. คำถามที่ต้องการกำหนดขอบเขตของคำตอบ (Restricted-Response Questions) คำถามนี้จะมีการจำกัดคำตอบที่นักเรียนต้องตอบลงในข้อสอบ โดยทั่วไปแล้วขอบเขตของเนื้อหาที่ใช้พิจารณาจะถูกกำหนดขึ้นอย่างแคบ ๆ ด้วยปัญหาและรูปแบบเฉพาะของคำตอบก็จะระบุไว้ด้วยเช่นกัน (โดยคำว่า "เขียนเป็นข้อ ๆ" "ให้คำนิยาม" และ "ให้เหตุผล" เป็นต้น) ในบางกรณีการตอบอาจจะถูกจำกัดมากขึ้นไปอีก ด้วยการใช้ส่วนเกริ่นนำหรือการใช้คำสั่งพิเศษ

ตัวอย่าง

"จงอธิบายข้อดีที่สัมพันธ์กันของข้อสอบปรนัยและข้อสอบแบบความเรียง เพื่อวัดผลการเรียนรู้ในระดับความเข้าใจ ให้เขียนตอบมา 1 หน้ากระดาษ"

ตัวอย่าง

นายโรเจอร์ส ซึ่งเป็นครูวิทยาศาสตร์ชั้น ม. 3 ต้องการที่จะวัดความสามารถในการตีความข้อมูลทางวิทยาศาสตร์ของนักเรียนของเราโดยใช้แบบทดสอบข้อเขียน

1. *จงบรรยายขั้นตอนที่นายโรเจอร์ต้องปฏิบัติ*
2. *จงให้เหตุผลเพื่อสนับสนุนความเหมาะสมของแต่ละขั้นตอน*

คำถามรูปแบบจำกัดคำตอบนี้มีทั้งข้อดีและข้อเสีย คำถามแบบนี้สามารถเตรียมได้ง่ายกว่า สัมพันธ์กับผลการเรียนรู้เฉพาะโดยตรงมากกว่า และให้คะแนนง่ายกว่า อย่างไรก็ตามคำถามแบบนี้ก็ให้โอกาสกับนักเรียนน้อยในการที่จะแสดงความสามารถของพวกเขาในด้านการจัดระเบียบ การรวบรวม และ การพัฒนารูปแบบในการตอบแบบใหม่ ๆ หากต้องการวัดผลในระดับการสังเคราะห์ และประเมินผล ก็ควรใช้คำถามแบบไม่จำกัดคำตอบ จะเหมาะสมกว่า

2. คำถามแบบไม่จำกัดคำตอบ (Extended – Response Questions) คำถามนี้จะให้อิสระในการตอบกับนักเรียน ไม่จำกัดในการกำหนดรูปแบบและขอบเขตการตอบ แม้ว่าจะมีการจำกัดในทางปฏิบัติบ้าง เช่นจำกัดด้านเวลา หรือจำกัดหน้ากระดาษ นักเรียนมีอิสระในการแสดงทักษะในการวิเคราะห์และประเมินผล ดังนั้นปริมาณโครงสร้างจึงหลากหลายกันไปในแต่ละข้อโดยขึ้นอยู่กับผลการเรียนรู้ที่ต้องการวัด

ตัวอย่าง ระดับการสังเคราะห์ :

จงเตรียมแผนที่สมบูรณ์สำหรับหลักสูตรที่คุณกำลังสอนหรือคาดว่าจะสอนเพื่อประเมินผลความสัมฤทธิ์ของนักเรียน โดยจะต้องมีกระบวนการที่คุณจะปฏิบัติ , เครื่องมือที่คุณจะใช้และเหตุผลในการเลือกของคุณ

ตัวอย่าง ระดับการประเมินผล :

(นักเรียนจะได้รับข้อสอบวัดผลสัมฤทธิ์ทางการเรียน ที่มีข้อผิดพลาดตรงคำสั่ง , หัวข้อ , และการจัดเรียงหัวข้อ)

จงเขียนการประเมินเชิงวิเคราะห์เกี่ยวกับข้อสอบฉบับนี้ โดยใช้เกณฑ์การประเมิน กฎและมาตรฐานในการสร้างข้อสอบที่บรรยายไว้ในตำราของคุณ รวมทั้งกล่าวถึงการวิเคราะห์โดยละเอียดเกี่ยวกับจุดเด่นและจุดด้อยของข้อสอบ และการประเมินคุณภาพโดยรวมและความมีประสิทธิผลที่เป็นไปได้

คำถามที่ไม่จำกัดคำตอบนี้จะทำให้เกิดการรวบรวมความคิดในเชิงสร้างสรรค์ การประเมินเนื้อหาโดยรวมและวิธีการแก้ปัญหาอย่างกว้าง ๆ สิ่งเหล่านี้เป็นผลการเรียนรู้ที่สำคัญซึ่งไม่สามารถวัดได้ด้วยรูปแบบข้อสอบประเภทอื่น ปัญหาใหญ่ที่สุดของข้อสอบลักษณะนี้ คือ การประเมินคำตอบด้วยความเชื่อมั่นที่สุด เพื่อให้ได้ผลการเรียนรู้ที่มีประโยชน์ สิ่งเหล่านี้เป็นงานที่ยากและใช้เวลานาน แต่ผลที่ได้ก็นับว่าคุ้มค่า

ฮอบกินส์ และ แอนเตส กับ กลอนลันด์ (Hopkins and Antes. 1979 : 94 ; Gronlund. 1982 : 74 – 75) นำเสนอประเภทของแบบทดสอบความเรียงไว้ 2 ประเภท คือ

1. แบบไม่จำกัดคำตอบ (Extended – Response Questions) ข้อสอบแบบนี้ให้เสรีภาพแก่ผู้ตอบอย่างมากในการกำหนดรูปแบบหรือขอบเขตของการตอบ เน้นความลึก และขอบเขตของความรู้ เน้นเสรีภาพของการแสดงออก ยั่วยุให้ผู้ตอบเกิดความคิดริเริ่มสร้างสรรค์ ข้อเสีย คือ ความเชื่อมั่นของคะแนนที่ได้ค่อนข้างต่ำ ข้อสอบอัตนัยแบบไม่จำกัดคำตอบเหมาะที่จะวัดการสังเคราะห์ และการประเมินค่า

2. แบบจำกัดคำตอบ (Restricted –Response Questions) ข้อสอบแบบนี้ต้องการคำตอบเฉพาะเจาะจงภายในขอบเขตของคำถาม ดังนั้นนักเรียนจะต้องจัดระเบียบความคิดเป็นอย่างดีเพื่อตอบให้ตรงกับประเด็นที่ถาม และไม่จำเป็นต้องตอบยาว ข้อสอบแบบจำกัดคำตอบเหมาะกับการวัด ความเข้าใจ การนำไปใช้ ส่วนดีของข้อสอบแบบจำกัดคำตอบอีกประการหนึ่ง คือ มีความเชื่อมั่นของคะแนนสูงกว่าแบบไม่จำกัดคำตอบ

ฮอบกินส์ และแอนเตส (Hopkins and Antes. 1990 : 231-232) ได้เสนอรูปแบบของแบบทดสอบความเรียงอีกครั้ง ว่าข้อสอบประเภทนี้ต้องการคำตอบที่เป็นประโยคหลายประโยคต่อเนื่องกัน ซึ่งแสดงความสมเหตุสมผล ผู้ตรวจความถูกต้องและคุณภาพของข้อสอบแบบนี้ ต้องเป็นผู้ที่มีความรู้ในวิชาที่สอบ ซึ่งแบ่งออกเป็น 2 ลักษณะคือ

1. แบบที่มีขอบเขตกว้างแต่ให้เวลาจำกัด (Extended-response) ข้อสอบลักษณะนี้จะทดสอบทักษะของผู้เขียนในการเลือกว่าจะเขียนอะไรและจะเขียนมากน้อยแค่ไหนในแต่ละส่วน รวมทั้งความสามารถในการจัดระเบียบและแสดงความคิดในเชิงตรรกะ ข้อสอบเขียนบรรยายแบบขยายความนี้เป็นข้อสอบประเภทปลายเปิด (Open-ended) และไม่ได้จำกัดความคิดของนักเรียน ข้อสอบแบบนี้มีประโยชน์ในการประเมินการเขียนโดยเฉพาะศิลปะเกี่ยวกับการใช้ภาษา

2. แบบจำกัดคำตอบ (Limited-response) จะมีขอบเขตแคบหรือกว้างก็ได้ แต่นักเรียนควรจะรู้อย่างแน่ชัดว่าขีดจำกัดนั้นคืออะไร ข้อสอบแบบนี้จะให้ขีดจำกัดที่ระบุได้อย่างชัดเจนเพื่อสร้างขอบเขตในการตอบให้นักเรียน กล่าวคือนักเรียนต้องตอบในกรอบเนื้อหาที่จำกัด ข้อสอบแบบนี้มีประโยชน์ต่อครูผู้สอนที่ต้องการทดสอบผลสัมฤทธิ์ทางการเรียนซึ่งต่างจากแบบทดสอบความเรียงแบบที่มีขอบเขตกว้างที่ต้องการวัดด้านทักษะการเขียน ผู้ออกข้อสอบควรพิจารณาถึงจุดมุ่งหมายในการวัดผลก่อนตัดสินใจเลือกว่าจะใช้รูปแบบใด

คูบิสซ์ และ บอร์ริช (Kubiszyn and Borich. 1996 : 105-106) ได้นำเสนอรูปแบบของข้อสอบมี 2 ประเภท คือ

1. ข้อสอบที่ไม่จำกัดคำตอบ (Extended-Response Items) ข้อสอบประเภทความเรียงจะปล่อยให้นักเรียนกำหนดความยาวและความซับซ้อนของคำตอบ ข้อสอบประเภทนี้มีประโยชน์มากที่สุดในระดับของการสังเคราะห์หรือประเมินผล เนื่องจากความยาวของข้อสอบประเภทนี้ และเวลาที่ต้องใช้ในการเรียบเรียงคำตอบ บางครั้งจึงอาจมอบหมายให้ทำเป็นรายงานหรือข้อสอบแบบนำกลับไปทำที่บ้าน (A take-home test) ข้อสอบประเภทนี้มีคุณค่าในการประเมินความสามารถในการสื่อสารมากกว่าการประเมินผลสัมฤทธิ์ทางการเรียน

ตัวอย่าง :

จงเปรียบเทียบความเหมือนและความแตกต่างระหว่างการบริหารประเทศของประธานาธิบดีคาร์เตอร์กับเรแกน โดยพิจารณาในด้านนโยบายทางเศรษฐกิจ สังคมและการทหาร พยายามอย่างเขียนในเชิงสนับสนุนฝ่ายใดฝ่ายหนึ่ง คำตอบของคุณจะได้คะแนนตามความจุดมุ่งหมาย ความถูกต้อง การจัดระเบียบ และความชัดเจน

2. ข้อสอบที่ต้องการจำกัดคำตอบ (Restricted-Response Items) ข้อสอบประเภทนี้ การตั้งคำถามเฉพาะขึ้นมาอย่างใดอย่างหนึ่ง ให้นักเรียนดึงข้อมูลที่เหมาะสมมาใช้เรียบเรียงให้ดี สรุปอย่างเหมาะสมแล้วเขียนออกมาภายใต้ขีดจำกัดของคำถามที่ตั้งขึ้น

ตัวอย่าง :

จงบอกความเหมือนและความแตกต่างทางการเมืองหลัก ๆ ระหว่างการเข้าร่วมในสงครามเกาหลีกับสงครามโลกครั้งที่ 2 ของประเทศสหรัฐอเมริกา เขียนตอบไม่เกิน 1 หน้า การให้คะแนนจะขึ้นอยู่กับความถูกต้อง การเรียบเรียง และความสั้นกระชับ

นอกจากนักวัดผลการศึกษาของต่างประเทศแล้ว บุญเชิด ภิญโญอนันตพงษ์ (2525 : 33) ก็ยังได้แบ่งประเภทของข้อสอบอัตนัย ไว้ 2 รูปแบบเช่นเดียวกัน ดังนี้

1. แบบไม่จำกัดคำตอบ หรือแบบขยายความ ข้อสอบแบบนี้จะถามความรู้ความสามารถต่าง ๆ โดยให้ผู้ตอบแสดงความคิดเห็นอย่างอิสระ ซึ่งจะสามารถวัดสมรรถภาพด้านความคิดริเริ่ม สร้างสรรค์ทัศนคติ การประเมินค่า ได้อย่างกว้างขวาง ปริมาณคำตอบจึงขึ้นอยู่กับคำถามและความรู้ที่สั่งสมไว้ว่ามีมากน้อยเพียงใด การกำหนดเวลาให้เขียนตอบจึงต้องกำหนดให้เหมาะสม

ข้อสอบแบบนี้ส่งเสริมให้นักเรียนรู้จักการรวบรวมความคิดต่าง ๆ การประเมินคุณค่าของสิ่งเหล่านั้นและการใช้วิธีการต่าง ๆ ในการแก้ปัญหา ซึ่งข้อสอบแบบอื่น ๆ ไม่สามารถจะนำมาใช้วัดได้ แต่จุดอ่อนของการถามแบบนี้ อยู่ที่การให้คะแนน เพราะยากที่จะหาเกณฑ์ในการให้คะแนนได้ถูกต้อง เนื่องจากผู้ตอบมีอิสระในการคิดและเขียนตอบ

2. แบบจำกัดคำตอบ ข้อสอบแบบนี้จะถามแบบจำเพาะเจาะจง และต้องการคำตอบเฉพาะเรื่องซึ่งผู้ตอบต้องจัดเรียงความคิดให้เป็นระเบียบ เพื่อให้ตรงประเด็นของคำถามเพียงสั้น ๆ ดังนั้น จึงต้องระวังเรื่องคำสั่งของโจทย์ ขอบเขตของเนื้อหา เวลาที่ให้นักเรียนเขียนตอบแบบนี้ มีความสะดวกในการให้คะแนนมากกว่าแบบแรก เพราะมีเกณฑ์ต่าง ๆ ที่จะตัดสินให้คะแนนมีมากขึ้น

พวงรัตน์ ทวีรัตน์ (2530 : 108) ได้กล่าวเพิ่มเติมเกี่ยวกับประเภทของข้องสอบอัตนัยว่า มี 2 ประเภท คือ

ก. แบบไม่กำหนดขอบเขตของการตอบหรือการตอบยาว (Extended Response) ข้อสอบอัตนัยแบบนี้จะเปิดโอกาสให้ผู้สอบตอบแสดงความรู้ ความคิดเห็นในสิ่งที่ถามอย่างเต็มที่และอย่างอิสระ เป็นข้อสอบอัตนัยที่สามารถวัดสมรรถภาพสมองด้านความคิดสร้างสรรค์ การประเมินค่า ตลอดจนความคิดเห็นและทัศนคติได้อย่างกว้างขวาง ปริมาณคำตอบขึ้นอยู่กับความรู้ความสามารถและความคุ้นเคยในสิ่งที่ถามของผู้ตอบว่า มีมากน้อยเพียงใด การกำหนดเวลาในการตอบจึงต้องกำหนดให้เหมาะสม ข้อสอบอัตนัยแบบนี้เป็นการวัดผู้สอบในคุณลักษณะ 4 อย่าง คือ

1. ความสามารถในการระลึกได้ถึงความรู้ที่ได้เรียนไปแล้ว
2. ความสามารถในการประเมินค่าความรู้ที่จำได้
3. ความสามารถในการรวบรวมความรู้ และความคิดให้เป็นระบบ
4. ความสามารถในการนำความรู้ ความคิดที่ได้จัดระบบแล้วมาใช้แก้ปัญหา หรือตอบคำถามในสิ่งที่ถามอย่างมีเหตุผล

ข. แบบกำหนดขอบเขตของคำตอบ (Restricted Response)

ข้อสอบอัตนัยแบบนี้ ลักษณะการถามจะเฉพาะเจาะจง และต้องการคำตอบที่เฉพาะเจาะจงภายในขอบเขตที่ถาม ดังนั้น ผู้ตอบจึงต้องจัดเรียงความคิดให้เป็นระเบียบ เพื่อตอบให้ตรงกับประเด็นที่ถาม และไม่จำเป็นต้องตอบยาว ความสำคัญอยู่ที่ว่าต้องเข้าใจคำถามและตอบตามประเด็นที่ถาม ข้อสอบอัตนัยแบบนี้ให้คะแนนง่ายกว่าแบบแรก ข้อสอบอัตนัยแบบกำหนดขอบเขตของคำตอบนี้ ใช้วัดความสามารถของนักเรียนในการสังเคราะห์ความคิด และบรรยายความคิดนั้นออกมาอย่างมีเหตุมีผล ดังนั้นจึงเหมาะที่ใช้วัดความสามารถของสมองด้านความเข้าใจ การนำไปใช้ การวิเคราะห์ การสังเคราะห์ และการประเมินค่า เป็นอย่างดี

นอกจากทั้ง 4 ท่านนี้ ก็มีแนวคิดของนักวัดผลการศึกษาอีกหลายท่านที่กำหนดประเภทของข้อสอบอัตนัยหรือแบบทดสอบความเรียง ในทำนองเดียวกัน คือ มี 2 ประเภท คือ แบบไม่จำกัดคำตอบ และแบบจำกัดคำตอบ อาทิเช่น บุญชม ศรีสะอาด (2520 : 97) บุญเชิด ภิญโญอนันตพงษ์ (2521 : 126) บุญเชิด ภิญโญอนันตพงษ์ (2519 : 31) และ บุญธรรม กิจปรีดาบริสุทธิ์ (2524 : 28)

ในที่นี้ ผู้วิจัยพบว่าการดำเนินการวิจัยครั้งนี้ ควรตั้งคำถามแบบกำหนดขอบเขตของคำตอบ เพราะเหมาะสมกับนักเรียนระดับประถมศึกษา

1.2.4 ลักษณะของแบบทดสอบความเรียง

นักวัดผลการศึกษาได้เสนอลักษณะของแบบทดสอบความเรียงไว้หลายท่าน อาทิเช่น

บุญธรรม กิจปรีดาบริสุทธิ์ (2524 : 28) กล่าวถึงลักษณะของข้อสอบอัตนัยว่ามีลักษณะเด่นอยู่ที่การให้อิสระแก่ผู้ตอบ ผู้ตอบจะต้องเรียบเรียงความรู้ ความเข้าใจ และความคิดเห็น แล้วเขียนคำตอบเองตามที่ตนถนัด คำตอบที่เขียนอาจจะมีตั้งแต่หนึ่งประโยคจนถึงหลาย ๆ หน้ากระดาษ ส่วนการตอบ ผู้ตอบก็ต้องใช้เวลาแบบทั้งหมดในการ คิด และ เขียน ถ้ามองในลักษณะคำถามหรือโจทย์ ก็เป็นคำถามที่กำหนดสถานการณ์ขึ้น เพื่อให้ผู้ตอบพรรณนา หรือแสดงความรู้ ความเข้าใจ และ ความคิดเห็นตั้งแต่กว้างที่สุด เข้าทำนอง

เขียนอะไรก็ได้ จนถึง แคบเฉพาะเจาะจงที่สุด ตามที่โจทย์กำหนด การตรวจข้อสอบแบบนี้ต้องใช้ผู้มีความรู้ความสามารถเกี่ยวกับปัญหาที่ถามเป็นอย่างดี จะใช้ผู้อื่นตรวจแทนไม่ได้

บุญเชิด ภิญโญอนันตพงษ์ (2525 : 33) กล่าวถึงลักษณะของแบบทดสอบความเรียง มี 2 ส่วน คือ ส่วนที่เป็นคำถามจะกำหนดเป็นสถานการณ์หรือปัญหาในรูปใดรูปหนึ่งอย่างกว้าง ๆ หรือเฉพาะเจาะจง เพื่อให้ผู้ตอบได้แสดงความรู้ ความเข้าใจ ความคิดเห็น ทัศนคติ ได้อย่างไม่จำกัด ดังนั้น โดยทั่วไปแบบทดสอบความเรียงจึงมีจำนวนข้อไม่มากนัก ส่วนมากมีจำนวนข้อไม่เกิน 10 ข้อ คำหรือข้อความที่ใช้เขียนถามมีความตื้น ลึก แคบ กว้าง แตกต่างกันไป เช่นคำว่า "การเปรียบเทียบ" ก็ต้องมีทั้งสิ่งที่แตกต่างกันหรือเหมือนกัน ส่วนคำว่า "ความแตกต่าง" นั้น มีแต่ความแตกต่างกันเพียงอย่างเดียว มีผู้เขียนหลายท่านได้เน้นว่า การเตรียมนักเรียนเพื่อตอบคำถามเป็นสิ่งจำเป็นและสำคัญมาก ซึ่งทำได้โดยการฝึกหัดให้นักเรียนได้รู้และเข้าใจขอบเขตของคำถามและให้รู้จักแบ่งเวลาเพื่อจะไม่ให้ใช้เวลาทำข้อใดข้อหนึ่งมากเกินไปจะช่วยให้การวัดและการตรวจให้คะแนนยุติธรรมและแม่นตรงยิ่งขึ้น

ส่วนที่เป็นคำตอบจะมีลักษณะและปริมาณไม่แน่นอน เนื่องจากผู้ตอบต้องเขียนตอบ โดยเรียบเรียงสำนวนจากความคิด ความเข้าใจอย่างอิสระ ดังนั้น คำตอบจึงอาจมีเพียงประโยคเดียวจนถึงหลาย ๆ หน้ากระดาษ การตอบแบบทดสอบความเรียงจะมีต้องจัดระเบียบคำตอบภายในเวลาที่กำหนดให้ ใช้สำนวนภาษาและแบบฉบับของตนเองเขียนตอบ โดยเขียนตอบจำนวนน้อยข้อ และต้องเขียนคำตอบให้ครอบคลุมคำถามอย่างสมบูรณ์และระมัดระวัง

พวงรัตน์ ทวีรัตน์ (2530 : 108) กล่าวว่าลักษณะของแบบทดสอบแบบความเรียงนี้ มี 2 ส่วน คือ ส่วนที่เป็นคำถาม ซึ่งอาจกำหนดในรูปแบบสถานการณ์หรือคำถามในรูปใดรูปหนึ่งอย่างกว้าง ๆ หรือบางครั้งก็เฉพาะเจาะจงเพื่อให้ผู้ตอบได้แสดงความรู้ ความเข้าใจ ความคิดเห็น หรือทัศนคติอย่างอิสระ ลักษณะการถามจะมีความตื้นลึกแตกต่างกันแล้วแต่ผู้ออกข้อสอบต้องการ แต่โดยทั่วไปมักเป็นที่เข้าใจว่า คำถามของแบบทดสอบความเรียงจะเป็นคำถามที่แคบลึกในประเด็นใดประเด็นหนึ่ง ผู้ตอบต้องมีความรู้ในเรื่องนั้นเป็นอย่างดีจึงจะสามารถเขียนตอบได้ การเตรียมตัวเพื่อตอบข้อสอบอัตนัยจึงเป็นสิ่งจำเป็นมาก โดยปรกติแบบทอสอบความเรียงจะมีจำนวนคำถามไม่มากนัก และคำถามแต่ละข้ออาจแบ่งเป็นคำถามย่อย ๆ ได้อีก และบางครั้งก็ระบุการให้คะแนนไว้ด้วย เพื่อช่วยให้การตรวจมีความเชื่อถือได้มากยิ่งขึ้น

ส่วนที่เป็นคำตอบ ซึ่งมีลักษณะและปริมาณไม่แน่นอน และโดยทั่วไปมักจะอยู่ในรูปของการเขียนบรรยาย โดยผู้ตอบจะเขียนเรียบเรียงสำนวนจากความคิดของตนเอง

1.2.5 รูปแบบคำถามของแบบทดสอบความเรียง

มอนโร และคาร์เตอร์ (บุญเชิด ภิญโญอนันตพงษ์. 2525 : 35 – 37 ; citing Monroe and Carter. 1923) ได้เสนอชนิดคำถามไว้ 20 แบบ ดังนี้

1. ถามให้เลือกระลึกประสบการณ์ออกมา เช่น "นายกรัฐมนตรีไทย ที่เคยเป็นทหารมาก่อน คือใครบ้าง ให้บอกมา 2 คน"

2. ถามให้ประเมินสิ่งที่ระลึกออกมา เช่น "จงบอกชื่อบุคคลที่มีอิทธิพลมากที่สุด ในการพัฒนาข้อสอบวัดสติปัญญามาสามคน"

3. ถามให้เปรียบเทียบของสองสิ่งในแง่ใดแง่หนึ่ง เช่น " จงเปรียบเทียบระหว่าง ข้อสอบอัตนัย กับข้อสอบปรนัย ในแง่การให้คะแนน"

4. ถามให้เปรียบเทียบของสองสิ่งในแง่ทั่ว ๆ ไป เช่น "จงเปรียบเทียบข้อสอบมาตรฐาน กับข้อสอบที่ครูสร้างขึ้นเอง"

5. ถามให้ตัดสินใจเพื่อสนับสนุนหรือคัดค้าน เช่น "ตามความคิดเห็นของท่าน ระหว่างการสอบแบบปากเปล่ากับการสอบ โดยใช้การเขียนตอบ ท่านสามารถทำข้อสอบแบบใดดีกว่า เพราะเหตุใด"

6. ถามถึงสาเหตุหรือผลลัพธ์ เช่น "จงบอกเหตุผลในการที่ข้อสอบปรนัยได้รับความนิยมใช้โดยทั่วไปในปัจจุบัน"

7. ถามให้อธิบายการใช้ หรือความหมายที่แท้จริงของคำ วลี หรือประโยค เช่น "จงอธิบายความหมายของคำว่า ปรนัย"

8. ถามให้สรุปความ หรือย่อความจากตำรา หรือจากบทความที่อ่าน เช่น "ให้สรุปประโยชน์และขีดจำกัดของข้อสอบอัตนัยมาเพียงหนึ่งหน้ากระดาษ

9. ถามให้วิเคราะห์ เช่น "ทำไมผู้มีความรอบรู้ทั้งหลายจึงได้สงสัยประสิทธิภาพของข้อสอบมาตรฐาน"

10. ถามหาความสัมพันธ์ เช่น "ทำไมความรู้ด้านพฤกษศาสตร์ จึงมีส่วนช่วยในการศึกษาด้านเกษตรศาสตร์"

11. ถามให้แสดงหรือยกตัวอย่างของหลักการทางวิทยาศาสตร์ โครงสร้างทางภาษา เช่น "จงยกตัวอย่างการใช้คาร์บอนบริสุทธิ์ ในงานอุตสาหกรรมมาสองอย่าง"

12. ถามให้จำแนกประเภท เช่น "คำถามต่อไปนี้ มีความคลาดเคลื่อนอะไรเกิดขึ้น ?"

13. ถามเกี่ยวกับการประยุกต์ กฎ หรือ หลักการในสถานการณ์ใหม่ เช่น "ถ้านำข้อสอบที่สร้างขึ้นในอเมริกามาใช้ในประเทศไทย จะเกิดข้อบกพร่องในเรื่องใดบ้าง จงอธิบาย"

14. ถามให้อภิปราย เช่น "จงอภิปรายบทบาทของศาสตราจารย์ ดร.ชวาล แพรัตกุล ในการพัฒนาข้อสอบ"

15. ถามให้ระบุจุดมุ่งหมายในการเลือกหรือจัดเรียบเรียงความรู้ เช่น "ทำไมหนังสือเล่มนี้จึงไม่อธิบายรายละเอียดของข้อสอบวัดสติปัญญาแบบรายบุคคล"

16. ถามให้วิจารณ์เพื่อความถูกต้องและมีเหตุผล เช่น "จงวิจารณ์หรือสนับสนุนคำกล่าวที่ว่าข้อสอบอัตนัยไม่เหมาะที่จะใช้กับนักเรียนชั้นประถมศึกษา"

17. ถามเกี่ยวกับโครงร่าง เช่น "จงเสนอโครงร่างของการสร้างข้อสอบแบบครูสร้างขึ้นเอง"

18. ถามให้จัดรวบรวมข้อเท็จจริงใหม่ เช่น "จากการเรียนวิชาวัดผลการศึกษาไปแล้ว ท่านคิดว่าจะนำเอาความรู้หรือข้อเสนอแนะอะไรไปใช้กับวิชาที่ท่านกำลังสอนอยู่บ้าง จงบอกมา 10 ประการ"

19. ถามโดยสร้างปัญหาขึ้นมาใหม่ เช่น "ปัญหาการใช้ข้อสอบอัตนัยที่ควรจะศึกษาเพิ่มเติมมีอะไรบ้าง"

20. ถามให้บอกวิธีการใหม่ เช่น "การเสนอแผนงานการเลื่อนชั้นโดยไม่ต้องสอบ"

ส่วนเวียตแมน (บุญเชิด ภิญโญอนันตพงษ์. 2521 :130-131 ; citing Weidman. 1933) ได้ปรับปรุงรูปแบบคำถามให้รัดกุม และได้จำแนกชนิดคำถามออกเป็น 11 แบบ โดยเรียงลำดับจากความซับซ้อนน้อยไปหามาก ดังนี้

1. คำถามประเภท ใคร อะไร เมื่อไร สิ่งใด และ ที่ไหน
2. ถามให้จัดเรียบเรียงรายการต่าง ๆ
3. ถามให้กำหนดโครงร่าง
4. ถามให้บรรยาย
5. ถามให้เปรียบเทียบความแตกต่าง
6. ถามให้เปรียบเทียบความคล้ายคลึง
7. ถามให้อธิบายหาเหตุผลประกอบ
8. ถามให้อภิปรายอย่างกว้างขวาง
9. ถามให้พัฒนาหรือทำให้ดีขึ้น
10. ถามให้สรุปย่อเรื่อง
11. ถามให้ประเมินค่า

เคอร์ติส (บุญเชิด ภิญโญอนันตพงษ์. 2521 : 131 ; citing Curtis.1943) ได้เสนอรูปแบบคำถามของแบบทดสอบความเรียงขึ้นมา 16 แบบ แต่ส่วนใหญ่ก็มีลักษณะที่คล้ายคลึงกับที่เสนอไว้ก่อนแล้ว ซึ่งมีรายละเอียด ดังนี้

1. ถามให้เปรียบเทียบความคล้ายคลึงและความแตกต่าง
2. ถามให้ตัดสินใจสนับสนุนและคัดค้าน
3. ถามให้ประยุกต์ กฎความจริงหรือหลักการไปสู่สถานการณ์ใหม่
4. ถามให้จัดประเภท
5. ถามหาความสัมพันธ์ที่เกี่ยวข้องกับสาเหตุและผลลัพธ์
6. ถามให้ยกตัวอย่างหรืออธิบายโดยใช้ภาพประกอบ
7. ถามให้ระบุจุดมุ่งหมายในการเลือกหรือการจัดเตรียมสิ่งที่จำเป็น
8. ถามให้วิจารณ์ ความถูกต้อง ความตรงของสถานการณ์ ข้อความหรือแผนภาพ
9. ถามให้อนุมานจากข้อมูล
10. ถามให้อภิปราย
11. ถามเกี่ยวกับโครงร่าง
12. ถามให้อธิบายหรือนิยาม
13. ถามให้ระลึกอย่างง่าย
14. ถามให้สรุป
15. ถามให้สังเกต
16. ถามโดยให้สร้างสถานการณ์ใหม่

เวียสม่า และ เจอร์ส (Wiersma and Jurs. 1985 :168) กล่าวว่าจุดเด่นของแบบทดสอบความเรียงอยู่ที่สามารถวัดผลการเรียนรู้ระดับสูงได้ดี เปิดโอกาสให้นักเรียนรวมรวม , วิเคราะห์ , และสังเคราะห์ความคิด นักเรียนจะต้องพัฒนาความคิดของตนเองและเขียนได้อย่างถูกหลักการ จึงได้เสนอวิธีการเขียนข้อสอบที่สามารถวัดพฤติกรรมการวัดผลการเรียนรู้ขั้นสูง ดังนี้

1. จงรวบรวมความคิดในการป้องกันตัวจากสิ่งใดสิ่งหนึ่ง
2. จงเปรียบเทียบของ A กับ B ในหัวข้อใดหัวข้อหนึ่ง
3. จงหาความแตกต่างของ A กับ B ในหัวข้อใดหัวข้อหนึ่ง
4. จงประเมินค่าของ A กับ B ในหัวข้อใดหัวข้อหนึ่ง
5. จงให้ข้อสรุปจากข้อมูลที่กำหนดให้
6. จงปรับปรุงแผนสำหรับสถานการณ์ที่เป็นปัญหา
7. จงอธิบายประโยชน์ของกฎทางชีววิทยา
8. จงพรรณนา หรือบรรยาย ความสัมพันธ์ของเหตุและผลระหว่าง A กับ B

เมอเรนส์ และ เลห์มานน์ (Mehrens and Lehmann. 1972) ได้เสนอรูปแบบของคำถามของแบบทดสอบความเรียงที่สำคัญ ๆ ไว้ 10 แบบดังนี้

1. ถามให้เปรียบเทียบ
2. ถามให้ตัดสินใจ
3. ถามให้หาสาเหตุหรือผล
4. ถามให้อธิบายความหมายของคำหรือข้อสอบ
5. ถามให้วิเคราะห์
6. ถามให้หาความสัมพันธ์ของคำ หรือประโยค
7. ถามให้อภิปราย
8. ถามให้ระลึกประสบการณ์ความจริง
9. ถามให้สร้างปัญหาขึ้นมาใหม่
10. ถามให้วิพากษ์วิจารณ์

และนักการศึกษาของประเทศไทย นำเสนอลักษณะของคำถามที่ควรใช้ในแบบทดสอบความเรียง ไว้หลายท่าน เช่น

บุญธรรม กิจปรีดาบริสุทธิ์ (2524 : 28) ได้นำเสนอลักษณะคำถามของแบบทดสอบความเรียง ซึ่งโจทย์หรือคำถามจะใช้ข้อความต่าง ๆ เหล่านี้ เช่น อะไร วิจารณ์ อย่างไร สรุปผล เท่าใด เสนอความคิดเห็น ให้คำจำกัดความ ยกตัวอย่างประกอบ เปรียบเทียบ บอกวิธี อภิปราย เรียงลำดับความสำคัญ บรรยาย หาความสัมพันธ์ อธิบาย หาเหตุผล พิสูจน์ และ ประยุกต์ อยู่ในตัวข้อคำถาม

ศิริชัย กาญจนวาสี (2542 : 178-179) เสนอรูปแบบคำถามสำหรับข้อสอบแบบความเรียงที่นิยมใช้ได้แก่

1. จงเปรียบเทียบความคล้ายคลึงและความแตกต่างระหว่าง “....” กับ “....”
2. จงอธิบายความสัมพันธ์ระหว่าง “....” กับ “....”
3. จงเชื่อมโยงความสัมพันธ์เชิงสาเหตุของ “....”
4. จงแสดงความคิดเห็นว่าท่านเห็นด้วยหรือไม่เห็นด้วย พร้อมทั้งเหตุผลเกี่ยวกับ “....”
5. จงสรุปประเด็นสำคัญของ “.....”
6. จงประยุกต์หลักการของ “....” เพื่อแก้ปัญหาเกี่ยวกับ “....”
7. จงวิเคราะห์ผลกระทบของ “....” ต่อ “.....”

8. จงวิเคราะห์จุดเด่นจุดด้อยของ "...."
9. จงสร้างกฎเกณฑ์ทั่วไปจาก "...."
10. จงสร้างระบบจำแนกของ "...."
11. จงสร้างหรือพัฒนาแผน/โครงสร้างเพื่อแก้ปัญหา
12. จงประเมินผลของ "...."

และได้เสนอข้อแนะนำการใช้แบบสอบถามความเรียง ไว้ว่า ควรใช้เมื่อต้องการวัดผลการเรียนรู้ในระดับสูงและซับซ้อน ควรใช้คำถามที่ชัดเจน มากข้อ ถามความสามารถหลายลักษณะสำคัญ ๆ ไม่ควรมีข้อสอบไว้ให้เลือก ควรกำหนดเวลาตอบอย่างเพียงพอ มีการเตรียมคำเฉลยที่ถูกต้องสมบูรณ์พร้อมเกณฑ์การตรวจให้คะแนน ควรตรวจข้อสอบจากกระดาษคำตอบที่เรียงอย่างสุ่มทีละข้อของทุกคนโดยไม่ดูรายชื่อ ควรอ่านคำตอบและประเมินคุณภาพของคำตอบโดยจำแนกเป็นกลุ่ม ๆ เช่น ดี ปานกลาง ยังใช้ไม่ได้ เป็นต้น แล้วตรวจให้คะแนนอย่างละเอียดของแต่ละคนในแต่ละกลุ่ม โดยควรเริ่มจากกลุ่มที่ดีที่สุดไปยังกลุ่มที่อ่อนที่สุด

1.3 การสร้างแบบทดสอบความเรียง

เอกสารที่เกี่ยวกับการสร้างแบบทดสอบความเรียง ได้มีนักวัดผลการศึกษาทั้งในและต่างประเทศนำเสนอไว้หลาย ๆ ท่าน ซึ่งผู้วิจัยได้รวบรวมใน 2 หัวข้อ คือ คำแนะนำในการสร้างแบบทดสอบความเรียง และขั้นตอนการสร้างแบบทดสอบความเรียง

1.3.1 คำแนะนำในการสร้างแบบทดสอบความเรียง

มาแชล และ เฮลส์ (Marshall and Hales. 1972 : 31-32) เสนอแนะแนวทางการวางแผนการเขียนแบบทดสอบความเรียงไว้ดังนี้

1. ให้มีเวลาพอเพียงในการสร้างข้อสอบแบบความเรียง การเขียนข้อสอบที่ดีนั้นต้องการเวลา ก่อนจะนำข้อสอบไปใช้ควรจะมั่นใจว่าข้อสอบนั้นมีคำสั่ง และคำถามชัดเจนไม่คลุมเครือ มีขอบเขตของการตอบและเป็นข้อสอบที่วัดสิ่งที่ครูต้องการวัด

2. ข้อสอบแบบความเรียง จะประกอบด้วยปัญหาที่ต้องการจะให้แก้ไข เนื่องจากข้อสอบความเรียงเหมาะที่จะวัดพฤติกรรมสมองขั้นสูง และเป็นกระบวนการที่เกี่ยวข้องกับการแก้ปัญหา

3. ควรจะนิยามปัญหาให้กระจ่างชัด ข้อสอบแบบความเรียงจะขาดความเที่ยงตรง ถ้าหากนักเรียนไม่สามารถตีความหมายของปัญหา นักเรียนทุกคนจะต้องตีความปัญหาให้ออกมาเหมือนกันเพราะนักเรียนจะต้องตอบ "ข้อสอบเดียวกัน"

4. ควรจะมีขอบเขตของปัญหา เป็นการยากที่นักเรียนจะตอบปัญหาที่คำถามคลุมเนื้อหามากเกินไป การไม่จำกัดขอบเขตปัญหาจะทำให้นักเรียนตอบอะไรก็ได้ที่เขารู้และเกี่ยวข้องกับปัญหานั้น ปัญหาที่ไม่มีขอบเขตจะยั่วยุให้นักเรียนเดาและทำให้การทดสอบมีความเที่ยงตรงต่ำ

5. ควรระบุคำแนะนำในการตอบข้อสอบให้ชัดเจน นักเรียนจะต้องเข้าใจตรงกันว่าข้อสอบแต่ละข้อคะแนนเท่าไร ควรจะตอบข้อสอบแต่ละข้อโดยใช้เวลาเท่าไร ควรจะใช้ข้อมูลอะไรในการตอบ รูปแบบในการเขียนตอบ ควรจะเป็นอย่างไร ควรจะหลีกเลี่ยงคำถามที่นักเรียนต้องเดาว่าครูต้องการอะไร

6. ไม่ควรจะมีข้อเลือกตอบ หากมีข้อเลือกตอบที่มีความสามารถต่างกัน ก็จะเลือกตอบคำถามที่ต่างกัน ตัวครูเองก็มีความโน้มเอียงในการให้คะแนนบางข้อมากกว่าบางข้อ ดังนั้นจึงเป็นการยากที่จะเปรียบเทียบความสามารถของนักเรียนแต่ละคน ยิ่งไปกว่านั้น นักเรียนอาจเลือกดูหนังสือเพียงบางส่วนเพื่อการสอบก็ได้

7. ควรจะทำคำเฉลยไว้ล่วงหน้า หลังจากสร้างคำถามแล้วครูควรจะเขียนคำตอบที่ตนเองเห็นว่าเหมาะสมที่สุด วิธีการนี้จะทำให้ครูสามารถพบข้อบกพร่องของคำถาม เช่น ความคลุมเครือ ความไม่รัดกุมของคำถาม เป็นต้น

กลอนลันด์ (Gronlund. 1977 : 81-83) เสนอกฎการตั้งคำถามแบบทดสอบความเรียงเพื่อให้คุณภาพของข้อสอบมีคุณภาพสูงไว้ดังนี้

1. ให้ใช้แบบทดสอบความเรียงเพื่อวัดผลการเรียนที่ซับซ้อนเท่านั้น
2. เชื่อมโยงคำถามกับผลการเรียนรู้ที่ต้องการวัดโดยตรงที่สุดเท่าที่จะเป็นไปได้
3. สร้างคำถามที่นักเรียนเข้าใจได้อย่างชัดเจน
4. ไม่มีตัวเลือกของคำถามให้นักเรียนเลือกตอบ ยกเว้นว่าผลการเรียนรู้ต้องการให้เป็นไปเช่นนั้น
5. ให้เวลาในการตอบอย่างเพียงพอ และระบุเวลาไว้ในแต่ละคำถามด้วย

ต่อมา ฮอบกินส์ และแอนเตส (Hopkins and Antes. 1990 : 233-234) เสนอข้อแนะนำในการสร้างแบบทดสอบความเรียงไว้ดังนี้

1. ใช้แบบทดสอบความเรียงเมื่อต้องการคำตอบของนักเรียนที่เป็นการอธิบาย การเปรียบเทียบความเหมือน การเปรียบเทียบความแตกต่าง การบรรยาย การวิเคราะห์ การสังเคราะห์หรือประเมินผล
2. ไม่ใช้แบบทดสอบแบบความเรียงในการวัดความจำ
3. ใช้ถ้อยคำในการตั้งคำถามที่ชัดเจนและกระชับ
4. ตั้งคำถามด้วยสถานการณ์ใหม่ ๆ ที่นักเรียนไม่คุ้นเคย ไม่เช่นนี้นักเรียนก็จะจำสิ่งที่เรียนมาแล้วนั้นมาตอบแบบเดิม ๆ
5. กำหนดขอบเขตในการตอบที่ต้องการไว้ในข้อสอบด้วย ขีดจำกัดในการตอบควรกำหนดตามเนื้อหาที่นักเรียนได้เรียน นอกจากนี้ความยากของคำตอบที่ต้องการก็อาจจะกำหนดโดยเวลา จำนวนข้อ หรือปริมาณที่ว่างในกระดาษคำตอบสำหรับเขียนตอบ
6. บอกให้นักเรียนรู้ว่าจะจัดการกับศัพท์เฉพาะหรือศัพท์เทคนิคในเนื้อหาที่จะสอบอย่างไร นักเรียนควรรู้ว่าจะต้องให้คำจำกัดความหรืออธิบายคำเหล่านี้ในการตอบด้วยหรือไม่
7. เขียนข้อสอบเป็นข้อ ๆ เพื่อวัดด้านต่าง ๆ ที่ได้ศึกษามา ไม่ควรออกข้อสอบให้นำข้อมูลทั่วไปมาตอบ
8. ถ้อยคำที่ใช้ในข้อสอบหรือคำสั่งจะต้องมีความหมายชัดเจน
9. ให้นักเรียนทุกคนตอบข้อสอบด้วยชุดข้อสอบชุดเดียวกันและเหมือนกัน
10. เลือกใช้แบบทดสอบแบบความเรียงแบบขยายความ เพื่อวัดทักษะการเขียน การสื่อสารหรือการจัดองค์ประกอบเท่านั้น

วอร์ด (Ward. 1999 : 130) ก็ได้เสนอคำแนะนำสำหรับการเขียนแบบทดสอบความเรียงไว้ 5 รูปแบบ ดังนี้

1. ตั้งคำถามเกี่ยวกับความคิดรวบยอดที่สำคัญ
2. ให้นักเรียนแสดงความรู้ของเขา , ไม่ใช่รายงานข้อมูลที่มีอยู่แล้ว
3. ให้นักเรียนแสดงความสามารถของตนอย่างสมบูรณ์ที่สุดเท่าที่จะทำได้

4. คำถามต้องสั้นและชัดเจนมากกว่าที่จะเป็นคำถามทั่ว ๆ ไป
5. แจ้งหัวข้อที่จะให้คะแนนให้ผู้สอบทราบก่อนทำข้อสอบทุกครั้ง

นักการศึกษาของประเทศไทย เช่น บุญเชิด ภิญโญอนันตพงษ์ (2525 : 38-41) ได้ให้ข้อแนะนำในการสร้างข้อสอบอัตนัยให้มีคุณภาพ ต้องคำนึงถึงจุดอ่อน 2 ประการ คือ ทำอย่างไรจึงจะสร้างคำถามให้สามารถวัดในสิ่งที่ต้องการวัดได้ และ ทำอย่างไรจึงจะทำให้การตรวจข้อสอบอัตนัยมีความเชื่อถือได้มากที่สุด ดังนั้นในการสร้างข้อสอบจะต้องเขียนข้อคำถาม และกำหนดแนวการตอบและเกณฑ์สำหรับตัดสินให้คะแนนอย่างละเอียดด้วย สาเหตุหนึ่งที่ทำให้คะแนนในการตรวจเชื่อถือไม่ค่อยได้ เพราะการละเลยส่วนนี้ไป ดังนั้นการสร้างข้อสอบอัตนัยจึงควรคำนึงถึงข้อแนะนำต่อไปนี้

1. วิเคราะห์จุดประสงค์ของวิชาที่จะออกข้อสอบให้ละเอียดครอบคลุม และเลือกรูปแบบคำถามให้เหมาะสมกับจุดประสงค์นั้น ๆ
2. เขียนคำถามให้ชัดเจนและตรงจุด ระบุให้แน่ชัดว่าต้องการให้ผู้ตอบทำอย่างไร
3. เขียนคำชี้แจงให้ละเอียดและชัดเจน เช่น

<u>คำชี้แจง</u> ข้อสอบมีทั้งหมด 5 ข้อ แต่ละข้อให้คะแนน 20 คะแนน ผู้ตอบต้องแบ่งเวลาในการตอบแต่ละข้อให้เท่า ๆ กัน การให้คะแนนจะพิจารณาถึงความสามารถในการเรียบเรียงความคิด ความถูกต้องของเนื้อหา ตัวสะกดและหลักไวยากรณ์ และความสะอาดเรียบร้อย

4. ควรเขียนข้อสอบแบบจำกัดคำตอบ มากกว่าข้อสอบแบบไม่จำกัดคำตอบ เพื่อให้การตรวจมีความเชื่อมั่นสูงขึ้น
5. หลีกเลี่ยงคำถามประเภทวัดความจำอย่างเดียว เพราะการวัดความจำสามารถใช้ข้อสอบปรนัยอื่น ๆ วัดได้ดีกว่า
6. ควรเขียนข้อสอบให้มากข้อ และหลายรูปแบบ เพื่อให้ครอบคลุมเนื้อหา และสิ่งที่ต้องการจะวัด ซึ่งช่วยให้การวัดมีความเที่ยงตรงตามเนื้อหา (Content Validity) สูงขึ้น
7. ควรกำหนดเวลาที่ใช้ตอบแต่ละข้ออย่างคร่าว ๆ ดังในข้อ 3. การคาดคะเนเวลาที่ควรจะใช้เขียนตอบ จะเป็นการช่วยให้ผู้ตอบสามารถกำหนดขอบเขตในการตอบได้เหมาะสม และสามารถทำข้ออื่น ๆ ได้ทัน
8. ควรเฉลยคำตอบไว้ทั้งคำตอบที่ถูกต้อง และแนวคำตอบอื่น ๆ ที่ควรจะเป็น เพื่อสามารถให้คะแนนได้เชื่อมั่นยิ่งขึ้น และยังเป็นการตรวจสอบดูว่าโจทย์นั้นชัดเจนพอที่จะให้ตอบตามคำเฉลยไว้หรือไม่ อีกทั้งเป็นการคาดคะเนคะแนนและเวลาให้ตอบที่เหมาะสมอีกด้วย
9. ถ้าข้อสอบมีมากข้อ และยากไม่เท่ากัน ควรจัดเรียงลำดับจากง่ายไปหายาก เพื่อยั่วยุให้ผู้ตอบเกิดกำลังใจในการตอบ การให้ข้อยากไว้ตอนต้นอาจทำให้เด็กตกใจหรือขวัญเสีย ซึ่งจะกระทบกระเทือนคะแนนข้ออื่น ๆ ได้ ดังนั้นควรเขียนคำชี้แจงให้นักเรียนเลือกตอบข้อใดก่อนก็ได้ เพื่อให้นักเรียนเลือกทำข้อง่ายตามความรู้สึกของนักเรียนเอง
10. ถ้าเป็นไปได้ ควรออกข้อสอบให้มีความยากใกล้เคียงกัน
11. ให้ผู้อื่นที่มีความรู้ในสาขาวิชาที่ออกข้อสอบ อ่านวิพากษ์วิจารณ์ เพื่อตรวจสอบข้อบกพร่องต่าง ๆ เช่น เขียนคำถามให้ครอบคลุมสิ่งสำคัญที่ต้องการวัดหรือไม่ คำถามยากพอเหมาะหรือไม่

12. หลีกเลี่ยงวิธีการสอบโดยให้เปิดตำราสอบ เพราะการสอบโดยให้เปิดตำราดูได้นี้ เด็กที่เรียนเก่งอยู่แล้วมีความชำนาญในการใช้หนังสือ มีทักษะในการอ่าน มีความคล่องแคล่ว มีความพยายาม และความอดทนสูงกว่าเด็กที่เรียนอ่อน ดังนั้นเด็กเก่งจึงมีแนวโน้มที่จะได้คะแนนสูงขึ้นกว่าปกติ ส่วนเด็กอ่อนจะมีแนวโน้มที่จะได้คะแนนต่ำลงกว่าเดิม เพราะมัวแต่ค้นหาคำตอบจนไม่สามารถจะตอบได้ทันตามกำหนดเวลา

13. ไม่ควรออกข้อสอบให้เลือกตอบเพียงบางข้อ เพราะนักเรียนอาจจะเลือกตอบข้อต่างกัน ซึ่งจะทำให้เกิดความแตกต่างในตัวข้อสอบเอง คะแนนของแต่ละคน จึงขึ้นอยู่กับข้อที่เลือกทำ

เยาวดี วิบูลย์ศรี (2528 : 124) และพวงรัตน์ ทวีรัตน์ (2530 : 113) ให้ข้อควรพิจารณาในการสร้างแบบทดสอบความเรียงไว้คล้าย ๆ กัน ดังนี้

1. ต้องมีการพัฒนาวัตถุประสงค์ของการเรียนการสอนที่ชัดเจน และ พิจารณาว่าวัตถุประสงค์และเนื้อหาของวิชานั้น ๆ หรือเนื้อหาตอนนั้น ๆ เหมาะสมกับแบบทดสอบถามความเรียงหรือไม่

2. จุดมุ่งหมายในการวัดควรให้ผู้ตอบแสดงความรู้ขั้นสูงของบลูม (Bloom,1977) คือ การนำไปใช้ การวิเคราะห์ การสังเคราะห์ การประเมินค่า ซึ่งมีความมุ่งหมายพื้นฐานที่เหมือนกัน คือ ความมุ่งหมายที่จะให้ผู้ตอบแสดงทักษะในการจัดระบบระเบียบความรู้ต่าง ๆ ที่เคยมีอยู่ออกมา

3. พิจารณาว่า ผู้สอบมีพื้นฐานด้านความเรียง และเนื้อหาวิชาเพียงพอที่จะเขียนตอบเป็นความเรียงได้ เช่น ไม่ใช้สอบกับเด็กเล็ก ๆ

4. พิจารณาว่า เวลาที่จะใช้ในการสอบมีเพียงพอหรือไม่ เพราะต้องใช้เวลาในการคิด และเขียนตอบ

ล้วน สายยศ และอังคณา สายยศ (2539 : 86-88) ก็ได้เสนอหลักการในการสร้างข้อคำถามแบบความเรียง ไว้ 5 ข้อ ดังนี้

1. ผู้สร้างควรศึกษาจุดประสงค์ของการสอบก่อน แล้วจึงเขียนข้อคำถาม

2. ควรใช้คำถามที่มีความกระจ่างชัด ด้วยหลักการถามและหลักภาษา ทั้งนี้เพื่อไม่ให้ผู้สอบเข้าใจผิดพลาด หรือมัวตีความหมายของคำถามที่เป็นไปได้หลายแง่หลายมุม

3. คำถามหนึ่ง ๆ ควรเป็นเรื่องเดียว ทั้งนี้เพื่อให้ผู้ตอบสามารถตอบตรงเป้าหมายที่ผู้ถามต้องการ

4. คำถามควรคำนึงถึงเวลาที่ให้ผู้ตอบทำการตอบ

5. คำถามทุกคำถามผู้สอบควรทำเฉลยไว้ และ วางแผนการให้คะแนนแต่ละส่วนว่าเป็นเท่าไรในเวลาที่จำกัดไว้ คำตอบที่สมบูรณ์ที่สุดควรเป็นอย่างไร เพื่อใช้เปรียบเทียบ

1.3.2 ขั้นตอนการสร้างแบบทดสอบความเรียง

การสร้างแบบทดสอบความเรียงให้มีคุณภาพ มีหลักการที่สำคัญ คือ ทำอย่างไรจึงจะให้ข้อสอบหรือข้อคำถามสะท้อนความจริงของสิ่งที่วัด เช่น ความรู้ ความสามารถ ความรู้สึก หรือทักษะที่แท้จริงของผู้ตอบ ซึ่งการสร้างข้อสอบเพื่อสะท้อนความจริงนั้น อยู่ที่การวางแผนและการดำเนินการสร้างที่เป็นระบบ การสร้างข้อสอบที่เป็นระบบต้องอาศัยผู้สร้างที่มีความรู้ในเนื้อหา ตลอดจนการวางแผนการออกข้อสอบที่ดี เขียนข้อสอบที่ตรง ตลอดจนตรวจสอบคุณภาพของข้อสอบแต่ละข้อ (อุทุมพร จามรมาน. 2535 : 29) จากการศึกษาขั้นตอนการสร้างข้อสอบความเรียงจาก เยาวดี วิบูลย์ศรี (2528 : 127-135) สรุปขั้นตอนการสร้างแบบทดสอบความเรียงได้ ดังนี้

ขั้นที่ 1 ตัดสินใจเลือกข้อมูลต่าง ๆ ที่ต้องการโดยพิจารณาด้วยว่าสิ่งที่กำลังวัดนั้นเหมาะกับข้อสอบความเรียงหรือไม่ เช่น ระบุจุดมุ่งหมายในการทดสอบ ระบุเนื้อหาให้ชัดเจน ทำตารางเนื้อหากับจุดมุ่งหมายในการทดสอบ กำหนดน้ำหนักของข้อสอบ จำนวนข้อ

ขั้นที่ 2 พิจารณาความเป็นอิสระในการตอบที่ยอมให้ ในการตอบข้อสอบความเรียงนี้ มีการกำหนดขอบเขตการตอบซึ่งอาจทำโดยการจำกัดเนื้อหา จำกัดความยาว จำกัดเวลา จำกัดทรัพยากรที่ใช้ เช่น ให้อภิปรายเฉพาะเหตุผลในตำรา หรือจำกัดรูปแบบการตอบ เช่น ให้บอกรายการ เขียนแผนภูมิ เขียนโครงร่าง

ขั้นที่ 3 กำหนดรูปแบบหรือลักษณะของข้อคำถาม กำหนดเวลาสอบ กำหนดคะแนน เขียนข้อคำถามเฉลยคำตอบ เพื่อประโยชน์ในการตัดสินให้คะแนน

ขั้นที่ 4 ตรวจสอบข้อสอบที่เขียนขึ้น ทดลองใช้ แก้ไขปรับปรุง

เยาวดี วิบูลย์ศรี (2528 : 124-135) ได้กล่าวถึงแนวทางในการเขียนแบบทดสอบความเรียงที่ดีจะต้องใช้วัดผลการเรียนรู้ขั้นสูงของบลูมได้ใน 4 พฤติกรรมดังนี้

<u>การนำไปใช้</u> (Application)

การวัดในระดับนี้ หมายถึง การนำความรู้ที่เป็นนามธรรม มาใช้แก้ปัญหาในสถานการณ์ที่เป็นรูปธรรม ความรู้นามธรรมนั้น อาจอยู่ในรูปแนวคิด กฎต่าง ๆ แบบแผนทั่วไป หรืออาจจะเป็นหลักการที่เกี่ยวกับเทคนิคต่าง ๆ ความคิดทางทฤษฎี ฯลฯ ดังนั้น คำถามที่จะวัดในระดับนี้จึงต้องเป็นคำถามที่จะให้ผู้สอบได้นำความรู้ต่าง ๆ มาประยุกต์กับสถานการณ์ให้เหมาะสม การกำหนดปัญหาจึงควรเป็นปัญหาใหม่ ๆ แต่ก็มีความสัมพันธ์กับความรู้ที่เรียนมา

การเขียนข้อคำถาม เพื่อให้นักเรียนได้ใช้ความรู้ที่เคยเรียนมาไปบรรยายถึงการแก้ปัญหาในสถานการณ์ที่เป็นรูปธรรมมีข้อควรคำนึงถึงอยู่ 3 ประการ คือ (1) ข้อคำถามที่วัดต้องเสนอในรูปสถานการณ์ที่เป็นรูปธรรม และจะต้องเป็นสถานการณ์ที่นักเรียนอาจประสบด้วยตนเองได้ หรือ มีส่วนเกี่ยวข้องกับสถานการณ์นั้น ๆ ได้ (2) สถานการณ์ที่เสนอขึ้นจำต้องมีพฤติกรรมบางอย่างให้แสดงเพื่อที่จะให้ปัญหาที่กำหนดขึ้น สามารถแก้ให้ลุล่วงไปได้ (3) พฤติกรรมที่จะให้แสดงนั้นควรตั้งอยู่บนพื้นฐาน ของการถ่ายทอดความรู้ที่เคยเรียนมาแล้ว

<u>การวิเคราะห์</u> (Analysis)

การวัดในระดับการวิเคราะห์ หมายถึง ความสามารถในการแยกแยะองค์ประกอบ หรือ ส่วนย่อยต่าง ๆ หรือแยกส่วนของแนวความคิด หรือแยกส่วนของวัตถุ (Objects) ตลอดจนการหาความสัมพันธ์ระหว่างส่วนต่าง ๆ หรือชี้ให้เห็นถึงการจัดระเบียบของหลักการต่าง ๆ เป็นต้น

การเขียนข้อคำถาม การเขียนข้อคำถามในการวัดการวิเคราะห์ คล้ายกับการเขียนข้อคำถามในการวัดการนำไปใช้ แต่ที่แตกต่างกัน คือ ข้อคำถามที่วัดการวิเคราะห์นั้น จะไม่มีส่วนที่เป็นปัญหา ส่วนที่เป็นสถานการณ์จะเป็นเรื่องที่นักเรียนคุ้นเคย แต่ไม่มีการถกเถียงในชั้นเรียนมาก่อนและสถานการณ์นั้น ๆ ก็จำเป็นต้องมีองค์ประกอบต่าง ๆ หรือส่วนย่อยต่าง ๆ ที่มีความสัมพันธ์กัน หรือ การจัดระเบียบของหลักการต่าง ๆ รวมเป็นสถานการณ์ซึ่งสามารถที่จะวิเคราะห์ได้ ดังนั้นข้อคำถามที่ถามถึง จึงมักจะให้แสดงการเปรียบเทียบข้อที่เหมือนกัน หรือต่างกัน

<u>การสังเคราะห์</u> (Synthesis)

การวัดในระดับการสังเคราะห์ หมายถึง ความสามารถในการรวมส่วนย่อยเข้าด้วยกัน ให้เป็นเรื่องเดียวกัน จัดเป็นระดับของการใช้สติปัญญาที่จะนำไปสู่พฤติกรรมที่สร้างสรรค์ของผู้เรียน แต่เป็นพฤติกรรมสร้างองค์ประกอบต่าง ๆ หรือส่วนย่อยต่าง ๆ ที่มีความสัมพันธ์กัน หรือการจัดระเบียบของหลักการต่าง ๆ

รวมเป็นสถานการณ์ซึ่งสามารถที่จะวิเคราะห์ได้ ดังนั้นข้อคำถามที่ถามถึง จึงมักจะให้แสดงการเปรียบเทียบข้อที่เหมือนกัน หรือต่างกัน

การสังเคราะห์ (Synthesis)

การวัดในระดับการสังเคราะห์ หมายถึง ความสามารถในการรวมส่วนย่อยเข้าด้วยกัน ให้เป็นเรื่องเดียวกัน จัดเป็นระดับของการใช้สติปัญญาที่จะนำไปสู่พฤติกรรมที่สร้างสรรค์ของผู้เรียน แต่เป็นพฤติกรรมสร้างสรรค์ที่ไม่สมบูรณ์ เพราะการสังเคราะห์นั้นยังต้องอยู่ในขอบเขตของปัญหาที่จำกัด หรืออยู่ภายในกรอบของทฤษฎี หรือแบบแผนวิธีการบางส่วนเท่านั้น บลูม (Bloom. 1977) ได้เสนอแนะแนวการสังเคราะห์ไว้ 3 ลักษณะ คือ (1) การสังเคราะห์เรื่องที่มีสาระเดียวกันหลายเรื่องให้เป็นเรื่องเดียวกัน (2) การสังเคราะห์แผนการหรือข้อเสนอ ให้เป็นชุดการปฏิบัติงานชิ้นหนึ่ง (3) การสังเคราะห์สิ่งที่มีความสัมพันธ์เชิงนามธรรม เช่น ทฤษฎีหนึ่ง เป็นต้น

การเขียนข้อคำถาม ในการวัดการสังเคราะห์ ต้องไม่มีข้อแนะนำรายละเอียดของสถานการณ์มากเกินไป เพราะการจะให้เกิดความคิดสร้างสรรค์นั้นผู้ตอบควรมีอิสระที่จะทำกิจกรรมต่าง ๆ รวมทั้งอิสระในการกำหนดเป้าหมาย และตัดสินคุณค่าว่าสิ่งใดหรือองค์ประกอบใดที่ทำให้เกิดผลิตผล แสดงถึงการสังเคราะห์ได้ ข้อคำถามควรเป็นปัญหาที่แปลกใหม่ ที่ผู้ตอบไม่เคยมีประสบการณ์ในชั้นเรียนมาก่อน เช่น การกำหนดหลักเกณฑ์ ทฤษฎี ออกแบบเครื่องมือ หรือ แผนการปฏิบัติ

การประเมินค่า (Evaluation)

การวัดในระดับการประเมินค่า หมายถึง การตัดสินคุณค่า (Value) สำหรับเป้าหมาย แนวความคิด แนวการแก้ปัญหา วิธีการ หรือ ผลงาน เป็นต้น การประเมินค่านั้นรวมถึงเกณฑ์หรือมาตรฐานที่ใช้ด้วย โดยบลูม (Bloom, 1977) ได้กล่าวถึง เกณฑ์ 2 ประเภท ในการประเมิน คือ (1) การประเมินซึ่งใช้เกณฑ์มาตรฐานภายใน (Internal Standards) เช่น การประเมินตามเกณฑ์ ในลักษณะของความถี่ (Consistency) ความแม่นยำในการใช้เหตุผล รวมทั้งเกณฑ์ในลักษณะที่เป็นแนวคิดของบุคคล เป็นการพิจารณาตัดสินกระบวนการก่อนที่งานนั้นจะบรรลุถึงจุดหมาย เช่น การบ่งชี้ถึงความถูกต้องของข้อโต้แย้งอย่างมีเหตุผล (2) การประเมินซึ่งใช้เกณฑ์มาตรฐานภายนอก (External Standards) เช่น การประเมินตามเกณฑ์ในลักษณะของความมีประสิทธิภาพ (Efficiency) ความประหยัด (Economy) ประโยชน์ที่ได้รับ (Utility of Means for Ends) รวมทั้งความดีเยี่ยมตามเกณฑ์มาตรฐานที่กำหนด เป็นการพิจารณาตัดสินเมื่องานนั้นสำเร็จลุล่วงแล้ว เช่น การเปรียบเทียบทฤษฎีที่เกี่ยวกับวัฒนธรรมบางอย่าง

การเขียนข้อคำถาม ข้อคำถามในการวัดระดับการประเมินจะประกอบด้วย 2 ส่วน คือ (1) ส่วนที่แสดงถึงสิ่งที่จะประเมิน ซึ่งควรมีหลักเหตุผลในการตรวจสอบได้ คือสามารถที่จะประเมินได้ เป็นต้นว่า บทกวี การบริหารขององค์การ นอกจากนี้ ก็มีการตั้งคำถาม เพื่อให้ผู้สอบได้พิจารณาว่า จะเลือกใช้เกณฑ์ใดที่เหมาะสมสำหรับการประเมินนั้น ซึ่งปกติแล้ว เกณฑ์ที่ใช้ในการประเมินนั้น อาจมีบางเกณฑ์ที่เคยเรียนมาแล้ว หรือบางเกณฑ์ต้องคิดขึ้นเอง (2) ส่วนของการชี้แจงคำตอบ ซึ่งจำเป็นต้องมี เช่น ความยาวของคำตอบ ตลอดจนรายละเอียดของคำตอบที่ต้องการ โดยทั่วไปแล้ว ลักษณะเฉพาะของการประเมินจะต้องมีความรับผิดชอบซึ่งแตกต่างจากการตัดสินด้วยความเห็นตามอำเภอใจ โดยการประเมินต้องอธิบายให้เห็นความสัมพันธ์ของวัตถุประสงค์กับเกณฑ์ ดังนั้น การชี้แจงคำตอบจึงเป็นส่วนที่จะให้ข้อมูลเพื่อให้ผู้ประเมินทราบวัตถุประสงค์ของการประเมินที่ชัดเจน ตลอดจนเกณฑ์ที่ใช้ในการประเมินด้วย

1.4 วิธีการตรวจให้คะแนนแบบทดสอบความเรียง

นักวัดผลการศึกษาหลายท่านได้ให้ความหมายและวิธีการตรวจให้คะแนนแบบทดสอบความเรียง วิธีวิเคราะห์ย่อย และวิธีประเมินรวม ไว้มากมาย เช่น

อีเบล (Ebel. 1972 : 149-152) ได้ให้ข้อเสนอแนะสำหรับการตรวจให้คะแนนไว้ 2 รูปแบบ คือ การให้คะแนนด้วยวิธีวิเคราะห์ และวิธีการประเมินคุณภาพโดยรวม (Analytic Scoring or Global-quality Scaling)

1.1 วิธีวิเคราะห์ วิธีนี้ส่วนประกอบที่สำคัญต่าง ๆ ของคำตอบจะเป็นตัวกำหนดการให้คะแนนมากน้อยอย่างอิสระ การพิจารณาจะพิจารณาทั้งส่วนประกอบที่สำคัญของคำตอบ และความสัมพันธ์ระหว่างส่วนประกอบเหล่านั้นด้วยว่าเป็นองค์ประกอบเดียวกันของคำตอบหรือไม่ แต่ถ้าความสัมพันธ์นั้นซับซ้อนและเข้าใจยาก วิธีการเทียบกับเกณฑ์ก็อาจจะทำได้ไม่สะดวกและใช้เวลานาน

1.2 วิธีประเมินคุณภาพโดยรวม วิธีนี้ผู้ตรวจให้คะแนนจะอ่านคำตอบโดยรวม และพิจารณาความประทับใจ แล้วจึงเปลี่ยนความประทับใจนั้นเป็นระดับคะแนนบันทึกไว้ ทำแบบนี้ทุก ๆ ข้อ วิธีการที่ดีกว่า เพื่อควบคุมความสม่ำเสมอ หรือความมีมาตรฐานในการตรวจกระดาษคำตอบที่ต่างกัน คือ การแบ่งกลุ่มคำตอบเป็นหลาย ๆ กองตามระดับคุณภาพที่ต่างกัน โดยในแต่ละกองจะมีลักษณะคำตอบที่คล้าย ๆ กัน เมื่อแบ่งกลุ่มแล้ว ผู้ตรวจให้คะแนนจึงมาพิจารณาตัดสินใหม่อีกครั้งว่าแต่ละกองเหมาะสมหรือยัง

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 ควรปิดบังชื่อของนักเรียนที่ทำข้อสอบไม่ให้ผู้ตรวจให้คะแนนทราบ
4. ถ้าเป็นไปได้ ควรจัดให้มีการตรวจให้คะแนนอย่างอิสระ

เมอร์เรน และเลห์มาน์น (Mehrens and Lehmann. 1973 : 229-238) ก็ได้อธิบายถึงการตรวจให้คะแนนด้วยวิธีประเมินรวม (Holistic Method)ว่า วิธีนี้คำตอบจะไม่ถูกแบ่งเป็นส่วน ๆ แต่ผู้ตรวจจะอ่านคำตอบอย่างรวดเร็วแล้วใช้ความประทับใจและใช้มาตรฐานบางอย่างกำหนดระดับของคำตอบ การตรวจคำตอบจะขึ้นอยู่กับระดับของการแบ่ง อาจจะแบ่งข้อสอบเป็น 2 กลุ่ม คือ "กลุ่มที่ยอมรับได้ – กลุ่มที่ยอมรับไม่ได้" หรือ 5 กลุ่ม คือ "ดีมากจนถึงต่ำกว่ามาตรฐาน" โดยมากจะแบ่งประมาณ 4 หรือ 5 กลุ่ม และอธิบายถึงการตรวจโดยวิธีวิเคราะห์ย่อย (Analytical Method) ว่าการให้คะแนนวิธีวิเคราะห์เป็นวิธีที่มีรูปแบบคำตอบประกอบด้วยประเด็นเฉพาะที่กำหนดไว้ก่อนแล้ว คะแนนของนักเรียนที่ได้จะขึ้นอยู่กับจำนวนประเด็นที่เขาตอบ รวมไปถึงส่วนอื่น ๆ เช่น แสดงความคิดเห็นได้ชัดเจน การให้เหตุผล และการยกตัวอย่างสนับสนุน

ในประเด็นคำตอบ และการกำหนดคะแนนในแต่ละประเด็นจะขึ้นอยู่กับเวลาที่ใช้ในการตอบ ความซับซ้อนของคำถาม และเนื้อหาที่ครูสอน

วิทท์เน่ย์ และ เซเบอร์ (วัญญา วิศาลาภรณ์. 2542 :16 ; citing Whitnay and Sabers. 1976 : 5) ได้นำเสนอการตรวจให้คะแนนข้อสอบแบบความเรียงหรือแบบอัตนัย จำแนกเป็น 2 วิธี คือ วิธีเทียบกับเกณฑ์กับวิธีจัดอันดับคุณภาพ ซึ่งมีรายละเอียด ดังนี้

1. การตรวจโดยวิธีเทียบกับเกณฑ์ (Analytical Meted หรือ Point Method) การตรวจให้คะแนนวิธีนี้ครูจะต้องกำหนดแนวการตอบไว้ล่วงหน้า โดยแยกแนวคำตอบออกเป็นตอนย่อย ๆ ตามความสำคัญ แล้วกำหนดคะแนนเต็มของแต่ละตอนย่อย ๆ หรือครูอาจแยกแนวคำตอบออกเป็นตอนย่อย ๆ ตามองค์ประกอบของการตอบ เช่น การจัดเรียงความคิด หลักฐานที่ยกมาอ้าง ตัวอย่างประกอบ เป็นต้น จากนั้นก็กำหนดคะแนนเต็มของแต่ละตอนย่อย เมื่อครูอ่านข้อสอบของนักเรียนก็จะให้คะแนนแต่ละตอนย่อย ๆ มารวมกันเป็นคะแนนที่ได้รับทั้งข้อ

2. การตรวจโดยวิธีการจัดอันดับคุณภาพ (Rating Method หรือ Holistic Method หรือ Scoring Method หรือ Global Scoring) การตรวจให้คะแนนวิธีนี้ ครูจะอ่านคำตอบของนักเรียนทุกคนทีละคน เมื่ออ่านกระดาษคำตอบแล้วก็จะแยกกระดาษคำตอบเป็นกลุ่มหรือเป็นกอง ตามระดับคุณภาพของการตอบ เช่น แยกกระดาษคำตอบออกเป็น 5 กอง ดังเช่น ดีมาก ดี พอใช้ เกือบพอใช้ อ่อน เมื่อครูอ่านกระดาษคำตอบแล้วก็จะจัดเข้ากองใดกองหนึ่งในห้ากองนี้ หลังจากนั้นครูก็จะพิจารณากระดาษคำตอบในแต่ละกอง โดยพิจารณาว่าใครตอบดีกว่ากัน แล้วเรียงกระดาษคำตอบตามลำดับของคุณภาพ แล้วให้คะแนนตามลำดับของคุณภาพอีกทีหนึ่ง

กลอนลันด์ (Gronlund. 1977 : 83-85) ได้สรุปกฎของการให้คะแนนแบบทดสอบความเรียง ไว้ 6 ข้อ ดังนี้

1. ประเมินคำตอบในแง่ของผลการเรียนรู้ที่ต้องการวัด
2. ให้คะแนนคำตอบที่ต้องการการตอบที่จำกัด โดยใช้วิธีการให้คะแนนเป็นประเด็น ๆ โดยนำคำตอบต้นแบบมาเทียบ
3. ให้เกรดกับคำตอบโดยใช้วิธีการจัดอันดับ โดยนำเกณฑ์ที่กำหนดไว้มาเทียบ
4. ประเมินคำตอบทั้งหมดของนักเรียนต่อหนึ่งคำถาม ก่อนที่จะตรวจคำตอบข้อต่อไป
5. ประเมินคำตอบโดยไม่รู้ว่าผู้ตอบเป็นใคร
6. ถ้าเป็นไปได้ ควรมีผู้ตรวจให้คะแนน 2 คนหรือมากกว่า สำหรับแต่ละคำตอบ

เวียสม่า และ เจอร์ส (Wiersma and Jurs. 1985 :175-177) กล่าวถึงการให้คะแนนแบบทดสอบความเรียง ว่าเป็นแบบทดสอบที่มีความเคร่งครัดในการให้เกณฑ์การให้คะแนนอย่างมาก เพราะว่าเป็นการยากที่จะให้คะแนน กล่าวคือ จะมีความเชื่อถือได้ยาก ซึ่งมีจุดสำคัญอยู่ที่กระบวนการให้คะแนนของผู้ตรวจ วิธีการนี้ก็จะช่วยเพิ่มความเป็นปรนัยในการให้คะแนนเพิ่มขึ้น อย่างไรก็ตามก็ต้องอาศัยเวลาที่มากขึ้นในการตรวจ ที่ก่อนจะให้คะแนนครูจะต้องมีวิธีการหรือกำหนดกรอบของคำตอบไว้ล่วงหน้า เรียกว่า โมเดลคำตอบ ซึ่งเป็นสิ่งจำเป็นที่จะเป็นการกำหนดค่าให้ประเด็นสำคัญ ๆ ที่จะทำให้การตรวจแตกต่างกันน้อยที่สุด วิธีการให้คะแนน มี 2 วิธี คือ

1. การให้คะแนนวิธีวิเคราะห์ย่อย (Analysis Scaling) จุดสำคัญของการให้คะแนนคำตอบอยู่ที่การจำแนกและกำหนดขอบเขตของคำตอบเป็นรายละเอียด มีความชัดเจน มาก จึงง่ายต่อการตรวจและมีความเชื่อถือได้

2. การให้คะแนนวิธีการให้คุณภาพโดยรวม หรือวิธีประเมินรวม (Holistic Scaling) เป็นวิธีที่อาศัยความประทับใจกับคำตอบ คำตอบจะถูกมองในภาพรวมมากกว่าจะแยกออกมาเป็นส่วยย่อย โดยจะแยกกระดาษคำตอบออกเป็น 2 กองหรือมากกว่า เช่น กองดีมาก กองดี กองพอใช้ กองแย่มาก การตรวจจะรวดเร็วและง่ายกว่า แต่ทำให้ไม่มีความเป็นปรนัย มีความเชื่อถือได้น้อย

ฮอบกินส์ และแอนเตส (Hopkins and Antes. 1990 : 237-238) กล่าวว่า มีนักการศึกษาหลายท่านเสนอวิธีการตรวจไว้ 2 วิธีให้เลือกใช้ตามความเหมาะสม สรุปได้ดังนี้

1. วิธีการวิเคราะห์ย่อย (Analytical Procedure) หลังจากที่ออกข้อสอบแบบความเรียงแล้ว ครูจะดำเนินการเขียนคำตอบต้นแบบ (Model Response) ทันที เพื่อให้มีโครงร่างขององค์ประกอบหลักซึ่งจะระบุไว้ในกระบวนการให้คะแนน ในขั้นตอนการตอบครูจะพบว่าหัวข้อที่จะให้นักเรียนทำนั้นจำเป็นต้องมีการเปลี่ยนแปลงหรือเขียนใหม่เพื่อให้ได้คำตอบที่ครูคาดหวัง การพัฒนาแนวทางในการให้คะแนนแบบทดสอบแบบความเรียงช่วยให้คะแนนมีความคงที่มากขึ้นโดยผ่านการประเมินของครูในแต่ละประเด็น การจัดองค์ประกอบเฉพาะที่คาดหวังจะช่วยผู้ให้คะแนนกำหนดค่าคะแนนหรือเกรดได้ ปัจจัยเพิ่มเติมอื่น ๆ ที่อาจจะนำมาใช้ในการตรวจข้อสอบ อาทิเช่น การจัดระเบียบในการเขียน , วิธีที่ใช้เขียนคำตอบ , ความถูกต้อง , ความสมบูรณ์ของคำตอบ , ความถูกต้องในการใช้ถ้อยคำ และลักษณะอื่น ๆ ที่ครูคิดว่ามีความสำคัญ ปัจจัยหรือลักษณะที่แตกต่างกันของคำตอบที่คาดหวัง อาจจะเขียนไว้ในแบบโครงร่างโดยมีคะแนนที่จะให้กับปัจจัยแต่ละอย่างในแง่ของคุณภาพของคำตอบ การรวมคะแนนทั้งหมดควรจะลดบทบาทของข้อมูลที่ไม่เกี่ยวข้องในคำตอบที่จะมีต่อผลรวมของคะแนน ส่วนใหญ่แล้วพบว่าวิธีการให้คะแนนแบบนี้ใช้เวลานาน และสร้างความเหนื่อยล้าถ้าหากข้อสอบมีจำนวนมาก อย่างไรก็ตามการให้คะแนนแบบวิเคราะห์ทำให้ง่ายต่อการถกเถียงเมื่อคืนข้อสอบให้กับนักเรียน คะแนนขั้นสุดท้ายสามารถจะดูความถูกต้องได้โดยพิจารณาว่าคำตอบนั้นตรงกับสิ่งที่คาดหวังไว้ได้ดีแค่ไหน

2. วิธีการประเมินภาพโดยรวม (Global Procedure) วิธีการนี้จะตรวจคำตอบ โดยอาศัยความประทับใจโดยภาพรวม ไม่มีการระบุประเด็นที่จะต้องมีในคำตอบไว้ล่วงหน้าเป็นการเฉพาะ วิธีการนี้ต้องอาศัยการอ่านข้อสอบในครั้งแรกอย่างรวดเร็วเพื่อสร้างความประทับใจอย่างกว้าง ๆ ต่อคำตอบแต่ละคำตอบ ทดลองแบ่งข้อสอบออกเป็นกลุ่ม ๆ ตามคุณภาพของคำตอบที่แตกต่างกันไป จำนวนกลุ่มนั้นไม่แน่นอน ขึ้นอยู่กับความต้องการของผู้ให้คะแนนและระดับการแยกความแตกต่างที่ต้องการ โดยทั่วไปแล้วผู้ให้คะแนนจะอ่านจนกระทั่งมีคำตอบหนึ่งที่ได้รับการประเมินว่าเหนือกว่าอย่างแน่นอน และใช้คำตอบนี้เป็นคะแนนอ้างอิงในการให้คะแนนคำตอบอื่น ๆ การอ่านอย่างรอบคอบขึ้นในรอบที่สองสามารถทำให้เกิดการเปลี่ยนแปลงจากกลุ่มที่กำหนดไว้ได้ กล่าวคือ การจัดกลุ่มคำตอบของนักเรียนอาจจะได้รับการกลั่นกรองอีกครั้งถ้าผู้ตรวจรู้สึกว่าการกำหนดประเภทขั้นแรกผิดพลาด ในขั้นนี้ข้อสอบที่แยกไว้ในแต่ละกลุ่มควรจะอยู่ในระดับเดียวกัน ถ้าข้อสอบฉบับใดมีความโดดเด่นกว่าฉบับอื่น ๆ ในกลุ่มก็ควรจะแยกไปอยู่ในกลุ่มที่ดีกว่า หรือถ้าแย่กว่าก็แยกไปกลุ่มแย่กว่า การแยกกลุ่มก่อนที่จะให้คะแนนนี้จะลดความเป็นไปได้ที่ว่าข้อสอบที่อยู่ในระดับเดียวกันได้คะแนนขั้นสุดท้ายแตกต่างกัน หลังจากที่ผู้อ่านพอใจกับการแยกกลุ่มอย่างเหมาะสมแล้ว ก็จะมีการให้คะแนนขั้นสุดท้ายหรือการประเมินอื่น ๆ การตรวจแบบทดสอบความเรียงโดยภาพรวมนี้จะมีประสิทธิภาพสำหรับข้อสอบที่มีจำนวนมาก อย่างไรก็ตามแม้ว่าวิธีการนี้จะทำให้การให้คะแนนมีความเชื่อมั่นสูง แต่ก็ยากที่จะอธิบายเหตุผล

ในการให้คะแนนให้นักเรียนฟังเมื่อคืนข้อสอบให้นักเรียนแล้ว คำตอบที่เป็นแบบอย่างควรจะนำไปอ่านให้นักเรียนในชั้นฟังเพื่อเป็นคะแนนอ้างอิงให้กับนักเรียนแต่ละคน

เกย์ (Gay. 1991 : 226-227) กล่าวถึงการตรวจให้คะแนนแบบทดสอบความเรียง เป็นกระบวนการที่ยากและใช้เวลานาน เราสามารถลดอัตวิสัยและความเชื่อมั่นของผู้ให้คะแนนแบบทดสอบความเรียงที่ต่ำได้โดยการวางแผนและการให้คะแนนอย่างรอบคอบ ซึ่งแบ่งเป็น 2 วิธี คือ วิธีวิเคราะห์ย่อย (Analytical Approach) และวิธีดูภาพโดยรวม (Global Approach)

วิธีวิเคราะห์ย่อย (Analytical Approach) คือ การระบุแง่มุมหรือองค์ประกอบทั้งหมดของคำตอบที่สมบูรณ์ และกำหนดคะแนนให้แต่ละส่วน คำตอบที่สมบูรณ์สำหรับส่วนต่าง ๆ อาจจำเป็นต้องคำนวณหาค่าเฉลี่ยมาเพียง 1 ค่า หรือมากกว่านั้น โดยมีวิธีการคือ คำตอบจะถูกนำมาอ่านเป็นรอบที่สอง และให้คะแนน คะแนนจากการอ่านทั้ง 2 รอบ จะถูกนำมาเปรียบเทียบกันถ้ามีความแตกต่างกันก็อาจจะอ่านใหม่ หรือ นำคะแนนจากการอ่าน 2 รอบนั้นมาเฉลี่ย การทำตารางข้อกำหนดเฉพาะในขั้นต้นจะช่วยให้การใช้วิธีแบบวิเคราะห์นี้สะดวกขึ้น

วิธีดูภาพโดยรวม (Global Approach) หรืออาจเรียกว่าวิธีการจัดอันดับ จะมีการให้คะแนนที่มีภาวะอัตวิสัยมากกว่า และมีการให้คะแนนที่เชื่อถือได้น้อยกว่า แต่เนื่องจากใช้เวลาน้อยกว่า จึงนิยมนำวิธีนี้มาใช้บ่อย เมื่อมีจำนวนนักเรียนที่สอบมาก วิธีนี้จะมีการระบุถึงแง่มุมและองค์ประกอบทั้งหมดของคำตอบที่สมบูรณ์ แต่ไม่ต้องกำหนดคะแนนให้แต่ละส่วน แต่ละคำตอบจะได้รับการตรวจให้คะแนนโดยดูภาพรวม แล้วทำการจัดอันดับจากนั้นจึงทำการอ่านซ้ำและให้คะแนนใหม่

ออสเทอร์ฮอฟ (Oosterhof. 1996 : 77-78) ก็ได้ทำการเปรียบเทียบการให้คะแนนวิธีวิเคราะห์ย่อยกับวิธีมองภาพโดยรวม ไว้ว่า การให้คะแนนวิธีวิเคราะห์ย่อยจะใช้แผนการให้คะแนนโดยละเอียดเพื่อประเมินคำตอบ แผนการให้คะแนนนั้นจะระบุถึงลักษณะเฉพาะที่ใช้ตรวจคำตอบของนักเรียน และบ่งบอกจำนวนคะแนนมาพร้อม ๆ กับลักษณะเฉพาะแต่ละอย่าง เหมือนเป็นรายการสำหรับตรวจสอบ (Checklist) ให้ผู้อ่านดูว่าคำตอบของนักเรียนมีส่วนใดบ้างที่ตรงกับลักษณะเฉพาะที่กำหนดไว้

ถ้าคำถามมีความซับซ้อนขึ้นและมีความยืดหยุ่นในการตอบมากขึ้น ก็มักจะใช้วิธีมองภาพโดยรวมในการตรวจ วิธีนี้คือการอ่านคำตอบทั้งหมดแล้วให้คะแนนตามคุณภาพโดยรวม วิธีการโดยทั่วไปก็คือ การแบ่งกระดาษคำตอบของนักเรียนออกเป็น 3 กลุ่ม คือ ดีมาก ดีปานกลาง และดีพอใช้ จากนั้นก็อ่านคำตอบในแต่ละกลุ่มอีกรอบ เพื่อแบ่งย่อยลงไปอีก จำนวนสุดท้ายของกลุ่มที่แบ่งออกมาจะสอดคล้องกับจำนวนคะแนนที่ให้

ข้อดีของวิธีแบบดูโดยรวม คือ 1. ค่อนข้างเร็ว 2. ใช้กับข้อสอบที่มีคำตอบที่ไม่สามารถแบ่งออกเป็นส่วน ๆ ได้ ข้อเสียของวิธีนี้คือ 1. คะแนนที่ได้มักมีความเชื่อถือได้น้อยกว่า 2. มีข้อมูลจำกัดที่จะอธิบายกับนักเรียนว่าทำไมจึงได้คะแนนเท่านั้นเท่านี้ ข้อเสียแรกอาจแก้ได้โดยการให้มีผู้อ่าน 2 คน มาเฉลี่ย

กระบวนการอีกอย่างหนึ่งในการปรับปรุงการให้คะแนนวิธีมองภาพโดยรวมก็คือสร้างคำตอบต้นแบบ (Model Answer) สำหรับแต่ละคำถาม ครูจะอ่านคำตอบต้นแบบของนักเรียนบางส่วนก่อนตัดสินใจที่จะให้คะแนน

กระบวนการสุดท้ายในการปรับปรุงวิธีการนี้ ก็คือ เกณฑ์ในการให้คะแนน ซึ่งจะประกอบด้วยข้อบรรยายที่แตกต่างกันหลายข้อ แต่ละข้อใช้กับระดับคุณภาพที่แตกต่างกัน ตั้งแต่การตอบที่แย่จนถึงการตอบที่ดีเยี่ยม เมื่อทำการให้คะแนน คำตอบของนักเรียนจะถูกนำไปเทียบหาข้อบรรยายภายในรูบริกที่ตรงกับคุณภาพของคำตอบของนักเรียนมากที่สุด โดยทั่วไปจะมีการกำหนดตัวเลขให้กับข้อบรรยายต่าง ๆ ในเกณฑ์

การให้คะแนน เช่น ถ้ามี 6 ข้อ ก็จะกำหนดข้อ 1-6 โดยที่ข้อ 1 เป็นข้อบรรยายสำหรับคำตอบที่แย่ และข้อ 6 สำหรับคำตอบที่เยี่ยม

ทัคแมน (Kubiszyn and Borich. 1996 : 112 ; citing Tuckman. 1975) ระบุถึงเกณฑ์การให้คะแนนสำหรับข้อสอบแบบความเรียงระดับสูง ว่ามีส่วนประกอบ 3 ส่วนที่มีประโยชน์ในการประเมินข้อสอบแบบความเรียงระดับสูง คือ

1. เนื้อหา (Content) ถึงแม้ว่าข้อสอบแบบความเรียงมักจะไม่ใช้เพื่อวัดความรู้ตามความข้อเท็จจริงมากเท่ากับใช้วัดกระบวนการคิด แต่ก็สามารถใช้และควรจะให้คะแนนกับเนื้อหาที่เขียนมาในคำตอบ

2. การเรียบเรียง (Organization) ดูว่าคำตอบนั้นมีส่วนที่เป็นบทนำ เนื้อความ และส่วนสรุปหรือไม่ หรืออาจจะตั้งเกณฑ์เฉพาะอื่น ๆ ขึ้นมา รวมถึงการสะกดคำและไวยากรณ์

3. กระบวนการ (Process) เกณฑ์ของกระบวนการนี้ ควรทำการประเมินทั้งความเพียงพอในการแก้ปัญหาหรือตัดสินใจ และเหตุผลสนับสนุน

วอร์ด (Ward. 1999 : 153-154) ก็ได้สรุปวิธีการตรวจให้คะแนนผลงานไว้ 4 วิธี ดังนี้ วิธีประเมินรวม (Holistic) วิธีวิเคราะห์ย่อย (Analytic) วิธีดูคุณลักษณะเบื้องต้น (Primary trait) และวิธีดูตามจุดประสงค์ (Domain) ส่วนใหญ่การศึกษาวิจัยเกี่ยวกับการให้คะแนนจะพิจารณาถึงการประเมินการเขียน คำแนะนำสำหรับวิธีการให้คะแนนทั้งหมดนี้จะรวมไปถึงเกณฑ์ที่กำหนดให้ต่าง ๆ

วิธีการวิเคราะห์ย่อย จะมีข้อมูลให้มากที่สุดแต่ก็ใช้เวลามากที่สุดเช่นกัน การให้คะแนนวิธีประเมินรวม นั้นง่ายพอสมควร แต่ก็ไม่มีข้อมูลเชิงวิเคราะห์ที่ครูส่วนใหญ่ต้องการ ส่วนการให้คะแนนแบบดูลักษณะเบื้องต้นจะมุ่งเน้นไปที่จุดมุ่งหมาย และความสำเร็จในการบรรลุจุดมุ่งหมายนั้น การให้คะแนนแบบดูจุดประสงค์เป็นการให้คะแนนแยกส่วนตามประเด็นหลัก ๆ ที่มีจำนวนจำกัด เกณฑ์การให้คะแนนสามารถใช้ได้กับงานทุกแบบ เช่น เกณฑ์การให้คะแนนสำหรับงานเขียน (Writing) ก็ใช้สำหรับงานเขียน ในที่นี้ผู้วิจัยพบว่า ลักษณะการให้คะแนนวิธีประเมินรวม จะสอดคล้องกับการตรวจให้คะแนนแบบทดสอบความเรียงนี้ โดยสรุปดังนี้

การให้คะแนนวิธีประเมินรวม

การให้คะแนนแบบนี้ตั้งอยู่บนสมมุติฐานที่ว่า ส่วนประกอบของงานที่จะตรวจนั้นมีส่วนสัมพันธ์กันมากจนกระทั่งยากที่จะแยกองค์ประกอบต่าง ๆ ออกจากกัน เพื่อให้คะแนนเป็นประเด็น ๆ ได้

คำตอบแต่ละคำตอบจะได้คะแนนเดียวซึ่งครอบคลุมเกณฑ์ต่าง ๆ ทั้งหมด โดยปกติมักจะเขียนคำบรรยายสั้น ๆ เกี่ยวกับผลงานในเกณฑ์คะแนนแต่ละจุด แล้วเลือกผลงานสำหรับแต่ละจุด เช่น ในการให้คะแนนตัวอย่างข้อเขียน ข้อเขียนที่มีอยู่จะถูกเลือกเพื่อเป็นตัวอย่างสำหรับแต่ละคะแนน และใช้เป็นตัวอย่างอ้างอิง เพื่อเป็นมาตรฐานให้กับผู้ตรวจคะแนน

อีกวิธีหนึ่งที่เป็นไปได้คือ การคัดแยกผลงานออกมาตามระดับคะแนนที่ต่างกันเป็นกลุ่ม ๆ เมื่อให้คะแนนข้อสอบทุกฉบับแล้ว ก็มาดูทวนอีกรอบว่าอยู่ในกลุ่มของระดับคะแนนที่เหมาะสมกับคุณภาพของงานแล้วหรือยัง

วอร์ด (Ward. 1999 : 132) ก็ได้ข้อเสนอแนะสำหรับการให้คะแนนแบบทดสอบความเรียงไว้ 7 หัวข้อ ดังนี้

1. กำหนดน้ำหนักให้แต่ละคำถาม
2. กำหนดจุดสำคัญที่สุดของคำถามนั้น ๆ
3. กำหนดคำย่อย ๆ ที่สำคัญรองลงมา
4. กำหนดน้ำหนักของจุดสำคัญ และจุดย่อย ๆ ทั้งหมด
5. เตรียมตัวอย่างคำตอบที่สมบูรณ์แบบ
6. เตรียมกำหนดทางเลือกที่จะเริ่มให้คะแนนที่ถูกต้องของแต่ละข้อ
7. เตรียมกำหนดคำตอบที่ผิด หรือคลาดเคลื่อนไปจากคำตอบ

วอร์ด (Ward. 1999 : 133) ก็ได้ยกตัวอย่างการให้คะแนนแบบทดสอบความเรียง ดังนี้

ตัวอย่างข้อสอบ

1. *จงอธิบายพายุเฮอริเคนที่ทำลายชายฝั่งตะวันออกของอเมริกาว่าพัฒนาไปอย่างไร และทำไมจึงพัดไปทางนั้น*
2. *ทำไมทุก ๆ คนจึงควรจะได้รับการห่วงใยเกี่ยวกับการใช้ปุ๋ยของชาวนาในการเพาะปลูก จงอธิบายว่า เกิดอะไรขึ้นที่ในการใช้ปุ๋ยและมีวิธีการอย่างไรถ้าเกิดผลกระทบต่อประชาชนที่อยู่ บริเวณใกล้เคียงกัน*

ตัวอย่างการให้คะแนน (Scoring Rubrics)

ให้ 5 คะแนน เมื่อ	*แสดงความเข้าใจเกี่ยวกับทุก ๆ เนื้อหาสำคัญ ๆ และกระบวนการต่าง ๆ* *บอกหัวข้อที่เกี่ยวข้องที่สำคัญอย่างถูกตามกฎเกณฑ์* *แสดงแนวความคิดและสื่อสารออกมาอย่างดี*
ให้ 4 คะแนน เมื่อ	*รู้ความคิดรวบยอดและจุดเน้น* *บอกข้อเกี่ยวข้องที่สำคัญอย่างถูกต้องกฎเกณฑ์แต่อาจมีบางอย่างตกหล่น* *แสดงแนวคิดที่แสดงว่านักเรียนเข้าใจ*
ให้ 3 คะแนน เมื่อ	*เข้าใจความคิดรวบยอดที่สำคัญ ๆ และรู้กระบวนการ* *รู้กฎเกณฑ์และบอกบางอย่างได้ถูกต้อง* *แสดงแนวคิดที่ถูกต้องแต่ไม่ทั้งหมด*
ให้ 2 คะแนน เมื่อ	*เข้าใจความคิดรวบยอดบางอย่าง* *รู้กฎบางอย่างแต่ไม่สามารถนำเสนอออกมาได้* *ยากที่จะแสดงความคิดออกมา*
ให้ 1 คะแนน เมื่อ	*มีความเข้าใจน้อยและมีแสดงความคิดรวบยอด* *ไม่บอกกฎต่าง ๆ* *ไม่แสดงความคิด*

กรมวิชาการ (2539 :54-59) ได้อธิบายถึงการให้คะแนนแบบวิเคราะห์ (Analytic Score) ว่าเพื่อให้การมองคุณภาพงานหรือความสามารถของนักเรียนได้อย่างชัดเจน จึงได้มีการแยกองค์ประกอบของการให้

คะแนน และอธิบายคุณภาพของงานในแต่ละองค์ประกอบเป็นระดับ โดยทั่วไปแล้วจะมีการแยกองค์ประกอบของงานเป็น 4 ด้าน คือ

1. ความเข้าใจในความคิดรวบยอด ข้อเท็จจริง เป็นการแสดงให้เห็นว่านักเรียนเข้าใจในความคิดรวบยอด หลักการในการแก้ปัญหาที่ถามอย่างกระจ่างชัด

2. การสื่อความหมาย สื่อสาร คือ ความสามารถในการอธิบาย นำเสนอ การบรรยาย เหตุผล แนวคิดให้ผู้อื่นเข้าใจได้ดี มีความคิดสร้างสรรค์

3. การใช้กระบวนการและยุทธวิธี สามารถเลือกใช้ยุทธวิธี กระบวนการนำไปสู่ความสำเร็จอย่างมีประสิทธิภาพ

4. ผลสำเร็จของงาน ความถูกต้องแม่นยำในผลสำเร็จของงาน หรืออธิบายที่มา และ ตรวจสอบผลงาน

ส่วนการให้คะแนนวิธีประเมินรวม (Holistic Method) คือ การให้คะแนนงานชิ้นใดชิ้นหนึ่งโดยดูภาพรวมของชิ้นงานว่า มีความเข้าใจในความคิดรวบยอด การสื่อความหมาย กระบวนการที่ใช้ และผลงานเป็นอย่างไร แล้วเขียนอธิบายคุณภาพของงาน หรือความสำเร็จของงานเป็นชิ้น ๆ โดยอาจจะแบ่งระดับคุณภาพ ตั้งแต่ 0-4 หรือ 0-6 สำหรับขั้นตอนการให้คะแนน อาจจะแบ่งวิธีการให้คะแนนหลายวิธี เช่น

วิธีที่ 1 แบ่งงานตามคุณภาพเป็น 3 กอง คือ

กองที่ 1 ได้แก่งานที่มีคุณภาพเป็นพิเศษ และเขียนอธิบายลักษณะของงานที่มีคุณภาพเป็นพิเศษ

กองที่ 2 ได้แก่งานที่ยอมรับได้และเขียนอธิบายลักษณะของงานที่ยอมรับได้

กองที่ 3 ได้แก่งานที่ยอมรับได้น้อยหรือยอมรับไม่ได้ และเขียนอธิบายลักษณะของงานที่ยอมรับได้น้อย

จากนั้นก็นำงานแต่ละกองมาให้คะแนนเป็น 2 ระดับ คือ

งานกองที่ 1 จะให้ 6 หรือ 5 คะแนน

งานกองที่ 2 จะให้ 4 หรือ 3 คะแนน

งานกองที่ 3 จะให้ 2 หรือ 1 คะแนน

วิธีที่ 2 กำหนดระดับความผิดพลาด คือพิจารณาความบกพร่องจากคำตอบว่ามีมากน้อยเพียงใด โดยจะหักจากระดับคะแนนสูงสุดลงมาทีละระดับ ดังนี้

คะแนน 4 หมายถึง คำตอบถูก แสดงเหตุผลถูก แนวคิดชัดเจน

3 หมายถึง คำตอบถูก เหตุผลถูก แต่มีข้อผิดพลาดเล็กน้อย

2 หมายถึง เหตุผลหรือการคำนวณผิดพลาด แต่มีแนวทางที่จะนำไปสู่คำตอบ

1 หมายถึง แสดงวิธีคิดเล็กน้อยแต่ไม่ได้คำตอบ

0 หมายถึง ไม่ตอบหรือตอบไม่ถูกเลย

วิธีที่ 3 กำหนดระดับการยอมรับและคำอธิบาย เช่น เกณฑ์การให้คะแนนของความสามารถเข้าใจเนื้อหาสาระ เขียนได้ 4 ระดับดังนี้

คะแนน 4 หมายถึง การสาธิตหรือแสดงออกให้เห็นถึงการเข้าใจที่สมบูรณ์ ครบถ้วน ถูกต้องในหลักการ ความคิดรวบยอด ข้อเท็จจริงของงานหรือสถานการณ์ที่กำหนด รวมทั้งเสนอแนวคิดใหม่ที่แสดงถึงความเข้าใจอย่างลึกซึ้งถึงกฎเกณฑ์หรือลักษณะของข้อมูล

3 หมายถึง การแสดงออกให้เห็นถึงการเข้าใจที่สมบูรณ์ ครบถ้วน ถูกต้องในหลักการ ความคิดรวบยอด ข้อเท็จจริงของงานหรือสถานการณ์ที่กำหนด

2 หมายถึง การแสดงออกให้เห็นถึงการเข้าใจที่ไม่สมบูรณ์ ครบถ้วน ถูกต้องในหลักการ ความคิดรวบยอด ข้อเท็จจริงของงานหรือสถานการณ์ที่กำหนดในบางส่วน

1 หมายถึง การแสดงออกให้เห็นถึงการเข้าใจในหลักการความคิดรวบยอด ข้อเท็จจริงของงาน หรือสถานการณ์ที่กำหนดได้น้อยมาก และเข้าใจไม่ถูกต้องบางส่วน

0 หมายถึง ไม่แสดงความคิดเห็นใด ๆ

ซึ่งทั้ง 3 วิธีดังกล่าวนั้น พอที่จะสรุปได้ว่า มีส่วนที่เหมือนกันคือ เป็นการตรวจให้คะแนนที่มองภาพรวมของผลงาน (ในที่นี้หมายถึง การตอบคำถามในกระดาษคำตอบของนักเรียน) แล้วแยกเป็นกอง ๆ ส่วนที่แตกต่างกันของทั้ง 3 วิธี ก็คือ

วิธีที่ 1 แยกโดยใช้คุณภาพของผลงานเป็นหลัก

วิธีที่ 2 แยกโดยใช้ความผิดพลาดของผลงานเป็นหลัก

วิธีที่ 3 แยกโดยใช้ความสามารถในการอธิบายหรือการแสดงออกให้เห็นถึงการเข้าใจเป็นหลัก

บุญเชิด ภิญโญอนันตพงษ์ (2525: 33) กล่าวถึงการตรวจให้คะแนนแบบทดสอบความเรียงไว้ว่า ผู้ตรวจให้คะแนนต้องอ่านคำตอบแล้วประเมินค่าออกมาเป็นคะแนนโดยวิธีใดวิธีหนึ่ง การตรวจให้คะแนนบางครั้งไม่สามารถใช้คำเฉลยได้ เพราะการเขียนตอบโดยให้เสนอความคิดเห็นสามารถที่จะเขียนอย่างไรก็ได้ และไม่สามารถระบุว่าความคิดเห็นอย่างใดถูกอย่างใดผิดได้ ดังนั้นผู้ตรวจให้คะแนนต้องเป็นผู้มีความรู้ในสาขาวิชานั้น การให้คะแนนต้องอาศัยทักษะและความพยายามในการอ่าน ซึ่งจะกระทำโดยรีบด่วน และไม่คิดไม่ได้ ปัญหาใหญ่ของข้อสอบอัตนัยก็คือ ผู้ตรวจให้คะแนนไม่มั่นคงแน่นอน เพราะการให้คะแนนขึ้นอยู่กับตัวผู้ตรวจให้คะแนนเป็นสำคัญ จึงได้ชื่อว่าอัตนัย (อัตตา + นัย) ดังนั้น ความลำเอียงในการให้คะแนนจึงเกิดขึ้นได้ง่าย ผู้ตรวจให้คะแนนต้องทำใจให้บริสุทธิ์ โดยยึดคุณธรรมอันสูงส่งในการให้คะแนน

และวิธีการให้คะแนนข้อสอบอัตนัย ที่ บุญเชิด ภิญโญอนันตพงษ์ (2521 : 135) ให้รายละเอียดจากการศึกษาค้นคว้าวิธีการให้คะแนนอย่างเชื่อมั่นได้ ดังมีรายละเอียดของแต่ละวิธีมีดังนี้

1. วิธีประเมินรวม (Holistic Method) ผู้ตรวจให้คะแนนจะอ่านข้อสอบทีละข้อ แล้วประเมินค่าเป็นส่วนรวมว่าข้อนั้นควรให้คะแนนเท่าไร โดยไม่แยกพิจารณาคะแนนด้านอื่น ๆ ถ้ามีกระดาษคำตอบจำนวนมากก็ใช้วิธีแยกกระดาษคำตอบออกเป็นกลุ่ม ๆ เช่น ดีมาก ดี ปานกลาง อ่อน อ่อนมาก แล้วพิจารณาจัดเรียงใหม่จากดีที่สุดจนถึงเลวที่สุด

2. วิธีวิเคราะห์ย่อย (Analytic Method) ผู้ตรวจให้คะแนนจะแบ่งคะแนนของแต่ละข้อออกเป็นส่วน ๆ เช่น ให้คะแนนการจัดเรียบเรียงความคิด ความมีเหตุผล ความถูกต้องของเนื้อหา ความถูกต้องในการใช้ภาษา และกำหนดน้ำหนักคะแนนในแต่ละส่วนแล้วจึงตรวจให้คะแนนแต่ละข้อ

3. วิธีแบ่งกลุ่ม (Sorting Method) ผู้ตรวจให้คะแนนอ่านกระดาษคำตอบอย่างรวดเร็ว แล้วแบ่งกระดาษคำตอบตามชนิดของคำตอบ โดยแบ่งเป็น 5 กลุ่ม คือ ดีมาก ดี ปานกลาง ค่อนข้างเลว เลวมาก จากนั้นก็อ่านคำตอบซ้ำอีกครั้งตามกลุ่มที่จัดไว้ และปรับแก้ ย้ายคำตอบบางแผ่นที่รู้สึกว่าจัดไว้ผิดกลุ่มกระดาษ คำตอบในแต่ละกลุ่มจะมีจำนวน 9% 20% 42% 20% และ 9%

4. วิธีจัดอันดับ (Rating Method) ผู้ตรวจให้คะแนนจะต้องอ่านกระดาษคำตอบทุกแผ่นก่อน แล้วนำกระดาษคำตอบนั้นมาจัดเป็นกลุ่ม ๆ ตามคุณภาพ ดีมาก ดี ปานกลาง พอใช้ และ ใช้ไม่ได้ แล้วมาพิจารณาในแต่ละกลุ่มอีกทีว่าคำตอบใดตรงและใกล้เคียงกับแนวคำตอบมากกว่ากัน ต่อจากนั้นจึงให้คะแนน

5. วิธีกำหนดค่าคะแนน (Point Method) ผู้ตรวจให้คะแนนจะเอาคำตอบมาเทียบเคียงกับคำเฉลยว่าใกล้เคียงมากน้อยเพียงใด แล้วจึงให้คะแนนเป็นคน ๆ ไป

วิธีให้คะแนนดังกล่าวสามารถจำแนกเป็น 2 กลุ่ม คือ

1) วิธีประเมินรวม วิธีจัดกลุ่ม และ วิธีจัดอันดับกลุ่มนี้จะให้คะแนนโดยอาศัยเกณฑ์ภายใน โดยนำคำตอบของแต่ละคนในกลุ่มเปรียบเทียบกัน เป็นวิธีการแบบอิงกลุ่ม ซึ่งจะให้คะแนนที่เชื่อมั่นกว่า

2) วิธีวิเคราะห์ย่อยและวิธีกำหนดค่าคะแนน กลุ่มนี้จะให้คะแนนโดยอาศัยเกณฑ์ภายนอก โดยนำคำตอบของแต่ละคนไปเทียบกับคำเฉลย เป็นวิธีการแบบอิงเกณฑ์ กลุ่มนี้จะให้คะแนนได้เที่ยงตรงกว่ากลุ่มแรก

ซึ่งได้เสนอวิธีการตรวจให้คะแนน ในกลุ่มที่ 1 คือ กลุ่มวิธีประเมินรวม ไว้ 3 ขั้น ดังนี้

1. อ่านคำตอบทั้งหมดให้ทะลุปรุโปร่งโดยเร็ว แบ่งกระดาษคำตอบออกเป็นประเภท ๆ พร้อมทั้งกำหนดเกรดให้เป็นกอง ๆ ในแต่ละกระดาษคำตอบ หากไม่แน่ใจก็ให้เขียนสัญลักษณ์ไว้ที่หัวกระดาษ แล้วใส่กองใดกองหนึ่งไว้ก่อน

2. อ่านคำตอบซ้ำใหม่อีกครั้งหนึ่ง โดยเฉพาะอย่างยิ่งกระดาษคำตอบที่มีเครื่องหมายไม่แน่ใจ การอ่านครั้งที่สองนี้อาจเปลี่ยนกองกระดาษคำตอบก็ได้

3. กำหนดเกรดเป็นตัวอักษรให้แต่ละกองคำตอบซึ่งแบ่งไว้แล้ว ซึ่งเป็นผลมาจากการเปรียบเทียบคำตอบกันเองภายในกลุ่ม

วิธีตรวจให้คะแนน ในกลุ่มที่ 2 คือ กลุ่มวิธีวิเคราะห์ย่อย ประกอบด้วย 3 ขั้น ดังนี้

1. เฉลยคำตอบไว้ล่วงหน้า โดยให้ครอบคลุมลักษณะที่สำคัญ ๆ ซึ่งนักเรียนอาจจะตอบออกมาในแต่ละคำถาม และกำหนดค่าคะแนนรวมของแต่ละข้อคำถามไว้

2. อ่านคำตอบข้อเดียวกันของนักเรียนทุกคนให้หมด จากนั้นจึงให้คะแนนกระดาษคำตอบแต่ละแผ่นตามที่ได้อ่านไปตามความเหมาะสม จากนั้นก็ตรวจข้ออื่น ๆ ในลักษณะเดียวกันจนหมดทุกข้อ ข้อสำคัญที่พึงระวังก็คือจะต้องกำหนดเกณฑ์การตรวจให้ครอบคลุมก่อนที่จะทำการตรวจจริงเสมอ

3. รวมคะแนนทั้งหมดทุกข้อของแต่ละคน แล้วให้เกรดเป็นตัวอักษรโดยอาศัยคะแนนเกณฑ์ปกติตำแหน่งเปอร์เซ็นต์หรืออาศัยโค้งการกระจายของคะแนน

พวงรัตน์ ทวีรัตน์ (2530 : 116-119) ได้เสนอวิธีการตรวจให้คะแนนแบบทดสอบอัตนัย มี 2 วิธี คือ

ก. การตรวจโดยวิธีเทียบกับเกณฑ์ (Analytical Method หรือ Point Score Method) การตรวจโดยวิธีเทียบเกณฑ์นี้ มีขั้นตอนการตรวจดังนี้

1. กำหนดเฉลยหรือแนวคำตอบของข้อสอบแต่ละข้อไว้ล่วงหน้า โดยอาจแยกแนวคำตอบเป็นตอนย่อย ๆ ตามความสำคัญหรือตามประเด็นที่ถาม พร้อมทั้งกำหนดค่าคะแนนรวมของแต่ละข้อไว้ ถ้าข้อใดมีหลายประเด็นที่ถาม ก็อาจให้คะแนนแยกเป็นตอน ๆ ไป แล้วจึงนำมารวมกันทีหลัง

2. ทำการตรวจ โดยอ่านคำตอบข้อเดียวกันของแต่ละคน เทียบกับเกณฑ์หรือแนวคำตอบที่ได้กำหนดไว้ เพื่อดูว่าคำตอบของแต่ละคน ใกล้เคียงกับแนวคำตอบเพียงใด แล้วจึงให้คะแนนเป็นรายบุคคลตามความเหมาะสม จากนั้นก็ตรวจข้ออื่น ๆ ในลักษณะเดียวกันจนหมดทุกข้อ

3. รวมคะแนนทุกข้อของแต่ละคน แปลงคะแนนดิบเป็นคะแนนมาตรฐาน (นิยมแปลงเป็นคะแนนมาตรฐาน ที T-score) แล้วจึงทำการตัดเกรด พร้อมกับกำหนดเกรดของแต่ละคนโดยให้เป็นตัวอักษร เช่น A B C D เป็นต้น

ข้อสังเกต

1) การให้คะแนนโดยวิธีเทียบเกณฑ์นี้ สิ่งสำคัญที่ต้องคำนึงถึง คือ การกำหนดเกณฑ์คำตอบหรือแนวคำตอบจะต้องกำหนดไว้พร้อมกับการเขียนข้อคำถามเลย ไม่ควรมากำหนดตอนที่จะเริ่มลงมือตรวจ

2) การให้คะแนนโดยวิธีเทียบเกณฑ์นี้ เหมาะสำหรับข้อสอบอัตนัยแบบกำหนดขอบเขตของคำตอบ (Restricted Response)

ข้อดีของการตรวจข้อสอบโดยวิธีเทียบเกณฑ์ มีดังนี้

1) คะแนนที่ได้จากการตรวจมีความเชื่อถือมากขึ้น เพราะได้มีการป้องกันข้อผิดพลาดในการให้คะแนนไว้อย่างดี โดยใช้วิธีการเทียบกับคำเฉลย

2) การตรวจให้คะแนนในข้อเดียวกันของทุกคนจนหมดทีละข้อ จะช่วยลดความเปลี่ยนแปลงของเกณฑ์การให้คะแนนแต่ละคนได้ การตรวจเช่นนี้ช่วยให้ผู้ตรวจให้คะแนนสามารถแบ่งการตรวจเป็นช่วง ๆ ได้ ไม่ต้องตรวจให้เสร็จทีเดียวทุกข้อ เช่น ถ้าข้อสอบมี 5 ข้อ อาจแบ่งตรวจเป็นวันละ 2 ข้อ ตอนเช้า 1 ข้อ ตอนบ่ายอีก 1 ข้อ หรือเป็นตอนอื่น ๆ ตามแต่สะดวก การทำเช่นนี้ เป็นการช่วยลดความเหนื่อยอ่อน และทำให้ผลการตรวจเกิดการเปลี่ยนแปลงไม่มากนัก

3) การกำหนดแนวคำตอบไว้เป็นเกณฑ์ จะช่วยให้ครูสามารถให้เหตุผลกับผู้สอบแต่ละคนได้ว่าทำไมจึงได้คะแนนเช่นนั้น

4) ในกรณีที่ผู้ตอบส่วนใหญ่ตอบไม่ตรงกับแนวคำตอบที่กำหนดเป็นเกณฑ์ไว้ ครูผู้ออกข้อสอบก็จะได้มองเห็นข้อบกพร่องของการเขียนข้อคำถาม เช่น คำถามไม่ชัดเจน หรือกำกวม หรือคำถามยากเกินไป เป็นต้น

ข้อจำกัด ของการตรวจโดยวิธีเทียบกับเกณฑ์

1) ครูจะต้องทำงานหนักขึ้น และเสียเวลามาก

2) การที่ผู้ตรวจให้คะแนนต้องคอยอ่านคำตอบที่สอดคล้องกับเกณฑ์ อาจทำให้เกิดการมองข้ามหรือไม่ได้อ่านข้อความสำคัญของผู้ตอบไปบ้าง

ข. การตรวจโดยวิธีจัดอันดับคุณภาพ (Holistic Method)

เป็นวิธีการตรวจให้คะแนนที่นับว่าให้ผลเป็นที่เชื่อถือได้อีกวิธีหนึ่ง ถ้าผู้ตรวจให้คะแนนดำเนินการตรวจให้คะแนนอย่างระมัดระวัง มีขั้นตอนที่สำคัญดังนี้

1. อ่านคำตอบของผู้สอบทุกคนให้หมดในขั้นต้น แล้วนำกระดาษคำตอบของผู้สอบมาจัดแยกเป็นกลุ่ม ๆ ตามคุณภาพของคำตอบ เช่น กลุ่มดีมาก ดี ปานกลาง อ่อน และยังใช้ไม่ได้ในระหว่างการอ่านคำตอบถ้ายังไม่แน่ใจว่าควรจัดไว้ในกลุ่มใด ก็ให้ทำเครื่องหมายคำถามที่แสดงว่ายังไม่แน่ใจไว้ก่อน เช่น คำตอบของนักเรียนคนหนึ่ง ผู้ตรวจให้คะแนนไม่แน่ใจว่าจะจัดไว้ในกลุ่มดี หรือ กลุ่มปานกลาง แต่ความรู้สึกอยากจะจัดอยู่ในกลุ่มปานกลาง ก็ให้เขียนไว้ว่า “กลุ่มปานกลาง” ไว้ก่อน แล้วจัดอยู่ในกลุ่มปานกลางไปก่อน

2. อ่านคำตอบทั้งหมดซ้ำใหม่อีกครั้งหนึ่ง โดยเฉพาะอย่างยิ่งกระดาษคำตอบที่มีเครื่องหมายไม่แน่ใจอยู่ ต้องอ่านอย่างรอบคอบ ในการอ่านครั้งที่สองนี้ กระดาษคำตอบอาจเปลี่ยนกลุ่มได้ การอ่านครั้งนี้

ขึ้นอยู่กับการอ่านครั้งแรก ถ้าครั้งแรกอ่านอย่างคร่าว ๆ ครั้งนี้ก็ต้องอ่านอย่างละเอียด เมื่ออ่านแล้วก็พิจารณาจัดเรียงลำดับกระดาษทั้งหมดตามลำดับอย่างต่อเนื่องกันไป จากดีที่สุด จนถึงไม่ดีที่สุด

3. ทำการให้คะแนน โดยแปลงคะแนนให้เป็นตำแหน่งร้อยละ และนำไปเทียบกับตารางแปลงตำแหน่งร้อยละให้เป็นคะแนน ก็จะทำให้ทราบคะแนนของผู้สอบในแต่ละข้อของผู้สอบแต่ละคน

การตรวจให้คะแนนโดยวิธีจัดอันดับคุณภาพนี้ ถ้าจะให้สะดวกและรวดเร็วขึ้น ควรให้ผู้ตอบแยกกระดาษคำตอบเป็นข้อ ๆ

และจากผลการศึกษาของ อุษณีย์ บัวศิริพันธุ์ (2543 : 20) เกี่ยวกับวิธีการตรวจให้คะแนนคำถามแบบความเรียง หรือแบบปลายเปิด สรุปได้ 2 คือ

1. วิธีวิเคราะห์ (Analytic Method หรือ Point Method) ใช้การวิเคราะห์ด้วยประเด็นที่มีลักษณะแตกต่างกันของคำตอบ

2. วิธีประเมินรวม (Holistic Method) เป็นวิธีที่ผู้ประเมินพิจารณาคำตอบโดยรวมมากกว่าตรวจสอบรายละเอียดปลีกย่อยเฉพาะ

จากการศึกษาเอกสารงานวิจัย ที่กล่าวมาข้างต้น พอสรุปถึงการตรวจให้คะแนน ซึ่งจากการศึกษามีผู้ให้คำจำกัดความแตกต่างกันไปหลากหลาย ดังนั้นผู้วิจัยจึงเลือกวิธีตรวจให้คะแนน 2 วิธีดังนี้

1. วิธีวิเคราะห์ย่อย (Analytic Method) เป็นวิธีตรวจที่เหมาะกับการตรวจแบบทดสอบความเรียงที่มีลักษณะจำกัดคำตอบ ซึ่งครูสามารถที่จะกำหนดแนวคำตอบไว้ล่วงหน้าได้ หรือทำเฉลยไว้ก่อน โดยแยกเป็นประเด็นย่อย ๆ ตามองค์ประกอบของการตอบหรือประเด็นที่ถาม และยังกำหนดน้ำหนักคะแนนในแต่ละส่วนก่อนแล้วจึงตรวจให้คะแนนแต่ละข้อ นิยมตรวจข้อเดียวให้เสร็จทุก ๆ คนก่อน แล้วค่อยตรวจข้อต่อไป การตรวจลักษณะนี้ เหมาะกับการสอบที่มีจำนวนผู้สอบไม่มากนัก

2. วิธีประเมินรวม (Holistic Method) เป็นวิธีตรวจให้คะแนนโดยรวม จากการเขียนของเด็ก จะไม่แยกประเด็นย่อย ๆ การให้คะแนนขึ้นอยู่กับคุณภาพของคำตอบ และความคิดของผู้ตรวจให้คะแนนโดยเฉพาะสามารถตรวจได้รวดเร็ว ซึ่งผู้ตรวจให้คะแนนจะแยกกลุ่มกระดาษคำตอบออกเป็นกลุ่ม ๆ ตามคุณภาพการตอบ อาจแยกเป็น 3 กลุ่ม คือ ดี ปานกลาง ไม่ดี หรือ แยกเป็น 6 ระดับ คือ ดีมาก ดี ค่อนข้างดี พอใช้ อ่อน แล้วกลับมาพิจารณาจัดเรียงใหม่จาก ดีที่สุด ถึงเลวที่สุด

ซึ่งการวิจัยครั้งนี้ ผู้วิจัยเลือกใช้วิธีตรวจทั้ง 2 วิธี โดยใช้เกณฑ์การตรวจให้คะแนนของ วอร์ด (Ward. 1999 : 133) ที่เสนอวิธีตรวจไว้ 3 ประเด็น คือ การแสดงความเข้าใจในเนื้อหา การนำเสนอหลักการที่ถูกต้อง และการจัดเรียบเรียงความคิด และใช้ระยะการตรวจห่างกัน 1 สัปดาห์

1.5 องค์ประกอบที่ส่งผลต่อคะแนนแบบทดสอบความเรียง

การตรวจให้คะแนนของแบบทดสอบความเรียง มีสิ่งที่ทำให้เกิดความคลาดเคลื่อนได้ ซึ่ง บุญเชิด ภิญโญอนันพงษ์ (2521 : 125) ได้กล่าวเกี่ยวกับความลำเอียงในการให้คะแนนของผู้ตรวจให้คะแนนที่อาจจะเกิดขึ้นโดยไม่ได้ตั้งใจ เช่น

1. The Halo Effect เป็นความลำเอียงชนิดหนึ่งที่เป็นผลจากผู้ตรวจให้คะแนน ได้รับอิทธิพลจากผู้เขียนตอบอันเป็นความประทับใจของผู้ตรวจให้คะแนนที่ได้ประสบจากผู้เขียนตอบนั้น จะทำให้ผู้ตรวจให้คะแนนประเมินค่าคะแนนนั้นออกมาสูงขึ้นหรือต่ำลง เนื่องจากความประทับใจในลักษณะเด่น – ด้อย ของผู้เขียนตอบ ตัวอย่างเช่น เพียงพร เป็นนักเรียนที่ประพฤติดี เคารพนบนอบครู ขยันทำการบ้าน คุณลักษณะเหล่านี้จะมีอิทธิพลต่อครูในการตรวจให้คะแนนคำตอบเพียงพรสูงขึ้น ส่วนสันติเป็นนักเรียนที่ซน ดื้อกระด้าง เป็นที่ระอาใจของครูอยู่เสมอ การตรวจให้คะแนนคำตอบของสันติซึ่งตอบได้ดีเท่าเพียงพร มีแนวโน้มที่จะได้คะแนนต่ำกว่าของเพียงพร

2. อารมณ์ของผู้ตรวจให้คะแนน เป็นต้นเหตุอีกอย่างหนึ่งที่ทำให้คะแนนผิดเพี้ยนไป ตัวอย่างเช่น ครูคนหนึ่งกำลังตรวจข้อสอบนักเรียนอยู่ มีโทรเลขจากกรุงเทพมาบอกว่าสอบเข้าเรียนประสานมิตรได้ ครูคนนี้จะตื่นเต้นดีใจ คะแนนจะมีแนวโน้มสูงขึ้นกว่าปกติ ในทางตรงข้าม ถ้าครูคนนี้ได้รับหนังสือจากจังหวัดลงโทษลดขั้นเงินเดือน เพราะไม่แต่งเครื่องแบบข้าราชการ ในวันที่รัฐมนตรีศึกษาเดินทางมาตรวจราชการ อารมณ์ของครูคนนี้จะขุ่นมัวทันที คะแนนจะมีแนวโน้มต่ำลงกว่าปกติ

พวงรัตน์ ทวีรัตน์ (2530 : 119) กล่าวว่าองค์ประกอบที่ส่งผลต่อการตรวจให้คะแนนของผู้ตรวจมีสาเหตุ 4 ประการ คือ

ผลจากสิ่งอื่น (The Halo Effect) องค์ประกอบด้านนี้ ได้แก่ การที่ผู้ตรวจให้คะแนนอ่านคำตอบข้อแรกของผู้สอบคนใดคนหนึ่ง ถ้าผู้สอบตอบได้ดีมาก ก็จะให้คะแนนสูง และเมื่อตรวจข้อต่อไป ก็มีแนวโน้มจะให้คะแนนสูงตามไปด้วย แม้ว่าคำตอบข้ออื่น ๆ จะไม่ดีเท่าที่ควร และในทางตรงข้าม ถ้าคำตอบข้อแรกผู้สอบตอบไม่ค่อยดี ก็จะได้คะแนนต่ำ และข้อต่อ ๆ ไป ก็จะมีแนวโน้มได้คะแนนต่ำตามไปด้วยเช่นกัน นอกจากนั้นความรู้จักมักคุ้นก็เป็นผลต่อการให้คะแนน เช่น ถ้าครูทราบว่ากระดาษคำตอบนั้นเป็นของนักเรียนที่เรียนดี และขยัน ก็จะให้คะแนนสูงกว่าที่ควรจะเป็น และในทางตรงข้ามถ้าเป็นของนักเรียนที่เรียนอ่อน และเป็นนักเรียนเกเร ก็จะให้คะแนนต่ำกว่าเป็นจริง

ความใจดี (Generosity) ของครู ครูหลายคนมักจะให้คะแนนผู้สอบหรือนักเรียนสูงกว่าที่ควรจะเป็นเสมอ เช่น ให้เกรด A เกือบค่อนห้องทั้ง ๆ ที่ตามความเป็นจริงแล้วอาจมีคนทำคะแนนได้ถึงเกณฑ์ที่จะให้เกรด A เพียงครึ่งห้อง หรือ เพียงหนึ่งในสามของนักเรียนทั้งห้องเท่านั้น ความคลาดเคลื่อนที่เกิดจากองค์ประกอบข้อนี้คล้ายคลึงกับข้อแรก

การปรับโทษ (Penalty) การให้คะแนนกด หรือการตัดทอนคะแนนนักเรียนที่ไม่สามารถตอบได้ตามที่ครูได้ตั้งเป้าหมายไว้ ครูก็จะทำการปรับโทษ โดยให้คะแนนต่ำกว่าที่ควรจะเป็น ทั้ง ๆ ที่นักเรียนทำคะแนนอยู่ในระดับที่ใช้ได้ ทั้งนี้เพราะครูได้ตั้งเป้าหมายไว้สูงเกินไป เมื่อนักเรียนตอบได้ไม่ถึงเกณฑ์ที่ครูตั้งเป้าหมายก็จะเกิดความรู้สึกว่านักเรียนมีความสามารถต่ำ จึงให้คะแนนน้อย ถ้าเกิดกรณีเช่นนี้ขึ้น ก็ควรจะได้ขอให้มีการตรวจสอบคำตอบกับเกณฑ์มาตรฐานวิธีสอน ตลอดจนระดับสติปัญญาของนักเรียนอีกครั้ง ก็จะแก้ไขความคลาดเคลื่อนกรณีนี้ได้

องค์ประกอบแทรกซ้อน (Extraneous Factors) ความคลาดเคลื่อนของการให้คะแนนอาจเกิดจากองค์ประกอบแทรกซ้อนอื่น ๆ เช่น อาจเกิดจากตัวนักเรียนเอง เป็นต้นว่า ลายมือ ความถูกต้องของการเขียน ตัวสะกดการันต์ การเรียบเรียงถ้อยคำ สำนวนภาษา หลักไวยากรณ์ สิ่งเหล่านี้มีผลต่อการให้คะแนนเป็นอย่างมาก ซึ่งครูจะต้องชี้แจงให้นักเรียนทราบอย่างชัดเจนว่า การให้คะแนนนั้นคำนึงสิ่งใดบ้าง เพื่อให้นักเรียนได้เพิ่มความระมัดระวัง

วอร์ด (Ward. 1999 : 162-164) ได้แบ่งปัญหาในการให้คะแนนโดยผู้ตรวจออกเป็น 2 ด้าน คือ ด้านความคลาดเคลื่อนทางเนื้อหา และด้านความคลาดเคลื่อนจากผู้ตรวจ ความคลาดเคลื่อนนี้จะต้องทำให้เกิดขึ้นน้อยที่สุด วิธีที่ช่วยได้ก็คือ การเขียนกำหนดเกณฑ์การตรวจให้ชัดเจนที่สุด

ความคลาดเคลื่อนทางเนื้อหา แบ่งเป็น 2 ชนิด คือ

1. ความคลาดเคลื่อนจากความใจดี (Generosity Error) เป็นผลที่ผู้ตรวจจะให้คะแนนสูง ซึ่งจะพบในหลาย ๆ สถานการณ์ เช่น โดยทั่วไปครูมากกว่า 90 % มักจะให้คะแนนลูกศิษย์ของตนเองสูง

2. ความคลาดเคลื่อนจากความลำเอียง (Halo Effect) เป็นพฤติกรรมส่วนบุคคลและตัวแปรที่เป็นความประทับใจส่วนตัวต่อนักเรียน หรือลายมือที่สวยงาม จะเป็นผลให้คะแนนที่ได้สูงกว่า เช่น ซาร่า เป็นนักเรียนที่ดี (หรือที่น่ารัก , เป็นคนใจดี) ก็จะส่งผลให้คะแนนของเธอสูงขึ้นในทุก ๆ หัวข้อ ทั้ง ๆ ที่เธออาจจะไม่มีความสามารถในด้านนั้น ๆ ก็ได้

ความคลาดเคลื่อนจากผู้ตรวจ มี 2 ลักษณะ คือ

1. ความคลาดเคลื่อนภายในผู้ตรวจ (Intrajudge Errors) เป็นความคลาดเคลื่อนจากการใช้มาตรฐานของผู้ตรวจเอง ที่อาจเกิดจากความเหนื่อยล้า ความเบื่อ หรือความฉุนเฉียว จากตัวกระดาษคำตอบขณะทำการตรวจให้คะแนน ผู้ตรวจต่างคนก็จะมีมาตรฐานต่างกัน

2. ความคลาดเคลื่อนจากภายนอกผู้ตรวจ (Interjudge Errors) เป็นความคลาดเคลื่อนจากความแตกต่างระหว่างผู้ตรวจต่างคนกัน เช่น ผู้ตรวจคนที่ 1 อาจจะให้คะแนนสูงกับนักเรียนทุกคน และอีกคนอาจจะให้คะแนนต่ำกับนักเรียนทุกคน

1.6 ความเชื่อมั่นของแบบทดสอบความเรียง

1.6.1 ความหมายของความเชื่อมั่น

นักการศึกษาและนักจิตวิทยา ได้ให้ความหมายของความเชื่อมั่นต่าง ๆ ดังนี้

บราวน์ (Brown. 1976 : 54) ได้ให้ความหมายเชิงปฏิบัติว่า ความเชื่อมั่น เท่ากับอัตราส่วนของความแปรปรวนที่ได้จากคะแนนจริง (True Score) กับความแปรปรวนของคะแนนที่สังเกตได้ (Observed Score)

ฮอบกินส์ และ แอนเตส (Hopkins and Antes. 1990 : 549) ให้ความหมายของความเชื่อมั่นหมายถึง ความคงที่ของคะแนนที่สอบได้ ซึ่งคือ ข้อมูลที่ได้จากการวัดหลาย ๆ ครั้งก็ได้ผลเท่าเดิม

อีเบลและฟิสไบด์ (Ebel and Frisbie. 1986 : 71) กล่าวว่า ความเชื่อมั่น เป็นค่าสัมประสิทธิ์ระหว่างคะแนนชุดหนึ่งกับคะแนนอีกชุดหนึ่งของแบบทดสอบที่มีลักษณะคล้ายคลึงกัน 2 ชุด และเป็นอิสระจากกันที่ได้ข้อมูลมาจากกลุ่มตัวอย่างกลุ่มเดียวกัน

บุญเชิด ภิญโญอนันตพงษ์ (2521 : 269) กล่าวว่า ความเชื่อมั่น หมายถึง ความคงที่ของคะแนนซึ่งได้มาจากการวัดนักเรียนกลุ่มเดียวกันด้วยแบบทดสอบฉบับเดียวกันหลาย ๆ ครั้ง หรือด้วยแบบทดสอบสองฉบับที่มีลักษณะเหมือนกัน หรือภายใต้เงื่อนไขของตัวแปรอื่น ๆ ในการวัดนั้น

อนันต์ ศรีโสภา (2524 : 5) กล่าวว่าความเชื่อมั่น หมายถึง การที่เรานำแบบทดสอบหรือแบบสอบถามไปวัดสิ่งเดียวกัน สองครั้งจะให้ระดับไม่เปลี่ยนแปลง (Consistency) หรือมีความคงที่ของคะแนนสูง

ประคอง กรรณสูต (2542 : 37) กล่าวว่า ความเชื่อมั่น คือ ความคงที่ หรือ ความคงเส้นคงวา (Consistency) ที่ได้จากการวัดสิ่งที่ต้องการวัด หรือ ความคงที่ของคะแนนที่ได้จากคนกลุ่มเดียวกันสองครั้งด้วยแบบทดสอบเดิมในเวลาที่ต่างกัน หรือความคงที่ของคะแนน จากการทดสอบคนกลุ่มเดียวกันด้วยข้อสอบชุดที่มีข้อสอบเทียบเท่ากัน

1.6.2 ความเชื่อมั่นของแบบทดสอบความเรียง

เคอร์ตัน (เยาวดี วิบูลย์ศรี. 2540 : 256 ; citing Cureton. 1958) ได้อภิปรายถึงปัญหาของความเชื่อมั่นของแบบทดสอบความเรียงว่า การที่จะได้ค่าประมาณความเชื่อมั่นของแบบทดสอบความเรียงนั้น จะต้องมีแบบทดสอบสองชุดที่มีความเท่าเทียมกันในเนื้อหาที่ต้องการวัด แล้วให้ผู้สอบคนเดียวกันสอบภายในช่วงเวลาที่กำหนด และในการตรวจ ผู้ตรวจให้คะแนนแต่ละคนควรจะอ่านคำตอบโดยตลอดสำหรับคำถามข้อเดียวกันของแบบทดสอบทั้งสองชุด ถ้าหากสามารถดำเนินการได้เช่นนี้ ค่าสหสัมพันธ์ระหว่างคะแนนที่ได้รับจากแบบทดสอบชุดหนึ่งกับคะแนนที่ได้รับจากแบบทดสอบอีกชุดหนึ่งก็ย่อมสามารถที่จะใช้เป็นค่าความเชื่อมั่นได้โดยประมาณ

เพย์นเน่ (เยาวดี วิบูลย์ศรี. 2540 : 256 ; citing Payne. 1968) กล่าวว่า วิธีหนึ่งที่จะช่วยให้ค่าความเชื่อมั่นของแบบทดสอบความเรียงสูงขึ้นก็คือ การเพิ่มจำนวนคำถามให้มากขึ้น และจำกัดทิศทางของคำตอบให้แคบเข้า การเขียนคำถามให้มีลักษณะเฉพาะมากขึ้นและแคบเข้ามา ย่อมจะมีผลทำให้ลดความคลุมเครือของคำตอบลงไปได้ ซึ่งจะทำให้ค่าความเชื่อมั่นของเครื่องมือเพิ่มขึ้น หรือคะแนนที่ให้จะมีค่าความเชื่อมั่นมากขึ้น

อีกวิธีหนึ่งที่จะช่วยปรับปรุงค่าความเชื่อมั่นของแบบทดสอบความเรียง ก็คือการกำหนดคำถามให้มีรายละเอียดเพียงพอ เพื่อที่จะช่วยให้นักเรียนทราบได้ว่า อะไรคือสิ่งที่ต้องการแท้จริงของคำถาม ถ้ามิฉะนั้นแล้วนักเรียนบางคนจะให้ความสนใจต่อแนวคิดหนึ่งของคำถาม ส่วนคนอื่น ๆ ก็อาจจะตอบในทิศทางของแนวคิดที่ต่างออกไป ในสถานการณ์เช่นนี้ย่อมจะได้แบบทดสอบความเรียงที่มีความเชื่อมั่นสูงได้ยาก

บุญเชิด ภิญโญอนันตพงษ์ (2521 : 137) ได้กล่าวถึง ความเชื่อมั่นของแบบทดสอบความเรียงว่า สิ่งสำคัญในการศึกษาความเชื่อมั่นของแบบทดสอบความเรียงมักจะเกี่ยวกับการให้คะแนนของผู้ตรวจให้คะแนนมากกว่าจะเป็นความเชื่อมั่นของข้อสอบโดยตรง ผู้ตรวจให้คะแนนที่ไม่มีความคงที่แน่นอนในการให้คะแนน จะมีผลกระทบต่อความเชื่อมั่นของข้อสอบ

ล้วน สายยศ และอังคณา สายยศ (2539 : 233) เสนอว่าวิธีการหาความเชื่อมั่นแบบความสอดคล้องในการตัดสินใจ (Decision Consistency Reliability) เป็นการหาความสอดคล้องระหว่างการทดสอบ 2 ครั้ง จากแบบทดสอบฉบับเดียว หรือ แบบทดสอบที่คู่ขนานกัน

2. ความเชื่อมั่นของผู้ตรวจให้คะแนน (Rater Reliability)

จากการศึกษาเอกสารพบว่า มีนักวัดผลการศึกษาทั้งในและต่างประเทศได้กล่าวถึง ความหมาย และความสำคัญของความเชื่อมั่นของผู้ตรวจให้คะแนนแบบทดสอบความเรียงไว้หลายท่าน ดังนี้

ฮอบกินส์ และ แอนเตส (Hopkins and Ants, 1990 : 550) ให้ความหมายของความเชื่อมั่นของผู้ตรวจให้คะแนนที่มีตั้งแต่สองคนขึ้นไป จะมีความเห็นสอดคล้องกันเกี่ยวกับจุดเน้น หรือจุดสำคัญต่าง ๆ ของ

คำตอบที่ผู้ตรวจให้คะแนนตามความเห็นของผู้ตรวจแต่ละคน ค่าความเชื่อมั่นของผู้ตรวจสูงหมายถึง สัมประสิทธิ์สหสัมพันธ์สูงตามไปด้วย ซึ่งผู้ตรวจข้อสอบที่ดีได้ ควรเป็นผู้ที่เขียนข้อสอบหรือผู้ที่สอนวิชานั้น ๆ จริง ๆ (Hopkins and Ants. 1990 : 235)

อูเบอร์แซก (Uebersax. 2000) ได้เสนอดัชนีความสอดคล้องจากคะแนนดิบ (Raw Agreement Indices) ไว้ 3 กลุ่ม คือ

1. ผู้ตรวจ 2 คน ตรวจให้คะแนนแบบ 2 ขั้ว (Two Raters, Dichotomous Ratings)
2. ผู้ตรวจ 2 คน ตรวจให้คะแนนแบบหลายขั้ว (Two Raters, Polytomous Ratings)
3. กรณีทั่วไป (Generalized Case)

ในที่นี้ผู้วิจัยสนใจศึกษาวิธีการของ กลุ่ม 1 และ 2 ดังนี้

1. ผู้ตรวจ 2 คน ตรวจให้คะแนนแบบ 2 ขั้ว (Two Raters, Dichotomous Ratings)

การตัดสินใช้คะแนนจากผู้ตรวจจำนวน 2 คน (หรือ ผู้เชี่ยวชาญ , ผู้ตัดสิน , ผู้วินิจฉัย หรือ อื่น ๆ) มารวมกัน ดังตารางต่อไปนี้

ตาราง 3 ผลรวมของการตรวจให้คะแนนแบบ 2 ขั้ว จากผู้ตรวจ 2 คน

ผู้ตรวจคนที่ 1	ผู้ตรวจคนที่ 2		รวม
	-	+	
-	a	b	a + b
+	c	d	c + d
รวม	a + c	b + d	N

ค่า a , b , c และ d ในที่นี้เป็นผลรวมจากความถี่ของการให้คะแนนของผู้ตรวจแต่ละคน คือผู้ตรวจคนที่ 1 และคนที่ 2 แบ่งได้ 2 ลักษณะ คือ

สัดส่วนของความคิดเห็นที่สอดคล้องกันทั้งหมด (Proportion of Overall Agreement)

สัดส่วนของคะแนนจากความคิดเห็นที่สอดคล้องกันทั้งหมด กำหนด p_o เป็นสัดส่วนของกรณีที่มีผู้ตรวจคนที่ 1 และ 2 ดังนี้

$$p_o = \frac{a+d}{a+b+c+d} = \frac{a+d}{N} \quad \text{..... (1)}$$

ผลการคำนวณ ถ้าค่า p_o สูง คือ มากกว่า .99 ก็จะสรุปได้ว่าความสอดคล้องระหว่างความคิดเห็นของผู้ตรวจไม่แตกต่างกัน

สัดส่วนของความคิดเห็นที่เฉพาะเจาะจง (Proportion of Specific Agreement)

สัดส่วนของความคิดเห็นทางบวก (p_S+) และความคิดเห็นทางลบ (p_S-) จะคำนวณได้จากสูตร ดังนี้

$$p_S+ = \frac{2a}{2a+b+c}; p_S- = \frac{2d}{2d+b+c} \quad \text{..... (2)}$$

สัดส่วนนี้แปลความหมายได้คล้ายกับการประมาณความเป็นไปได้ที่ตั้งเงื่อนไขไว้ ตัวอย่างคือ $p_S +$ เป็นค่าประมาณจากผู้ตรวจที่ให้คะแนนทางบวกได้คล้าย ๆ กับผู้ตรวจคนอื่น

2. ผู้ตรวจ 2 คน ตรวจให้คะแนนแบบหลายขั้ว (Two Raters, Polytomous Ratings)

ใช้การตัดสินจากผู้ตรวจ 2 คน ที่ให้คะแนนแบบหลายขั้ว โดยให้ C เป็นจำนวนของกลุ่มหรือระดับในการให้คะแนน ผลของการให้คะแนนจากผู้ตรวจ 2 คน จะถูกรวมตามวิธี C x C ดังตารางต่อไปนี้

ตาราง 4 ผลรวมของการให้คะแนนแบบหลายขั้ว โดยผู้ตรวจ 2 คน

ผู้ตรวจคนที่ 1	ผู้ตรวจคนที่ 2			รวม
	1	2	...C	
1	n_{11}	n_{12}	n_{1C}	$n_{1.}$
2	n_{21}	n_{22}	n_{2C}	n_{2}
.	.	.	.	.
.	.	.	.	.
.	.	.	.	.
C	n_{C1}	n_{C2}	n_{C3}	$n_{C.}$
รวม	$n_{.1}$	$n_{.2}$	$n_{.C}$	N

ให้ n_{ij} แทน จำนวนของกรณีที่ได้วางแผนให้คะแนนชุด i จากผู้ตรวจคนที่ 1
และ ชุด j โดยผู้ตรวจคนที่ j เมื่อ i , j = 1, ..., C

. แทน ค่าผลรวม (magical Sum)

เช่น $n_{j.}$ เป็นผลรวมของ n_{ij} ที่ j = 1,...,C

แบ่งได้ 2 ลักษณะ คือ

สัดส่วนของความคิดเห็นที่สอดคล้องกันทั้งหมด (Proportion of Overall Agreement)

การคำนวณ p_o ซึ่งก็คือผลรวมของความถี่จากการตรวจให้คะแนน ซึ่งคำนวณด้วยสูตร ดังนี้

$$p_o = \frac{1}{N}\sum_{i=1}^{C} n_{ij} \qquad \text{..... (3)}$$

สัดส่วนของความคิดเห็นที่แตกต่าง (Proportion of Specific Agreement)

ใช้คำตอบจากตาราง 2 สัดส่วนของความคิดเห็นที่แตกต่างกัน จากข้อมูลชุด i ใช้สูตรดังนี้

$$p_S(i) = \frac{2n_{ij}}{n_{i.} + n_{.j}} \qquad \text{.....(4)}$$

เจมส์ , และคนอื่น ๆ (James and others. 1984 ; 1993) ได้ทำการพัฒนาสัมประสิทธิ์ความเชื่อมั่นระหว่างผู้ตรวจ (Interrater reliability Coefficient (r_{wg}) ซึ่งจะประเมินความสอดคล้องระหว่างผู้ตรวจ โดยกำหนดให้ r_{wg} เป็นการลดสัดส่วนความแปรปรวนที่ผิดพลาดสำหรับการให้คะแนนที่ได้มา เพื่อเทียบกับการจัดแบ่งที่แสดงถึงการตอบแบบสุ่ม (ตัวอย่าง เช่น การแบ่งในกรณีที่แต่ละคะแนนในเกณฑ์ของอันดับคะแนนมีความเป็นไปได้เท่า ๆ กันในการนำไปใช้) กล่าวคือ

$$r_{wg} = 1 - (S_x^2 / \sigma_E^2) \quad \text{.....(1)}$$

โดยที่ S_X^2 เท่ากับค่าความแปรปรวนของผลการตรวจให้คะแนน และ σ_E^2 เท่ากับค่าความแปรปรวนของประชากร ในการจัดแบ่งอันดับคะแนนที่ได้จากการสุ่ม (เช่น การตรวจให้คะแนนที่เท่ากันของคำตอบที่เป็นไปได้แต่ละประเด็น) ค่าของ r_{wg} มีค่าตั้งแต่ 0.0 จนถึง 1.0 เมื่อความแปรปรวนของคะแนนที่ได้มาเป็นแบบสุ่ม (เช่น $S_X^2 = \sigma_E^2$) ค่า $r_{wg} = 0$ ซึ่งชี้ให้เห็นว่า ไม่มีความสอดคล้องกันเลย หรือ หาก ผลการตรวจให้คะแนนเหมือนกัน (เช่น $S_X^2 = 0$), ค่า $r_{wg} = 1$ ซึ่งชี้ให้เห็นว่ามีความสอดคล้องโดยสมบูรณ์ระหว่างผู้ตรวจทั้งหมด

ค่า r_{wg} ที่ได้ควรจะถือเป็นดัชนีบ่งบอกความสอดคล้อง (Index of Agreement) ไม่ใช้ค่าความคงที่ (Consistency) เจมส์ (James. 1993) อธิบายว่า การให้คะแนนบางประเภท อาจมีระดับความคงที่สูง (เช่น ผู้ตรวจ 2 คน หรือมากกว่านั้นให้คะแนนออกมาคล้ายคลึงกัน เมื่อให้คะแนนตัวอย่างหลาย ๆ ตัวอย่างบนมาตรฐานเดียวกัน) แต่ค่าการให้คะแนนที่แท้จริงที่ให้ผู้ตรวจคนเดียวอาจจะต่างกันมาก ให้พิจารณาถึงกรณีที่ผู้ตรวจคนที่ 1 ให้คะแนน “4” ทุกครั้ง ผู้ตรวจคนที่ 2 ให้คะแนน “2” และทุกครั้งที่ผู้ตรวจคนที่ 1 ให้คะแนน “3” ผู้ตรวจคนที่ 2 ให้คะแนน “1” ค่าสหสัมพันธ์ระหว่างความคงที่ระหว่างผู้ตรวจ 2 คน นั้นจะสูงมาก แต่ ระดับความสอดคล้องระหว่างผู้ตรวจนี้จะต่ำ เพราะว่าพวกเขาจะไม่ให้ค่าที่เป็นคะแนนซ้ำกันเหมือนเดิม ข้อดีของสัมประสิทธิ์ r_{wg} ก็คือ สัมประสิทธิ์ที่ได้มาจะแปรผันได้ตามค่าสมบูรณ์ในการให้คะแนน เช่นเดียวกับการจัดอันดับคะแนนจากการตรวจให้คะแนน

อนันต์ ศรีโสภา (2524 : 58) กล่าวว่าโดยทั่วไป การให้คะแนนแบบทดสอบความเรียง (Essay Test) เรามักจะไม่ค่อยคิดถึงเรื่องความเชื่อมั่นของการให้คะแนนของครูมากนัก ซึ่งข้อสอบแบบความเรียง มักจะมีความคลาดเคลื่อนเนื่องจากการให้คะแนนของครูอยู่เป็นอันมาก อาทิเช่น ถ้านำกระดาษคำตอบ ไปให้ครูสองคนตรวจให้คะแนน คะแนนที่ได้ในแต่ละข้ออาจแตกต่างกัน หรือแม้ให้ครูคนเดียวกันตรวจ ถ้าตรวจต่างเวลากัน คะแนนที่ได้ในแต่ละข้อก็แตกต่างกันอีก เป็นต้น ดังนั้น จึงมีความจำเป็นอย่างยิ่งจะต้องคำนวณหาความเชื่อมั่นของผู้ให้คะแนนในกรณีที่ข้อสอบนั้นเป็นอัตนัย ซึ่งจะช่วยให้การตีความหมายของคะแนนที่ให้นั้นมีความถูกต้องมากยิ่งขึ้น ดังนั้น วิธีการหาความเชื่อมั่นของผู้ตรวจให้คะแนน 2 คน ที่ต้องหาค่าเฉลี่ยของผู้ตรวจให้คะแนนทั้ง 2 คน ก่อน แล้วใช้สูตรของ Spearman – Brown และถ้ามีผู้ตรวจให้คะแนนมากกว่า 2 คน และ การให้คะแนนเป็นแบบจัดตำแหน่ง (Rank) ให้ใช้สูตรของ ของ Guilford (1954)

$$\bar{r} = 1 - \frac{k(4N+2)}{(k-1)(N-1)} + \frac{12\sum S^2}{k(k-1)N(N^2-1)}$$

เมื่อ $\bar{r}$ แทน ความเชื่อมั่นเฉลี่ยระหว่างผู้ตรวจให้คะแนนแต่ละคน

k แทน จำนวนผู้ให้คะแนน

N แทน จำนวนนักเรียนที่เข้าสอบ

S แทน ผลบวกของตำแหน่งที่คะแนนของนักเรียนแต่ละคน

โกวิท ประวาลพฤกษ์ และ สมศักดิ์ สินธุระเวชญ์ (2525 : 173) กล่าวว่า การให้คะแนนแบบทดสอบความเรียง โดยผู้ตรวจให้คะแนนเพียงคนเดียว มักจะมีความคลาดเคลื่อนจากการให้คะแนน หรือถ้าจะใช้ผู้ตรวจให้คะแนน 2 คน การให้คะแนนของผู้ตรวจให้คะแนนทั้ง 2 คน ในเรื่องเดียวกันก็ยังคงมีความแตกต่างอยู่อีก ดังนั้น การคำนวณหาค่าความเชื่อมั่นของผู้ตรวจให้คะแนนให้คะแนนย่อมจะช่วยให้การตีความหมายของคะแนนมีความถูกต้องมากยิ่งขึ้น

3. วิธีการหาความสอดคล้องระหว่างผู้ตรวจให้คะแนน แบบ RAI

เบอร์รี่ สตอกค์ และคนอื่น ๆ (Burry-Stock and others. 1996) ได้ศึกษาดัชนีความสอดคล้องระหว่างผู้ตรวจให้คะแนน (Rater Agreement Indexes : RAI) ได้ผลดังนี้

<u>ความสอดคล้องระหว่างผู้ตรวจให้คะแนน</u> (Interrater Agreement)

การลงความเห็นระหว่างผู้ตรวจให้คะแนนถูกอ้างถึงระดับการตัดสินในการให้คะแนน พฤติกรรมที่ต้องการศึกษา เป็นความแตกต่างทางความคิดตามวิธีการทางการวัดทางจิตวิทยาเรื่องความเชื่อมั่น ในความคิดความเข้าใจของการลงความเห็นระหว่างผู้ตรวจให้คะแนน เป็นความจำเป็นทางการวัดผลทางจิตวิทยาที่นำข้อมูลเป็นรายบุคคลและคะแนนรวมในวิชาต่าง ๆ ที่มีจุดมุ่งหมายกำหนดไว้ตามวิธีการของการแสดงความคิดเห็นของผู้ตรวจให้คะแนน มีความตั้งใจในการนำมาใช้กับผู้ฝึกหัด หรือนักศึกษา หรือ นักจิตวิทยาวัดผล และนักวิจัย

การศึกษาสามารถสรุปวิธีการคำนวณได้จากผลการสังเกตพฤติกรรมในระดับ I points หรือระดับช่วงชั้น ที่มีพิสัยของดัชนีจากค่าใหญ่ที่สุด คือ 1 ไปหาค่าน้อยที่สุด คือ 0 จากความคิดเห็นของผู้ตรวจให้คะแนนหลาย ๆ คนได้ 6 วิธี คือ

1. RAI สำหรับข้อสอบ 1 ข้อ , ผู้ตรวจให้คะแนน 2 คน , สำหรับ 1 เนื้อหา เป็นสูตรคำนวณจากผลการให้คะแนนของผู้ตรวจให้คะแนน 2 คน ที่ให้คะแนนในพฤติกรรมเดียวด้วยกัน ที่มีข้อมูลแบบ I ระดับ (I points) หรือช่วงชั้น ถ้าผู้ตรวจให้คะแนน 2 คน ให้คะแนนเหมือนกัน ค่า RAI = 1 ถ้าผู้ตรวจให้คะแนน 2 คนให้คะแนนแตกต่างกัน /ไม่เหมือนกันเลย ค่า RAI=0

2. RAI ที่ใช้ผลรวมของคะแนนที่มีผู้ตรวจให้คะแนน 2 คน , 1 เนื้อหา และ K พฤติกรรม ใช้คะแนนที่ได้จากการสังเกตสำหรับ k พฤติกรรมของผู้ตรวจให้คะแนน 2 คน สูตรนี้สามารถแสดงค่า RAI สำหรับพฤติกรรม k

3. RAI ที่ใช้ผลรวมของคะแนนที่มีผู้ตรวจให้คะแนน 2 คน , N เนื้อหา , K พฤติกรรม สำหรับผลการให้คะแนนจากผู้ตรวจให้คะแนน 2 คนที่ให้คะแนน N เนื้อหา ใน K พฤติกรรมที่เหมือนกัน และต่อเนื่องกัน

สำหรับการตรวจที่มีจำนวนผู้ให้คะแนนมีมากกว่า 2 คน หรือมีผู้ให้คะแนน M คน ให้ $\overline{R}$ แสดงค่าเฉลี่ยจากผู้ให้คะแนน M คน

4. RAI สำหรับข้อสอบข้อเดียว , ผู้ตรวจให้คะแนนหลายคน , 1 เนื้อหา มีที่มาของคะแนนจากผู้ตรวจให้คะแนน M คน ที่ให้คะแนนที่เหมือนกันใน 1 พฤติกรรม ที่ต่อเนื่องกันที่ I ระดับ

5. RAI สำหรับผู้ตรวจให้คะแนน M คน , K พฤติกรรม , 1 เนื้อหา ซึ่งหาค่าเฉลี่ยของพฤติกรรมที่ k ของ RAIs ของผู้ตรวจให้คะแนน M คน ใน 1 พฤติกรรม (นั่นคือ RAI ของผู้ตรวจให้คะแนน M คน ใน 1 พฤติกรรม สามารถหาค่าเฉลี่ยของผู้ตรวจให้คะแนน 2 คน) RAIs ที่ผู้ตรวจให้คะแนน 2 คน หรือแบบคู่ แสดงได้ถึงข้อเปรียบเทียบกับผู้ตรวจให้คะแนนคนเดียวกับค่าเฉลี่ยกับผู้ให้คะแนนอื่น ๆ

6. RAI สำหรับผู้ตรวจให้คะแนน M คน, K พฤติกรรม , N เนื้อหา ให้ R_{mnk} เป็นคะแนนของผู้ให้คะแนน M คน ในการตรวจ k พฤติกรรม จากจำนวน N^{th} เนื้อหา

จากข้อมูลการศึกษาค้นคว้าเอกสารทั้งหมดนี้ ผู้วิจัยได้นำสูตรที่ 6 มาใช้ในการวิเคราะห์ข้อมูล ดังมีรายละเอียดของสูตร ดังนี้

$$RAI = 1 - \frac{\sum^{K}\sum^{N}\sum^{M}\left|R_{mnk} - \overline{R}_{nk}\right|}{KN(M-1)(I-1)} \quad \text{.....(6)}$$

โดยที่ $\overline{R}_k = \frac{1}{M}\sum^{M} R_{mk}$

เมื่อ R แทน ผลการตรวจของผู้ตรวจให้คะแนน
M แทน จำนวนผู้ตรวจให้คะแนนทั้งหมด
N แทน จำนวนนักเรียนในกลุ่มตัวอย่าง
K แทน จำนวนพฤติกรรมย่อย
I แทน จำนวนช่วงคะแนน (0,1,2,3,4,5)

ข. งานวิจัยที่เกี่ยวข้อง

1. งานวิจัยในต่างประเทศ

งานวิจัยเกี่ยวกับการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจกำลังเริ่มเป็นที่สนใจในวงการศึกษาของต่างประเทศ ซึ่งผู้วิจัยได้รวบรวมตามหัวข้อต่อไปนี้

1.1 งานวิจัยที่เกี่ยวข้องกับความเชื่อมั่นของผู้ตรวจ (Rater Reliability)

เทสส์ และคนอื่น ๆ (Tasse and others. 1994) ศึกษางานวิจัยเกี่ยวกับ ความสอดคล้องของผู้ตรวจ เรื่อง แบบทดสอบซ้ำ และ ความเชื่อมั่นระหว่างผู้ตรวจแบบมาตรวัดพฤติกรรมการปรับตัวของ ดิวเบก (Quebec) ซึ่งได้รับการนำไปใช้อย่างแพร่หลายในเมืองดิวเบก (ประเทศแคนาดา) เพื่อประเมินพฤติกรรมของผู้ที่มีระดับสติปัญญาต่ำในสภาพแวดล้อมเกี่ยวกับการศึกษา การทำงาน การอยู่อาศัย หรือ สภาพแวดล้อม

ในโรงพยาบาล การศึกษานี้ประเมินความเชื่อมั่นของผู้ตรวจและความเชื่อมั่นโดยวิธีสอบซ้ำของ QABS และเพื่อวัดความเชื่อมั่นของการสอบซ้ำ QABS จึงได้รับการประเมินโดยผู้ตรวจจำนวน 27 คน กับกลุ่มตัวอย่างจำนวน 45 คน (อายุระหว่าง 6 ถึง 47 ปี) จากบ้านเรือน สถานที่พักอาศัย โรงเรียน และโรงพยาบาลต่าง ๆ และเพื่อให้ได้ค่าความเชื่อมั่นของผู้ตรวจ ได้ใช้กลุ่มตัวอย่างอีกกลุ่มจำนวน 44 คน (อายุระหว่าง 3 ถึง 54 ปี) แต่ละคนจะได้รับการประเมินจากผู้ตรวจ 2 คน ค่าสัมประสิทธิ์ความเชื่อมั่นระหว่างการตรวจโดยใช้วิธีประเมินภาพรวม (Global Score) กับระดับพฤติกรรมการปรับตัว มีค่าเท่ากับ .92 คำนวณโดยวิธีสอบซ้ำ และระหว่างผลรวมของคะแนนดิบ มีค่าเท่ากับ .99 ค่าสัมประสิทธิ์ความสอดคล้องของผู้ตรวจที่ใช้การตรวจแบบประเมินภาพรวม (Global Score) กับระดับพฤติกรรมการปรับตัว มีค่าเท่ากับ .83 และ ระหว่างผลรวมของคะแนนดิบ เท่ากับ .99 พบว่า QABS เป็นมาตรวัดพฤติกรรมการปรับตัวแบบอิงเกณฑ์ที่มีความเชื่อถือได้สูง มีเกณฑ์ปกติ และแบบเกณฑ์การให้คะแนนที่ชัดเจนง่ายต่อการใช้งาน

ไอโซนิโอ (Isonio. 1991) ศึกษางานวิจัยเกี่ยวกับความเชื่อมั่นของผู้ตรวจ เรื่อง ผู้ตัดสินตัวอย่างงานเขียนของวิทยาลัยโกลเดนเวสท์ : การประเมินความเชื่อมั่นของผู้ตรวจ งานวิจัยนี้ต้องการศึกษาตัวอย่างงานเขียนของนักเรียน โดยให้คะแนนแบบประเมินภาพรวม (Holistically) โดยใช้ผู้ตรวจเป็นรายคู่ ด้วยเกณฑ์การวัดแบบ 6 ระดับ (Six-point Scale) ผู้ตรวจต้องทำเครื่องหมาย + หรือ - วิธีนี้เป็นการเปลี่ยนจากการให้คะแนนแบบจำนวนเต็มมาเป็นแบบทศนิยม ในปี 1991 ตัวอย่างงานเขียนจำนวน 499 ตัวอย่าง ที่นำมาจากส่วนหนึ่งของการทดสอบเพื่อจัดอันดับสำหรับนักเรียนในโครงการ SOAR ของ Coast Community College District (เขตวิทยาลัยชุมชนแถบชายฝั่ง) ที่วิเคราะห์การให้คะแนนที่น่าเชื่อถือของผู้ตรวจ งานวิจัยนี้ใช้ 2 วิธี เพื่อประเมินความคงที่ในการตัดสินตัวอย่างงานเขียน วิธีแรกที่ใช้เป็นการคำนวณความแตกต่างระหว่างค่าบ่งบอกระดับตัวอย่างงานเขียน 2 ค่า มีค่าเท่ากับ 34.4 % ของคู่ที่ตัดสินทั้งหมด ค่าบ่งบอกระดับจากผู้ตัดสินแต่ละคนจะเหมือนกัน แต่ละค่าจะแตกต่างกันไป 1 ใน 3 ระดับ มี 30.9 % ของตัวอย่าง และ 2 ใน 3 ระดับ มี 15.4 % ของตัวอย่าง และ 1 ระดับ มี 13.4 % โดยรวม แล้วค่าบ่งชี้ระดับจะแตกต่างออกไป 1 ระดับหรือน้อยกว่า 94.1 % ของตัวอย่าง ซึ่งเกินมาตรฐานต่ำสุดที่ 90 % ที่แนะนำโดย California Community Colleges (CCC) วิธีที่สองที่ใช้ก็คือ สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน แสดงความสัมพันธ์ระหว่างการตรวจเป็นคู่ สำหรับตัวอย่างข้อเขียน ค่าความสัมพันธ์มีค่าเป็นบวก (r = .76) ซึ่งเกินมาตรฐานต่ำสุดที่ .75 ตามคำแนะนำของ CCC และ ข้อมูลที่ได้ยอมรับสมมติฐานที่ว่าการตรวจจากผู้ตรวจแต่ละคนได้ผลสอดคล้องกัน

1.2 งานวิจัยที่เกี่ยวข้องกับแบบทดสอบความเรียง

บริดแมน และคนอื่น ๆ (Bridgeman and others. 1997) ได้ศึกษางานวิจัยที่เกี่ยวข้องกับแบบทดสอบความเรียง เรื่อง การเลือกหัวข้อในการเขียนความเรียง : ผลกระทบของการวัดภาคปฏิบัติ และค่าความเที่ยงตรง งานวิจัยนี้มีจุดมุ่งหมายต้องการประเมินความสามารถของนักเรียนเอกประวัติศาสตร์ ในการเลือกหัวข้อการเขียนความเรียงที่เขาคิดว่าจะทำคะแนนได้สูงสุด 2 คำถามที่ใกล้เคียงกัน จะมีคะแนนสัมพันธ์กับการสอบโดยไม่มีโอกาสในการเลือกหัวข้อในการตอบคำถามทั้งหมดมาก่อน กลุ่มตัวอย่างเป็นนักเรียนเอกประวัติศาสตร์อเมริกา และเอกประวัติศาสตร์ยุโรป คะแนนที่ได้มา 1 ใน 3 ให้ค่าความเบี่ยงเบนมาตรฐานสูง นักเรียนในกลุ่มประวัติศาสตร์อเมริกา ประมาณ 32 % เลือกหัวข้อผิดพลาด เพราะทั้ง 32 % นั้นกลับไปได้คะแนนในหัวข้ออื่นที่ไม่ได้เลือกเองสูงกว่า ส่วนนักเรียนกลุ่มประวัติศาสตร์ยุโป 29 % ก็เช่นเดียวกัน กลุ่มประวัติศาสตร์อเมริกามีค่าสหสัมพันธ์ = .40 กับเกณฑ์ภายนอก และเมื่อเปรียบเทียบการเขียน

ความเรียงหัวข้ออื่น ๆ ให้ค่าสหสัมพันธ์ = .34 ส่วนในกลุ่มประวัติศาสตร์ยุโรป ให้ค่าสหสัมพันธ์ = .52 กับเกณฑ์ภายนอก และเมื่อเปรียบเทียบกับการเขียนความเรียงหัวข้ออื่น ๆ ให้ค่าสหสัมพันธ์ = .44

การ์เนอร์ และ เอนเกลฮาร์ด (Garner and Engelhard. 1999) ได้ศึกษางานวิจัยเรื่อง ความแตกต่างของเพศในการสอบปฏิบัติโดยใช้แบบทดสอบปรนัย และแบบทดสอบความเรียงวิชาคณิตศาสตร์ งานวิจัยนี้มุ่งศึกษาผลของความแตกต่างของเพศในการสอบปฏิบัติที่ใช้แบบทดสอบปรนัยและความเรียงวิชาคณิตศาสตร์ ระดับมัธยมศึกษาตอนปลาย โดยสร้างขึ้นตามหลักสูตรที่กำหนดขึ้น จุดมุ่งหมายที่สองเพื่อแสดงบทพิสูจน์วิธีการแบบฟาเซทแบบราชท์โมเดล (1980) สำหรับเปรียบเทียบการปฏิบัติที่มีรูปแบบแตกต่างกันและค้นหารูปแบบที่มีฟังชั่นต่างกัน (DIF) กลุ่มตัวอย่างที่ใช้เป็นนักเรียนระดับ 11 จำนวน 3,952 คน ที่เข้าเรียนในปี 1994 และผ่านการทดสอบด้วยแบบทดสอบ Georgia High School Graduation Test ในการวิเคราะห์ผลจากทั้งสองกลุ่มได้ค่าเฉลี่ยและค่า DIF ของเพศหญิง แสดงความแตกต่างอย่างมีนัยสำคัญ และมีความคงที่ภายในสูงกว่านักเรียนชาย เมื่อใช้แบบทดสอบแบบปรนัยวิชาพีชคณิตและการวัด , ตัวเลข และการคำนวณ ,การวิเคราะห์ข้อมูล และเหตุผลเฉพาะ คะแนนเฉลี่ยของเพศชายให้ค่าแตกต่างอย่างมีนัยสำคัญสูงกว่าเพศหญิง 2 ใน 8 ข้อของข้อสอบแบบอัตนัย อย่างไรก็ตาม เมื่อเพศชายและเพศหญิงก็ให้ค่าสอดคล้องกัน ส่วนด้านความสามารถ ให้ค่าแตกต่างกันอย่างมีนัยสำคัญเมื่อปฏิบัติโดยใช้แบบทดสอบอัตนัย สรุปได้ว่า ความแตกต่างของเพศในการทดสอบคณิตศาสตร์มีคุณภาพดีหรือไม่ขึ้นอยู่กับเนื้อหาและรูปแบบของข้อสอบ

ปอมพลุน และ ซันบาย (Pomplun and Sundbye. 1999) ได้ศึกษางานวิจัย เรื่อง ความแตกต่างของเพศ ของแบบทดสอบแบบขยายความ วิชาการอ่าน (Constructed Response Reading Items) การศึกษาครั้งนี้ต้องการค้นหาความแตกต่างของเพศในการตอบแบบทดสอบการอ่านจากโครงการประเมินของรัฐบาลแยกออกเป็นองค์ประกอบที่สัมพันธ์เชิงโครงสร้างและไม่สัมพันธ์กัน จัดให้เหมือนกัน โดยผู้เชี่ยวชาญทางเนื้อหาวิชา ที่เกี่ยวกับวิชาวรรณคดี และวิชานำร่อง (A Pilot Study) ใช้ผู้ตรวจ 4 คน ที่ได้รับการฝึกอบรมสำหรับการให้คะแนน ใช้กระดาษคำตอบจำนวน 500 แผ่น แบ่งออกเป็น 2 ระดับ และ แบ่งเป็นระหว่างเพศประกอบด้วย ส่วนคำตอบ (เฉลย) , ส่วนคำตอบที่ไม่สัมพันธ์กัน , คำตอบผิด , จำนวนคำ ที่เขียน และวิธีการวัดที่สับสน แบบฟอร์มการให้คะแนนของรัฐบาล ความสัมพันธ์ระหว่างสิ่งต่าง ๆ ที่ต้องการศึกษาเป็นการตรวจแบบประเมินรวม , การให้คะแนนย่อย , ความแตกต่างของเพศที่ต้องการค้นหาความสัมพันธ์ ความคิดและการวิเคราะห์แบบถดถอย ผลที่ได้คือ จำนวนคำและจำนวนคำตอบที่ไม่สัมพันธ์กัน กับความแตกต่างระหว่างเพศ เมื่อใช้วิธีการตรวจแบบประเมินรวม แสดงค่าทางสถิติแตกต่างกันอย่างมีนัยสำคัญ และค่าทางสถิติแตกต่างกันอย่างมีนัยสำคัญ เมื่อศึกษาตัวแปรควบคุม ซึ่งงานนี้ต้องการศึกษาสิ่งที่ส่งผลกระทบต่อความแตกต่างของผู้ตรวจที่มีต่อนักเรียนชายและนักเรียนหญิง

เพาเวอร์ และโฟวเลส (Powers and Fowles. 1998) ได้ทำการศึกษางานวิจัย เรื่อง ผลของการเปิดเผยหัวข้อของการสอบแบบความเรียงก่อนการสอบ เพื่อวัดผลของแบบวัดภาคปฏิบัติและความเที่ยงตรงของเครื่องมือ โดยการบอกหัวข้อของข้อสอบก่อนการทดสอบจริง ผู้ถูกวิจัยที่ตอบรับเข้าร่วมจำนวน 300 คน ทั้งหมดเป็นนักเรียนที่ใกล้จบการศึกษามาตอบข้อสอบ 2 ลักษณะ คือ 1 ตอบโดยไม่เคยรู้หัวข้อข้อสอบมาก่อน และอีก 1 ตอบโดยรู้หัวข้อของข้อสอบก่อนสอบ 2 สัปดาห์ รวมแล้วแต่ละคนจะต้องข้อสอบ 4 ข้อ การวิจัยนี้ผลการวิเคราะห์ไม่ปรากฏความแตกต่างอย่างมีนัยสำคัญทางสถิติ ระหว่างผลการเปิดเผยหัวข้อก่อนปฏิบัติกับการปฏิบัติโดยมีทักษะการเขียน ข้อจำกัดของการศึกษาถูกอภิปรายในภาพของกลุ่มทฤษฎีสรุปอ้างอิง ที่มีสถานการณ์เสี่ยงสูงและขอบเขตที่จะให้ค่าที่ได้เป็นบวกและลบได้

1.3 งานวิจัยต่างประเทศที่เกี่ยวข้องกับวิธีการตรวจให้คะแนน

ไครท และสทีสเชอร์ (Klien and Stecher. 1998) ศึกษางานวิจัยเกี่ยวกับวิธีการตรวจให้คะแนน เรื่อง การตรวจแบบวิเคราะห์ย่อยกับแบบประเมินรวมสำหรับการทดสอบภาคปฏิบัติวิชาวิทยาศาสตร์ งานวิจัยนี้ต้องการศึกษา 2 ส่วน คือ หาความสอดคล้องภายในของผู้ตรวจ (Interreader) , ความเชื่อมั่นของการตรวจให้คะแนน (Score Reliability) และความต้องการเวลาในการตรวจของผู้ตรวจจำนวน 3 คน สำหรับการสอบภาคปฏิบัติวิชาวิทยาศาสตร์ การศึกษาแรกเกี่ยวกับการให้คะแนนคำตอบที่ได้จากนักเรียนเกรด 5 , 8 และ 10 ใน 3 จุดมุ่งหมาย ("หลักสูตรมาตรฐาน") โดยวิธีการปฏิบัติ การศึกษาต่อมา คือต้องการคำนวณคะแนนทั้ง 3 ระดับที่ได้มา การศึกษาทั้ง 2 ข้อแรกใช้การตรวจแบบวิเคราะห์ย่อยและการประเมินรวม เพื่อให้ระดับคำตอบที่ได้แตกต่างกันในด้านเกณฑ์ย่อยของการให้คะแนนเท่านั้น การตรวจแบบวิเคราะห์ย่อยจะใช้เวลามากแต่ก็ยังให้ค่าความสอดคล้องภายในของผู้ตรวจสูง แต่กระนั้นเมื่อทำการเฉลี่ยจากทุก ๆ คำถามของเครื่องมือแล้ว คะแนนของนักเรียนที่ตรวจแบบประเมินรวมให้ค่าความเชื่อมั่นเหมือนกับคะแนนที่ได้จากการตรวจแบบวิเคราะห์ย่อย แสดงว่า ทั้งสองวิธีมีความสัมพันธ์กันสูงเมื่อลดความสอดคล้องภายในระหว่างผู้ตรวจ โดยใช้ผู้ตรวจ 2 คน ต่อ 1 คำถาม จะไม่ปรากฏผลต่อค่าที่คำนวณได้

2. งานวิจัยภายในประเทศ

งานวิจัยภายในประเทศเกี่ยวกับ ความเชื่อมั่นของผู้ตรวจให้คะแนน นั้นยังพบน้อยมาก ที่ผู้วิจัยรวบรวมได้ มีดังนี้

พรรณี เจียมสุบุตร (2543) ได้ศึกษาเปรียบเทียบความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่มีจำนวนผู้ตรวจและวิธีการตรวจต่างกัน การศึกษาครั้งนี้มีจุดมุ่งหมายเพื่อเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่มีจำนวนผู้ตรวจ 2 คน และ 3 คน โดยใช้วิธีการตรวจวิธีประเมินรวม และวิธีวิเคราะห์ย่อย โดยศึกษากับกลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2542 ของโรงเรียนในสังกัดกรมสามัญศึกษา จังหวัดกรุงเทพฯ กลุ่ม 7 จำนวน 70 คน ซึ่งเลือกมาโดยการสุ่มแบบหลายขั้นตอน เครื่องมือที่ใช้ในการศึกษาเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 เรื่อง สมการและอสมการ ผลการศึกษาได้ผลดังนี้

1. ค่าความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ โดยการตรวจให้คะแนนด้วยวิธีการวิเคราะห์ที่มีจำนวนผู้ตรวจ 2 คน และ 3 คน มีค่าเท่ากับ 0.92 และ 0.96 ตามลำดับ ซึ่งแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนการตรวจให้คะแนนด้วยวิธีรวม ที่มีจำนวนผู้ตรวจ 2 และ 3 คน มีค่าเท่ากับ 0.97 และ 0.98 ตามลำดับ ซึ่งแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

2. ค่าความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ ที่ใช้วิธีการตรวจให้คะแนนแบบวิธีรวม และวิธีวิเคราะห์ของผู้ตรวจให้คะแนน 2 คน มีค่าเท่ากับ 0.97 และ 0.92 ตามลำดับ ซึ่งแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบที่ใช้วิธีการตรวจให้คะแนนวิธี

ประเมินรวม และวิธีการวิเคราะห์ ของผู้ตรวจให้คะแนน 3 คน มีค่าเท่ากับ 0.98 และ 0.96 ตามลำดับ ซึ่งแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .05

ไพรัตน์ วงษ์นาม (2533 : 105-123) ได้ศึกษาเปรียบเทียบผลของวิธีตรวจ การชี้แจงและการรู้ผลการเรียนของผู้ตอบ ที่มีต่อสัมประสิทธิ์การอ้างอิงสรุป (Generalizability Coefficient) ของการสอบด้วยแบบทดสอบความเรียง ที่วัดความสามารถในการแสดงความคิดเห็นเกี่ยวกับข่าว และเหตุการณ์ ของนักเรียนชั้นประถมปีที่ 6 จำนวน 30 คน ข้อสอบความเรียง 5 ข้อ และครูผู้ตรวจให้คะแนน จำนวน 20 คน สุ่มแบ่งผู้ตรวจให้คะแนนเป็น 2 กลุ่ม ๆ ละ 10 คน สำหรับการตรวจ 2 วิธี คือวิธีประเมินรวม และวิธีวิเคราะห์ ในแต่ละวิธีทำการตรวจ 3 ครั้ง คือ ครั้งแรก ตรวจโดยใช้ประสบการณ์เดิมของผู้ตรวจให้คะแนน ครั้งที่สอง ตรวจตามเกณฑ์ที่ได้รับการชี้แจงจากผู้วิจัยโดยไม่รู้ผลการเรียนของผู้ตอบ ครั้งที่สาม ตรวจตามเกณฑ์ที่ได้รับการชี้แจงจากผู้วิจัยเหมือนครั้งที่สอง แต่รู้ผลการเรียนของผู้ตอบด้วย พบว่า เมื่ออ้างอิงสรุปไปยังเอกภาพของข้อสอบและผู้ตรวจให้คะแนนพร้อมกัน ค่าสัมประสิทธิ์การอ้างอิงสรุป (ความเชื่อมั่น) ของแบบสอบความเรียงที่ตรวจโดยวิธีประเมินรวม มีค่าอยู่ระหว่าง 0.3328 – 0.4782 และวิธีวิเคราะห์ มีค่าอยู่ระหว่าง 0.3348 – 0.5895 และใช้ข้อสอบ 5 ข้อ ผู้ตรวจให้คะแนนอย่างน้อย 5 คน ใช้การตรวจวิธีวิเคราะห์ ตามที่ได้รับการชี้แจงและไม่รู้ผลการเรียนของผู้ตอบ จึงจะให้ค่าสัมประสิทธิ์การอ้างอิงสรุปไม่น้อยกว่า 0.50 เมื่ออ้างอิงสรุปไปยังเอกภพของผู้ตรวจให้คะแนนอย่างเดียว ค่าสัมประสิทธิ์การสรุปอ้างอิง (ความเชื่อมั่น) ที่ตรวจโดยวิธีประเมินรวม มีค่าอยู่ระหว่าง 0.4743 –0.6865 และวิธีวิเคราะห์ มีค่าอยู่ระหว่าง 0.5985 – 0.7761 และใช้ข้อสอบ 5 ข้อ ผู้ตรวจให้คะแนน 5 คน ในการตรวจวิธีวิเคราะห์ตามที่ได้รับการชี้แจง และ ไม่รู้ผลการเรียนของผู้ตอบ จึงจะให้ค่าสัมประสิทธิ์การสรุปอ้างอิง มีค่าไม่น้อยกว่า 0.50 เมื่ออ้างอิงไปยังเอกภพของผู้ตรวจให้คะแนนอย่างเดียว พบว่า เฉพาะกรณีที่ผู้ตรวจให้คะแนนได้รับการชี้แจงและไม่รู้ผลการเรียนของผู้ตรวจให้คะแนนเท่านั้น ที่ให้ค่าสัมประสิทธิ์การสรุปอ้างอิง สูงกว่าการตรวจโดยใช้ประสบการณ์เดิม ถ้ารู้ผลการเรียนของผู้ตอบค่าสัมประสิทธิ์การอ้างอิงสรุปจะไม่แตกต่างกัน

ดำเกิง วงศ์สว่าง (2528) ได้ศึกษาผลของการทดสอบย่อยโดยใช้แบบทดสอบอัตนัย เติมคำ และเลือกตอบ ที่มีต่อความสามารถในการเขียนภาษาไทย ของนักเรียนชั้นประถมศึกษาปีที่ 3 โรงเรียนสามเสนนอก จำนวน 144 คน ผลการวิจัยพบว่า ในส่วนการตรวจให้คะแนนแบบทดสอบวัดความสามารถในการเขียนภาษาไทย 1 ฉบับ จำนวน 3 ข้อ ของผู้ตรวจ 2 คน โดยกำหนดเกณฑ์การให้คะแนนยึดตามแนวของ ประสิทธิ์ กาพย์กลอน นั้น ข้อสอบข้อ 1 ข้อ 2 และข้อ 3 มีค่าสัมประสิทธิ์สหสัมพันธ์เท่ากับ 0.8682 0.8849 และ 0.9454 ตามลำดับ และเมื่อทดสอบนัยสำคัญ พบว่า มีนัยสำคัญทางสถิติที่ระดับ .01 ซึ่งแสดงว่าการตรวจให้คะแนนของผู้ตรวจ 2 คน มีความเชื่อถือได้

จิราพร อัครสมพงศ์ (2536) ได้ศึกษาเปรียบเทียบความสอดคล้องของคะแนนแบบสอบความเรียงจากผู้ตรวจให้คะแนนและวิธีการตรวจต่างกัน ที่มีวัตถุประสงค์เพื่อ ศึกษาเปรียบเทียบความสอดคล้องในการให้คะแนนของผู้ตรวจให้คะแนน 2 คน ที่ตรวจแบบสอบความเรียงโดยใช้วิธีเดียวกัน และความสอดคล้องในการให้คะแนนระหว่างวิธีการตรวจที่แตกต่าง พบว่าความสัมพันธ์ในการให้คะแนนระหว่างผู้ตรวจให้คะแนนด้วยวิธีเดียวกัน ผู้ตรวจให้คะแนนที่ตรวจด้วยวิธีเดียวกันมีความสัมพันธ์กันสอดคล้องกัน ในทิศทางเดียวกัน และ มีค่าสัมประสิทธิ์สหสัมพันธ์ของการตรวจแบบวิธีวิเคราะห์ให้เสร็จเป็นรายข้อ กับการตรวจแบบวิธีประเมินรวมให้เสร็จเป็นรายข้อมีค่าสูงสุด คือ .8116

อุษณีย์ บัวศิริพันธุ์ (2543) ศึกษาเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจ จำนวนผู้ตรวจให้คะแนน และประสบการณ์ของผู้ตรวจให้คะแนน แตกต่างกัน กรณีผลการศึกษาเกี่ยวกับวิธีการตรวจให้คะแนน การตรวจแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนนของผู้ตรวจให้คะแนนเอง เมื่อมีจำนวนผู้ตรวจให้คะแนน 3 , 4 , 5 และ 6 คน มีคะแนนเฉลี่ยอยู่ระหว่าง 8.79 – 9.05 โดยการตรวจให้คะแนนเมื่อมีจำนวนผู้ตรวจให้คะแนน 4 คน คะแนนเฉลี่ยมีแนวโน้มสูงที่สุด รองลงมาคือ มีจำนวนผู้ตรวจให้คะแนน 5 คน และเมื่อมีจำนวนผู้ตรวจให้คะแนน 6 คน คะแนนเฉลี่ยมีแนวโน้มต่ำที่สุด ส่วนการตรวจให้คะแนนแบบประเมินรวมโดยใช้เกณฑ์การให้คะแนน เมื่อมีจำนวนผู้ตรวจให้คะแนน 3 , 4 , 5 และ 6 คน มีคะแนนเฉลี่ยอยู่ระหว่าง 9.82 –10.05 โดยการตรวจให้คะแนน เมื่อมีจำนวนผู้ตรวจให้คะแนน 4 คน คะแนนเฉลี่ยมีแนวโน้มสูงที่สุด รองลงมาคือ จำนวนผู้ตรวจให้คะแนน 3 และ 6 คน และ เมื่อมีจำนวนผู้ตรวจให้คะแนน 5 คน คะแนนเฉลี่ยมีแนวโน้มต่ำที่สุด

สุพานี กอหมั่นศิลป์ (2543) ศึกษาความเชื่อมั่นของแบบประเมินทักษะกีฬาบาสเกตบอล ที่ใช้ผู้ประเมินจำนวนต่างกัน ประเมินนักเรียนจำนวนต่างกัน ที่มีจุดมุ่งหมายเพื่อเปรียบเทียบค่าความเชื่อมั่นของแบบประเมินทักษะกีฬาบาสเกตบอลที่ใช้ผู้ประเมินจำนวนต่างกัน ประเมินนักเรียนจำนวนต่างกัน โดยกลุ่มตัวอย่างที่ใช้เป็นนักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 2 ปีการศึกษา 2542 จาก 4 โรงเรียน จำนวน 180 คน ในจังหวัดราชบุรี จากการสุ่มแบบสองขั้นตอน โดยแบ่งเป็น 3 กลุ่ม คือ กลุ่มนักเรียนเข้ารับการประเมินครั้งละ 1 คน จำนวน 30 คน กลุ่มเข้ารับการประเมินครั้งละ 2 คน จำนวน 60 คน กลุ่มเข้ารับการประเมินครั้งละ 3 คน จำนวน 90 คน ผลการวิจัยพบว่า การเปรียบเทียบค่าความเชื่อมั่นแบบประเมินทักษะกีฬาบาสเกตบอลที่ใช้ผู้ประเมินจำนวนต่างกัน คือ จำนวนหนึ่งคน สองคน และสามคน ประเมินนักเรียนครั้งละ 1 คน 2 คน และ 3 คน ค่าความเชื่อมั่นเฉลี่ยของผู้ประเมินจำนวน 1 คน มีค่าต่ำสุด และ ค่าความเชื่อมั่นเฉลี่ยของผู้ประเมิน จำนวน 3 คน มีค่าสูงสุด เมื่อทดสอบความแตกต่างของค่าความเชื่อมั่น พบว่า มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และการเปรียบเทียบค่าความเชื่อมั่นของแบบประเมินทักษะกีฬาบาสเกตบอลที่ประเมินนักเรียนจำนวนต่างกัน คือ ครั้งละหนึ่งคน สองคน และสามคน เมื่อใช้ผู้ประเมินจำนวน 1 คน 2 คน และ 3 คน ค่าความเชื่อมั่นเฉลี่ยของนักเรียนที่เข้ารับการประเมินครั้งละ 1 คน มีค่าต่ำสุด และค่าความเชื่อมั่นเฉลี่ยของนักเรียนที่เข้ารับการประเมินครั้งละ 3 คน มีค่าสูงสุด เมื่อทดสอบความแตกต่างของค่าความเชื่อมั่น พบว่า มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

จากการศึกษางานวิจัยที่เกี่ยวข้องทั้งภายในประเทศและต่างประเทศ จะเห็นได้ว่า แบบทดสอบความเรียงสามารถวัดทักษะด้านการคิด ความเข้าใจ การให้เหตุผล การวิเคราะห์ และ ความสามารถในการแก้ปัญหาได้ดี ซึ่งเหมาะที่จะนำมาใช้วัดและประเมินผลวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) นอกจากนี้จำนวนผู้ตรวจให้คะแนนก็มีผลให้ความเชื่อมั่นระหว่างผู้ตรวจแตกต่างกันไป ดังนั้น ผู้วิจัยจึงสนใจที่จะศึกษาหาความเชื่อมั่นของผู้ตรวจว่าจะแปรเปลี่ยนไปอย่างไรถ้ามีจำนวนผู้ตรวจให้คะแนนและใช้วิธีการตรวจต่างกัน

สมมุติฐานการวิจัย

1. ค่าความเชื่อมั่นของผู้ตรวจให้คะแนนแบบทดสอบความเรียงที่มีจำนวนผู้ตรวจสองคน สามคน และ สี่คน เมื่อใช้การตรวจวิธีเดียวกัน มีค่าแตกต่างกัน

2. ค่าความเชื่อมั่นของผู้ตรวจให้คะแนนแบบทดสอบความเรียงที่ใช้การตรวจวิธีวิเคราะห์ย่อย และการตรวจวิธีประเมินรวม เมื่อมีผู้ตรวจจำนวนเท่ากัน มีค่าแตกต่างกัน

บทที่ 3

วิธีดำเนินการวิจัย

การกำหนดประชากรและเลือกกลุ่มตัวอย่าง

ประชากรที่ใช้ในการวิจัย

ประชากรที่ใช้ในการวิจัย เป็นนักเรียนชั้นประถมศึกษาปีที่ 5 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาเอกชน สาขากลุ่มกรุงเทพฯ ตะวันออก เขตสาทร เขตบางรัก และเขตบางคอแหลม กรุงเทพมหานคร ในปีการศึกษา 2543 จำนวน 3,325 คน

กลุ่มตัวอย่างที่ใช้ในการวิจัย

กลุ่มตัวอย่างที่ใช้ในการวิจัย เป็นนักเรียนชั้นประถมศึกษาปีที่ 5 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาเอกชน สาขากลุ่มกรุงเทพฯ ตะวันออก เขตสาทร เขตบางรัก และเขตบางคอแหลม กรุงเทพมหานคร ในปีการศึกษา 2543 จำนวน 60 คน ที่ได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi – Stage Random Sampling) และเนื่องจากผู้วิจัยมีจุดมุ่งหมายในการวิจัย เพื่อศึกษาผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง เมื่อมีจำนวนผู้ตรวจและการตรวจให้คะแนนต่างวิธีกัน ดังนั้นการกำหนดขนาดกลุ่มตัวอย่างจึงคำนึงถึงความเหมาะสมเพื่อจะให้ได้ข้อมูลที่น่าเชื่อถือมาศึกษาวิเคราะห์ กลุ่มตัวอย่างผู้ตอบจึงควรมีขนาดเล็ก แต่ไม่ควรต่ำกว่า 30 หน่วย (ประคอง กรรณสูต. 2528:11) ผู้วิจัยจึงกำหนดขนาดของกลุ่มตัวอย่างที่เป็นผู้ตอบโดยเลือกโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาเอกชน สาขากลุ่มกรุงเทพ ฯ ตะวันออก มา 3 เขต คือ เขตสาทร เขตบางรัก และ เขตบางคอแหลม ซึ่งเป็นกลุ่มตัวอย่างที่มีคุณลักษณะใกล้เคียงกัน โดยเลือกแบบเจาะจงก่อนทำการสุ่มกลุ่มตัวอย่างแบบหลายขั้นตอน ซึ่งมีขั้นตอนการสุ่มกลุ่มตัวอย่าง ดังนี้

ขั้นที่ 1 สุ่มโรงเรียน จากทั้ง 3 เขต แต่ละเขตคิดเป็นร้อยละ 40 ได้โรงเรียนแต่ละเขต คือ เขตสาทร 2 โรงเรียน เขตบางรัก 2 โรงเรียน และเขตบางคอแหลม 1 โรงเรียน โดยวิธีการสุ่มอย่างง่าย ทั้งหมด 5 โรงเรียน

ขั้นที่ 2 สุ่มห้องเรียนจากโรงเรียนที่สุ่มได้ โรงเรียนละ 1 ห้องเรียน เป็น 5 ห้องเรียน แบ่งเป็นห้องเรียนที่ใช้ทดลองเครื่องมือ 2 โรงเรียน คือโรงเรียนสัจจพิทยา โรงเรียนเปรมฤดีศึกษา และห้องเรียนที่เป็นกลุ่มตัวอย่าง 3 โรงเรียน คือ โรงเรียนเซนต์หลุยส์ศึกษา โรงเรียนอัสสัมชัญแผนกประถม และ โรงเรียนอัสสัมชัญศึกษา

ขั้นที่ 3 สุ่มนักเรียนจากห้องเรียนที่เป็นกลุ่มตัวอย่าง ห้องเรียนละ 20 คน ได้จำนวนนักเรียนที่เป็นกลุ่มตัวอย่างทั้งหมด 60 คน ดังแสดงในตาราง 5

ตาราง 5 จำนวนนักเรียนที่ใช้เป็นกลุ่มตัวอย่างในการเก็บรวบรวมข้อมูล

เขต	โรงเรียน	กลุ่มตัวอย่าง
สาทร	โรงเรียนเซนต์หลุยส์ศึกษา	20
	โรงเรียนอัสสัมชัญแผนกประถม	20
บางรัก	โรงเรียนอัสสัมชัญศึกษา	20
	รวม	**60**

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยนี้ เป็นแบบทดสอบความเรียง 1 ฉบับ จำนวน 5 ข้อ ข้อละ 5 คะแนน เพื่อสอบวัดความรู้เนื้อหาในกลุ่มวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ชั้นประถมศึกษาปีที่ 5 ครอบคลุมเนื้อหาดังต่อไปนี้

หน่วยที่ 5 เรื่อง หัตถกรรม และอุตสาหกรรม จำนวน 2 ข้อ
หน่วยที่ 8 เรื่อง ประเทศเพื่อนบ้าน จำนวน 1 ข้อ
หน่วยที่ 12 เรื่อง ข่าว และเหตุการณ์ จำนวน 2 ข้อ

การสร้างเครื่องมือที่ใช้ในการวิจัย

ผู้วิจัยได้สร้างแบบทดสอบความเรียงเพื่อใช้ในการเก็บรวบรวมข้อมูล 1 ฉบับ จำนวน 5 ข้อ ตามขั้นตอนการสร้างแบบทดสอบความเรียง และพร้อมทั้งจัดทำคู่มือการตรวจ(เฉลย) มีรายละเอียดตามลำดับขั้นดังนี้



ภาพประกอบ ลำดับขั้นการสร้างแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต (สังคม)

จากภาพประกอบ แสดงลำดับขั้นการสร้างแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) มีรายละเอียดการดำเนินการสร้างตามขั้นตอน ดังนี้

1. กำหนดจุดมุ่งหมายในการสร้างแบบทดสอบ

1.1 เพื่อสร้างแบบทดสอบวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ที่มีลักษณะเป็นแบบทดสอบความเรียง จำนวน 5 ข้อ

1.2 เพื่อนำคะแนนจากแบบทดสอบไปใช้ในการศึกษาความเชื่อมั่นของผู้ตรวจที่มีจำนวนผู้ตรวจและวิธีการตรวจต่างกัน

2. ศึกษาทฤษฎีและเอกสารงานวิจัยที่เกี่ยวข้อง

2.1 ศึกษาทฤษฎีและเอกสารงานวิจัยที่เกี่ยวข้องกับความหมาย ลักษณะ ประเภทการสร้าง และวิธีการตรวจแบบทดสอบความเรียง

2.2 ศึกษาโครงสร้างหลักสูตร คู่มือครู กำหนดการสอน แผนการสอน และเอกสารการสอนวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ชั้นประถมศึกษาปีที่ 5 ตามหลักสูตรประถมศึกษา พ.ศ. 2521 (ฉบับปรับปรุง พ.ศ. 2533) ของกระทรวงศึกษาธิการ

2.3 ศึกษาวิธีการหาค่าความเชื่อมั่นของผู้ตรวจที่มีจำนวนตั้งแต่ 2 คนและมากกว่า 2 คน ตามวิธีของดัชนีความสอดคล้องของผู้ตรวจ (RAI)

3. วิเคราะห์หลักสูตร โดยสร้างตารางวิเคราะห์หลักสูตร คัดเลือกเนื้อหา และระบุจุดมุ่งหมาย ของการสร้างตามพฤติกรรมการวัดความรู้ขั้นสูงของบลูม (Bloom, 1977) คือ ความเข้าใจ นำไปใช้ วิเคราะห์ สังเคราะห์ และประเมินค่า และกำหนดน้ำหนักในแต่ละเนื้อหาว่าจะวัดตรงกับพฤติกรรมด้านใด ดังแสดงในตาราง 6

ตาราง 6 แสดงการวิเคราะห์หลักสูตรตามรายละเอียดของเนื้อหา

เนื้อหา	พฤติกรรม						รวม	อันดับความสำคัญ
	ความรู้-จำ	ความเข้าใจ	นำไปใช้	วิเคราะห์	สังเคราะห์	ประเมินค่า		
หน่วยที่ 5 เรื่อง หัตถกรรม และอุตสาหกรรม	3	6	5	8	5	6	33	2
หน่วยที่ 8 เรื่อง ประเทศเพื่อนบ้าน	6	6	6	6	4	4	32	3
หน่วยที่12 เรื่องข่าวและเหตุการณ์	2	4	7	8	6	8	35	1
รวม	11	16	18	22	15	18	100	
อันดับความสำคัญ	6	4	2	1	5	2		

4. สร้างเครื่องมือตามตารางวิเคราะห์หลักสูตร จำนวน 10 ข้อเพื่อทดสอบและหาคุณภาพรายข้อ และคัดเลือกให้เหลือเพียง 5 ข้อ และกำหนดให้แต่ละข้อมีคะแนนเต็ม 5 คะแนนเท่ากัน พร้อมทั้งจัดทำคู่มือการตรวจ (เฉลย) และสร้างเกณฑ์การตรวจให้คะแนนรายข้อ ทั้งวิธีการวิเคราะห์ กับวิธีการประเมินรวม เพื่อให้ผู้ตรวจใช้เป็นคู่มือการตรวจทั้ง 2 วิธี

5. นำข้อคำถามที่ผู้วิจัยสร้างขึ้นให้ผู้เชี่ยวชาญ จำนวนทั้งหมด 5 คน ตรวจสอบความเที่ยงตรงเชิงเนื้อหา (Content Validity) ของข้อคำถาม และคู่มือการตรวจ ดังรายชื่อในภาคผนวก ซึ่งผู้เชี่ยวชาญทั้งหมดต้องทำการพิจารณาความสอดคล้องในด้านต่าง ๆ ต่อไปนี้ คือ

1) ความสอดคล้องของข้อคำถามกับเนื้อหาที่ต้องการวัด

2) การเขียนข้อคำถามได้ชัดเจน ตรงกับสิ่งที่ต้องการถาม

3) ความยากและความเหมาะสมทางภาษากับระดับชั้นที่วัด

4) คู่มือการตรวจ ตามวิธีการตรวจให้คะแนนทั้ง 2 วิธี

นำผลมาวิเคราะห์หาค่าความเที่ยงตรงเชิงเนื้อหาโดยใช้ดัชนีความสอดคล้อง (IOC) โดยรวมความเห็นในช่อง เห็นด้วย (+1) ไม่แน่ใจ (0) และ ไม่เห็นด้วย (-1) แล้วหาค่าเฉลี่ยตามวิธีของ โรวิเนลลี่ และแฮมมินตัน คัดเลือกข้อที่มีค่า IOC ตั้งแต่ 0.5 ขึ้นไป ปรากฏผลการคำนวณค่า IOC มีค่าอยู่ระหว่าง 0.6 –1.00 ดังแสดงในภาคผนวก

6. นำแบบทดสอบที่ผ่านการตรวจคุณภาพขั้นต้นแล้ว ไปทดสอบครั้งที่ 1 กับกลุ่มตัวอย่างที่ใช้ทดลองเครื่องมือ คือ นักเรียนชั้นประถมศึกษาปีที่ 5 โรงเรียนเปรมฤดีศึกษา จำนวน 45 คน โดยจะให้นักเรียนทำข้อสอบทีละข้อ เพื่อตรวจสอบเวลาที่เหมาะสมสำหรับใช้ในการทำข้อสอบแต่ละข้อ โดยพิจารณาจากจำนวนนักเรียนที่ทำเสร็จประมาณ 80 % ของนักเรียนทั้งห้อง

7. นำผลการทดสอบครั้งที่ 1 มาตรวจให้คะแนนตามเกณฑ์ที่กำหนดไว้โดยวิธีประเมินรวม (Holistic Method) โดยใช้ผู้ตรวจเป็นครูผู้สอนวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ระดับชั้นประถมศึกษาปีที่ 5 จำนวน 2 คน นำผลการตรวจให้คะแนน มาพิจารณาหาค่าความยาก โดยใช้สูตรการกระจายของค่าเฉลี่ยรายข้อ (Pooled Mean) ซึ่งมีค่าความยากอยู่ระหว่าง 0.086 – 0.464 และ หาค่าอำนาจจำแนกเป็นรายข้อ โดยใช้สูตร t-test เลือกข้อที่มีค่า t ที่นัยสำคัญ .05 ได้ข้อสอบจำนวน 5 ข้อ คือ ข้อ 2 , ข้อ 4 , ข้อ 5 , ข้อ 8 และ ข้อ 10 โดยมีค่าความยากอยู่ระหว่าง 0.204-0.291 และค่าอำนาจจำแนก(t) มีค่าอยู่ระหว่าง 1.700 – 4.573

8. นำแบบทดสอบไปทดสอบครั้งที่ 2 กับนักเรียนชั้นประถมศึกษาปีที่ 5 ที่โรงเรียนสัจจพิทยา กับการทดลองครั้งที่ 2 จำนวน 40 คน

9. หาคุณภาพของแบบทดสอบทั้งฉบับ โดยหาความเชื่อมั่นของแบบทดสอบ ด้วยสูตรหาสัมประสิทธิ์สรุปอ้างอิง (ρ^2) ของเบรนนอน และเคน ที่วิเคราะห์โดยใช้โปรแกรมสำเร็จรูป GENOVA (Generalizability Analysis of Variance System) (Brennan. 1983 : 17) ปรากฏค่าสัมประสิทธิ์สรุปอ้างอิง มีค่าเท่ากับ 0.687 แสดงว่าแบบทดสอบชุดนี้มีค่าความเชื่อมั่นค่อนข้างสูง

10. นำแบบทดสอบไปทดสอบครั้งที่ 3 กับกลุ่มตัวอย่าง คือ โรงเรียนเซนต์หลุยส์ศึกษา โรงเรียนอัสสัมชัญแผนกประถม และโรงเรียนอัสสัมชัญศึกษา จำนวน 60 คน นำผลการเก็บรวบรวมข้อมูลมาทำการวิเคราะห์ตามจุดมุ่งหมายของการวิจัยต่อไป

ตัวอย่างเครื่องมือที่ใช้ในการวิจัย

0) ให้นักเรียนอธิบายปัญหาที่เกิดขึ้นในการผลิตสินค้าหัตถกรรมและสินค้าอุตสาหกรรมในประเทศไทยมาให้ชัดเจน (5 คะแนน)

แนวคำตอบ

ปัญหาที่เกิดขึ้นในการผลิตสินค้าหัตถกรรม มีตัวอย่างดังนี้

1. ไม่ได้รับการสนับสนุนจากรัฐบาลเท่าที่ควร
2. ไม่ตลาดรองรับสินค้าที่ผลิตได้
3. ไม่มีการวางแผนสำหรับการผลิตในระยะยาว
4. ผู้ชำนาญไม่นิยมถ่ายทอดความรู้แก่ชนรุ่นหลัง
5. คนไทยไม่นิยมใช้สินค้าหัตถกรรม
6. ขาดแคลนวัตถุดิบ
7. ราคาสินค้าแพง เพราะระยะเวลาในการผลิตนาน

ปัญหาที่เกิดขึ้นในการผลิตสินค้าอุตสาหกรรม มีตัวอย่างดังนี้

1. วัตถุดิบไม่เพียงพอในการผลิต
2. ขาดแคลนแรงงาน
3. ขาดผู้เชี่ยวชาญทางด้านเทคนิคการผลิต
4. มีการพัฒนาทางเทคโนโลยีชั้นสูงล่าช้ากว่าประเทศคู่แข่ง
5. ไม่มีการแก้ไขปัญหาจากหน่วยงานของรัฐอย่างจริงจัง
6. ส่งผลกระทบต่อสภาพแวดล้อม เกิดปัญหามลพิษต่าง ๆ
7. ขาดการประสานงานระหว่างประเทศ เพื่อหาตลาดรองรับ
8. ขาดแคลนเงินทุนในการตั้งโรงงานผลิตสินค้าอุตสาหกรรมขนาดใหญ่

เกณฑ์การตรวจโดยวิธีวิเคราะห์ย่อย (Analytic Method)

เกณฑ์ที่ใช้ในการตรวจให้คะแนนโดยวิธีวิเคราะห์ จำแนกไว้ 3 เกณฑ์ (Ward. 1999 : 133) ดังนี้

1. ความถูกต้องของเนื้อหา 2 คะแนน
2. การนำเสนอหลักการที่ถูกต้อง 2 คะแนน
3. การจัดเรียบเรียงความคิด 1 คะแนน

1. ความถูกต้องของเนื้อหา 2 คะแนน

ให้ 2 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามแนวคำตอบและมีปัญหานอกเหนือจากแนวคำตอบ

1 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามแนวคำตอบ

0 คะแนน เมื่อ ตอบคำถามได้น้อยกว่าครึ่งหนึ่งของปัญหาที่ระบุในแนวคำตอบ

2. การนำเสนอหลักการที่ถูกต้อง 2 คะแนน

ให้ 2 คะแนน เมื่อ นำเสนอหลักการที่สำคัญ ๆ ได้ทั้ง 2 ปัญหา คือปัญหาจากการผลิตสินค้าหัตถกรรม และปัญหาจากการผลิตสินค้าอุตสาหกรรม

1 คะแนน เมื่อ นำเสนอหลักการโดยรวมกันไม่แยกออกเป็น 2 ปัญหา

0 คะแนน เมื่อ ไม่สามารถนำเสนอปัญหาจากคำถาม

3. การจัดเรียบเรียงความคิด 1 คะแนน

ให้ 1 คะแนน เมื่อ สามารถจัดเรียบเรียงความคิดได้ชัดเจน เข้าใจง่าย

0 คะแนน เมื่อ ไม่สามารถจัดเรียบเรียงความคิดให้เข้าใจได้

เกณฑ์การตรวจโดยวิธีประเมินรวม (Holistic Method)

เกณฑ์การให้คะแนนโดยวิธีประเมินรวม จำแนกไว้ 5 ระดับ (Ward. 1999 : 132) ดังนี้

ให้ 5 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามแนวคำตอบและมีปัญหาเพิ่มเติมนอกเหนือจากแนวคำตอบ
นำเสนอหลักการได้ทั้ง 2 ปัญหา คือ ปัญหาจากการผลิตสินค้าหัตถกรรมและปัญหาจากการผลิตสินค้าอุตสาหกรรม
จัดเรียบเรียงความคิดได้ชัดเจน เข้าใจง่าย

ให้ 4 คะแนน เมื่อ ตอบคำถามถูกต้องตามแนวคำตอบ
นำเสนอหลักการได้ครบทั้ง 2 ปัญหา คือ ปัญหาจากการผลิตสินค้าหัตถกรรม และปัญหาจากการผลิตสินค้าอุตสาหกรรม
จัดเรียบเรียงแนวคิดได้ชัดเจน

ให้ 3 คะแนน เมื่อ ตอบคำถามได้ถูกต้องมากกว่าครึ่งหนึ่งของแนวคำตอบ
นำเสนอปัญหาได้โดยรวมไม่แยกออกเป็น 2 ปัญหา
จัดเรียบเรียงความคิดได้แต่ไม่ชัดเจนทั้งหมด

ให้ 2 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามแนวคำตอบน้อยกว่าครึ่งหนึ่ง
นำเสนอปัญหาได้อย่างใดอย่างหนึ่ง
แสดงการจัดเรียบเรียงความคิดออกมาอย่างกำกวม

ให้ 1 คะแนน เมื่อ ตอบคำถามไม่ถูกต้องตามแนวคำตอบแต่ยังถูกต้องตามเนื้อหา
ไม่แยกแยะตามหลักการของปัญหาใดปัญหาหนึ่ง
ไม่สามารถเรียบเรียงความคิดได้

ให้ 0 คะแนน เมื่อ ตอบคำถามใด ๆ ไม่ถูกต้องตามเนื้อหา หรือไม่ตอบเลย

วิธีดำเนินการเก็บรวบรวมข้อมูล

ผู้วิจัยได้ดำเนินการเก็บรวบรวมข้อมูลในการวิจัย โดยใช้แบบทดสอบความเรียงวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ดังนี้

1. นำหนังสือจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อขอความร่วมมือกับโรงเรียนที่ใช้เป็นกลุ่มตัวอย่าง

2. ติดต่อประสานงาน และขออนุญาตครูใหญ่ และครูผู้สอนของโรงเรียนที่ใช้เป็นกลุ่มตัวอย่าง เพื่อกำหนดวัน เวลา สถานที่ และวิธีดำเนินการสอบ และจัดประชุมครูผู้ตรวจที่ร่วมในการวิจัยครั้งนี้ ให้เกิดความเข้าใจเป็นอย่างดี

3. ชี้แจงให้นักเรียนที่เป็นกลุ่มตัวอย่าง และ ครูผู้ตรวจให้ทราบถึงจุดมุ่งหมายของการวิจัย และความสำคัญของการวิจัยนี้ ก่อนดำเนินการสอบ

4. นำกระดาษคำตอบที่นักเรียนทำทั้งหมด มาถ่ายเอกสารเป็น 8 ชุด เพื่อจัดให้แก่ผู้ตรวจ (ผู้ตรวจแต่ละคนต้องตรวจให้คะแนนแบบทดสอบฉบับเดียวกัน 2 ครั้ง) ดำเนินการตรวจตามจุดมุ่งหมายของการวิจัย ดังนี้

4.1 จัดกระดาษคำตอบ 4 ชุด สำหรับการตรวจครั้งที่ 1 ด้วยผู้ตรวจจำนวนสี่คน ลงมือตรวจโดยวิธีวิเคราะห์ย่อย (Analytic Method) ตามคู่มือการตรวจที่เตรียมไว้ให้กับผู้ตรวจทุกคน พร้อมทั้งชี้แจงให้ตรวจกระดาษคำตอบทีละข้อให้ครบทุกคน และจนครบทุกข้อ

4.2 หลังจากนั้น 1 สัปดาห์ จัดกระดาษคำตอบ 4 ชุด สำหรับการตรวจครั้งที่ 2 ด้วยผู้ตรวจจำนวนสี่คน ตรวจโดยวิธีประเมินรวม (Holistic Method) ตามคู่มือการตรวจที่เตรียมไว้ให้กับผู้ตรวจทุกคน พร้อมทั้งชี้แจงให้ตรวจกระดาษคำตอบทีละข้อให้ครบทุกคน และจนครบทุกข้อ

5. นำผลที่ได้จากการตรวจมาทำการวิเคราะห์ข้อมูล ดังนี้

5.1 คำนวณหาค่าสถิติพื้นฐาน ($\bar{X}$, $S.D.$ และ $C.V.$)

5.2 คำนวณหาค่าดัชนีความสอดคล้องของผู้ตรวจ โดยใช้ดัชนีความสอดคล้องระหว่างผู้ตรวจ (RAI) มีรูปแบบการคำนวณตามลำดับ ดังนี้

5.2.1 ผู้ตรวจจำนวนสองคน คือ นำผลการตรวจของผู้ตรวจคนที่ 1 กับคนที่ 2 , ผู้ตรวจคนที่ 1 กับคนที่ 3 , ผู้ตรวจคนที่ 1 กับคนที่ 4 , ผู้ตรวจคนที่ 2 กับคนที่ 3 , ผู้ตรวจคนที่ 2 กับคนที่ 4 , ผู้ตรวจคนที่ 3 กับคนที่ 4 ที่ตรวจโดยวิธีวิเคราะห์ย่อย (Analytic Method)

5.2.2 ผู้ตรวจจำนวนสามคน คือ นำผลการตรวจของผู้ตรวจคนที่ 1 , 2 กับ คนที่ 3 , ผู้ตรวจคนที่ 1 , 2 กับคนที่ 4 , ผู้ตรวจคนที่ 1 , 3 กับคนที่ 4 และผู้ตรวจคนที่ 2 , 3 กับคนที่ 4 ที่ตรวจโดยวิธีวิเคราะห์ย่อย (Analytic Method)

5.2.3 ผู้ตรวจจำนวนสี่คน คือ นำผลการตรวจของผู้ตรวจคนที่ 1 , 2 , 3 กับ คนที่ 4 ที่ตรวจโดยวิธีวิเคราะห์ย่อย (Analytic Method)

5.2.4 ผู้ตรวจจำนวนสองคน คือ นำผลการตรวจของผู้ตรวจคนที่ 1 กับคนที่ 2 , ผู้ตรวจคนที่ 1 กับคนที่ 3 , ผู้ตรวจคนที่ 1 กับคนที่ 4 , ผู้ตรวจคนที่ 2 กับคนที่ 3 , ผู้ตรวจคนที่ 2 กับคนที่ 4 , ผู้ตรวจคนที่ 3 กับคนที่ 4 ที่ตรวจโดยวิธีประเมินรวม (Holistic Method)

5.2.5 ผู้ตรวจจำนวนสามคน คือ นำผลการตรวจของผู้ตรวจคนที่ 1 , 2 กับคนที่ 3 , ผู้ตรวจคนที่ 1 , 2 กับคนที่ 4 , ผู้ตรวจคนที่ 1 , 3 กับคนที่ 4 และผู้ตรวจคนที่ 2 , 3 กับคนที่ 4 ที่ตรวจโดยวิธีประเมินรวม (Holistic Method)

5.2.6 ผู้ตรวจจำนวนสี่คน คือ นำผลการตรวจของผู้ตรวจคนที่ 1 , 2 , 3 กับคนที่ 4 ที่ตรวจโดยวิธีประเมินรวม (Holistic Method)

5.3 นำค่า RAI ที่ตรวจโดยวิธีวิเคราะห์ย่อย มาหาค่าเฉลี่ย ($_A\bar{R}_M$) ความเชื่อมั่นของผู้ตรวจสองคน , สามคน และ สี่คน ตามลำดับ และ นำค่า RAI ที่ตรวจโดยวิธีประเมินรวม มาหาค่าเฉลี่ย ($_H\bar{R}_M$) ความเชื่อมั่นของผู้ตรวจสองคน สามคน และ สี่คน เช่นเดียวกัน

5.4 ทดสอบสมมุติฐานของการวิจัย

ทดสอบความแตกต่างของค่าดัชนีความสอดคล้องของผู้ตรวจ ตามสมมุติฐานของการวิจัย โดยใช้สูตร UX_1 ของวูดรัฟฟ์ และเฟลดท์ (Woodruff and Feldt. 1986 : 393-413) ถ้ามีความแตกต่างอย่างมีนัยสำคัญทางสถิติ จะต้องทำการทดสอบเพื่อเปรียบเทียบความแตกต่างเป็นรายคู่ ด้วยสูตรของ วูดรัฟฟ์ และเฟลดท์ (Woodruff and Feldt. 1986 : 393-413) อีกครั้ง

การจัดกระทำและการวิเคราะห์ข้อมูล

สถิติที่ใช้ แบ่งออกเป็น 3 ลักษณะ คือ

1. สถิติพื้นฐาน

1.1 หาค่าเฉลี่ย (Mean) และค่าเฉลี่ยร่วม (Pooled Mean) (ประคอง กรรณสูต. 2540: 42)

1.2 หาส่วนเบี่ยงเบนมาตรฐาน (Standard deviation) และส่วนเบี่ยงเบนมาตรฐานร่วม (Polled Standard deviation) (ประคอง กรรณสูต, 2540: 49-55)

1.3 หาค่าสัมประสิทธิ์ความแปรผัน (Coefficient of Variation) (ประคอง กรรณสูต. 2540: 56 - 57)

2. สถิติที่ใช้หาคุณภาพของเครื่องมือวัด

2.1 สถิติใช้ในการวิเคราะห์ข้อสอบแบบทดสอบความเรียง

2.1.1 หาค่าความยาก โดยใช้วิธีการกระจายของค่าเฉลี่ยรายข้อ โดยการใช้สัดส่วนของค่าเฉลี่ยร่วม (Pooled Mean) กับจำนวนข้อสอบ สูตรมีดังนี้ (ประคอง กรรณสูต. 2540: 42)

$$P = \frac{1}{K}\sum_{i=1}^{k} N_i \bar{X}_i$$

เมื่อ P แทน ค่าความยากของแบบทดสอบ

$\bar{X}_i$ แทน ค่าเฉลี่ยของคะแนนรายข้อ

N แทน จำนวนนักเรียนทั้งหมด

K แทน จำนวนข้อสอบ

2.1.2 หาค่าอำนาจจำแนก (t) คำนวณจากสูตร ดังนี้ (ประคอง กรรณสูต. 2542 : 100)

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \qquad df = n_1 + n_2 - 2$$

เมื่อ $\overline{X}_1$ แทน ค่าเฉลี่ยของคะแนนจากกลุ่มสูง

$\overline{X}_2$ แทน ค่าเฉลี่ยของคะแนนจากกลุ่มต่ำ

S_1^2 แทน ความแปรปรวนของคะแนนกลุ่มสูง

S_2^2 แทน ความแปรปรวนของคะแนนกลุ่มต่ำ

n_1, n_2 แทน จำนวนนักเรียนของกลุ่มสูงและกลุ่มต่ำ

2.2 หาค่าดัชนีความสอดคล้องของผู้เชี่ยวชาญ (Index of Consistency :IOC) (บุญเชิด ภิญโญอนันตพงษ์. 2527 : 69) ดังนี้

$$IOC = \frac{\sum R}{N}$$

เมื่อ IOC แทน ดัชนีความสอดคล้องมีค่าอยู่ระหว่าง –1 ถึง +1

$\sum R$ แทน ผลรวมของการพิจารณาของผู้เชี่ยวชาญ

N แทน จำนวนผู้เชี่ยวชาญ

2.3 หาความเชื่อมั่นของแบบทดสอบ โดยใช้สัมประสิทธิ์การสรุปอ้างอิง (ρ^2) ของเบรนนอน และเคน ที่วิเคราะห์โดยใช้โปรแกรมสำเร็จรูป GENOVA (Generalizability Analysis of Variance System) (Brennan. 1983 : 17)

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

เมื่อ ρ^2 แทน สัมประสิทธิ์การสรุปอ้างอิง

$\sigma^2(p)$ แทน ความแปรปรวนของคะแนนเอกภพ

$\sigma^2(\delta)$ แทน ความแปรปรวนของคะแนนความคลาดเคลื่อน

3. สถิติที่ใช้ทดสอบสมมุติฐาน

3.1 หาค่าเฉลี่ยของคะแนนที่ได้จากการตรวจของผู้ตรวจทั้งหมด โดยใช้สูตรจากดัชนีความสอดคล้องระหว่างผู้ตรวจ ของเบอร์รี่ – สต๊อก และคนอื่น ๆ (Burry-Stock and others. 1996) ดังนี้

$$\overline{R}_k = \frac{1}{M}\sum^{M} R_{mk}$$

เมื่อ $\overline{R}$ แทน ค่าเฉลี่ยของคะแนนที่ได้จากผู้ตรวจทั้งหมด

M แทน จำนวนผู้ตรวจ

$\sum R$ แทน ผลรวมของคะแนนจากผู้ตรวจทั้งหมด

3.2 ใช้สูตรดัชนีความสอดคล้องระหว่างผู้ตรวจให้คะแนน (RAI) ของ เบอร์รี่ – สต๊อก และคนอื่น ๆ (Burry-Stock and others. 1996) แบบที่ 6 คือ ผู้ตรวจ M คน ,k พฤติกรรม , N เนื้อหา

$$RAI = 1 - \frac{\sum^{K}\sum^{N}\sum^{M}\left|R_{mnk} - \overline{R}_{nk}\right|}{KN(M-1)(I-1)}$$

เมื่อ R แทน คะแนนจากผู้ตรวจแต่ละคน แต่ละข้อ

$\overline{R}$ แทน ค่าเฉลี่ยของคะแนนที่ได้จากผู้ตรวจทั้งหมด

M แทน จำนวนผู้ตรวจ

K แทน จำนวนข้อคำถาม

I แทน ระดับคะแนน

$\sum R$ แทน ผลรวมของคะแนนจากผู้ตรวจทั้งหมด

3.4 สถิติทดสอบนัยสำคัญความแตกต่างของค่าความเชื่อมั่นหลายค่า และรายคู่ ซึ่งประมาณจากกลุ่มตัวอย่างกลุ่มเดียว โดยสูตร UX_1 (บุญเชิด ภิญโญอนันตพงษ์. 2537 ; citing Woodruff and Feldt. 1986 : 393-413)

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

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

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

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

$$C_u = \frac{4}{9m(m-1)(N_c - 1)} \sum_{i=2}^{m} \sum_{j=1}^{i=1} \left[r_{ij}\right]^2 u_i u_j$$

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

$$\overline{n}_h = \frac{M}{\sum(1/n_i)}$$

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

N แทน จำนวนนักเรียนที่ทดสอบ

r แทน ค่าความเชื่อมั่นที่คำนวณได้

r_{ij} แทน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างตัวแปร

S_u^2 แทน ความแปรปรวนของ u_i

C_u แทน ความแปรปรวนร่วมของ u_i

$\overline{u}$ แทน ค่าเฉลี่ยของคะแนน u_i

m แทน จำนวนค่า RAI ที่ใช้ทดสอบ

บทที่ 4

ผลการวิเคราะห์ข้อมูล

ในการวิเคราะห์ข้อมูลและการแปลความหมายของการวิเคราะห์ข้อมูล ในการวิจัยครั้งนี้ เพื่อให้เกิดความเข้าใจตรงกัน ผู้วิจัยจึงได้กำหนดสัญลักษณ์และอักษรย่อที่ใช้ในการวิเคราะห์ข้อมูลดังต่อไปนี้

สัญลักษณ์และอักษรย่อในการวิเคราะห์ข้อมูล

$\overline{X}_M$ แทน ค่าเฉลี่ย

S.D. แทน ค่าเบี่ยงเบนมาตรฐานของคะแนน

C.V. แทน ค่าสัมประสิทธิ์ความแปรผันของคะแนน

RAI แทน ดัชนีความสอดคล้องระหว่างผู้ตรวจ ของเบอร์รี่-สต๊อก และคนอื่น ๆ (Burry-Stock and others.1996)

$\overline{R}$ แทน ค่าเฉลี่ยของความเชื่อมั่นของผู้ตรวจ

${}_A\overline{R}_M$ แทน ค่าเฉลี่ยของความเชื่อมั่นของผู้ตรวจ ที่ตรวจด้วยวิธีวิเคราะห์ย่อย เมื่อมีจำนวนผู้ตรวจ M คน

${}_H\overline{R}_M$ แทน ค่าเฉลี่ยของความเชื่อมั่นของผู้ตรวจ ที่ตรวจด้วยวิธีประเมินรวม เมื่อมีจำนวนผู้ตรวจ M คน

M แทน จำนวนผู้ตรวจ

UX_1 แทน สถิติทดสอบนัยสำคัญที่มีการแจกแจงแบบไคกำลังสอง (χ^2)

p แทน ระดับนัยสำคัญ

การเสนอผลการวิเคราะห์ข้อมูล

การศึกษาค้นคว้าครั้งนี้ ผู้วิจัยได้เสนอผลการวิเคราะห์ข้อมูลตามลำดับ ดังนี้

1. ค่าสถิติพื้นฐานของแบบทดสอบความเรียง
2. ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง
3. การเปรียบเทียบค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง

3.1 เปรียบเทียบค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง ที่มีผู้ตรวจจำนวนสองคน สามคน และ สี่คน เมื่อใช้การตรวจวิธีเดียวกัน

3.2 เปรียบเทียบค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง ที่ตรวจด้วยวิธีวิเคราะห์ย่อย (Analytic Method) และวิธีประเมินรวม (Holistic Method) เมื่อมีผู้ตรวจจำนวนเท่ากัน

ผลการวิเคราะห์ข้อมูล

1. ค่าสถิติพื้นฐานของแบบทดสอบความเรียง

ในการวิจัยครั้งนี้ ผู้วิจัยนำคะแนนของแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ระดับชั้นประถมศึกษาปีที่ 5 จำนวน 5 ข้อ ที่ได้จากการตรวจให้คะแนนสองวิธี คือ วิธีวิเคราะห์ย่อย (Analytic Method) และวิธีประเมินรวม (Holistic Method) มาวิเคราะห์หาค่าสถิติพื้นฐานของแบบทดสอบจำแนกตามผู้ตรวจจำนวนสองคน สามคน และสี่คน ปรากฏผลดังตาราง 7

ตาราง 7 ค่าสถิติพื้นฐานของแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต(สังคม) จำแนกตามวิธีการตรวจให้คะแนน และจำนวนผู้ตรวจ

วิธีการตรวจ	จำนวนผู้ตรวจ	ผู้ตรวจคนที่	คะแนนเต็ม	$\overline{X}_M$	$S.D.$	$C.V.$ (%)
วิธีวิเคราะห์ย่อย (Analytic Method)	หนึ่งคน	คนที่ 1	25	14.583	3.341	22.910
		คนที่ 2	25	12.567	2.560	20.373
		คนที่ 3	25	11.950	3.721	31.137
		คนที่ 4	25	12.983	2.703	20.817
		$\overline{X}_{t1}$	**25**	**13.021**	**3.266**	**25.083**
	สองคน	คนที่ 1 กับ 2	25	13.575	2.677	19.722
		คนที่ 1 กับ 3	25	13.275	3.215	24.216
		คนที่ 1 กับ 4	25	13.783	2.791	20.251
		คนที่ 2 กับ 3	25	12.266	2.753	22.444
		คนที่ 2 กับ 4	25	12.775	2.455	19.220
		คนที่ 3 กับ 4	25	12.475	3.048	24.435
		$\overline{X}_{t2}$	**25**	**13.025**	**5.912**	**45.389**
	สามคน	คนที่ 1 , 2 กับ 3	25	13.039	2.764	21.197
		คนที่ 1 , 2 กับ 4	25	13.378	2.559	19.128
		คนที่ 1 , 3 กับ 4	25	13.178	2.931	22.245
		คนที่ 2 , 3 กับ 4	25	12.505	2.681	21.304
		$\overline{X}_{t3}$	**25**	**13.025**	**6.218**	**47.739**
	สี่คน	คนที่ 1 , 2 , 3 กับ 4	25	13.025	2.681	20.584
		$\overline{X}_{t4}$	**25**	**13.025**	**6.423**	**49.315**

ตาราง 7 (ต่อ)

วิธีการตรวจ	จำนวนผู้ตรวจ	ผู้ตรวจคนที่	คะแนนเต็ม	$\bar{X}_M$	$S.D.$	$C.V.$(%)
วิธีประเมินรวม (Holistic Method)	**หนึ่งคน**	คนที่ 1	25	13.600	2.644	19.441
		คนที่ 2	25	11.767	2.302	19.563
		คนที่ 3	25	12.467	2.977	23.880
		คนที่ 4	25	12.817	2.771	21.619
		$\bar{X}_{t1}$	**25**	**12.663**	**2.765**	**21.835**
	สองคน	คนที่ 1 กับ 2	25	12.683	2.264	17.852
		คนที่ 1 กับ 3	25	13.033	2.541	19.496
		คนที่ 1 กับ 4	25	13.292	2.508	18.869
		คนที่ 2 กับ 3	25	12.117	2.462	20.321
		คนที่ 2 กับ 4	25	12.375	2.346	18.960
		คนที่ 3 กับ 4	25	12.725	2.705	21.257
		$\bar{X}_{t2}$	**25**	**12.704**	**7.064**	**55.600**
	สามคน	คนที่ 1 , 2 กับ 3	25	12.611	2.344	18.583
		คนที่ 1 , 2 กับ 4	25	12.783	2.308	18.058
		คนที่ 1 , 3 กับ 4	25	13.017	2.516	19.328
		คนที่ 2 , 3 กับ 4	25	12.407	2.448	19.733
		$\bar{X}_{t3}$	**25**	**12.704**	**7.335**	**57.739**
	สี่คน	คนที่ 1 , 2 , 3 กับ 4	25	12.704	2.369	18.653
		$\bar{X}_{t4}$	**25**	**12.704**	**10.059**	**79.183**

ผลการวิเคราะห์ตามตาราง 7 พบว่าสถิติพื้นฐานของแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต(สังคม) ที่มีวิธีการตรวจให้คะแนนวิธีวิเคราะห์ย่อย (Analytic Method) ที่มีผู้ตรวจจำนวนหนึ่งคน สองคน สามคน และ สี่คน มีค่าเฉลี่ยร่วม($\bar{X}_t$) เท่ากับ 13.021 คะแนน และ 13.025 คะแนน ซึ่งกลุ่มที่มีผู้ตรวจมากกว่าหนึ่งคนมีค่าเท่ากันทั้ง 3 กลุ่ม แต่ส่วนเบี่ยงเบนมาตรฐานร่วมมีค่าเท่ากับ 3.266 , 5.912 , 6.218 และ 6.423 ตามลำดับ และ ค่าสัมประสิทธิ์ความแปรผัน (Coefficient of Variation) ของค่าเฉลี่ยร่วม มีค่าเท่ากับ 25.083 % , 45.389 % , 47.739 % และ 49.315 % ตามลำดับ

และเมื่อตรวจให้คะแนนวิธีประเมินรวม (Holistic Method) ที่มีผู้ตรวจจำนวนหนึ่งคน สองคน สามคน และ สี่คน มีค่าเฉลี่ยร่วม($\bar{X}_t$)เท่ากับ 12.663 คะแนน และ 12.704 คะแนน ซึ่งกลุ่มที่มีผู้ตรวจมากกว่าหนึ่งคนมีค่าเท่ากันทั้ง 3 กลุ่ม แต่ส่วนเบี่ยงเบนมาตรฐานร่วมมีค่าเท่ากับ 2.765 , 7.064 , 7.335 และ 10.059

ตามลำดับ และ ค่าสัมประสิทธิ์ความแปรผัน (Coefficient of Variation) ของค่าเฉลี่ยร่วม มีค่าเท่ากับ 21.835 % , 55.600 % , 57.739 % และ 79.183 % ตามลำดับ

2. ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง

ผู้วิจัยได้คำนวณค่าความเชื่อมั่นของผู้ตรวจ โดยใช้ดัชนีความสอดคล้องระหว่างผู้ตรวจ ของเบอร์รี่-สต๊อก และคนอื่น ๆ (Burry-Stock and others. 1996) กรณีผู้ตรวจสองคน สามคน และสี่คน นำค่า RAI ที่ได้มาหาค่าเฉลี่ย ($\overline{R}$) ตามจำนวนผู้ตรวจ และวิธีการตรวจทั้งสองแบบ ปรากฏผลดัง ตาราง 8

ตาราง 8 ค่าเฉลี่ย ($\overline{R}$) ความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง จำแนกตามวิธีการตรวจให้คะแนน และจำนวนผู้ตรวจ

จำนวนผู้ตรวจ	ผู้ตรวจคนที่	การตรวจวิธีวิเคราะห์ย่อย (Analytic Method)		การตรวจวิธีประเมินรวม (Holistic Method)	
		RAI	$_A\overline{R}_M$	RAI	$_H\overline{R}_M$
สองคน	คนที่ 1 กับ 2	0.800	**0.797**	0.833	**0.829**
	คนที่ 1 กับ 3	0.763		0.819	
	คนที่ 1 กับ 4	0.774		0.782	
	คนที่ 2 กับ 3	0.815		0.867	
	คนที่ 2 กับ 4	0.839		0.838	
	คนที่ 3 กับ 4	0.792		0.834	
สามคน	คนที่ 1 , 2 กับ 3	0.814	**0.818**	0.831	**0.839**
	คนที่ 1 , 2 กับ 4	0.826		0.838	
	คนที่ 1 , 3 กับ 4	0.799		0.830	
	คนที่ 2 , 3 กับ 4	0.834		0.858	
สี่คน	คนที่ 1 , 2 , 3 กับ 4	0.828	**0.828**	0.851	**0.851**

ผลการวิเคราะห์ตามตาราง 8 พบว่า ค่าเฉลี่ย ($_A\overline{R}_M$)ความเชื่อมั่นของผู้ตรวจ เมื่อใช้การตรวจวิธีวิเคราะห์ย่อย ที่มีผู้ตรวจจำนวนสองคน สามคน และ สี่คน มีค่าเท่ากับ 0.797 , 0.818 และ 0.828 ตามลำดับ กรณีผู้ตรวจสองคน ผู้ตรวจคนที่ 2 กับคนที่ 4 มีค่า RAI สูงสุด คือ 0.839 และผู้ตรวจคนที่ 1 กับคนที่ 3 มีค่า RAI ต่ำสุด คือ 0.763 กรณีผู้ตรวจสามคน ผู้ตรวจคนที่ 2 , 3 และ 4 มีค่า RAI สูงสุด คือ 0.834 และผู้ตรวจคนที่ 1 , 3 กับ 4 มีค่า RAI ต่ำสุด คือ 0.799 และกรณีผู้ตรวจสี่คน มีค่า RAI สูงสุด คือ 0.828 แสดงว่า จำนวนผู้ตรวจมากกว่ามีแนวโน้มตรวจให้ความเชื่อมั่นสูงกว่าจำนวนผู้ตรวจน้อยกว่า

ส่วน ค่าเฉลี่ย (${}_H\overline{R}_M$)ความเชื่อมั่นของผู้ตรวจ เมื่อใช้การตรวจวิธีประเมินรวม ที่มีผู้ตรวจจำนวนสองคน สามคน และ สี่คน มีค่าเท่ากับ 0.829 , 0.839 และ 0.851 ตามลำดับ กรณีผู้ตรวจสองคน ผู้ตรวจคนที่ 2 กับคนที่ 3 มีค่า RAI สูงสุด คือ 0.867 และผู้ตรวจคนที่ 1 กับคนที่ 4 มีค่า RAI ต่ำสุด คือ 0.782 กรณีผู้ตรวจสามคน ผู้ตรวจคนที่ 2 , 3 และ 4 มีค่า RAI สูงสุด คือ 0.858 และผู้ตรวจคนที่ 1 , 3 กับ 4 มีค่า RAI ต่ำสุด คือ 0.830 และกรณีผู้ตรวจสี่คน มีค่า RAI สูงสุด คือ 0.851 แสดงว่าการตรวจวิธีประเมินรวม มีแนวโน้มให้ค่าความเชื่อมั่นของผู้ตรวจค่าสูงกว่าวิธีตรวจแบบวิเคราะห์ย่อย ทุกกรณี

3. การเปรียบเทียบค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง

3.1 เปรียบเทียบความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง ที่มีผู้ตรวจจำนวนสองคน สามคน และ สี่คน เมื่อใช้การตรวจวิธีเดียวกัน

การวิเคราะห์ผลขั้นนี้ ผู้วิจัยนำค่าเฉลี่ย($\overline{R}$) ความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง ที่มีผู้ตรวจจำนวนสองคน สามคน และ สี่คน เมื่อใช้การตรวจวิธีเดียวกัน มาเปรียบเทียบกันโดยใช้สูตร UX_1 ปรากฏผลดังตาราง 9

ตาราง 9 ผลการเปรียบเทียบค่าเฉลี่ย ($\overline{R}$)ความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง ที่มีผู้ตรวจจำนวนสองคน สามคน และ สี่คน เมื่อใช้การตรวจวิธีวิเคราะห์ย่อย กับ การตรวจวิธีประเมินรวม

วิธีการตรวจ	จำนวนผู้ตรวจ	$\overline{R}$	UX_1	p
แบบวิเคราะห์ย่อย (Analytic Method)	2	0.797	0.263	0.594
	3	0.818		
	4	0.828		
แบบประเมินรวม (Holistic Method)	2	0.829	0.189	0.513
	3	0.839		
	4	0.851		

ผลการวิเคราะห์ตามตาราง 9 ปรากฏว่า ค่าเฉลี่ยความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียงที่มีผู้ตรวจจำนวนสองคน สามคน และ สี่คน เมื่อใช้การตรวจวิธีวิเคราะห์ย่อย มีค่าเท่ากับ 0.797 , 0.818 และ 0.828 ตามลำดับ มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ (UX_1 = 0.263 , p = .594) และเมื่อใช้การตรวจวิธีประเมินรวม มีค่าเท่ากับ 0.829 , 0.839 และ 0.851 ตามลำดับ มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ (UX_1 = 0.189 , p = .513)

3.2 เปรียบเทียบค่าเฉลี่ย ($\overline{R}$)ความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง ที่ตรวจด้วยวิธีวิเคราะห์ย่อย (Analytic Method) และวิธีประเมินรวม (Holistic Method) เมื่อมีผู้ตรวจจำนวนเท่ากัน

การวิเคราะห์ผลขั้นนี้ ผู้วิจัยนำค่าเฉลี่ย($\overline{R}$) ความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง เมื่อใช้การตรวจต่างวิธีกัน คือ ที่ตรวจด้วยวิธีวิเคราะห์ย่อย (Analytic Method) และวิธีประเมินรวม (Holistic Method) เมื่อมีผู้ตรวจจำนวนเท่ากัน มาเปรียบเทียบกันโดยใช้สูตร UX_1 ปรากฏผลดังตาราง 10

ตาราง 10 แสดงการเปรียบเทียบค่าเฉลี่ย ($\overline{R}$)ความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง เมื่อใช้การตรวจวิธีวิเคราะห์ย่อย กับ วิธีประเมินรวม ที่มีผู้ตรวจจำนวนเท่ากัน

จำนวนผู้ตรวจ	ผู้ตรวจคนที่	การตรวจวิธีวิเคราะห์ย่อย (Analytic Method)		การตรวจวิธีประเมินรวม (Holistic Method)		UX_1	p
		RAI	$_A\overline{R}_M$	RAI	$_H\overline{R}_M$		
สองคน	1 กับ 2	0.800	**0.797**	0.833	**0.829**	3.059	0.764
	1 กับ 3	0.763		0.819			
	1 กับ 4	0.774		0.782			
	2 กับ 3	0.815		0.867			
	2 กับ 4	0.839		0.838			
	3 กับ 4	0.792		0.834			
สามคน	1,2 กับ 3	0.814	**0.818**	0.831	**0.839**	1.466	0.224
	1,2 กับ 4	0.826		0.838			
	1,3 กับ 4	0.799		0.830			
	2,3 กับ 4	0.834		0.858			
สี่คน	1,2,3 กับ 4	0.828	**0.828**	0.851	**0.851**	2.079	0.145

ผลการวิเคราะห์ตามตาราง 10 ปรากฏว่า ค่าเฉลี่ยความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง เมื่อใช้ผู้ตรวจสองคน ตรวจวิธีวิเคราะห์ย่อย กับ ตรวจวิธีประเมินรวม มีค่าเท่ากับ 0.797 และ 0.829 ตามลำดับ มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ (UX_1 = 3.059 , p = 0.764) เมื่อใช้ผู้ตรวจสามคน มีค่าเท่ากับ 0.818 และ 0.839 ตามลำดับ มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ (UX_1 = 1.466 , p = 0.224) และเมื่อใช้ผู้ตรวจสี่คน มีค่าเท่ากับ 0.828 และ 0.851 ตามลำดับ มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ (UX_1 = 2.079 , p = 0.145)

บทที่ 5

สรุปผล อภิปราย และข้อเสนอแนะ

ความมุ่งหมายของการวิจัย

การวิจัยครั้งนี้ มีจุดมุ่งหมายสำคัญ เพื่อศึกษาและเปรียบเทียบค่าความเชื่อมั่นของผู้ตรวจให้คะแนนแบบทดสอบความเรียง ที่มีผู้ตรวจจำนวนต่างกัน และตรวจให้คะแนนวิธีวิเคราะห์ย่อย และวิธีประเมินรวม โดยมีจุดมุ่งหมายเฉพาะของการวิจัย ดังนี้

1. เพื่อเปรียบเทียบผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง ที่มีผู้ตรวจจำนวนสองคน สามคน และ สี่คน โดยแยกเป็น

   1.1 เมื่อตรวจให้คะแนนวิธีวิเคราะห์ย่อย

   1.2 เมื่อตรวจให้คะแนนวิธีประเมินรวม

2. เพื่อเปรียบเทียบผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง ที่ใช้การตรวจให้คะแนนวิธีวิเคราะห์ย่อย และให้คะแนนวิธีประเมินรวม โดยแยกเป็น

   2.1 เมื่อใช้ผู้ตรวจสองคน

   2.2 เมื่อใช้ผู้ตรวจสามคน

   2.3 เมื่อใช้ผู้ตรวจสี่คน

สมมุติฐานของการวิจัย

1. ค่าความเชื่อมั่นของผู้ตรวจให้คะแนนแบบทดสอบความเรียงที่มีผู้ตรวจจำนวนสองคน สามคน และ สี่คน เมื่อใช้การตรวจวิธีเดียวกัน มีค่าแตกต่างกัน

2. ค่าความเชื่อมั่นของผู้ตรวจให้คะแนนแบบทดสอบความเรียงที่ใช้การตรวจวิธีวิเคราะห์ย่อย และการตรวจวิธีประเมินรวม เมื่อมีผู้ตรวจจำนวนเท่ากัน มีค่าแตกต่างกัน

ประชากรที่ใช้ในการวิจัย

ประชากรที่ใช้ในการวิจัย เป็นนักเรียนชั้นประถมศึกษาปีที่ 5 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาเอกชน สาขากลุ่มกรุงเทพฯ ตะวันออก เขตสาทร เขตบางรัก และเขตบางคอแหลม กรุงเทพมหานคร ในปีการศึกษา 2543 จำนวน 3,325 คน

กลุ่มตัวอย่างที่ใช้ในการวิจัย

กลุ่มตัวอย่างที่ใช้ในการวิจัย เป็นนักเรียนชั้นประถมศึกษาปีที่ 5 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาเอกชน สาขากลุ่มกรุงเทพฯ ตะวันออก เขตสาทร เขตบางรัก และเขตบางคอแหลม กรุงเทพมหานคร ในปีการศึกษา 2543 จำนวน 60 คน ที่ได้มาโดยการสุ่มแบบหลายขั้นตอน (Multi – Stage Random Sampling)

เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยนี้ เป็นแบบทดสอบความเรียง 1 ฉบับ จำนวน 5 ข้อ ข้อละ 5 คะแนน เพื่อสอบวัดความรู้ในกลุ่มวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ชั้นประถมศึกษาปีที่ 5 ครอบคลุมเนื้อหาดังต่อไปนี้

หน่วยที่ 5	เรื่อง หัตถกรรม และอุตสาหกรรม	จำนวน 2	ข้อ
หน่วยที่ 8	เรื่อง ประเทศเพื่อนบ้าน	จำนวน 1	ข้อ
หน่วยที่ 12	เรื่อง ข่าว และเหตุการณ์	จำนวน 2	ข้อ

วิธีดำเนินการเก็บรวบรวมข้อมูล

ผู้วิจัยได้ดำเนินการเก็บรวบรวมข้อมูลในการวิจัย โดยใช้แบบทดสอบความเรียงวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ดังนี้

1. นำหนังสือจากบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อขอความร่วมมือกับโรงเรียนที่ใช้เป็นกลุ่มตัวอย่าง

2. ติดต่อประสานงาน และขออนุญาตครูใหญ่ และครูผู้สอนของโรงเรียนที่ใช้เป็นกลุ่มตัวอย่าง เพื่อกำหนดวัน เวลา สถานที่ และวิธีดำเนินการสอบ และจัดประชุมครูผู้ตรวจที่ร่วมในการวิจัยครั้งนี้ ให้เกิดความเข้าใจเป็นอย่างดี

3. ชี้แจงให้นักเรียนที่เป็นกลุ่มตัวอย่าง และ ครูผู้ตรวจให้ทราบถึงจุดมุ่งหมายของการวิจัย และความสำคัญของการวิจัยนี้ ก่อนดำเนินการสอบ

4. นำกระดาษคำตอบที่นักเรียนทำทั้งหมด มาถ่ายเอกสารเป็น 8 ชุด เพื่อจัดให้แก่ผู้ตรวจ (ผู้ตรวจแต่ละคนต้องตรวจให้คะแนนแบบทดสอบฉบับเดียวกัน 2 ครั้ง) ดำเนินการตรวจตามจุดมุ่งหมายของการวิจัยดังนี้

4.3 จัดกระดาษคำตอบ 4 ชุด สำหรับการตรวจครั้งที่ 1 ด้วยผู้ตรวจจำนวนสี่คน ลงมือตรวจโดยวิธีวิเคราะห์ย่อย (Analytic Method) ตามคู่มือการตรวจที่เตรียมไว้ให้กับผู้ตรวจทุกคน พร้อมทั้งชี้แจงให้ตรวจกระดาษคำตอบทีละข้อให้ครบทุกคน และจนครบทุกข้อ

4.4 หลังจากนั้น 1 สัปดาห์ จัดกระดาษคำตอบ 4 ชุด สำหรับการตรวจครั้งที่ 2 ด้วยผู้ตรวจจำนวนสี่คน ตรวจโดยวิธีประเมินรวม (Holistic Method) ตามคู่มือการตรวจที่เตรียมไว้ให้กับผู้ตรวจทุกคน พร้อมทั้งชี้แจงให้ตรวจกระดาษคำตอบทีละข้อให้ครบทุกคน และจนครบทุกข้อ

5.3 คำนวณหาค่าสถิติพื้นฐาน ($\bar{X}$, $S.D$ และ $C.V.$)

5.4 คำนวณหาค่าดัชนีความสอดคล้องของผู้ตรวจ โดยดัชนีความสอดคล้องระหว่างผู้ตรวจ (RAI) โดยมีรูปแบบการคำนวณตามลำดับ ดังนี้

5.2.1 ผู้ตรวจจำนวนสองคน คือ นำผลการตรวจของผู้ตรวจคนที่ 1 กับคนที่ 2 , ผู้ตรวจคนที่ 1 กับคนที่ 3 , ผู้ตรวจคนที่ 1 กับคนที่ 4 , ผู้ตรวจคนที่ 2 กับคนที่ 3 , ผู้ตรวจคนที่ 2 กับคนที่ 4 , ผู้ตรวจคนที่ 3 กับคนที่ 4 ที่ตรวจโดยวิธีวิเคราะห์ย่อย (Analytic Method)

5.2.2 ผู้ตรวจจำนวนสามคน คือ นำผลการตรวจของผู้ตรวจคนที่ 1 , 2 กับ คนที่ 3 , ผู้ตรวจคนที่ 1 , 2 กับคนที่ 4 , ผู้ตรวจคนที่ 1 , 3 กับคนที่ 4 และผู้ตรวจคนที่ 2 , 3 กับคนที่ 4 ที่ตรวจโดยวิธีวิเคราะห์ย่อย (Analytic Method)

5.2.3 ผู้ตรวจจำนวนสี่คน คือ นำผลการตรวจของผู้ตรวจคนที่ 1 , 2 , 3 กับ คนที่ 4 ที่ตรวจโดยวิธีวิเคราะห์ย่อย (Analytic Method)

5.2.4 ผู้ตรวจจำนวนสองคน คือ นำผลการตรวจของผู้ตรวจคนที่ 1 กับคนที่ 2 , ผู้ตรวจคนที่ 1 กับคนที่ 3 , ผู้ตรวจคนที่ 1 กับคนที่ 4 , ผู้ตรวจคนที่ 2 กับคนที่ 3 , ผู้ตรวจคนที่ 2 กับคนที่ 4 , ผู้ตรวจคนที่ 3 กับคนที่ 4 ที่ตรวจโดยวิธีประเมินรวม (Holistic Method)

5.2.5 ผู้ตรวจจำนวนสามคน คือ นำผลการตรวจของผู้ตรวจคนที่ 1 , 2 กับคนที่ 3 , ผู้ตรวจคนที่ 1 , 2 กับคนที่ 4 , ผู้ตรวจคนที่ 1 , 3 กับคนที่ 4 และผู้ตรวจคนที่ 2 , 3 กับคนที่ 4 ที่ตรวจโดยวิธีประเมินรวม (Holistic Method)

5.2.6 ผู้ตรวจจำนวนสี่คน คือ นำผลการตรวจของผู้ตรวจคนที่ 1 , 2 , 3 กับคนที่ 4 ที่ตรวจโดยวิธีประเมินรวม (Holistic Method)

5.3 นำค่า RAI ที่ตรวจโดยวิธีวิเคราะห์ย่อย มาหาค่าเฉลี่ย ($_{A}\bar{R}_{M}$) ความเชื่อมั่นของผู้ตรวจสองคน , สามคน และ สี่คน ตามลำดับ และ นำค่า RAI ที่ตรวจโดยวิธีประเมินรวม มาหาค่าเฉลี่ย ($_{H}\bar{R}_{M}$) ความเชื่อมั่นของผู้ตรวจสองคน สามคน และ สี่คน เช่นเดียวกัน

5.4 ทดสอบสมมุติฐานของการวิจัย

ทดสอบความแตกต่างของค่าดัชนีความสอดคล้องของผู้ตรวจ ตามสมมุติฐานของการวิจัย โดยใช้สูตร UX_1 ของวูดรัฟฟ์ และเฟลดท์ (Woodruff and Feldt. 1986 : 393-413)

สรุปผลการวิเคราะห์ข้อมูล

การศึกษาครั้งนี้มุ่งที่จะศึกษาและเปรียบเทียบผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียงที่มีผู้ตรวจจำนวนสองคน สามคน และ สี่คน เมื่อใช้การตรวจวิธีเดียวกัน และผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียงที่ใช้การตรวจต่างวิธีกัน เมื่อมีผู้ตรวจจำนวนเท่ากัน โดยคำนวณหาค่าความเชื่อมั่นของผู้ตรวจด้วยสูตรของ เบอร์รี่-สต็อก และคนอื่น ๆ และ ทดสอบความแตกต่างค่าความเชื่อมั่นของผู้ตรวจด้วยสูตร UX_1 ผลการวิจัยปรากฏผล ดังนี้

1. สถิติพื้นฐานของแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต(สังคม) เมื่อตรวจให้คะแนนวิธีวิเคราะห์ย่อย (Analytic Method) ที่มีผู้ตรวจจำนวนหนึ่งคน สองคน สามคน และ สี่คน มีค่าเฉลี่ยร่วม ($\bar{X}_t$) เท่ากับ 13.021 คะแนน และ 13.025 คะแนน ซึ่งกลุ่มที่มีผู้ตรวจมากกว่าหนึ่งคนมีค่าเท่ากันทั้ง 3 กลุ่ม แต่ส่วนเบี่ยงเบนมาตรฐานร่วมมีค่าเท่ากับ 3.299 , 5.912 , 6.218 และ 6.423 ตามลำดับ และ ค่าสัมประสิทธิ์

ความแปรผัน (Coefficient of Variation) ของค่าเฉลี่ยร่วม มีค่าเท่ากับ 25.083 % , 45.389 % , 47.739 % และ 49.315 % ตามลำดับ

และเมื่อตรวจให้คะแนนวิธีประเมินรวม (Holistic Method) ที่มีผู้ตรวจจำนวนหนึ่งคน สองคน สามคน และ สี่คน มีค่าเฉลี่ยร่วม($\overline{X}_t$)เท่ากับ 12.663 คะแนน และ 12.704 คะแนน ซึ่งกลุ่มที่มีผู้ตรวจมากกว่าหนึ่งคน มีค่าเท่ากันทั้ง 3 กลุ่ม แต่ส่วนเบี่ยงเบนมาตรฐานร่วมมีค่าเท่ากับ 2.765 , 7.064 , 7.335 และ 10.059 ตามลำดับ และ ค่าสัมประสิทธิ์ความแปรผัน (Coefficient of Variation) ของค่าเฉลี่ยร่วม มีค่าเท่ากับ 21.835 %, 55.600 % , 57.739 % และ 79.183 % ตามลำดับ

2. ค่าเฉลี่ย (${}_A\overline{R}_M$)ความเชื่อมั่นของผู้ตรวจ เมื่อใช้การตรวจวิธีวิเคราะห์ย่อย ที่มีผู้ตรวจจำนวนสองคน สามคน และ สี่คน มีค่าเท่ากับ 0.797 , 0.818 และ 0.828 ตามลำดับ กรณีผู้ตรวจสองคน ผู้ตรวจคนที่ 2 กับคนที่ 4 มีค่า RAI สูงสุด คือ 0.839 และผู้ตรวจคนที่ 1 กับคนที่ 3 มีค่า RAI ต่ำสุด คือ 0.763 กรณีผู้ตรวจสามคน ผู้ตรวจคนที่ 2 , 3 และ 4 มีค่า RAI สูงสุด คือ 0.834 และผู้ตรวจคนที่ 1 , 3 กับ 4 มีค่า RAI ต่ำสุด คือ 0.799 และกรณีผู้ตรวจสี่คน มีค่า RAI สูงสุด คือ 0.828 แสดงว่า จำนวนผู้ตรวจมากกว่ามีแนวโน้มตรวจให้ความเชื่อมั่นสูงกว่าจำนวนผู้ตรวจน้อยกว่า

ส่วน ค่าเฉลี่ย (${}_H\overline{R}_M$)ความเชื่อมั่นของผู้ตรวจ เมื่อใช้การตรวจวิธีประเมินรวม ที่มีผู้ตรวจจำนวนสองคน สามคน และ สี่คน มีค่าเท่ากับ 0.829 , 0.839 และ 0.851 ตามลำดับ กรณีผู้ตรวจสองคน ผู้ตรวจคนที่ 2 กับคนที่ 3 มีค่า RAI สูงสุด คือ 0.867 และผู้ตรวจคนที่ 1 กับคนที่ 4 มีค่า RAI ต่ำสุด คือ 0.782 กรณีผู้ตรวจสามคน ผู้ตรวจคนที่ 2 , 3 และ 4 มีค่า RAI สูงสุด คือ 0.858 และผู้ตรวจคนที่ 1 , 3 กับ 4 มีค่า RAI ต่ำสุด คือ 0.830 และกรณีผู้ตรวจสี่คน มีค่า RAI สูงสุด คือ 0.851 แสดงว่าการตรวจวิธีประเมินรวม มีแนวโน้มให้ค่าความเชื่อมั่นของผู้ตรวจค่าสูงกว่าวิธีตรวจแบบวิเคราะห์ย่อย ทุกกรณี

3. ค่าเฉลี่ยความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียงที่มีผู้ตรวจจำนวนสองคน สามคน และสี่คน เมื่อใช้การตรวจวิธีวิเคราะห์ย่อย มีค่าเท่ากับ 0.797 , 0.818 และ 0.828 ตามลำดับ มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ (UX_1 = 0.263 , p = .594) และเมื่อใช้การตรวจวิธีประเมินรวม มีค่าเท่ากับ 0.829 , 0.839 และ 0.8510 ตามลำดับ มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ (UX_1 = 0.189 , p = .513)

4. ค่าเฉลี่ยความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง เมื่อใช้ผู้ตรวจสองคน ตรวจวิธีวิเคราะห์ย่อย กับ ตรวจวิธีประเมินรวม มีค่าเท่ากับ 0.797 และ 0.829 ตามลำดับ มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ (UX_1 = 3.059 , p = 0.764) เมื่อใช้ผู้ตรวจสามคน มีค่าเท่ากับ 0.818 และ 0.839 ตามลำดับ มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ (UX_1 = 1.466 , p = 0.224) และ เมื่อใช้ผู้ตรวจสี่คน มีค่าเท่ากับ 0.828 และ 0.851 ตามลำดับ มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ (UX_1 = 2.079 , p = 0.145)

อภิปรายผล

การศึกษาครั้งนี้มุ่งที่จะศึกษาและเปรียบเทียบผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียงที่มีผู้ตรวจจำนวนสองคน สามคน และ สี่คน เมื่อใช้การตรวจวิธีเดียวกัน และผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียงที่ใช้การตรวจต่างวิธีกัน เมื่อมีผู้ตรวจจำนวนเท่ากัน โดยคำนวณหาค่าความเชื่อมั่นของผู้ตรวจด้วยสูตร RAI ของเบอร์รี่-สต็อก และคนอื่น ๆ (Burry-Stock and others.1996) และทดสอบความแตกต่างค่าความเชื่อมั่นของผู้ตรวจด้วยสูตร UX_1 ว่าการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจที่มีผู้ตรวจจำนวนต่างกัน ได้ค่าใกล้เคียงกันหรือไม่ และเมื่อตรวจด้วยการตรวจต่างวิธีกัน ให้ค่าใกล้เคียงกันหรือไม่ ซึ่งผลการศึกษาสามารถนำมาอภิปรายได้ดังนี้

1. เมื่อพิจารณาเปรียบเทียบผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ชั้นประถมศึกษาปีที่ 5 ที่ใช้การตรวจวิธีเดียวกัน แต่มีผู้ตรวจจำนวนต่างกัน คือ สองคน สามคน และ สี่คน มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ซึ่งไม่สอดคล้องกับสมมุติฐานที่ตั้งไว้ แสดงว่า จำนวนผู้ตรวจไม่มีผลต่อคะแนนที่ได้จากการตรวจ ผู้ตรวจทั้งสี่คนมีความเชื่อถือได้ไม่แตกต่างกัน แต่ข้อมูลเบื้องต้นก็พบว่ามีแนวโน้มที่ผู้ตรวจจำนวนสี่คน จะให้ค่าความเชื่อมั่นของผู้ตรวจสูงกว่าผู้ตรวจจำนวนสองคน แต่เมื่อทดสอบสมมุติฐาน พบว่าค่าความเชื่อมั่นของผู้ตรวจทั้งหมดไม่แตกต่างกัน น่าจะมีสาเหตุมาจากการกำหนดให้ผู้ตรวจมีคุณสมบัติใกล้เคียงกัน คือ เป็นครูผู้สอนวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ระดับชั้นประถมศึกษาปีที่ 5 และมีอายุการทำงานมากกว่า 5 ปี ทั้งหมด ทุกคนมีความชำนาญในการตรวจให้คะแนน และ การศึกษาคู่มือการตรวจที่กำหนดให้ก่อนการตรวจทั้ง 2 วิธี ก็สามารถส่งผลต่อค่าความเชื่อมั่นของผู้ตรวจให้มีค่าใกล้เคียงกันมากที่สุด เช่นเดียวกับ ฮอบกินส์ และแอนเตส (Hopkins and Antes. 1990:235) กล่าวว่าค่าความเชื่อมั่นของผู้ตรวจจะสูงก็ต่อเมื่อผู้ตรวจเป็นผู้ที่เขียนข้อสอบหรือผู้ที่สอนวิชานั้น ๆ จริง ๆ ซึ่งสอดคล้องกับงานวิจัยของ อุษณีย์ บัวศิริพันธุ์ (2543 : 55-56) ที่กล่าวว่า ประสบการณ์ของผู้ตรวจที่มีอายุการทำงานมากกว่า 5 ปี มีความเชี่ยวชาญในเนื้อหาวิชา และผ่านการตรวจให้คะแนนแบบทดสอบความเรียงมามาก และมีการกำหนดเกณฑ์การให้คะแนนที่เป็นมาตรฐาน จะส่งผลต่อคะแนนที่ได้จากการตรวจดีกว่า นอกจากนี้ยังสอดคล้องกับ พรรณี เจียมสุบุตร (2543) ที่ศึกษาแล้วพบว่า การตรวจแบบทดสอบความเรียงที่ใช้จำนวนผู้ตรวจมากกว่า 2 คน ให้ความเชื่อมั่นที่ไม่แตกต่างกัน เช่นเดียวกับงานวิจัยนี้ รวมทั้ง ไอโซนิโอ (Isonio. Steven. 1991) ที่เสนอว่าการตรวจผลงานการเขียนของนักเรียน โดยวิธีประเมินภาพรวม ด้วยผู้ตรวจทีละคู่ ผลปรากฏว่าทั้งสองคนให้คะแนนสอดคล้องกันเมื่อมีคู่มือแนะนำการตรวจประกอบการตรวจให้คะแนน แสดงว่าการตรวจแบบทดสอบความเรียง สามารถตรวจให้คะแนนโดยใช้ผู้ตรวจจำนวนกี่คนก็ได้ โดยต้องมีคู่มือการตรวจให้คะแนนทุกครั้ง ซึ่งในที่นี้ผู้วิจัยเสนอให้ใช้ผู้ตรวจ 2 คน โดยควรคัดเลือกผู้ตรวจ ที่เป็นครูผู้สอนวิชานั้น ๆ และมีประสบการณ์การสอนมาพอสมควร

2. เมื่อพิจารณาเปรียบเทียบผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ชั้นประถมศึกษาปีที่ 5 ที่มีผู้ตรวจจำนวนเท่ากัน แต่ใช้การตรวจต่างวิธีกัน คือ ตรวจด้วยวิธีวิเคราะห์ย่อย และวิธีประเมินรวม มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ซึ่งไม่สอดคล้องกับสมมุติฐานที่ตั้งไว้เช่นกัน แสดงว่า การตรวจด้วยวิธีวิเคราะห์ย่อย กับวิธีประเมินรวม ให้ผลการตรวจให้คะแนนไม่แตกต่างกัน ซึ่งอาจเนื่องมาจาก คุณภาพของคู่มือการตรวจ การศึกษาคู่มือการตรวจ และการชี้แจงก่อนการตรวจแบบทดสอบความเรียง โดยกำหนดให้ตรวจให้เสร็จเป็นรายข้อแล้วจึงตรวจข้อต่อไป เหมือนกันทุกคน และทุกคนมีอิสระในการตรวจ ไม่ถูกการกำหนดเวลามาบีบรัด และตรวจโดยมิได้มีสิ่งกดดัน เช่น กลุ่มเพื่อนครู กลุ่มตัวอย่างไม่ได้เป็นนักเรียนที่ตนเองสอน การตรวจไม่มีผลต่อคะแนนของนักเรียน สิ่งต่าง ๆ เหล่านี้

มีผลต่อคะแนนที่ได้จากการตรวจแบบทดสอบความเรียง แม้ว่าจะมีจำนวนผู้ตรวจกี่คนก็ตาม ซึ่งสอดคล้องกับงานวิจัยของ จิราพร อัครสมพงษ์ (2536) ที่พบว่า การดำเนินการตรวจลักษณะนี้จะให้ค่าสัมประสิทธิ์ความเชื่อมั่นสูงสุด เช่นเดียวกับงานวิจัยของ ซูซาน เลน และคนอื่น ๆ (Zuzanne Lane and others. 1996 : 71-91) ที่ว่าความคลาดเคลื่อนของการตรวจจะลดลงเมื่อมีการตรวจโดยมีเกณฑ์การให้คะแนน ดังนั้นการวางแผนการตรวจให้คะแนนผลงานหรือแบบทดสอบความเรียงที่ให้มีค่าใกล้เคียงกันมากที่สุด จะใช้วิธีการตรวจแบบใดก็ได้ตามจุดมุ่งหมายของการสอบ กล่าวคือ ถ้าต้องการทดสอบแบบอิงเกณฑ์ ควรใช้การตรวจวิธีวิเคราะห์ย่อย ซึ่งผลที่ได้เมื่อตรวจด้วยวิธีนี้จะให้คะแนนที่มีความเที่ยงตรง ถ้ามีจุดมุ่งหมายทดสอบแบบอิงกลุ่ม ควรใช้การตรวจวิธีประเมินรวม ซึ่งจะให้คะแนนที่มีความเชื่อมั่นดีกว่า (บุญเชิด ภิญโญอนันตพงษ์. 2525 : 33) นอกจากนี้ควรออกแบบคู่มือการตรวจที่กำหนดเกณฑ์การให้คะแนน(Rubric) ทุกระดับคะแนนให้ชัดเจน พร้อมกับประชุมชี้แจงผู้ตรวจก่อนการตรวจทุกครั้งเสมอ

ข้อเสนอแนะ

ผลการวิจัยครั้งนี้ สามารถสรุปเป็นข้อเสนอแนะในการนำไปใช้และในการวิจัยได้ดังนี้

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 การตรวจแบบทดสอบความเรียง ผลงานการเขียน ผลงานภาคปฏิบัติใด ๆ ผู้สอนสร้างคู่มือการตรวจให้คะแนนก่อนทุกครั้ง โดยจะกำหนดการตรวจด้วยวิธีวิเคราะห์ย่อย หรือ วิธีประเมินรวม อย่างใดอย่างหนึ่งก็ได้

1.2 การตรวจแบบทดสอบความเรียง ผลงานการเขียน ผลงานภาคปฏิบัติใด ๆ ก็ตาม ควรกำหนดคุณสมบัติผู้ตรวจให้ชัดเจน คือ ต้องเป็นครูผู้สอนวิชานั้น ๆ และมีประสบการณ์การสอน เพื่อให้เกิดความเชื่อมั่นของผู้ตรวจได้อย่างแท้จริง โดยมิต้องคำนึงถึงจำนวนผู้ตรวจ

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 ควรศึกษาลักษณะเดียวกันนี้ โดยสร้างเครื่องมือการวิจัยเป็นการประกวดภาพวาดศิลปะ หรือประกวดงานเขียนใด ๆ โดยเลือกกลุ่มตัวอย่างเป็นนักเรียนระดับมัธยมศึกษาขึ้นไป หรือ ผู้ใหญ่ และให้ผู้ตรวจมีจำนวนสองคนขึ้นไป

2.2 ควรศึกษาลักษณะเดียวกันนี้ โดยนำข้อมูลมาศึกษาเปรียบเทียบโดยใช้สูตรในกลุ่มสัมประสิทธิ์ความสอดคล้องของผู้ตรวจ อื่น ๆ (Agreement Coefficient)

บรรณานุกรม

บรรณานุกรม

กรมวิชาการ. (2539). *การประเมินผลจากสภาพจริง(Authentic Assessment).* กรุงเทพฯ : คุรุสภาลาดพร้าว.

กนกวรรณ เอี่ยมชัย.(2539). *การศึกษาค่าความเชื่อมั่นของแบบทดสอบเอ็มอีคิว วิชาแนวคิดพื้นฐานและหลักการพยาบาลที่ตรวจให้คะแนนต่างกัน.* ปริญญานิพนธ์ กศ.ม. (การวัดผลการศึกษา).กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

โกวิท ประวาลพฤกษ์ และ สมศักดิ์ สินธุระเวชญ์. (2525). *การประเมินในชั้นเรียน.* กรุงเทพฯ : วัฒนาพาณิช.

คุรุสภา. (2542) *พิธีมอบเกียรติบัตรครูดีเด่น (ครั้งที่ 2)ประจำปี 2541.* กรุงเทพฯ : โรงเรียนวัฒนาวิทยาลัย.

จำนง พรายแย้มแข. (2530). *เทคนิคการวัดและประเมินผลการเรียนรู้กับการสอนซ่อมเสริม(ตามกระบวนการทางวิทยาศาสตร์).* กรุงเทพฯ : ไทยวัฒนาพานิช.

จิราพร อัครสมพงศ์.(2536). *การเปรียบเทียบความสอดคล้องของคะแนนแบบสอบความเรียงจากผู้ตรวจและวิธีการตรวจต่างกัน.* วิทยานิพนธ์ ค.ม. กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

ฉวีวรรณ ครองยุทธ. (2539). *การประมาณค่าความเชื่อมั่นของมาตรวัดประเมินค่าที่มีขนาดกลุ่มตัวอย่างต่างกัน.* ปริญญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

ชวาล แพรัตกุล.(2520). *เทคนิคการเขียนข้อทดสอบ.* กรุงเทพฯ : โรงพิมพ์พิทักษ์อักษร.

_____. (2508). *เทคนิคการวัดผล.* กรุงเทพฯ : วัฒนาพานิช.

ดำเกิง วงศ์สว่าง. (2528). *ผลของการทดสอบย่อยโดยใช้แบบทดสอบเลือกตอบ เติมคำ และเลือกตอบ ที่มีต่อความสามารถในการเขียนภาษาไทย ของนักเรียนชั้นมัธยมศึกษาปีที่ 3* . ปริญญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

บุญชม ศรีสะอาด. (2520). *การวัดและประเมินผลการศึกษา.* คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ มหาสารคาม.

_____. (2534). "การวิเคราะห์ข้อสอบแบบอื่นนอกเหนือจากแบบเลือกตอบ,". *วารสารวัดผลการศึกษา.* 12(36).

_____. (2540). *การวิจัยทางการวัดผลและประเมินผล.* กรุงเทพฯ : สุวีริยาสาส์น.

บุญเชิด ภิญโญอนันตพงษ์. (2521). *การวัดและประเมินผลการศึกษา : ทฤษฎีประยุกต์.* มหาวิทยาลัยศรีนครินทรวิโรฒ.

_____. (2525). *เอกสารประกอบการสอนวิชา วัดผล 301 การสร้างแบบทดสอบ 1.* มหาวิทยาลัยศรีนครินทร วิโรฒ.

_____ . (2527). *การทดสอบแบบอิงเกณฑ์ : แนวคิดและวิธีการ.* กรุงเทพฯ : โอเดียนสโตร์.

บุญเชิด ภิญโญอนันตพงษ์. (2538). "การประเมินผลแนวใหม่ : พอตโฟลิโอ," *ศึกษาศาสตร์.* 1(1) : 59-67.

_____ . (2544). *การประเมินการเรียนรู้เน้นผู้เรียนเป็นสำคัญ : แนวคิดและวิธีการ.* กรุงเทพฯ : สำนักงานคณะกรรมการการศึกษาแห่งชาติ.

บุญธรรม กิจปรีดาบริสุทธิ์. (2524). *คู่มืออาจารย์การวัดและประเมินผลการเรียนการสอน.* กรุงเทพฯ : การพิมพ์พระนคร.

บุศรินทร์ สิริปัญญาธร. (2541). *แนวโน้มและอัตราการเปลี่ยนแปลงพฤติกรรมร่วมมือของเด็กปฐมวัยที่เกิดจากกิจกรรมศิลปะแบบสื่อผสมเป็นกลุ่ม.* ปริญญานิพนธ์ กศ.ม. (การศึกษาปฐมวัย). กรุงเทพฯ : บัณฑิตวิทยาลัย : มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

ประคอง กรรณสูต. (2528). *สถิติเพื่อการวิจัยทางพฤติกรรมศาสตร์. ฉบับปรับปรุงแก้ไข.* กรุงเทพฯ : ศูนย์หนังสือ ดร. ศรีสง่า.

_____. (2540). *สถิติศาสตร์ประยุกต์สำหรับครู.* กรุงเทพฯ : ไทยวัฒนาพานิช .

_____. (2542). *สถิติเพื่อการวิจัยทางพฤติกรรมศาสตร์.* กรุงเทพฯ : สำนักพิมพ์จุฬาลงกรณ์มหาวิทยาลัย.

พรรณี เจียมสุบุตร. (2543). *การเปรียบเทียบความเชื่อมั่นของแบบทดสอบวัดความสามารถในการ แก้โจทย์ปัญหาทางคณิตศาสตร์ที่มีจำนวนผู้ตรวจและวิธีการตรวจต่างกัน.* ปริญญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

พนิดา ฟุ้งเกียรตินำสุข. (2540). *การประมาณค่าความเชื่อมั่นของมาตรวัดประเมินค่าที่มีรูปแบบการตอบแตกต่างกัน.* ปริญญานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

พวงแก้ว ปุณยกนก. (2531). "แบบสอบอัตนัยประยุกต์ (เอ็ม อี คิว) เพื่อใช้วัดทักษะการแก้ปัญหา," *รายงานผลการวิจัย.* กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย.

พวงรัตน์ ทวีรัตน์. (2530). *การสร้างและพัฒนาแบบทดสอบวัดผลสัมฤทธิ์.* สำนักทดสอบทางการศึกษาและจิตวิทยา, มหาวิทยาลัยศรีนครินทรวิโรฒ.

ไพรัตน์ วงษ์นาม. (2533). *สัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง.* วิทยานิพนธ์ ค.ม. (เอกวิจัยและสถิติ) กรุงเทพฯ : บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

เยาวดี วิบูลย์ศรี. (2528). *หลักการวัดผลและการสร้างข้อสอบ.* กรุงเทพฯ : ภาควิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

_____. (2540). *การวัดผลและการสร้างแบบสอบผลสัมฤทธิ์.* กรุงเทพฯ : จุฬาลงกรณ์มหาวิทยาลัย.

รุจิร์ ภู่สาระ. (2520). *เอกสารประกอบการบรรยายวิชา ED.337 .* กรุงเทพฯ : คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ.

ล้วน สายยศ และ อังคณา สายยศ. (2539). *เทคนิคการวัดผลการเรียนรู้.* กรุงเทพฯ : ชมรมเด็ก .

_____. (2540). *สถิติวิทยาทางการวิจัย.* กรุงเทพฯ : สุวีริยาสาส์น.

ล้วน สายยศ และ อังคณา สายยศ. (2541). *เทคนิคการสร้างและสอบข้อสอบความถนัดทางการเรียน.* กรุงเทพฯ : ชมรมเด็ก.

ศิริชัย กาญจนวาสี. (2542). *ทฤษฎีการทดสอบแบบดั้งเดิม.* พิมพ์ครั้งที่ 1. กรุงเทพฯ : ภาควิชาวิจัยการศึกษา คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย.

ส. วาสนา ประวาลพฤกษ์. (2533). "การพิจารณาความเชื่อมั่นของแบบทดสอบการปฏิบัติ," *วารสารวัดผลทางการศึกษา.* 12(35).

สมพงษ์ อัศวศุภฤกษ์. (2542). *การเปรียบเทียบความเชื่อมั่นของแบบทดสอบที่มีวิธีการตรวจ และความยาวของแบบทดสอบต่างกัน.* ปริญญานิพนธ์ กศ.ม.(การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

สุพานี กอหมั่นศิลป์. (2543). *การศึกษาความเชื่อมั่นของแบบประเมินทักษะกีฬาบาสเกตบอลที่ใช้ผู้ประเมินจำนวนต่างกันประเมินนักเรียนจำนวนต่างกัน.* ปริญญานิพนธ์ กศ.ม.(การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

สุมณฑา พรหมบุญ. (2543). "เอนทรานซ์... ปรนัยหรืออัตนัย," *สานปฏิรูป.* 3 (24).

สำนักงานคณะกรรมการการศึกษาแห่งชาติ. (2542). *พระราชบัญญัติการศึกษาแห่งชาติ พ.ศ.2542.* สำนักนายกรัฐมนตรี.

วัญญา วิศาลาภรณ์ (2542). "การวางแผนสร้างแบบทดสอบ," *เอกสารประกอบการเรียนวิชา วัดผล 502.* ถ่ายเอกสาร.

_____. (2533). *การสร้างแบบทดสอบเพื่อการวิจัยและพัฒนาการเรียนการสอน.* กรุงเทพฯ : อักษราพิพัฒน์.

อรอนงค์ บำรุง. (2542). *การแสดงหลักฐานความเที่ยงตรงและความเชื่อมั่นของแบบทดสอบปลายเปิดวิชาคณิตศาสตร์.* ปริญญานิพนธ์ กศ.ม.(การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

อัจฉรา อรุณรักษ์สมบัติ. (2542). *การสร้างแบบทดสอบวัดภาคปฏิบัติโดยวิธีเขียนตอบและวิธีปฏิบัติจริงวิชางานประดิษฐ์ กลุ่มการงานและพื้นฐานอาชีพ ของนักเรียนชั้นประถมศึกษาปีที่ 6.* ปริญญานิพนธ์ กศ.ม.(การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

อุทุมพร จามรมาน. (2531). "ข้อสอบปรนัย : คุณค่าต่อระบบการศึกษาไทย," *วารสารวัดผลการศึกษา.* หน้า 71 - 74.

อุษณีย์ บัวศิริพันธุ์. (2543). *การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจ จำนวนผู้ตรวจ และประสบการณ์ของผู้ตรวจ แตกต่างกัน.* ปริญญานิพนธ์ กศ.ม.(การวัดผลการศึกษา). กรุงเทพฯ : บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

อนันต์ ศรีโสภา. (2524). *การวัดและประเมินผลการศึกษา.* กรุงเทพฯ : ไทยวัฒนาพานิช.

Bloom, B.S. Ed. (1977). *Taxonomy of Education Objective: The Classification of Educational Goals., Handbook 1, Cognitive Domain.* New York : Longman.

Brennan, R.L. (1983). *Eliminates of Generalizability Theory.* Iowa : The American College Testing Program.

Bridgeman and others. (1997). "Choice Among Essay Topics: Impact on Performance and Validity," *Journal of Educational Measurement.* 34(3) : 273-286.

Brown,F.G. (1976). *Principles of Educational and Psychological Testing.* New York : HolT. Rinehart and Winston.

Burry-Stock ,Judith A., Laurie ,Cecelia, Chissom, Brad S. (1996). "Rater Agreement Indexes For Performance Assessment," *Educational and Psychological measurement.* 2 : 251- 261.

Coffman, W.E. (1971). "Essay Examination," *Educational Measurement.* 2nd Ed.

Ebel, R.L. and Frisbie, D.A. (1986). *Essentials of Educational Measurement.* New Jersey : Prentice Hall Inc.

Garner and Engelhard, Jr. (1999). "Gender differences in Performance on multiple-choice and Constructed Response Mathematics Items," *Applied Measurement in Education.* 12(1) : 29 – 51.

Gay, L.R. (1991). *Educational Evaluation and Measurement Competencies for Analysis and Application.* 2nd Ed. Merrill.

Gronlund, Norman E. (1971). *Measurement and Evaluation in Teaching.* 2nd Ed. Macmillan, 218-233.

Gronlund, Norman E. (1977). *Constructing Achievement Test.* 2nd Ed. Prentice-Hall, Inc.

Hopkins, C.D and Antes, R.L. (1990). *Classroom Measurement and Evaluation.* 3rd Ed. Itasca, IL.

Hoyt, C. (1941). "Test reliability obtained by analysis of Variance," *Psychometrika.* 6 :153-160.

Isonio, Steven. (1991). "Judgments of Placement Writing Samples at Golden West College : an Evaluation of Inter-rater Reliability," *NISC DISC REPORT.* ERIC 1985-1999.

J.Randolphlaw & Paul, J. Sherman. (2000). "Do Raters Agree ? Assessing Inter-Rater Agreement in the Evaluation of Air Crew Resource Management Skills," *www.psy.utexas.edu.*

James, L.R., Demaree, R.G., Wolfe, G. (1993). "r_{wg} : An assessment of within-group interrater agreement," *Journal of applied psychology.* 78 : 306-309.

Kaasa, Terry Rae. (1998). "Inter-rater Reliability of Formally Trained and Self – Trained Raters using the Edmonton Functional Assessment tool," *NISC DISC REPORT.* ERIC 1985-1999.

Klein and Stecher. (1998). "Analytic Versus Holistic Scoring of Science Performance Tasks," *Applied Measurement in Education.* 11(2) : 121-137.

Kubiszyn, T. and Borich, G. (1996). *Educational Testing and Measurement : Classroom application and practice.* 5th Ed. Haper Collins College Publishers.

Lane, S. and others. (1996, Spring). "Generalizability and Validity of a Mathematics Performance Assessment," *Journal of Educational Measurement.* 33(1) : 71 –91.

Mehrens, W.A. and Lehmann, I.J. (1973). *Measurement and Evaluation in Education and Psychology.* New York : Holt, Rinehart and Winston.

Oosterhof, Albert. (1996). *Developing and Using Classroom Assessments.* Prentice-Hall, Inc.

Paul, A. Bott. (1996). *Testing and Assessment.* Allyn and Bacon.

Pomplun and Sundbye. (1999). "Gender Differences in Constructed Response Reading Items," *Applied Measurement in Education.* 12(1) : 95-109.

Powers, Donald E. and Bennett, Randy Elliot. (1999). "Effect of Allowing Examinees to Select Questions on a test of Divergent Thinking," *Applied Measurement in Education.* 12(3) : 257-279.

Powers and Fowles. (1998). "Effects of Preexamination Disclosure of Essay Topics," *Applied Measurement in Education.* 11(2) : 139-157.

Robert F. Fagot. (1994). "An Ordinal Coefficient of Relational Agreement For Multiple Judges," *Psychometrika.* 59 (2) : 241-251.

Tasse, Marc J. and others. (1994). "Test-Retest and Inter-rater Reliability of the Quebec Adaptive Behavior Scale." *NISC DISC REPORT.* ERIC 1985-1999.

Torrance, Harry. (1995). *Evaluation Authentic Assessment : Problem and Possibilities in new approaches to assessment.* Biddles Ltd.

Uebersax, John. (2000). "Raw Agreement Indices," *http://ourworld.compuserve.com.*

Ward, Annie W. (1999). *Assessment in the Classroom.* p. 421. Wadsworth Publishing Company.

Wiersma and Jurs. (1985). *Educational Measurement and testing.* Allyn and Bacon,Inc.

Woodruff, D.J. & Feldt, L.S. (1986). "Test for equality of several alpha coefficients when their sample estimates are dependent," *Psychometrika.* 51 : 393-413.

ภาคผนวก

ภาคผนวก ก

1. รายนามผู้เชี่ยวชาญในการตรวจสอบคุณภาพเครื่องมือ
2. ดัชนีความสอดคล้องของผู้เชี่ยวชาญ (IOC)
3. ดัชนีความง่าย ดัชนีค่าอำนาจจำแนกของแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต (สังคม)
4. คู่มือการตรวจ แบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ระดับชั้นประถมศึกษาปีที่ 5 การตรวจวิธีวิเคราะห์ย่อย (Analytic Method)
5. คู่มือการตรวจ แบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ระดับชั้นประถมศึกษาปีที่ 5 การตรวจวิธีประเมินรวม (Holistic Method)

รายนามผู้เชี่ยวชาญตรวจสอบคุณภาพเครื่องมือ

ผู้วิจัยได้เลือกผู้เชี่ยวชาญที่มีวุฒิการศึกษาตั้งแต่ระดับปริญญาโท สาขาวิชาเอกการวัดผลการศึกษา และ ครูผู้สอน ที่มีวุฒิปริญญาตรี วิชาเอกสังคมศึกษา และมีประสบการณ์การสอนวิชาสังคมศึกษามาเป็นเวลา 5 ปี ขึ้นไป ซึ่งมีรายนามผู้เชี่ยวชาญในการตรวจสอบคุณภาพเครื่องมือ ดังนี้

1. ผู้เชี่ยวชาญทางการวัดผล

1. รศ.วัญญา วิศาลาภรณ์	ภาควิชาการวัดผลและวิจัยทางการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ
2. นางสาวยุพาวดี รัชดาพรรณาธิกุล	อาจารย์ฝ่ายวิชาการ โรงเรียนอัสสัมชัญ กรุงเทพ

2. ผู้เชี่ยวชาญทางการสอนวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม)

3. นางปราณี วงศ์สมบุญ	ผู้ช่วยครูใหญ่ฝ่ายปกครอง โรงเรียนเปรมฤดีศึกษา
4. นางสุกัญญา วิเชียรดิลก	ครูผู้สอนวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) โรงเรียนอัสสัมชัญแผนกประถม
5. นางสาวเพชรรัตน์ ชมจันทร์	ครูผู้สอนวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) โรงเรียนอัสสัมชัญแผนกประถม

ตาราง 11 ค่าดัชนีความสอดคล้องของผู้เชี่ยวชาญตรวจสอบคุณภาพเครื่องมือ (IOC)

ข้อ	ผู้เชี่ยวชาญคนที่					ค่า IOC	ผลการคัดเลือก
	1	2	3	4	5		
1	1	1	0	1	1	0.8	ใช้ได้
2	1	1	0	1	1	0.8	ใช้ได้
3	1	1	0	0	1	0.6	ใช้ได้
4	1	0	0	1	1	0.6	ใช้ได้
5	1	1	0	1	1	0.8	ใช้ได้
6	1	1	0	1	1	0.8	ใช้ได้
7	1	1	0	0	1	0.6	ใช้ได้
8	1	1	0	1	1	0.6	ใช้ได้
9	1	1	0	0	1	0.6	ใช้ได้
10	1	1	1	1	1	1.0	ใช้ได้

ตาราง 12 ค่าความยาก (P) ค่าอำนาจจำแนก (t) ของแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต (สังคม)

ข้อ	ค่าความยาก (P)	ค่าอำนาจจำแนก (t)	ผลการคัดเลือก
1	0.464	1.486	
2	0.291	1.700	คัดเลือกไว้
3	0.171	3.180	
4	0.324	2.988	คัดเลือกไว้
5	0.236	4.573	คัดเลือกไว้
6	0.147	0.787	
7	0.086	1.494	
8	0.204	2.796	คัดเลือกไว้
9	0.173	1.244	
10	0.224	2.267	คัดเลือกไว้

คู่มือการตรวจแบบทดสอบความเรียง
วิชาสร้างเสริมประสบการณ์ชีวิต (สังคม)
ชั้นประถมศึกษาปีที่ 5
การตรวจวิธีวิเคราะห์ย่อย (Analytic Method)

คำชี้แจง

คู่มือตรวจนี้ประกอบด้วย

- แบบทดสอบแบบความเรียง จำนวน 5 ข้อ
- แนวคำตอบ
- เกณฑ์การตรวจให้คะแนนวิธีวิเคราะห์ย่อย

ขั้นตอนการตรวจ ดำเนินการดังนี้

1. ให้คุณครูอ่านคำตอบจากแบบทดสอบแต่ละข้อ ที่นักเรียนตอบคำถาม 1 รอบ
2. ใช้คู่มือการตรวจนี้เป็นแนวทางในการตรวจคำตอบของนักเรียน โดยพิจารณาว่าคำตอบของนักเรียน ถูกต้องตามแนวคำตอบหรือไม่ มากน้อยเพียงใด
3. ตรวจโดยวิธีวิเคราะห์ย่อย (Analytic Method) ซึ่งมีขั้นตอนการตรวจ ดังนี้

    4.1 พิจารณาคำตอบตามเกณฑ์การให้คะแนน ต่อไปนี้

        4.1.1 ความถูกต้องของเนื้อหา คะแนนเต็ม 2 คะแนน

        4.1.2 การนำเสนอหลักการที่ถูกต้อง คะแนนเต็ม 2 คะแนน

        4.1.3 การจัดเรียบเรียงความคิด คะแนนเต็ม 1 คะแนน

    4.2 ตรวจคำตอบทีละข้อ ทีละเกณฑ์ตามคู่มือการตรวจ ทีละข้อ หลังจากครบ 60 คนแล้วจึงเริ่มตรวจข้อต่อไปจนครบ 5 ข้อ

1. ให้นักเรียนอธิบายถึงผลดีของการส่งเสริมให้ชาวบ้านในชนบทประกอบอาชีพอุตสาหกรรมในครัวเรือนเพิ่มขึ้น และวิเคราะห์ถึงปัญหาที่จะเกิดขึ้นมา 2 ปัญหา

แนวคำตอบ

สิ่งที่นักเรียนควรกล่าวถึงผลดีของการส่งเสริมดังกล่าว มีตัวอย่างดังนี้

1. ช่วยให้คนในชนบทมีงานทำ
2. ลดอัตราการย้ายถิ่นเข้าเมือง
3. นำวัตถุดิบที่มีในท้องถิ่นมาแปรรูปทำให้ขายได้ราคาดี
4. ลดปริมาณการสูญเสียของวัตถุดิบที่มีในท้องถิ่น
5. สร้างความมั่นคงให้แก่ชุมชน

ฯลฯ

นักเรียนควรกล่าวถึงปัญหาที่เกิดขึ้น อาทิเช่น

1. ระยะเวลาในการผลิตใช้เวลานาน
2. ไม่มีหน่วยงานหาตลาดรองรับสินค้าที่ผลิต
3. ไม่มีเงินลงทุน หรือขยายกิจการ
4. ขาดบุคลากรที่เชี่ยวชาญทางด้านเทคนิคการผลิต
5. ปัญหาพ่อค้าคนกลาง รับซื้อสินค้าที่ผลิตได้ในราคาถูก

ฯลฯ

เกณฑ์การตรวจให้คะแนน

1. ความถูกต้องของเนื้อหา 2 คะแนน

ให้ 2 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามเนื้อหา มีผลดีและปัญหามากกว่าที่กำหนดให้นอกเหนือจาก แนวคำตอบ

1 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามแนวคำตอบ

0 คะแนน เมื่อ ตอบคำถามไม่ถูกต้องตามแนวคำตอบหรือไม่ถูกต้องตามเนื้อหา

2. การนำเสนอหลักการที่ถูกต้อง 2 คะแนน

ให้ 2 คะแนน เมื่อ นำเสนอหลักการที่สำคัญ ๆ ได้ทั้ง 2 หลักการ คือ ผลดีของการส่งเสริมและปัญหาที่จะเกิดขึ้น

1 คะแนน เมื่อ นำเสนอหลักการโดยรวม ไม่แยกออกเป็น 2 หลักการได้ชัดเจน

0 คะแนน เมื่อ ไม่สามารถนำเสนอหลักการจากคำถามได้

3. การจัดเรียบเรียงความคิด 1 คะแนน

ให้ 1 คะแนน เมื่อ สามารถจัดเรียบเรียงความคิดได้ชัดเจน เข้าใจง่าย

0 คะแนน เมื่อ ไม่สามารถจัดเรียบเรียงความคิดให้เข้าใจได้

2. ให้นักเรียนจัดกลุ่มประเทศเพื่อนบ้านต่อไปนี้ตามลักษณะการปกครองที่คล้ายคลึงกัน และบอกลักษณะการปกครองของแต่ละกลุ่ม มาให้ถูกต้อง

มาเลเซีย สาธารณรัฐสิงคโปร์ สาธารณรัฐสังคมนิยมเมียนม่า(พม่า)
สาธารณรัฐฟิลิปปินส์ สาธารณรัฐอินโดนีเซีย กัมพูชาประชาธิปไตย
สาธารณรัฐสังคมนิยมเวียดนาม ไทย สาธารณรัฐประชาธิปไตยประชาชนลาว
เนรากาบรูไนดารุสซาลาม

แนวคำตอบ

นักเรียนต้องจำแนกลักษณะการปกครองของประเทศต่าง ๆ เหล่านี้ออกเป็นกลุ่ม ดังนี้

กลุ่ม 1 มาเลเซีย ไทย กัมพูชาประชาธิปไตย
มีลักษณะการปกครองแบบประชาธิปไตย มีพระมหากษัตริย์เป็นประมุข

กลุ่ม 2 สาธารณรัฐสิงคโปร์ สาธารณรัฐอินโดนีเซีย สาธารณรัฐฟิลิปปินส์
มีลักษณะการปกครองแบบสาธารณรัฐ มีประธานาธิบดีเป็นประมุข

กลุ่ม 3 สาธารณรัฐสังคมนิยมเมียนม่า(พม่า) สาธารณรัฐประชาธิปไตยประชาชนลาว
สาธารณรัฐสังคมนิยมเวียดนาม
มีลักษณะการปกครองแบบสังคมนิยม มีประธานาธิบดีเป็นประมุข

กลุ่ม 4 เนรากาบรูไนดารุสซาลาม
มีลักษณะการปกครองแบบสมบูรณาญาสิทธิราชย์ มีสุลต่านเป็นประมุข

เกณฑ์การตรวจให้คะแนน

1. ความถูกต้องของเนื้อหา 2 คะแนน

ให้ 2 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามเนื้อหา ถูกต้องตามแนวคำตอบอย่างสมบูรณ์ จำแนกประเทศที่กำหนดให้เป็นกลุ่ม ๆ ตามลักษณะการปกครองได้ถูกต้อง

1 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามแนวคำตอบ แต่มีผิดไม่เกิน 3 ใน 10 ประเทศ

0 คะแนน เมื่อ ตอบคำถามไม่ถูกต้องตามแนวคำตอบหรือไม่ถูกต้องตามเนื้อหา

2. การนำเสนอหลักการที่ถูกต้อง 2 คะแนน

ให้ 2 คะแนน เมื่อ นำเสนอหลักการที่สำคัญ ๆ ได้ คือ จำแนกกลุ่มประเทศได้ถูกต้อง บอกชื่อประเทศที่อยู่ในกลุ่มได้ และบอกลักษณะการปกครองได้

1 คะแนน เมื่อ นำเสนอหลักการโดยรวม ไม่จัดกลุ่มประเทศตามแนวคำตอบ หรือจำแนกประเทศแต่บอกชื่อประเทศไม่ถูกต้อง หรือบอกลักษณะการปกครองไม่สมบูรณ์

0 คะแนน เมื่อ ไม่สามารถนำเสนอหลักการจากคำถามได้เลย

3. การจัดเรียบเรียงความคิด 1 คะแนน

ให้ 1 คะแนน เมื่อ สามารถจัดเรียบเรียงความคิดได้ชัดเจน เข้าใจง่าย

0 คะแนน เมื่อ ไม่สามารถจัดเรียบเรียงความคิดให้เข้าใจได้

3. ให้นักเรียนอ่านข่าวต่อไปนี้ แล้วอภิปรายถึงมาตรการที่เกิดขึ้นตามข่าวนี้ จะก่อให้เกิดผลเสียต่อประเทศเพื่อนบ้านดังที่ข่าวกล่าวถึงอย่างไร และเสนอวิธีการแก้ปัญหานี้มา 2 วิธี

กัวลาลัมเปอร์ : ตำรวจมาเลเซียจะตั้งด่านสกัดจับผู้ขับขี่ชาวต่างชาติซึ่งเคยถูกเจ้าหน้าที่ให้ใบสั่งฐานฝ่าฝืนการจราจรของมาเลเซียมานานแล้ว แต่ไม่ยอมมาเสียค่าปรับซะที ทั้งนี้ ตำรวจมาเลเซียระบุจะเอาจริงกับผู้ฝ่าฝืน ซึ่งมีมากกว่า 93,000 คน ซึ่งในจำนวนนี้มีเพื่อนบ้านอย่างบรูไนสิงคโปร์ และไทยรวมอยู่ในบัญชีรายชื่อด้วย ข่าวแจ้งว่า ชาวสิงคโปร์ถูกขึ้นบัญชีดำมากที่สุดถึง 85,194 คน ขณะที่ไทยมีเพียง 1,086 คน เท่านั้น

แนวคำตอบ

การอภิปรายของนักเรียน อาจกล่าวในลักษณะต่อไปนี้

1. กล่าวถึงผลกระทบต่อความสัมพันธ์ระหว่างประเทศ
2. กล่าวถึงการลดจำนวนนักท่องเที่ยวเข้าประเทศในประเทศมาเลเซีย
3. กล่าวถึงขั้นตอนการขอวีซ่าเข้าประเทศเข้มงวด ส่งผลต่อการค้าขายระหว่างชายแดนในแต่ละประเทศยากขึ้น

นักเรียนอาจเสนอแนวทางแก้ปัญหา อาทิเช่น

1. ผู้เข้าประเทศต้องศึกษาหาความรู้เกี่ยวกับกฎหมายของประเทศนั้น ๆ ก่อนเดินทาง
2. สถานฑูตต้องชี้แจงข้อบังคับหรือกฎหมายที่เข้มงวดแก่นักท่องเที่ยว
3. งดการเดินทางเข้าประเทศนั้น ๆ

เกณฑ์การตรวจให้คะแนน

1. ความถูกต้องของเนื้อหา 2 คะแนน

ให้ 2 คะแนน เมื่อ อภิปรายและเสนอวิธีการแก้ปัญหาได้ตามลักษณะที่กำหนดให้ มีประเด็นเพิ่มเติมนอกเหนือจากแนวคำตอบ

1 คะแนน เมื่อ อภิปรายและเสนอวิธีการแก้ปัญหาได้ตามลักษณะที่กำหนดให้เท่านั้น หรืออภิปรายและเสนอวิธีการแก้ปัญหาตามความคิดของตนเองนอกเหนือจากแนวคำตอบ และยังเกี่ยวข้องกับเนื้อหาตามที่คำถามกำหนด

0 คะแนน เมื่อ ไม่สามารถอภิปรายและเสนอวิธีการแก้ปัญหาตามที่คำถามกำหนดให้ หรืออภิปรายและเสนอวิธีการแก้ปัญหาแต่ไม่เกี่ยวข้องกับเนื้อหาตามที่คำถามกำหนด

2. การนำเสนอหลักการที่ถูกต้อง 2 คะแนน

ให้ 2 คะแนน เมื่อ นำเสนอหลักการที่สำคัญ ๆ ได้ คือ อภิปรายผลกระทบที่เกิดจากข่าวนี้ และ เสนอวิธีการแก้ปัญหาได้

1 คะแนน เมื่อ นำเสนอหลักการจากคำถามได้หลักการใดหลักการหนึ่งเท่านั้น

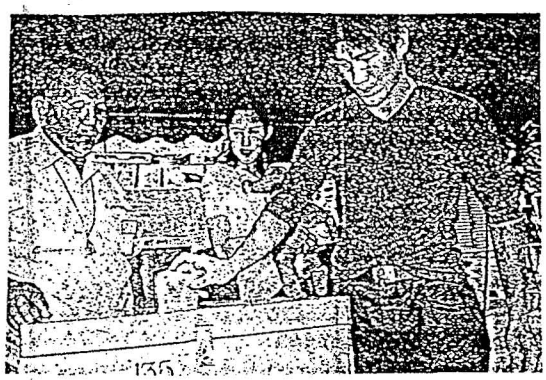
0 คะแนน เมื่อ ไม่สามารถนำเสนอหลักการจากคำถามได้เลย

3. การจัดเรียบเรียงความคิด 1 คะแนน

ให้ 1 คะแนน เมื่อ สามารถจัดเรียบเรียงความคิดในการอภิปรายและเสนอวิธีการแก้ปัญหาได้ชัดเจน เข้าใจง่าย

0 คะแนน เมื่อ ไม่สามารถจัดเรียบเรียงความคิดให้เข้าใจได้

4. จงดูภาพต่อไปนี้ แล้วจงเขียนข่าวสั้นจากภาพมาให้ถูกต้องตามวิธีการเขียนข่าว



แนวคำตอบ

นักเรียนจะต้องเขียนข่าวสั้นให้สอดคล้องกับภาพที่กำหนดให้ ตัวอย่างเช่น

หัวข้อข่าว "พลเมืองดีของชาติ" / "สมรักษ์เลือกตั้ง" / "ไม่นอนหลับทับสิทธิ์"

วัน เหตุการณ์ที่เกิดขึ้น สถานที่ที่ไป หรือ คำให้สัมภาษณ์)

เกณฑ์การตรวจให้คะแนน

1. ความถูกต้องของเนื้อหา 2 คะแนน

ให้ 2 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามเนื้อหา และวิธีการเขียนข่าวสั้นจากภาพ

1 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามวิธีการเขียนข่าว แต่ไม่สอดคล้องกับภาพ

0 คะแนน เมื่อ ไม่สามารถเขียนข่าวสั้นได้สอดคล้องกับภาพ

2. การนำเสนอหลักการที่ถูกต้อง 2 คะแนน

ให้ 2 คะแนน เมื่อ นำเสนอหลักการที่สำคัญ ๆ ได้ คือ หัวข้อข่าว และรายละเอียดของข่าว

1 คะแนน เมื่อ นำเสนอหลักการจากคำถามได้หลักการใดหลักการหนึ่งเท่านั้น แต่เขียนรายละเอียดของข่าวไม่ชัดเจน

0 คะแนน เมื่อ ไม่สามารถนำเสนอหลักการจากคำถามได้เลย

3. การจัดเรียบเรียงความคิด 1 คะแนน

ให้ 1 คะแนน เมื่อ สามารถจัดเรียบเรียงความคิดในการเขียนข่าวได้สอดคล้องกับภาพได้ชัดเจน

0 คะแนน เมื่อ ไม่สามารถจัดเรียบเรียงความคิดได้ตามวิธีการเขียนข่าวได้

5. ให้นักเรียนศึกษาข่าว แล้วสรุปข่าวตามประเด็นสำคัญต่อไปนี้ หัวข้อข่าว ประเภทของข่าว แหล่งข่าว และใจความสำคัญของข่าว ให้ครบถ้วน

ยันผงชูรสไม่อันตราย : นายแพทย์วิชัย ขัติวิทยากุล สาธารณสุข จ.ขอนแก่น เปิดเผยว่าจากการออกข่าวเผยแพร่ของ อย. ได้ชี้แจงให้ผู้บริโภคให้คลายความวิตกกังวลถึงอันตรายของผงชูรสอย่างเป็นทางการแล้ว ระบุว่า มีความเป็นพิษต่ำมากและไม่ก่อให้เกิดความเสี่ยงอันตรายต่อผู้บริโภค ยกเว้นในรายที่แพ้ผงชูรส แต่อาการมักไม่รุนแรง มักหายเอง ผงชูรสมีชื่อทางเคมีว่า โมโนโซเดียมกลูตาเมท เป็นอาหารเฉพาะ ซึ่งผู้ผลิตและผู้นำเข้าต้องได้รับใบอนุญาตผลิตและนำเข้า โดยขอขึ้นทะเบียนตำรับอาหารมีคุณภาพมาตรฐานตามที่กฎหมายกำหนดไว้ เพื่อเป็นหลักประกันความปลอดภัยให้กับผู้บริโภค และข้อสำคัญในการเลือกซื้อผงชูรสที่ได้มาตรฐาน

ที่มา นสพ.เดลินิวส์ วันที่ 6 มกราคม 2544

แนวคำตอบ

นักเรียนต้องตอบคำถามนี้ ตามแนวคำตอบอย่างสมบูรณ์ จึงจะถูกต้องตามเนื้อหา ดังนี้

หัวข้อข่าว ยันผงชูรสไม่อันตราย

ประเภทของข่าว ข่าวสุขภาพ / ข่าวทั่วไป

แหล่งข่าว จากหนังสือพิมพ์

ใจความสำคัญของข่าว "นายแพทย์วิชัย ขัติวิทยากุล สาธารณสุข จ.ขอนแก่น เปิดเผยว่าจากการออกข่าวเผยแพร่ของ อย. ได้ชี้แจงให้ผู้บริโภคให้คลายความวิตกกังวลถึงอันตรายของผงชูรสอย่างเป็นทางการแล้ว ระบุว่า มีความเป็นพิษต่ำมากและไม่ก่อให้เกิดความเสี่ยงอันตรายต่อผู้บริโภค ยกเว้นในรายที่แพ้ผงชูรส แต่อาการมักไม่รุนแรง"

เกณฑ์การตรวจให้คะแนน

1. ความถูกต้องของเนื้อหา 2 คะแนน

ให้ 2 คะแนน เมื่อ บอกหลักการสรุปข่าวจากข่าวที่กำหนดให้ได้ตามแนวคำตอบอย่างครบถ้วน

1 คะแนน เมื่อ บอกหลักการสรุปข่าวจากข่าวที่กำหนดให้ได้ ตามแนวคำตอบแต่ไม่สมบูรณ์

0 คะแนน เมื่อ ไม่สามารถสรุปข่าวจากข่าวที่กำหนดให้ได้

2. การนำเสนอหลักการที่ถูกต้อง 2 คะแนน

ให้ 2 คะแนน เมื่อ นำเสนอหลักการที่สำคัญ ๆ ได้ คือ หัวข้อข่าว ประเภทของข่าว แหล่งข่าว และใจความสำคัญของข่าว ได้ครบถ้วน

1 คะแนน เมื่อ นำเสนอหลักการจากคำถามได้ แต่ขาดหัวข้อใดหัวข้อหนึ่ง

0 คะแนน เมื่อ ไม่สามารถนำเสนอหลักการจากคำถามได้เลย

3. การจัดเรียบเรียงความคิด 1 คะแนน

ให้ 1 คะแนน เมื่อ สามารถจัดเรียบเรียงความคิดได้ตามวิธีการสรุปข่าวได้ชัดเจน

0 คะแนน เมื่อ ไม่สามารถจัดเรียบเรียงความคิดได้ตามวิธีการสรุปข่าวได้

คู่มือการตรวจแบบทดสอบความเรียง
วิชาสร้างเสริมประสบการณ์ชีวิต (สังคม)
ชั้นประถมศึกษาปีที่ 5
การตรวจวิธีประเมินรวม (Holistic Method)

คำชี้แจง

คู่มือตรวจนี้ประกอบด้วย

- แบบทดสอบแบบความเรียง จำนวน 5 ข้อ
- แนวคำตอบ
- เกณฑ์การตรวจให้คะแนนวิธีประเมินรวม

ขั้นตอนการตรวจ ดำเนินการดังนี้

1. ให้คุณครูอ่านคำตอบจากแบบทดสอบแต่ละข้อ ที่นักเรียนตอบคำถาม 1 รอบ
2. ใช้คู่มือการตรวจนี้เป็นแนวทางในการตรวจคำตอบของนักเรียน โดยพิจารณาว่าคำตอบของนักเรียน ถูกต้องตามแนวคำตอบหรือไม่ มากน้อยเพียงใด
3. (ดำเนินการหลังการตรวจวิธีวิเคราะห์ย่อยเสร็จแล้ว 1 สัปดาห์) ใช้การตรวจวิธีประเมินรวม (Holistic Method) ดำเนินการตรวจตามขั้นตอนต่อไปนี้

4.3 พิจารณาคำตอบตามเกณฑ์การให้คะแนน โดยมองตามภาพรวม และ แบ่งกระดาษคำตอบออกเป็น 6 กอง คือ กอง 5 คะแนน กอง 4 คะแนน กอง 3 คะแนน กอง 2 คะแนน กอง 1 คะแนน และ กอง 0 คะแนน

4.4 กระดาษคำตอบแผ่นใด ไม่แน่ใจว่าจะอยู่กลุ่มใด ให้คุณครูทำเครื่องหมายไว้และวางไว้กองที่น่าจะอยู่ในกองนั้น ๆ หลังจากนั้นกลับมาพิจารณาอีกครั้ง

4.5 ดำเนินการตรวจทีละข้อ จนครบ 60 คน จึงตรวจข้อต่อไปจนครบ 5 ข้อ

1. ให้นักเรียนอธิบายถึงผลดีของการส่งเสริมให้ชาวบ้านในชนบทประกอบอาชีพอุตสาหกรรมในครัวเรือนเพิ่มขึ้น และวิเคราะห์ถึงปัญหาที่จะเกิดขึ้นมา 2 ปัญหา

แนวคำตอบ

สิ่งที่นักเรียนควรกล่าวถึงผลดีของการส่งเสริมดังกล่าว มีตัวอย่างดังนี้

1. ช่วยให้คนในชนบทมีงานทำ
1. ลดอัตราการย้ายถิ่นเข้าเมือง
2. นำวัตถุดิบที่มีในท้องถิ่นมาแปรรูปทำให้ขายได้ราคาดี
3. ลดปริมาณการสูญเสียของวัตถุดิบที่มีในท้องถิ่น
4. สร้างความมั่นคงให้แก่ชุมชน

ฯลฯ

นักเรียนควรกล่าวถึงปัญหาที่เกิดขึ้น อาทิเช่น

6. ระยะเวลาในการผลิตใช้เวลานาน
7. ไม่มีหน่วยงานหาตลาดรองรับสินค้าที่ผลิต
8. ไม่มีเงินลงทุน หรือขยายกิจการ
9. ขาดบุคลากรที่เชี่ยวชาญทางด้านเทคนิคการผลิต
10. ปัญหาพ่อค้าคนกลาง รับซื้อสินค้าที่ผลิตได้ในราคาถูก

ฯลฯ

เกณฑ์การตรวจให้คะแนน

ให้ 5 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามเนื้อหา บอกผลดีและมีปัญหาเพิ่มเติมนอกเหนือจากแนวคำตอบ
นำเสนอหลักการได้ทั้ง 2 หลักการ คือ ผลดีของการส่งเสริม และ ปัญหาที่จะเกิดขึ้น
จัดเรียบเรียงความคิดได้ชัดเจน เข้าใจง่าย

ให้ 4 คะแนน เมื่อ ตอบคำถามถูกต้องตามแนวคำตอบ
นำเสนอหลักการได้ครบทั้ง 2 หลักการ คือ ผลดีของการส่งเสริม และปัญหาที่จะเกิดขึ้น
จัดเรียบเรียงแนวคิดได้ชัดเจน

ให้ 3 คะแนน เมื่อ ตอบคำถามได้ถูกต้องมากกว่าครึ่งหนึ่งของแนวคำตอบ
นำเสนอหลักการได้โดยรวมไม่แยกออกเป็น 2 หลักการ
จัดเรียบเรียงความคิดได้แต่ไม่ชัดเจนทั้งหมด

ให้ 2 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามแนวคำตอบน้อยกว่าครึ่งหนึ่ง
นำเสนอหลักการได้เพียงอย่างใดอย่างหนึ่ง
แสดงการจัดเรียบเรียงความคิดออกมาอย่างกำกวม

ให้ 1 คะแนน เมื่อ ตอบคำถามไม่ถูกต้องตามแนวคำตอบแต่ยังถูกต้องตามเนื้อหา
ไม่แยกแยะตามหลักการที่คำถามระบุไว้
ไม่สามารถเรียบเรียงความคิดได้

ให้ 0 คะแนน เมื่อ ตอบคำถามใด ๆ ไม่ถูกต้องตามเนื้อหา หรือไม่ตอบเลย

2. ให้นักเรียนจัดกลุ่มประเทศเพื่อนบ้านต่อไปนี้ตามลักษณะการปกครองที่คล้ายคลึงกัน และบอกลักษณะการปกครองของแต่ละกลุ่มมาให้ถูกต้อง

มาเลเซีย สาธารณรัฐสิงคโปร์ สาธารณรัฐสังคมนิยมเมียนม่า(พม่า)
สาธารณรัฐฟิลิปปินส์ สาธารณรัฐอินโดนีเซีย กัมพูชาประชาธิปไตย
สาธารณรัฐสังคมนิยมเวียดนาม ไทย สาธารณรัฐประชาธิปไตยประชาชนลาว
เนรากาบรูไนดารุสซาลาม

แนวคำตอบ

นักเรียนต้องจำแนกลักษณะการปกครองของประเทศต่าง ๆ เหล่านี้ออกเป็นกลุ่ม ดังนี้

กลุ่ม 1 มาเลเซีย ไทย กัมพูชาประชาธิปไตย
มีลักษณะการปกครองแบบประชาธิปไตย มีพระมหากษัตริย์เป็นประมุข

กลุ่ม 2 สาธารณรัฐสิงคโปร์ สาธารณรัฐอินโดนีเซีย สาธารณรัฐฟิลิปปินส์
มีลักษณะการปกครองแบบสาธารณรัฐ มีประธานาธิบดีเป็นประมุข

กลุ่ม 3 สาธารณรัฐสังคมนิยมเมียนม่า(พม่า) สาธารณรัฐประชาธิปไตยประชาชนลาว
สาธารณรัฐสังคมนิยมเวียดนาม
มีลักษณะการปกครองแบบสังคมนิยม มีประธานาธิบดีเป็นประมุข

กลุ่ม 4 เนรากาบรูไนดารุสซาลาม
มีลักษณะการปกครองแบบสมบูรณาญาสิทธิราชย์ มีสุลต่านเป็นประมุข

เกณฑ์การตรวจให้คะแนน

ให้ 5 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามเนื้อหา ถูกต้องตามแนวคำตอบอย่างสมบูรณ์
จำแนกประเทศที่กำหนดให้เป็นกลุ่ม ๆ ตามลักษณะการปกครองได้ถูกต้อง
นำเสนอหลักการที่สำคัญ ๆ ได้ คือ จำแนกกลุ่มประเทศได้ถูกต้อง บอก
ชื่อประเทศที่อยู่ในกลุ่มได้ และบอกลักษณะการปกครองได้
จัดเรียบเรียงความคิดได้ชัดเจน เข้าใจง่าย

ให้ 4 คะแนน เมื่อ ตอบคำถามถูกต้องตามแนวคำตอบ แต่ผิดไม่เกิน 1 ประเทศ ใน 10 ประเทศ
นำเสนอหลักการได้ครบตามคำถาม แต่มีบางประเด็นไม่สมบูรณ์
จัดเรียบเรียงแนวคิดได้ชัดเจน

ให้ 3 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามเนื้อหาคล้ายคลึงกับแนวคำตอบ มีผิดไม่เกิน 2 ใน 10 ประเทศ
นำเสนอหลักการได้ แต่ขาดประเด็นใดประเด็นหนึ่ง
จัดเรียบเรียงความคิดได้แต่ไม่ชัดเจนทั้งหมด

ให้ 2 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามเนื้อหาแต่มีผิดมากกว่า 3 ประเทศใน 10 ประเทศ
นำเสนอหลักการได้ประเด็นใดประเด็นหนึ่ง
แสดงการจัดเรียบเรียงความคิดออกมาอย่างกำกวม

ให้ 1 คะแนน เมื่อ ตอบคำถามไม่ถูกต้องตามแนวคำตอบแต่ยังถูกต้องตามเนื้อหา
ไม่แยกแยะตามหลักการที่คำถามระบุไว้
ไม่สามารถเรียบเรียงความคิดได้

ให้ 0 คะแนน เมื่อ ตอบคำถามใด ๆ ไม่ถูกต้องตามเนื้อหา หรือไม่ตอบเลย

3. ให้นักเรียนอ่านข่าวต่อไปนี้ แล้วอภิปรายถึงมาตรการที่เกิดขึ้นตามข่าวนี้ จะก่อให้เกิดผลเสียต่อประเทศเพื่อนบ้านดังที่ข่าวกล่าวถึงอย่างไร และ เสนอวิธีการแก้ปัญหานี้มา 2 วิธี

กัวลาลัมเปอร์ : ตำรวจมาเลเซียจะตั้งด่านสกัดจับผู้ขับขี่ชาวต่างชาติซึ่งเคยถูกเจ้าหน้าที่ให้ใบสั่งฐานฝ่าฝืนการจราจรของมาเลเซียมานานแล้ว แต่ไม่ยอมมาเสียค่าปรับซะที ทั้งนี้ ตำรวจมาเลเซียระบุจะเอาจริงกับ ผู้ฝ่าฝืน ซึ่งมีมากกว่า 93,000 คน ซึ่งในจำนวนนี้มีเพื่อนบ้านอย่างบรูไนสิงคโปร์และไทย รวมอยู่ในบัญชีรายชื่อด้วย ข่าวแจ้งว่า ชาวสิงคโปร์ถูกขึ้นบัญชีดำมากที่สุดถึง 85,194 คน ขณะที่ไทยมีเพียง 1,086 คน เท่านั้น

แนวคำตอบ

การอภิปรายของนักเรียน อาจกล่าวในลักษณะต่อไปนี้

1. กล่าวถึงผลกระทบต่อความสัมพันธ์ระหว่างประเทศ
2. กล่าวถึงการลดจำนวนนักท่องเที่ยวเข้าประเทศในประเทศมาเลเซีย
3. กล่าวถึงขั้นตอนการขอวีซ่าเข้าประเทศเข้มงวด ส่งผลต่อการค้าขายระหว่างชายแดนในแต่ละประเทศยากขึ้น

นักเรียนอาจเสนอแนวทางแก้ปัญหา อาทิเช่น

1. ผู้เข้าประเทศต้องศึกษาหาความรู้เกี่ยวกับกฎหมายของประเทศนั้น ๆ ก่อนเดินทาง
2. สถานฑูตต้องชี้แจงข้อบังคับหรือกฎหมายที่เข้มงวดแก่นักท่องเที่ยว
3. งดการเดินทางเข้าประเทศนั้น ๆ

เกณฑ์การตรวจให้คะแนน

ให้	5 คะแนน เมื่อ	อภิปรายและเสนอวิธีการแก้ปัญหาได้ตามลักษณะที่กำหนดให้ มีประเด็นเพิ่มเติมนอกเหนือจากแนวคำตอบ นำเสนอหลักการที่สำคัญ ๆ ได้ คือ อภิปรายผลกระทบที่เกิดจากข่าวนี้ และ เสนอวิธีการแก้ปัญหาได้ จัดเรียบเรียงความคิดในการอภิปรายและเสนอวิธีการแก้ปัญหาได้ชัดเจนเข้าใจง่าย
ให้	4 คะแนน เมื่อ	อภิปรายและเสนอวิธีการแก้ปัญหาตามความคิดของตนเองนอกเหนือจากแนวคำตอบทั้งหมด นำเสนอหลักการได้ครบตามคำถาม แต่มีบางประเด็นไม่สมบูรณ์ จัดเรียบเรียงแนวคิดได้ชัดเจน
ให้	3 คะแนน เมื่อ	อภิปรายและเสนอวิธีการแก้ปัญหาได้ตามลักษณะที่กำหนดให้เท่านั้น นำเสนอหลักการได้ แต่ขาดประเด็นใดประเด็นหนึ่ง จัดเรียบเรียงความคิดได้แต่ไม่ชัดเจนทั้งหมด
ให้	2 คะแนน เมื่อ	อภิปรายถึงผลกระทบและเสนอวิธีการแก้ปัญหาได้ไม่สมบูรณ์ นำเสนอหลักการได้ประเด็นใดประเด็นหนึ่ง จัดเรียบเรียงความคิดออกมาอย่างกำกวม
ให้	1 คะแนน เมื่อ	ไม่สามารถอภิปรายถึงผลกระทบได้แต่เสนอวิธีการแก้ปัญหาได้เล็กน้อย ไม่สามารถแยกแยะหลักการตามที่คำถามระบุไว้ได้ ไม่สามารถเรียบเรียงความคิดได้
ให้	0 คะแนน เมื่อ	ไม่สามารถอภิปรายได้ตามข้อมูลที่กำหนดให้ หรือไม่ตอบคำถามใด ๆ เลย

4. จงดูภาพต่อไปนี้ แล้วจงเขียนข่าวสั้นจากภาพมาให้ถูกต้องตามวิธีการเขียนข่าว



แนวคำตอบ

นักเรียนจะต้องเขียนข่าวสั้นให้สอดคล้องกับภาพที่กำหนดให้ ตัวอย่างเช่น

หัวข้อข่าว "พลเมืองดีของชาติ" / "สมรักษ์เลือกตั้ง" / "ไม่นอนหลับทับสิทธิ์"

วัน เหตุการณ์ที่เกิดขึ้น สถานที่ที่ไป หรือ คำให้สัมภาษณ์)

เกณฑ์การตรวจให้คะแนน

ให้ 5 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามเนื้อหา และวิธีการเขียนข่าวสั้นจากภาพ
นำเสนอหลักการที่สำคัญได้ คือ หัวข้อข่าว และรายละเอียดของข่าวได้
สอดคล้องกับภาพ
จัดเรียบเรียงความคิดในการเขียนข่าวได้ชัดเจน

ให้ 4 คะแนน เมื่อ ตอบคำถามได้ถูกต้องตามเนื้อหา และวิธีการเขียนข่าวสั้นจากภาพ แต่ขาดรายละเอียดบางประการ
นำเสนอหลักการได้ครบตามคำถาม แต่มีบางประเด็นไม่สมบูรณ์
จัดเรียบเรียงแนวคิดในการเขียนได้

ให้ 3 คะแนน เมื่อ เขียนข่าวสั้นได้ตามวิธีการเขียนข่าวสั้นจากภาพ แต่สอดคล้องกับภาพไม่ค่อยสมบูรณ์
นำเสนอหลักการได้ แต่ขาดประเด็นใดประเด็นหนึ่ง

จัดเรียบเรียงความคิดได้แต่ไม่ชัดเจนทั้งหมด

ให้ 2 คะแนน เมื่อ เขียนข่าวสั้นได้แต่ไม่ชัดเจนตามภาพที่กำหนดให้
นำเสนอหลักการได้ประเด็นใดประเด็นหนึ่ง
จัดเรียบเรียงความคิดออกมาอย่างกำกวม

ให้ 1 คะแนน เมื่อ เขียนข่าวสั้นได้แต่ไม่ถูกต้องตามวิธีการ
ไม่สามารถแยกแยะหลักการตามที่คำถามระบุไว้ได้
ไม่สามารถเรียบเรียงความคิดได้

ให้ 0 คะแนน เมื่อ ไม่สามารถเขียนข่าวได้จากภาพที่กำหนดให้ หรือไม่เขียนข้อความใด ๆ เลย

5. ให้นักเรียนศึกษาข่าว แล้วสรุปข่าวตามประเด็นสำคัญต่อไปนี้ หัวข้อข่าว ประเภทของข่าว แหล่งข่าว และใจความสำคัญของข่าว ให้ครบถ้วน

ยันผงชูรสไม่อันตราย : นายแพทย์วิชัย ขัติวิทยากุล สาธารณสุข จ.ขอนแก่น เปิดเผยว่าจากการออกข่าวเผยแพร่ของ อย. ได้ชี้แจงให้ผู้บริโภคให้คลายความวิตกกังวลถึงอันตรายของผงชูรสอย่างเป็นทางการแล้ว ระบุว่า มีความเป็นพิษต่ำมากและไม่ก่อให้เกิดความเสี่ยงอันตรายต่อผู้บริโภค ยกเว้นในรายที่แพ้ผงชูรส แต่อาการมักไม่รุนแรง มักหายเอง ผงชูรสมีชื่อทางเคมีว่า โมโนโซเดียมกลูตาเมท เป็นอาหารเฉพาะ ซึ่งผู้ผลิตและผู้นำเข้าต้องได้รับใบอนุญาตผลิต และนำเข้า โดยขอขึ้นทะเบียนตำรับอาหารมีคุณภาพมาตรฐานตามที่กฎหมายกำหนดไว้ เพื่อเป็นหลักประกันความปลอดภัยให้กับผู้บริโภค และข้อสำคัญในการเลือกซื้อผงชูรสที่ได้มาตรฐาน

ที่มา นสพ.เดลินิวส์ วันที่ 6 มกราคม 2544

แนวคำตอบ

นักเรียนต้องตอบคำถามนี้ ตามแนวคำตอบอย่างสมบูรณ์ จึงจะถูกต้องตามเนื้อหา ดังนี้

หัวข้อข่าว ยันผงชูรสไม่อันตราย

ประเภทของข่าว ข่าวสุขภาพ / ข่าวทั่วไป

แหล่งข่าว จากหนังสือพิมพ์

ใจความสำคัญของข่าว "นายแพทย์วิชัย ขัติวิทยากุล สาธารณสุข จ.ขอนแก่น เปิดเผยว่าจากการออกข่าวเผยแพร่ของ อย. ได้ชี้แจงให้ผู้บริโภคให้คลายความวิตกกังวลถึงอันตรายของผงชูรสอย่างเป็นทางการแล้ว ระบุว่า มีความเป็นพิษต่ำมากและไม่ก่อให้เกิดความเสี่ยงอันตรายต่อผู้บริโภค ยกเว้นในรายที่แพ้ผงชูรส แต่อาการมักไม่รุนแรง"

เกณฑ์การตรวจให้คะแนน

ให้	5 คะแนน เมื่อ	สรุปข่าวจากข่าวที่กำหนดให้ได้ ตามแนวคำตอบอย่างครบถ้วน นำเสนอหลักการที่สำคัญได้ คือ หัวข้อข่าว ประเภทของข่าว แหล่งข่าว และใจความสำคัญของข่าว ได้ครบถ้วน จัดเรียบเรียงความคิดได้ตามวิธีการสรุปข่าวได้ชัดเจน
ให้	4 คะแนน เมื่อ	สรุปข่าวจากข่าวที่กำหนดให้ได้ ตามแนวคำตอบ แต่ยังไม่สมบูรณ์ ขาดเนื้อหาใดเนื้อหาหนึ่ง นำเสนอหลักการได้ครบตามคำถาม แต่มีบางหัวข้อไม่สมบูรณ์ จัดเรียบเรียงแนวคิดในการสรุปข่าวได้
ให้	3 คะแนน เมื่อ	สรุปข่าวจากข่าวที่กำหนดให้ได้ ตามแนวคำตอบ แต่ขาดความสมบูรณ์ของเนื้อหาไปมากกว่าครึ่ง นำเสนอหลักการได้ แต่ขาดบางหัวข้อไปแต่ไม่เกิน 2 หัวข้อ จัดเรียบเรียงความคิดได้แต่ไม่ชัดเจนทั้งหมด
ให้	2 คะแนน เมื่อ	สรุปข่าวจากข่าวที่กำหนดให้ได้ แต่ไม่ครบถ้วนตามเนื้อหา นำเสนอหลักการได้ น้อยกว่า 2 หัวข้อ จัดเรียบเรียงความคิดออกมาอย่างกำกวม
ให้	1 คะแนน เมื่อ	สรุปข่าวได้ แต่ไม่สอดคล้องกับข่าวที่กำหนดให้ ไม่สามารถแยกแยะหลักการตามที่คำถามระบุไว้ได้ ไม่สามารถเรียบเรียงความคิดได้
ให้	0 คะแนน เมื่อ	ไม่สามารถสรุปข่าวได้จากข่าวที่กำหนดให้ หรือไม่เขียนข้อความใด ๆ เลย

ภาคผนวก ข

1. รายนามผู้ตรวจแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต(สังคม) ชั้นประถมศึกษาปีที่ 5
2. เปรียบเทียบค่าสถิติพื้นฐาน จากการตรวจวิธีวิเคราะห์ย่อย และ วิธีประเมินรวม จำแนกเป็นรายบุคคล
3. ผลการตรวจแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต เมื่อใช้วิธีการตรวจแบบวิเคราะห์ย่อย และค่าสถิติพื้นฐาน
4. ผลการตรวจแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต เมื่อใช้วิธีการตรวจแบบประเมินรวม และค่าสถิติพื้นฐาน
5. ค่าความเชื่อมั่นของผู้ตรวจ (RAI) แบบทดสอบความเรียง เมื่อใช้วิธีการตรวจแบบวิเคราะห์ย่อย
6. ค่าความเชื่อมั่นของผู้ตรวจ (RAI) แบบทดสอบความเรียง เมื่อใช้วิธีการตรวจแบบประเมินรวม

รายนามผู้ตรวจแบบทดสอบความเรียง

ชื่อ นามสกุล นางปราณี วงศ์สมบุญ .....(R1)
ตำแหน่ง ผู้ช่วยครูใหญ่ฝ่ายปกครอง และ ครูผู้สอนวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ระดับชั้นประถมศึกษาปีที่ 5
สถานที่ทำงาน โรงเรียนเปรมฤดีศึกษา
วุฒิทางการศึกษา ศษ.บ. เอกสังคม มัธยมศึกษา มหาวิทยาลัยสุโขทัยธรรมมาธิราช
ประสบการณ์สอน 22 ปี

ชื่อ นามสกุล นางสุกัญญา วิเชียรดิลก .....(R2)
ตำแหน่ง ครูผู้สอนวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ระดับชั้นประถมศึกษาปีที่ 5
สถานที่ทำงาน โรงเรียนอัสสัมชัญแผนกประถม
วุฒิทางการศึกษา คบ. เอกสังคมศึกษา สถาบันราชภัฏบ้านสมเด็จเจ้าพระยา
ประสบการณ์สอน 8 ปี

ชื่อ นามสกุล นางสาวเพชรรัตน์ ชมจันทร์ .....(R3)
ตำแหน่ง ครูผู้สอนวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ระดับชั้นประถมศึกษาปีที่ 5
สถานที่ทำงาน โรงเรียนอัสสัมชัญแผนกประถม
วุฒิทางการศึกษา คบ. เอกสังคมศึกษา สถาบันราชภัฏธนบุรี
ประสบการณ์สอน 19 ปี

ชื่อ นามสกุล นายสุชิน บุญพิทักษ์กิจ .....(R4)
ตำแหน่ง หัวหน้าหมวดวิชาสังคมศึกษา และ ครูผู้สอนวิชาสร้างเสริมประสบการณ์ชีวิต (สังคม) ระดับชั้นประถมศึกษาปีที่ 6
สถานที่ทำงาน โรงเรียนเปรมฤดีศึกษา
วุฒิทางการศึกษา ศษ.บ. เอกประถมศึกษา มหาวิทยาลัยสุโขทัยธรรมมาธิราช
ประสบการณ์สอน 15 ปี

ตาราง 13 เปรียบเทียบค่าสถิติพื้นฐาน จากการตรวจวิธีวิเคราะห์ย่อย และ วิธีประเมินรวม จำแนกเป็นรายบุคคล

ผู้ตรวจ	วิธีวิเคราะห์ย่อย			วิธีประเมินรวม		
	$\bar{X}$	*S.D.*	*C.V.*	$\bar{X}$	*S.D.*	*C.V.*
คนที่ 1	14.583	3.341	22.910	13.600	2.644	19.566
คนที่ 2	12.567	2.560	20.373	11.767	2.302	19.566
คนที่ 3	11.950	3.721	31.137	12.467	2.977	23.881
คนที่ 4	12.983	2.703	20.817	12.817	2.771	21.619
เฉลี่ยรวม	**13.021**	**3.266**	**25.083**	**12.663**	**2.765**	**21.835**

ตาราง 14 ผลการตรวจแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต เมื่อใช้การตรวจวิธีวิเคราะห์ย่อย และ ค่าสถิติพื้นฐาน

คนที่	เต็ม	R1	R2	R3	R4	$\overline{X}_{R12}$	$\overline{X}_{R13}$	$\overline{X}_{R14}$	$\overline{X}_{R23}$	$\overline{X}_{R24}$	$\overline{X}_{R34}$	$\overline{X}_{R123}$	$\overline{X}_{R124}$	$\overline{X}_{R134}$	$\overline{X}_{R234}$	$\overline{X}_{R1234}$
1	25	18	15	21	18	16.50	19.50	18.00	18.00	16.50	19.50	18.00	17.00	19.00	18.00	18.50
2	25	19	14	22	17	16.50	20.50	18.00	18.00	15.50	19.50	18.33	16.67	19.33	18.00	18.50
3	25	12	10	11	10	11.00	11.50	11.00	10.50	10.00	10.50	11.00	10.67	11.00	10.75	10.67
4	25	18	15	20	19	16.50	19.00	18.50	17.50	17.00	19.50	17.67	17.33	19.00	18.00	18.50
5	25	13	8	10	12	10.50	11.50	12.50	9.00	10.00	11.00	10.33	11.00	11.67	10.75	10.83
6	25	19	13	17	16	16.00	18.00	17.50	15.00	14.50	16.50	16.33	16.00	17.33	16.25	16.33
7	25	16	15	18	15	15.50	17.00	15.50	16.50	15.00	16.50	16.33	15.33	16.33	16.00	16.17
8	25	16	15	12	14	15.50	14.00	15.00	13.50	14.50	13.00	14.33	15.00	14.00	14.25	13.83
9	25	17	15	16	15	16.00	16.50	16.00	15.50	15.00	15.50	16.00	15.67	16.00	15.75	15.67
10	25	12	11	14	13	11.50	13.00	12.50	12.50	12.00	13.50	12.33	12.00	13.00	12.50	12.83
11	25	15	11	12	12	13.00	13.50	13.50	11.50	11.50	12.00	12.67	12.67	13.00	12.50	12.33
12	25	14	13	14	15	13.50	14.00	14.50	13.50	14.00	14.50	13.67	14.00	14.33	14.00	14.17
13	25	17	17	13	17	17.00	15.00	17.00	15.00	17.00	15.00	15.67	17.00	15.67	16.00	15.67
14	25	15	13	11	13	14.00	13.00	14.00	12.00	13.00	12.00	13.00	13.67	13.00	13.00	12.67
15	25	19	12	11	12	15.50	15.00	15.50	11.50	12.00	11.50	14.00	14.33	14.00	13.50	12.83
16	25	18	14	11	15	16.00	14.50	16.50	12.50	14.50	13.00	14.33	15.67	14.67	14.50	14.00
17	25	16	10	9	12	13.00	12.50	14.00	9.50	11.00	10.50	11.67	12.67	12.33	11.75	11.33
18	25	12	12	10	13	12.00	11.00	12.50	11.00	12.50	11.50	11.33	12.33	11.67	11.75	11.67
19	25	17	18	14	16	17.50	15.50	16.50	16.00	17.00	15.00	16.33	17.00	15.67	16.25	15.83
20	25	23	17	19	18	20.00	21.00	20.50	18.00	17.50	18.50	19.67	19.33	20.00	19.25	19.00
21	25	13	10	17	14	11.50	15.00	13.50	13.50	12.00	15.50	13.33	12.33	14.67	13.50	14.17
22	25	13	16	10	15	14.50	11.50	14.00	13.00	15.50	12.50	13.00	14.67	12.67	13.50	13.17
23	25	15	13	10	13	14.00	12.50	14.00	11.50	13.00	11.50	12.67	13.67	12.67	12.75	12.33
24	25	14	14	13	12	14.00	13.50	13.00	13.50	13.00	12.50	13.67	13.33	13.00	13.25	13.00
25	25	15	15	13	12	15.00	14.00	13.50	14.00	13.50	12.50	14.33	14.00	13.33	13.75	13.33
26	25	10	10	11	11	10.00	10.50	10.50	10.50	10.50	11.00	10.33	10.33	10.67	10.50	10.67
27	25	9	9	12	11	9.00	10.50	10.00	10.50	10.00	11.50	10.00	9.67	10.67	10.25	10.67
28	25	11	11	9	10	11.00	10.00	10.50	10.00	10.50	9.50	10.33	10.67	10.00	10.25	10.00
29	25	12	13	8	11	12.50	10.00	11.50	10.50	12.00	9.50	11.00	12.00	10.33	11.00	10.50
30	25	16	8	13	9	12.00	14.50	12.50	10.50	8.50	11.00	12.33	11.00	12.67	11.50	11.33
31	25	22	15	15	17	18.50	18.50	19.50	15.00	16.00	16.00	17.33	18.00	18.00	17.25	16.83

ตาราง 14 (ต่อ)

คนที่	เต็ม	R1	R2	R3	R4	$\overline{X}_{R12}$	$\overline{X}_{R13}$	$\overline{X}_{R14}$	$\overline{X}_{R23}$	$\overline{X}_{R24}$	$\overline{X}_{R34}$	$\overline{X}_{R123}$	$\overline{X}_{R124}$	$\overline{X}_{R134}$	$\overline{X}_{R234}$	$\overline{X}_{R1234}$
32	25	16	13	7	10	14.50	11.50	13.00	10.00	11.50	8.50	12.00	13.00	11.00	11.50	10.50
33	25	19	16	14	17	17.50	16.50	18.00	15.00	16.50	15.50	16.33	17.33	16.67	16.50	16.17
34	25	16	10	12	14	13.00	14.00	15.00	11.00	12.00	13.00	12.67	13.33	14.00	13.00	13.00
35	25	12	11	11	13	11.50	11.50	12.50	11.00	12.00	12.00	11.33	12.00	12.00	11.75	11.83
36	25	13	11	8	12	12.00	10.50	12.50	9.50	11.50	10.00	10.67	12.00	11.00	11.00	10.67
37	25	13	13	13	14	13.00	13.00	13.50	13.00	13.50	13.50	13.00	13.33	13.33	13.25	13.33
38	25	19	16	16	17	17.50	17.50	18.00	16.00	16.50	16.50	17.00	17.33	17.33	17.00	16.83
39	25	14	14	9	11	14.00	11.50	12.50	11.50	12.50	10.00	12.33	13.00	11.33	12.00	11.33
40	25	19	14	17	16	16.50	18.00	17.50	15.50	15.00	16.50	16.67	16.33	17.33	16.50	16.50
41	25	12	8	8	8	10.00	10.00	10.00	8.00	8.00	8.00	9.33	9.33	9.33	9.00	8.67
42	25	13	14	10	11	13.50	11.50	12.00	12.00	12.50	10.50	12.33	12.67	11.33	12.00	11.50
43	25	16	14	14	17	15.00	15.00	16.50	14.00	15.50	15.50	14.67	15.67	15.67	15.25	15.33
44	25	14	12	9	12	13.00	11.50	13.00	10.50	12.00	10.50	11.67	12.67	11.67	11.75	11.33
45	25	8	11	9	11	9.50	8.50	9.50	10.00	11.00	10.00	9.33	10.00	9.33	9.75	9.83
46	25	18	15	14	13	16.50	16.00	15.50	14.50	14.00	13.50	15.67	15.33	15.00	15.00	14.50
47	25	14	15	11	14	14.50	12.50	14.00	13.00	14.50	12.50	13.33	14.33	13.00	13.50	13.17
48	25	13	12	11	13	12.50	12.00	13.00	11.50	12.50	12.00	12.00	12.67	12.33	12.25	12.17
49	25	12	13	13	12	12.50	12.50	12.00	13.00	12.50	12.50	12.67	12.33	12.33	12.50	12.50
50	25	11	9	7	10	10.00	9.00	10.50	8.00	9.50	8.50	9.00	10.00	9.33	9.25	9.00
51	25	10	9	11	11	9.50	10.50	10.50	10.00	10.00	11.00	10.00	10.00	10.67	10.25	10.50
52	25	12	9	11	9	10.50	11.50	10.50	10.00	9.00	10.00	10.67	10.00	10.67	10.25	10.17
53	25	15	12	10	12	13.50	12.50	13.50	11.00	12.00	11.00	12.33	13.00	12.33	12.25	11.83
54	25	7	11	6	11	9.00	6.50	9.00	8.50	11.00	8.50	8.00	9.67	8.00	8.75	8.67
55	25	12	14	8	11	13.00	10.00	11.50	11.00	12.50	9.50	11.33	12.33	10.33	11.25	10.67
56	25	16	14	11	12	15.00	13.50	14.00	12.50	13.00	11.50	13.67	14.00	13.00	13.25	12.67
57	25	12	9	6	7	10.50	9.00	9.50	7.50	8.00	6.50	9.00	9.33	8.33	8.50	7.83
58	25	12	9	6	11	10.50	9.00	11.50	7.50	10.00	8.50	9.00	10.67	9.67	9.50	9.17
59	25	11	9	6	9	10.00	8.50	10.00	7.50	9.00	7.50	8.67	9.67	8.67	8.75	8.33
60	25	20	15	14	14	17.50	17.00	17.00	14.50	14.50	14.00	16.33	16.33	16.00	15.75	15.17
$\bar{X}$						13.58	13.28	13.78	12.27	12.78	12.48	13.04	13.38	13.18	13.03	12.84
S.D.						2.68	3.21	2.79	2.75	2.46	3.05	2.76	2.56	2.93	2.68	2.77
C.V.						19.72	24.22	20.25	22.44	19.22	24.43	21.20	19.13	22.25	20.58	21.55

ตาราง 15 ผลการตรวจแบบทดสอบความเรียง วิชาสร้างเสริมประสบการณ์ชีวิต เมื่อใช้การตรวจวิธีประเมินรวม และ ค่าสถิติพื้นฐาน

นร.	เต็ม	R1	R2	R3	R4	$\bar{X}_{R12}$	$\bar{X}_{R13}$	$\bar{X}_{R14}$	$\bar{X}_{R23}$	$\bar{X}_{R24}$	$\bar{X}_{R34}$	$\bar{X}_{R123}$	$\bar{X}_{R124}$	$\bar{X}_{R134}$	$\bar{X}_{R234}$	$\bar{X}_{R1234}$
1	25	17	14	15	18	15.50	16.00	17.50	14.50	16.00	16.50	15.33	16.33	16.67	16.00	16.17
2	25	18	16	19	17	17.00	18.50	17.50	17.50	16.50	18.00	17.67	17.00	18.00	17.50	17.67
3	25	12	9	9	9	10.50	10.50	10.50	9.00	9.00	9.00	10.00	10.00	10.00	9.75	9.50
4	25	18	15	21	19	16.50	19.50	18.50	18.00	17.00	20.00	18.00	17.33	19.33	18.25	18.83
5	25	11	9	12	12	10.00	11.50	11.50	10.50	10.50	12.00	10.67	10.67	11.67	11.00	11.33
6	25	15	14	17	16	14.50	16.00	15.50	15.50	15.00	16.50	15.33	15.00	16.00	15.50	15.83
7	25	15	14	15	15	14.50	15.00	15.00	14.50	14.50	15.00	14.67	14.67	15.00	14.75	14.83
8	25	14	12	14	14	13.00	14.00	14.00	13.00	13.00	14.00	13.33	13.33	14.00	13.50	13.67
9	25	19	13	14	15	16.00	16.50	17.00	13.50	14.00	14.50	15.33	15.67	16.00	15.25	15.00
10	25	11	12	14	13	11.50	12.50	12.00	13.00	12.50	13.50	12.33	12.00	12.67	12.50	12.83
11	25	11	11	12	12	11.00	11.50	11.50	11.50	11.50	12.00	11.33	11.33	11.67	11.50	11.67
12	25	14	14	12	15	14.00	13.00	14.50	13.00	14.50	13.50	13.33	14.33	13.67	13.75	13.67
13	25	17	13	16	17	15.00	16.50	17.00	14.50	15.00	16.50	15.33	15.67	16.67	15.75	16.00
14	25	13	12	9	12	12.50	11.00	12.50	10.50	12.00	10.50	11.33	12.33	11.33	11.50	11.17
15	25	16	11	11	13	13.50	13.50	14.50	11.00	12.00	12.00	12.67	13.33	13.33	12.75	12.50
16	25	16	13	11	15	14.50	13.50	15.50	12.00	14.00	13.00	13.33	14.67	14.00	13.75	13.50
17	25	11	11	10	12	11.00	10.50	11.50	10.50	11.50	11.00	10.67	11.33	11.00	11.00	11.00
18	25	11	12	11	14	11.50	11.00	12.50	11.50	13.00	12.50	11.33	12.33	12.00	12.00	12.17
19	25	14	15	15	15	14.50	14.50	14.50	15.00	15.00	15.00	14.67	14.67	14.67	14.75	14.83
20	25	20	16	16	18	18.00	18.00	19.00	16.00	17.00	17.00	17.33	18.00	18.00	17.50	17.33
21	25	12	10	14	13	11.00	13.00	12.50	12.00	11.50	13.50	12.00	11.67	13.00	12.25	12.67
22	25	15	15	13	16	15.00	14.00	15.50	14.00	15.50	14.50	14.33	15.33	14.67	14.75	14.67
23	25	14	14	16	13	14.00	15.00	13.50	15.00	13.50	14.50	14.67	13.67	14.33	14.25	14.33
24	25	11	12	12	12	11.50	11.50	11.50	12.00	12.00	12.00	11.67	11.67	11.67	11.75	11.83
25	25	16	12	11	12	14.00	13.50	14.00	11.50	12.00	11.50	13.00	13.33	13.00	12.75	12.33
26	25	13	10	11	11	11.50	12.00	12.00	10.50	10.50	11.00	11.33	11.33	11.67	11.25	11.17
27	25	9	9	10	11	9.00	9.50	10.00	9.50	10.00	10.50	9.33	9.67	10.00	9.75	10.00
28	25	14	10	10	12	12.00	12.00	13.00	10.00	11.00	11.00	11.33	12.00	12.00	11.50	11.33
29	25	13	10	9	12	11.50	11.00	12.50	9.50	11.00	10.50	10.67	11.67	11.33	11.00	10.83
30	25	13	10	14	9	11.50	13.50	11.00	12.00	9.50	11.50	12.33	10.67	12.00	11.50	11.50
31	25	18	15	15	17	16.50	16.50	17.50	15.00	16.00	16.00	16.00	16.67	16.67	16.25	16.17

ตาราง 15 (ต่อ)

คนที่	เต็ม	R1	R2	R3	R4	$\overline{X}_{R12}$	$\overline{X}_{R13}$	$\overline{X}_{R14}$	$\overline{X}_{R23}$	$\overline{X}_{R24}$	$\overline{X}_{R34}$	$\overline{X}_{R123}$	$\overline{X}_{R124}$	$\overline{X}_{R134}$	$\overline{X}_{R234}$	$\overline{X}_{R1234}$
32	25	14	9	10	10	11.50	12.00	12.00	9.50	9.50	10.00	11.00	11.00	11.33	10.75	10.50
33	25	17	14	14	17	15.50	15.50	17.00	14.00	15.50	15.50	15.00	16.00	16.00	15.50	15.50
34	25	15	9	11	14	12.00	13.00	14.50	10.00	11.50	12.50	11.67	12.67	13.33	12.25	12.33
35	25	14	9	10	13	11.50	12.00	13.50	9.50	11.00	11.50	11.00	12.00	12.33	11.50	11.50
36	25	11	8	11	12	9.50	11.00	11.50	9.50	10.00	11.50	10.00	10.33	11.33	10.50	10.83
37	25	13	9	13	14	11.00	13.00	13.50	11.00	11.50	13.50	11.67	12.00	13.33	12.25	12.67
38	25	16	15	17	17	15.50	16.50	16.50	16.00	16.00	17.00	16.00	16.00	16.67	16.25	16.50
39	25	14	13	13	11	13.50	13.50	12.50	13.00	12.00	12.00	13.33	12.67	12.67	12.75	12.50
40	25	16	13	17	16	14.50	16.50	16.00	15.00	14.50	16.50	15.33	15.00	16.33	15.50	15.83
41	25	11	9	10	8	10.00	10.50	9.50	9.50	8.50	9.00	10.00	9.33	9.67	9.50	9.33
42	25	13	10	10	11	11.50	11.50	12.00	10.00	10.50	10.50	11.00	11.33	11.33	11.00	10.83
43	25	17	14	18	17	15.50	17.50	17.00	16.00	15.50	17.50	16.33	16.00	17.33	16.50	16.83
44	25	11	13	14	12	12.00	12.50	11.50	13.50	12.50	13.00	12.67	12.00	12.33	12.50	12.67
45	25	11	12	10	11	11.50	10.50	11.00	11.00	11.50	10.50	11.00	11.33	10.67	11.00	10.83
46	25	17	14	12	13	15.50	14.50	15.00	13.00	13.50	12.50	14.33	14.67	14.00	14.00	13.50
47	25	14	14	15	14	14.00	14.50	14.00	14.50	14.00	14.50	14.33	14.00	14.33	14.25	14.33
48	25	13	11	13	13	12.00	13.00	13.00	12.00	12.00	13.00	12.33	12.33	13.00	12.50	12.67
49	25	12	11	13	12	11.50	12.50	12.00	12.00	11.50	12.50	12.00	11.67	12.33	12.00	12.17
50	25	12	11	9	9	11.50	10.50	10.50	10.00	10.00	9.00	10.67	10.67	10.00	10.25	9.83
51	25	12	11	10	11	11.50	11.00	11.50	10.50	11.00	10.50	11.00	11.33	11.00	11.00	10.83
52	25	12	9	11	9	10.50	11.50	10.50	10.00	9.00	10.00	10.67	10.00	10.67	10.25	10.17
53	25	12	11	12	12	11.50	12.00	12.00	11.50	11.50	12.00	11.67	11.67	12.00	11.75	11.83
54	25	9	9	11	11	9.00	10.00	10.00	10.00	10.00	11.00	9.67	9.67	10.33	10.00	10.33
55	25	15	11	9	10	13.00	12.00	12.50	10.00	10.50	9.50	11.67	12.00	11.33	11.25	10.67
56	25	11	15	12	12	13.00	11.50	11.50	13.50	13.50	12.00	12.67	12.67	11.67	12.50	12.33
57	25	11	10	7	7	10.50	9.00	9.00	8.50	8.50	7.00	9.33	9.33	8.33	8.75	8.17
58	25	9	7	7	11	8.00	8.00	10.00	7.00	9.00	9.00	7.67	9.00	9.00	8.50	8.67
59	25	9	8	7	9	8.50	8.00	9.00	7.50	8.50	8.00	8.00	8.67	8.33	8.25	8.17
60	25	14	14	14	14	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00	14.00
$\overline{X}$						12.68	13.03	13.29	12.12	12.38	12.73	12.61	12.78	13.02	12.70	12.71
S.D.						2.26	2.54	2.51	2.46	2.35	2.71	2.34	2.31	2.52	2.37	2.46
C.V.						17.85	19.50	18.87	20.32	18.96	21.26	18.58	18.06	19.33	18.65	19.35

ตาราง 16 ค่าความเชื่อมั่นของผู้ตรวจ (RAI) แบบทดสอบความเรียง เมื่อใช้การตรวจวิธีวิเคราะห์ย่อย

จำนวนผู้ตรวจ	กรณีผู้ตรวจคนที่	ข้อ	$\lvert R_{mnk} - \overline{R}_{nk} \rvert$	$\sum \lvert R_{mnk} - \overline{R}_{nk} \rvert$	RAI
สองคน	1 กับ 2	1	58.000	299.750	0.800
		2	44.000		
		3	56.000		
		4	91.750		
		5	50.000		
	1 กับ 3	1	66.000	355.000	0.763
		2	49.000		
		3	90.000		
		4	110.000		
		5	40.000		
	1 กับ 4	1	56.000	339.000	0.774
		2	70.000		
		3	47.000		
		4	96.000		
		5	70.000		
	2 กับ 3	1	66.000	277.000	0.815
		2	37.000		
		3	64.000		
		4	58.000		
		5	52.000		
	2 กับ 4	1	44.000	241.000	0.839
		2	42.000		
		3	39.000		
		4	44.000		
		5	72.000		
	3 กับ 4	1	44.000	312.000	0.792
		2	65.000		
		3	77.000		
		4	58.000		
		5	68.000		

ตาราง 16 (ต่อ)

จำนวนผู้ตรวจ	กรณีผู้ตรวจคนที่	ข้อ	$\left\|R_{mnk}-\overline{R}_{nk}\right\|$	$\sum\left\|R_{mnk}-\overline{R}_{nk}\right\|$	RAI
สามคน	1 , 2 กับ 3	1	108.690	557.990	0.814
		2	81.960		
		3	127.380		
		4	151.330		
		5	88.630		
	1 , 2 กับ 4	1	95.330	523.170	0.826
		2	92.580		
		3	87.320		
		4	135.320		
		5	112.620		
	1 , 3 กับ 4	1	101.300	603.920	0.799
		2	110.600		
		3	129.310		
		4	155.380		
		5	107.330		
	2 , 3 กับ 4	1	93.300	498.400	0.834
		2	86.600		
		3	112.630		
		4	97.220		
		5	108.650		
สี่คน	1 , 2 , 3 กับ 4	1	144.000	776.500	0.827
		2	131.000		
		3	163.000		
		4	191.000		
		5	147.500		

ตาราง 17 ค่าความเชื่อมั่นของผู้ตรวจ (RAI) แบบทดสอบความเรียง เมื่อใช้การตรวจวิธีประเมินรวม

จำนวนผู้ตรวจ	กรณีผู้ตรวจคนที่	ข้อ	$\lvert R_{mnk} - \overline{R}_{nk} \rvert$	$\sum \lvert R_{mnk} - \overline{R}_{nk} \rvert$	RAI
สองคน	1 กับ 2	1	63.000	250.000	0.833
		2	56.000		
		3	47.000		
		4	40.000		
		5	44.000		
	1 กับ 3	1	66.000	272.000	0.819
		2	54.000		
		3	48.000		
		4	54.000		
		5	50.000		
	1 กับ 4	1	51.000	327.000	0.782
		2	103.000		
		3	42.000		
		4	51.000		
		5	80.000		
	2 กับ 3	1	41.000	200.000	0.867
		2	32.000		
		3	37.000		
		4	46.000		
		5	44.000		
	2 กับ 4	1	44.000	243.000	0.838
		2	55.000		
		3	39.000		
		4	41.000		
		5	64.000		
	3 กับ 4	1	43.000	249.000	0.834
		2	59.000		
		3	54.000		
		4	45.000		
		5	48.000		

ตาราง 17 (ต่อ)

จำนวนผู้ตรวจ	กรณีผู้ตรวจคนที่	ข้อ	$\left\|R_{mnk}-\overline{R}_{nk}\right\|$	$\sum\left\|R_{mnk}-\overline{R}_{nk}\right\|$	RAI
สามคน	1 , 2 กับ 3	1	104.610	506.560	0.831
		2	88.620		
		3	80.650		
		4	86.000		
		5	85.300		
	1 , 2 กับ 4	1	94.580	486.990	0.838
		2	122.610		
		3	79.260		
		4	81.240		
		5	109.300		
	1 , 3 กับ 4	1	96.580	510.370	0.8299
		2	127.260		
		3	88.570		
		4	94.020		
		5	103.940		
	2 , 3 กับ 4	1	77.960	424.370	0.858
		2	89.260		
		3	82.610		
		4	82.580		
		5	91.960		
สี่คน	1 , 2 , 3 กับ 4	1	135.000	672.000	0.851
		2	152.000		
		3	121.000		
		4	125.000		
		5	139.000		

ประวัติผู้วิจัย

ประวัติย่อของผู้วิจัย

ชื่อ ชื่อสกุล	นางนิศารัตน์ คงสวัสดิ์
วัน เดือน ปีเกิด	28 กันยายน 2509
สถานที่เกิด	อ.แกลง จ.ระยอง
ที่อยู่ปัจจุบัน	73/31 ซ.อินทามระ 26 แขวงดินแดง เขตดินแดง กรุงเทพมหานคร 10320
ตำแหน่งหน้าที่การงานในปัจจุบัน	หัวหน้างานวัดผล ฝ่ายวิชาการ
สถานที่ทำงานปัจจุบัน	โรงเรียนอัสสัมชัญแผนกประถม 90/1 ซ.สาทร 11 ถนนสาทรใต้ แขวงยานนาวา เขตสาทร กรุงเทพมหานคร 10120
ประวัติการศึกษา	
พ.ศ. 2525	ระดับมัธยมศึกษาตอนต้น โรงเรียนแกลง "วิทยสถาวร" อ.แกลง จ.ระยอง
พ.ศ. 2528	ระดับมัธยมศึกษาตอนปลาย โรงเรียนศรีอยุธยา เขตพญาไท กรุงเทพมหานคร
พ.ศ. 2532	ระดับปริญญาตรี กศ.บ. สาขาวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ ประสานมิตร
พ.ศ. 2544	ระดับปริญญาโท กศ.ม. สาขาวัดผลการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ