

การใช้วิธีการตรวจคุณลักษณะ และสัดส่วนจำนวนผู้ตรวจให้คะแนนที่มีผลต่อความเที่ยงตรงของ
การวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3



เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาคุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา

ตุลาคม 2556

การใช้วิธีการตรวจคุณลักษณะ และสัดส่วนจำนวนผู้ตรวจให้คะแนนที่มีผลต่อความเที่ยงตรงของ
การวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3



ปริญญาานิพนธ์
ของ
บุษวรรษ์ แสนปลื้ม

เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาคุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา

ตุลาคม 2556

ลิขสิทธิ์เป็นของมหาวิทยาลัยศรีนครินทรวิโรฒ

การใช้วิธีการตรวจคุณลักษณะ และสัดส่วนจำนวนผู้ตรวจให้คะแนนที่มีผลต่อความเที่ยงตรงของ
การวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3



เสนอต่อบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ เพื่อเป็นส่วนหนึ่งของการศึกษา
ตามหลักสูตรปริญญาคุษฎีบัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา

ตุลาคม 2556

บุษวรรษ์ แสนปลื้ม. (2556). การใช้วิธีการตรวจ คุณลักษณะและสัดส่วนจำนวนผู้ตรวจให้คะแนนที่มีผลต่อความเที่ยงตรงของการวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3. ปรินทิพนิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. คณะกรรมการควบคุม: รองศาสตราจารย์ ดร.องอาจ นัยพัฒน์, อาจารย์ ดร.ชูศักดิ์ ชัมภลลิขิต, อาจารย์ ดร.สุวิมล กฤษศฤงคาร

การวิจัยในครั้งนี้ มีจุดมุ่งหมายเพื่อสร้างและพัฒนาคุณภาพของเครื่องมือวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 เพื่อศึกษาคุณลักษณะของผู้ตรวจให้คะแนน คือ ความเข้มงวด/ใจดี ความลำเอียงด้านเพศ และการทำหน้าที่ต่างกันของผู้ให้คะแนนเมื่อเวลาผ่านไป และเพื่อตรวจสอบความเที่ยงตรงของความสามารถในการเขียนของนักเรียน เครื่องมือที่ใช้ในการวิจัย คือ วิธีการให้คะแนนการเขียนเรียงความ 3 รูปแบบ และข้อสอบการเขียนจดหมายและวิธีการตรวจให้คะแนนการเขียนจดหมาย 3 รูปแบบ กลุ่มตัวอย่างที่ใช้ในการวิจัย คือ นักเรียนชั้นประถมศึกษาปีที่ 3 จำนวน 50 คน ที่ได้มาจากการสุ่มแบบหลายขั้นตอน และนักศึกษาคณะครุศาสตร์ ที่เรียนวิชาเอกภาษาไทย ชั้นปีที่ 4 มหาวิทยาลัยราชภัฏ 3 แห่ง ที่ได้มาจากการสุ่มอย่างง่าย จำนวน 130 คน

ผลการวิจัยสรุปได้ว่า

1. เครื่องมือวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 มีคุณภาพตามทฤษฎีมาตรฐานเดิม ทฤษฎีการสรุปอ้างอิง และทฤษฎีการตอบสนองข้อสอบ
2. คุณลักษณะของผู้ตรวจให้คะแนนการวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 พบว่า 1) ความเข้มงวด/ใจดี 1.1) แบบแยกองค์ประกอบ เข้มงวด 17 คน เป็นกลาง 89 คน ใจดี 9 คน 1.2) แบบรวมองค์ประกอบ เข้มงวด 10 คน เป็นกลาง 83 คน ใจดี 15 คน 1.3) แบบผสมผสาน เข้มงวด 17 คน เป็นกลาง 79 คน ใจดี 16 คน 2) การระบุคุณลักษณะระบุความเข้มงวด/ใจดี มีความสอดคล้องกัน ความลำเอียงด้านเพศ 2.1) แบบแยกองค์ประกอบ เข้าข้างเพศชาย 8 คน เป็นกลาง 92 คน เพศหญิง 15 คน 2.2) แบบรวมองค์ประกอบ เข้าข้างเพศชาย 5 คน เป็นกลาง 97 คน เพศหญิง 8 คน 2.3) แบบผสมผสาน เข้าข้างเพศชาย 10 คน เป็นกลาง 91 คน เพศหญิง 11 คน การระบุคุณลักษณะความลำเอียงด้านเพศมีความสอดคล้องกัน 3) การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป 3.1) แบบแยกองค์ประกอบ ไม่คงที่ 83 คน คงที่ 33 คน 3.2) แบบรวมองค์ประกอบ ไม่คงที่ 41 คน คงที่ 68 คน 3.3) แบบผสมผสาน ไม่คงที่ 71 คน คงที่ 41 คน การระบุคุณลักษณะการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไปไม่สอดคล้องกัน
3. ผลการหาความเที่ยงตรงของการวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 มีความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย และมีความเที่ยงตรงตามสภาพ โดยที่ 1) เมื่อผู้ตรวจมีคุณลักษณะลำเอียง การใช้วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบในสัดส่วน จำนวนผู้ตรวจให้คะแนน 3 คน ผู้ตรวจให้คะแนนมีความลำเอียงด้านเพศเข้าข้างเพศชาย 1 คน และลำเอียงเข้าข้างเพศหญิง 2 คน มีความความเที่ยงตรงตามสภาพสูงที่สุด และ 2) เมื่อผู้ตรวจมีคุณลักษณะเป็นกลาง

การใช้วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนจำนวนผู้ตรวจให้คะแนน 3 คน โดยผู้ตรวจให้คะแนนเป็นกลาง ไม่เข้มงวด/ใจดี มีความความเที่ยงตรงตามสภาพสูงที่สุด



THE EFFECT OF SCORING METHOD, RATER CHARACTERISTICS AND PROPORTION BY
THE RATERS INFLUENCING THE VALIDITY UNDER THE COMPETENCY OF COMPOSING
THE ESSAY BY THE STUDENTS IN ELEMENTARY SCHOOLING GRADE 3



Presented in Partial Fulfillment of the Requirements for the
Doctor of Education Degree in Testing and Measurement
at Srinakharinwirot University

October 2013

Butsawan Saenpluem. (2013). *The Effect of Scoring Method, Rater Characteristics and Proportion by the Raters Influencing the Validity Under the Competency of Students Composing the Essay by the Students in Elementary Schooling Grade 3*. Dissertation, Ed.D. (Testing and Measurement). Bangkok: Graduate School, Srinakharinwirot University. Advisor Committee: Assoc.Prof.Dr.Ong-Art Naiyapatana, Dr.Chusak Khumpalikit, Dr.Suwimon Kritkharuehart.

The purposes of this study were to create and develop the tools measuring the writing ability of the students in the elementary schooling grade 3 in order to study the raters' characteristics (severity/leniency, gender differential rater functioning (DRF) and Differential Rater Functioning Over Time (DRIFT)) on scoring the writing Ability differently and Validity Under the Competency of Composing the Essay. The instruments used in the study were the 3 types of scoring rubrics on the essay-type test and on the Writing Mail Test. The samples of this study were selected from 50 students from the elementary schooling grade 3 and 130 students majoring Thai language subject from Faculty of Education, 3 Rajabhat Universities through simple sampling.

The results of this study revealed that:

1. The instruments used for measuring the writing competency by the elementary schooling grade 3 identified the quality according to the following theories : Classical Test Theory Generalizability Theory and Item Response Theory.

2. According to the results of having identified the Rater Characteristic on measuring the writing ability of the students in the elementary schooling grade 3 as followings : The severity/leniency were : 1) 17 raters with the severity, 89 raters with centrality and 9 raters with leniency for the analytic scoring rubrics. 2) 10 raters with the severity, 83 raters with centrality and 15 raters with leniency for the holistic scoring rubrics. 3) 17 raters with the severity, 79 raters with centrality and 16 raters with leniency for the annotated scoring rubrics. The severity/leniency possessed no difference. DRF were : 1) 8 raters with bias relative to boy, 92 raters with centrality and 15 raters with bias relative to girl for the analytic scoring rubrics. 2) 5 raters with bias relative to boy, 93 raters with centrality and 8 raters with bias relative to girl for the holistic scoring rubrics. 3) 10 raters with bias relative to boy, 91 raters with centrality and 11 raters with bias relative to girl for the annotated scoring rubrics. DRF possessed no difference. DRIFT were : 1) 83 raters with inaccuracy and 33 raters with accuracy for the analytic scoring rubrics. 2) 41 raters with inaccuracy and 68 raters with

accuracy for the holistic scoring rubrics. 3) 71 raters with inaccuracy and 41 raters with accuracy for the annotated scoring rubrics. DRIFT possessed difference.

3. According to the results of the validity on measuring the writing competency by the elementary schooling grade 3, it indicated the Construct Validity in terms of Multitrait-Multimethod, and the Concurrent Validity as followings: 1) For the 3 raters with bias relative to gender based on the analytic scoring rubric, it identified 1 raters with bias relative to boys and 2 raters with bias relative to girls indicating the highest level of the convergent validity. 2) For the 3 raters with centrality based on the analytic scoring rubric, it identified no severity/leniency indicating the highest level of the convergent validity.



ปริญญานิพนธ์

เรื่อง

การใช้วิธีการตรวจ คุณลักษณะ และสัดส่วนจำนวนผู้ตรวจให้คะแนนที่มีผลต่อความเที่ยงตรงของ
การวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3

ของ

บุษวรรรษ์ แสนปลื้ม

ได้รับอนุมัติจากบัณฑิตวิทยาลัยให้นับเป็นส่วนหนึ่งของการศึกษาตามหลักสูตร

ปริญญาการศึกษาดุริยางค์บัณฑิต สาขาวิชาการทดสอบและวัดผลการศึกษา

ของมหาวิทยาลัยศรีนครินทรวิโรฒ

.....คณบดีบัณฑิตวิทยาลัย

(รองศาสตราจารย์.ดร.สมชาย สันติวัฒนกุล)

วันที่.....เดือน.....พ.ศ.2556

คณะกรรมการควบคุมปริญญานิพนธ์

คณะกรรมการสอบปากเปล่า

..... ประธาน

..... ประธาน

(รองศาสตราจารย์ ดร.องอาจ นัยพัฒน์)

(อาจารย์ ดร.สุวพร เข้มเฮง)

..... กรรมการ

..... กรรมการ

(อาจารย์ ดร.ชูศักดิ์ ชัมภลลิขิต)

(รองศาสตราจารย์ ดร.องอาจ นัยพัฒน์)

..... กรรมการ

..... กรรมการ

(อาจารย์ ดร.สุวิมล กฤษศยาสา)

(อาจารย์ ดร.ชูศักดิ์ ชัมภลลิขิต)

..... กรรมการ

(อาจารย์ ดร.สุวิมล กฤษศยาสา)

..... กรรมการ

(อาจารย์ ดร.น้ำผึ้ง อินทะเนตร)



งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัย
จาก
บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ

ประกาศคุณูปการ

ปริญญาโทฉบับนี้ สำเร็จได้ด้วยความกรุณาอย่างสูงจากคณะกรรมการควบคุมปริญญาโท
รองศาสตราจารย์ ดร.องอาจ นัยวัฒน์ ประธานควบคุมปริญญาโท อาจารย์ ดร.ชูศักดิ์ ชัมภลลิขิต
และอาจารย์ ดร.สุวิมล กฤษศยาสาห์ กรรมการควบคุมปริญญาโท ที่ได้ให้คำปรึกษา ชี้แนะแนวทาง
ในการดำเนินการวิจัยทุกขั้นตอน พร้อมทั้งได้แก้ไขข้อบกพร่องต่างๆ ด้วยความเอาใจใส่ ตลอดจน
ให้กำลังใจเสมอมา ผู้วิจัยขอกราบขอบพระคุณในความกรุณาของท่านทั้งสามด้วยความเคารพอย่างสูง

ขอกราบขอบพระคุณกรรมการที่แต่งตั้งเพิ่มเติมทั้งสองท่านคือ อาจารย์ ดร.สุวพร เข้มเฮง
และอาจารย์ ดร.น้ำผึ้ง อินทเนตร ที่ให้ความกรุณาเป็นประธานและกรรมการในการสอบปริญญาโท
และให้ข้อเสนอแนะที่เป็นประโยชน์เพิ่มเติมทำให้ปริญญาโทฉบับนี้สมบูรณ์ยิ่งขึ้น

ขอกราบขอบพระคุณคณาจารย์ภาควิชาการวัดผลและวิจัยการศึกษา มหาวิทยาลัยศรีนครินทรวิโรฒ
ทุกท่านที่ประสิทธิ์ประสาทวิชาความรู้แก่ผู้วิจัย และให้คำปรึกษาอย่างดียิ่งตลอดมา พร้อมทั้งขอขอบ
พระคุณบัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ ที่ได้สนับสนุนทุนอุดหนุนการวิจัยในครั้งนี้

ขอกราบขอบพระคุณผู้เชี่ยวชาญทุกท่านที่ให้ความช่วยเหลือตรวจสอบคุณภาพของเครื่องมือ
ที่ใช้ในการวิจัย ตลอดจนผู้บริหาร ครู และนักเรียนชั้นประถมศึกษาปีที่ 3 โรงเรียนเมืองสุรินทร์ ที่เสียสละ
เวลาในการทำแบบทดสอบวัดความสามารถในการเขียนจดหมาย ขอขอบพระคุณคณาจารย์ และ
นักศึกษาชั้นปีที่ 4 คณะครุศาสตร์ มหาวิทยาลัยราชภัฏเลย มหาวิทยาลัยราชภัฏสุรินทร์ มหาวิทยาลัย
ราชภัฏบุรีรัมย์ และมหาวิทยาลัยราชภัฏร้อยเอ็ด ที่ได้ให้ความร่วมมือในการเป็นกลุ่มตัวอย่างในการศึกษา
ครั้งนี้เป็นอย่างดี

ขอกราบขอบพระคุณบิดา มารดา ขอบคุณน้องสาว น้องชาย ญาติมิตร เพื่อนๆ พี่ๆ ร่วมรุ่น
ปริญญาเอกการทดสอบและวัดผลการศึกษาทุกคน ตลอดจนผู้เกี่ยวข้องทุกท่านที่ให้ความช่วยเหลือ
และให้กำลังใจตลอดมา

ท้ายนี้ประโยชน์และคุณค่าของปริญญาโทฉบับนี้ ผู้วิจัยขอมอบแต่บิดา มารดา คณาจารย์
และผู้เกี่ยวข้องทุกท่านที่มีส่วนร่วมในการทำให้ปริญญาโทฉบับนี้สำเร็จลุล่วงด้วยดี

บุษวรรษ์ แสนปลื้ม

สารบัญ

บทที่	หน้า
1 บทนำ	1
ภูมิหลัง	1
ความมุ่งหมายของการวิจัย	5
คำถามของการวิจัย	6
สมมติฐานของการวิจัย	6
ความสำคัญของการวิจัย	7
ขอบเขตของการวิจัย	8
ข้อจำกัดสำหรับการวิจัย	10
นิยามศัพท์เฉพาะ	10
กรอบแนวคิดในการวิจัย	13
2 เอกสารและงานวิจัยที่เกี่ยวข้อง	17
ตอนที่ 1 การเขียน	17
การเขียนเรียงความ	17
การเขียนจดหมาย	22
การวัดความสามารถในการเขียน	24
โมเดลการวัดความสามารถในการตรวจให้คะแนนการเขียน	35
ตอนที่ 2 คุณลักษณะของผู้ตรวจให้คะแนน	43
ความเข้มงวด/ใจดีของผู้ให้คะแนน	47
ความลำเอียงด้านเพศของผู้ให้คะแนน	50
การทำหน้าที่ต่างกันของผู้ให้คะแนนเมื่อเวลาผ่านไป	53
ตอนที่ 3 ความเที่ยงตรง	59
ความหมายของความเที่ยงตรง	59
ประเภทของความเที่ยงตรง	60
การแสดงผลฐานความเที่ยงตรง	61
ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง	64
การตรวจให้คะแนนการเขียน	64
คุณลักษณะของผู้ตรวจให้คะแนน	67

สารบัญ (ต่อ)

บทที่	หน้า
2(ต่อ)	
ความเที่ยงตรง	69
3 วิธีดำเนินการวิจัย	76
การกำหนดประชากรและกลุ่มตัวอย่าง	76
การสร้างเครื่องมือที่ใช้ในการวิจัย	77
การเก็บรวบรวมข้อมูล	84
การดำเนินการวิจัย	85
สถิติที่ใช้ในการวิจัย	91
เกณฑ์ที่ใช้ในการพิจารณา	95
4 ผลการวิเคราะห์ข้อมูล	97
ตอนที่ 1 การแสดงคุณภาพของเครื่องมือวัดความสามารถในการเขียนของ นักเรียนชั้นประถมศึกษาปีที่ 3	98
ตอนที่ 2 การศึกษาคูณลักษณะของผู้ตรวจให้คะแนน	110
ตอนที่ 3 การตรวจสอบความเที่ยงตรงของความสามารถในการเขียนของนักเรียน ชั้นประถมศึกษาปีที่ 3	118
5 สรุป อภิปรายผลและข้อเสนอแนะ	124
ความมุ่งหมายของการวิจัย	124
คำถามของการวิจัย	124
สมมติฐานของการวิจัย	125
การดำเนินการวิจัย	126
สรุปผลการวิจัย	128
อภิปรายผล	130
ข้อเสนอแนะ	133
บรรณานุกรม	135

สารบัญ (ต่อ)

บทที่	หน้า
ภาคผนวก	141
ภาคผนวก ก.....	142
ภาคผนวก ข.....	192
ภาคผนวก ค.....	209
ประวัติย่อผู้วิจัย	222



บัญชีตาราง

ตาราง	หน้า
1 เกณฑ์การตรวจให้คะแนนแบบรวมองค์ประกอบ	33
2 เกณฑ์การตรวจให้คะแนนแบบแยกองค์ประกอบ	33
3 คำอธิบายคุณภาพ “การพูดอภิปราย”	34
4 การกำหนดคะแนนการตัดสินระดับคุณภาพ	35
5 การสังเคราะห์งานวิจัยที่เกี่ยวข้อง	72
6 แสดงจำนวนประชากรและกลุ่มตัวอย่าง	77
7 ตัวอย่างคะแนนแบบแยกองค์ประกอบ	78
8 ตัวอย่างวิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ	79
9 ตัวอย่างคะแนนแบบรวมองค์ประกอบ	79
10 ตัวอย่างวิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ	81
11 ตัวอย่างวิธีการตรวจให้คะแนนแบบผสมผสาน	82
12 แบบแผนการทดลองแบบ Counterbalanced Designs ที่ใช้ในการวิจัย	86
13 แสดงรูปแบบการดำเนินการหาความเที่ยงตรงเชิงโครงสร้างโดยใช้ผู้ตรวจที่มี คุณลักษณะลำเอียง	88
14 แสดงรูปแบบการดำเนินการหาความเที่ยงตรงเชิงโครงสร้างโดยใช้ผู้ตรวจที่มี คุณลักษณะเป็นกลาง	91
15 ค่าสถิติพื้นฐานของคะแนนการเขียนเรียงความ จำแนกตามวิธีการตรวจให้คะแนน .	99
16 ค่าสถิติพื้นฐานของคะแนนการเขียนจดหมาย จำแนกตามวิธีการตรวจให้คะแนน ...	100
17 ค่าความแปรปรวนจากแหล่งต่าง ๆ ของการเขียนเรียงความ	101
18 ค่าความแปรปรวนจากแหล่งต่าง ๆ ของการเขียนจดหมาย	102
19 แสดงค่าพารามิเตอร์ของนักเรียน จำแนกตามคุณลักษณะผู้ตรวจให้คะแนน	103
20 แสดงค่าพารามิเตอร์ของข้อสอบการเขียนเรียงความ จำแนกตามคุณลักษณะผู้ตรวจให้ คะแนน	105
21 แสดงค่าพารามิเตอร์ของผู้ตรวจให้คะแนน จำแนกตามคุณลักษณะผู้ตรวจให้คะแนน	108
22 แสดงเส้นภาพแสดงความเข้มงวด/ใจดีของผู้ตรวจให้คะแนน จากโปรแกรม FAC	111
23 แสดงการระบุความเข้มงวด/ใจดีของผู้ตรวจให้คะแนน จำแนกตามวิธีการตรวจให้ คะแนน	112
24 แสดงการระบุความลำเอียงด้านเพศ จำแนกตามวิธีการตรวจให้คะแนน	114

บัญชีตาราง (ต่อ)

ตาราง		หน้า
25	แสดงการระบุการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป จำแนกตามวิธีการตรวจให้คะแนน	116
26	ค่าความสอดคล้องในการระบุคุณลักษณะของผู้ตรวจให้คะแนน เมื่อใช้วิธีการตรวจให้คะแนนเรียงความแตกต่างกัน	117
27	ความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย	119
28	แสดงค่าสหสัมพันธ์และค่าสหสัมพันธ์แปลงของความเที่ยงตรงตามสภาพ เมื่อผู้ตรวจมีคุณลักษณะลำเอียง	120
29	แสดงค่าสหสัมพันธ์และค่าสหสัมพันธ์แปลงของความเที่ยงตรงตามสภาพ เมื่อผู้ตรวจมีคุณลักษณะเป็นกลาง	123



บัญชีภาพประกอบ

ภาพประกอบ	หน้า
1 กรอบแนวคิดในการศึกษาความเที่ยงตรงเชิงโครงสร้าง	13
2 กรอบแนวคิดในการศึกษาความเที่ยงตรงตามสภาพ	15
3 รูปแบบการวัดในการสอบเรียงความของ Community English Program (CEP)	43



บทที่ 1

บทนำ

ภูมิหลัง

ภาษาไทยเป็นเอกลักษณ์ประจำชาติไทย เป็นวัฒนธรรมของการใช้ถ้อยคำเพื่อสื่อสารให้เกิดความเข้าใจที่ตรงกัน คนไทยจึงมีความจำเป็นต้องอ่านและเขียนภาษาไทยได้จึงจะอยู่ในสังคมอย่างปกติสุข ในปัจจุบันทุกคนที่เกิดเป็นคนไทย จะมีสิทธิขั้นพื้นฐานในการรับการศึกษาในระบบโรงเรียน โดยมีจุดประสงค์หลักเพื่อให้ทุกคนอ่านออกเขียนได้และสามารถนำไปต่อยอดเพื่อการศึกษาในระดับสูงขึ้นไป หรือประกอบอาชีพในอนาคต (ชำนาญ สุภณิตย์; และ สุทิน พูลสวัสดิ์. 2547: 257) การศึกษาในระดับประถมศึกษาตอนต้นนั้น ผู้เรียนจะต้องมีผลการประเมินการอ่าน คติวิเคราะห์ และเขียนในระดับผ่านเกณฑ์ การประเมินตามที่สถานศึกษากำหนด ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ที่ได้กำหนดคุณภาพการเขียนของนักเรียน ไว้ในมาตรฐานการเรียนรู้ในกลุ่มสาระการเรียนรู้ภาษาไทย สาระที่ 2 การเขียน มาตรฐาน ท 2.1 คือ ใช้กระบวนการเขียน เขียนสื่อสาร เขียนเรียงความ ย่อความ และเขียนเรื่องราวในรูปแบบต่างๆ เขียนรายงานข้อมูลสารสนเทศ และรายงานการศึกษาค้นคว้าอย่างมีประสิทธิภาพ (กระทรวงศึกษาธิการ. 2552ก) ซึ่งจะเห็นได้ว่าการเขียนเป็นวิธีการขั้นพื้นฐานที่สำคัญยิ่งในการวัดการอ่านออกเขียนได้ โดยช่วงที่สำคัญที่สุดในการวัดการเขียนเพื่อให้เกิดการพัฒนาอย่างสูงสุด คือช่วงการเริ่มต้นในระดับประถมศึกษานั้นเอง

การเขียนเรียงความ เป็นการเขียนร้อยแก้วที่เรียบเรียงข้อความเป็นเรื่องราวถ่ายทอดออกมาเป็นภาษาเขียน โดยใช้กลวิธีทั้งทางด้านกรเขียนและภาษาให้เหมาะสมกับเนื้อหา ตลอดจนการเลือกสรรถ้อยคำ สำนวนโวหาร และการใช้ภาษาให้ถูกต้อง เพื่อแสดงความรู้ ความคิด ความรู้สึกของผู้เขียนที่เขียนถึงบุคคล สถานที่ เรื่อง หรือสิ่งใดสิ่งหนึ่ง โดยอาศัยความรู้ ความเข้าใจ ความคิดเห็น หรือประสบการณ์ของผู้เขียนที่เกี่ยวข้องกับเรื่องนั้นๆ โดยที่ผู้เขียนต้องการถ่ายทอดความรู้ ความคิด และความรู้สึกมายังผู้อ่าน (เปลื้อง ณ นคร. 2512; ถวัลย์ มาศจรัส. 2545; ประพนธ์ เรืองณรงค์. 2545; ผจงวาด พูลแก้ว; พัชณี ปี่เพราะ; และ สุวรรณ โสประดิษฐ์. 2551; กระทรวงศึกษาธิการ. 2552ก; และ ชูลีพร ดัดงาม. 2552) ส่วนการเขียนจดหมายเป็นวิธีสื่อสารที่เป็นลายลักษณ์อักษรแทนการพูด ถึงแม้ในปัจจุบันเทคโนโลยีสื่อสารมีให้เลือกสื่อสารได้หลายช่องทาง แต่จดหมายก็ยังคงมีความจำเป็นอยู่มาก (บุณยลักษณ์ เอี่ยมสำอางค์; เกื้อกมล พฤษประมุข; และ โสภิต พิทักษ์. 2551) โดยเฉพาะการติดต่อกิจกรรมระหว่างหน่วยงานต่างๆ ทั้งนี้อาจเนื่องมาจาก การเขียนจดหมายสามารถทำให้ผู้รับจดหมายได้รับข้อความที่มีความสมบูรณ์ชัดเจน ทำให้เกิดความเข้าใจที่ตรงกัน อีกทั้งจดหมายยังสามารถใช้เป็นหลักฐานของการติดต่อระหว่างกันได้ ซึ่งการเขียนเรียงความแตกต่างกับการเขียนจดหมายในส่วนที่เป็นองค์ประกอบของรูปแบบของการเขียน โดยที่รูปแบบการเขียนเรียงความประกอบไปด้วย คำนำ เนื้อเรื่อง สรุป แต่รูปแบบการเขียนจดหมายจะประกอบไปด้วย ที่อยู่ วันที่ การขึ้นต้น และการลงท้าย แต่เนื้อเรื่องในจดหมายนั้นมีส่วนที่คล้ายกับ

การเขียนเรียงความ คือ เมื่อจบการเขียนส่วนที่เป็นการขึ้นต้นแล้ว เนื้อเรื่องที่เขียนจะประกอบไปด้วย คำนำ เนื้อเรื่อง และสรุป แล้วจึงจบการเขียนจดหมายด้วยการเขียนคำลงท้าย

โดยทั่วไปแบบทดสอบมีจุดมุ่งหมายเพื่อทำความเข้าใจ เพื่ออธิบาย และเพื่อทำนายพฤติกรรมต่าง ๆ ของผู้สอบ จุดมุ่งหมายของแบบทดสอบจะบรรลุผลหรือไม่ย่อมขึ้นอยู่กับความสามารถในการอ้างอิงคะแนนนั้นไปยังผู้สอบได้อย่างแท้จริง ความเที่ยงตรงเป็นความสามารถของเครื่องมือ ในการวัดในสิ่งที่ต้องการตามจุดมุ่งหมายที่กำหนด นอกจากนี้ความเที่ยงตรงยังเป็นระดับความสอดคล้อง หรือความสัมพันธ์ระหว่างคะแนนที่ได้จากการวัดกับสิ่งที่ต้องการวัดอีกด้วย (ล้วน สายยศ; และ อังคณา สายยศ. 2543; มหาวิทยาลัยสุโขทัยธรรมาธิราช. 2549; เตือนใจ เกตุษา. 2549; เยาวดี วิบูลย์ศรี. 2552; Linderman; & Merenda. 1979; Carmines; & Zeller. 1987; Lyman. 1991; & Gruijter; & Van Der Kamp. 2008) ความเที่ยงตรงของแบบทดสอบมีประโยชน์ทั้งในทางทฤษฎีและในทางปฏิบัติของการทดสอบ ในการแบ่งประเภทของความเที่ยงตรง มักแบ่งออกตามเกณฑ์ที่ใช้ในการอ้างอิง โดยส่วนใหญ่แล้วแบ่งออกเป็น 3 ประเภท คือ 1) ความเที่ยงตรงเชิงเนื้อหา (Content Validity) เป็นความสามารถของเครื่องมือที่วัดได้ตามเนื้อหาที่ต้องการจะวัด ซึ่งการพิจารณาความเที่ยงตรงชนิดนี้ จะใช้การวิเคราะห์อย่างมีเหตุผล (Rational Analysis) แบ่งออกเป็น 2 ประเภทย่อย คือ ความเที่ยงตรงเชิงเหตุผล (Logical Validity) และความเที่ยงตรงเชิงพินิจ (Face Validity) 2) ความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) เป็นความสามารถของเครื่องมือที่วัดได้ตรงตามลักษณะหรือตามทฤษฎีต่าง ๆ ของโครงสร้างนั้น และ 3) ความเที่ยงตรงตามเกณฑ์สัมพันธ์ (Criterion-Related Validity) เป็นความสามารถของเครื่องมือที่นำผลการวัดของแบบทดสอบไปหาความสัมพันธ์กับเกณฑ์ที่ต้องการ แบ่งได้เป็น 2 ประเภทย่อย คือ ความเที่ยงตรงตามสภาพ (Concurrent Validity) และความเที่ยงตรงเชิงพยากรณ์ (Predictive Validity) ปัญหาที่สำคัญของการขาดความเที่ยงตรงอยู่ที่ความพอเพียง ความเหมาะสม และความเป็นเหตุเป็นผลของหลักฐานที่จะนำไปใช้ในการอ้างอิง ความเที่ยงตรงแต่ละประเภทจะมีลักษณะแตกต่างกันไป ขึ้นอยู่กับเกณฑ์ที่ใช้ในการเปรียบเทียบนั่นเอง เนื่องจากสารสนเทศที่ได้จากการวิจัยในครั้งนี้ สามารถใช้เกณฑ์ในการเปรียบเทียบได้ 2 เกณฑ์ คือ เกณฑ์ที่เปรียบเทียบด้านโครงสร้าง และเกณฑ์ที่เปรียบเทียบในเวลาใกล้เคียงกัน ดังนั้นงานวิจัยชิ้นนี้จึงสนใจการแสดงผลหลักฐานความเที่ยงตรง 2 ประเภท คือ ความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย และความเที่ยงตรงตามสภาพ

การทดสอบทุกรูปแบบมีหน้าที่เหมือนกันคือมุ่งวัดผลสัมฤทธิ์ในการเรียนของนักเรียน การวัดความสามารถในการเขียนของนักเรียนก็เช่นกัน คะแนนที่ได้จากการตรวจให้คะแนนควรจะเป็นตัวแทนที่ดีของความสามารถของนักเรียนที่ผ่านมาพบว่ามี การนำทฤษฎีการตอบสนองข้อสอบ (Item-Response Theory : IRT) ซึ่งเป็นการนำโมเดลตอบสนองรายข้อมาใช้ให้เป็นประโยชน์ต่อการวัดผล โดยที่โมเดลการวัดที่เหมาะสมในการตรวจให้คะแนนการเขียน คือ พาเชียลเครดิตโมเดล (Partial Credit Model : PCM) ที่ได้รับการพัฒนาให้ดีขึ้นโดย Wright และ Masters (1982: 38) เนื่องจากพาเชียลเครดิตโมเดลมีลักษณะคล้ายโมเดลของราล์ซ ที่ใช้หลักการคำนวณความน่าจะเป็นของการตอบแต่ละระดับชั้น จากการตอบโดยตรงแบบขั้นตอนเดียว (Direct IRT Method) และลักษณะเกณฑ์ในแต่ละชุดนั้นจะถูกอธิบาย

ด้วยค่าพารามิเตอร์ระดับความยากประจำข้อ (Item Step Difficulty : δ_{ij}) ในทุกเกณฑ์ย่อยจะมีค่าพารามิเตอร์ความชันเท่ากัน (Equal Slope) และอยู่บนพื้นฐานที่ว่าเกณฑ์ทุกข้อจะต้องมีความสัมพันธ์ที่เท่าเทียมกันกับคุณลักษณะความสามารถของผู้สอบ (Ability of Examinee : θ) ดังนั้นโมเดลนี้จึงเหมาะสมสำหรับการวัดความสามารถในการเขียนที่ผู้วิจัยสร้างขึ้น

การวัดความสามารถในการเขียนของนักเรียนนั้น นอกจากจะต้องใช้โมเดลที่เหมาะสมแล้วยังมีปัจจัยอื่นที่เกี่ยวข้องอีกหลายประการ ซึ่งปัญหาที่นักเรียนทำคะแนนการเขียนได้น้อยนั้น อาจไม่ได้เกิดจากนักเรียนขาดการฝึกหัด หรือขาดทักษะการเขียนเพียงอย่างเดียวเท่านั้น แต่อาจเกิดจากคุณภาพของการตรวจให้คะแนนที่ขาดมาตรฐาน ไม่มีความคงเส้นคงวาและความเชื่อถือได้ หรือจากความลำเอียงเฉพาะตัวของผู้ตรวจ เช่น ความเข้มงวด ความใจดี ความคล้อยตามกัน การให้คะแนนที่มีแนวโน้มเป็นกลาง หรือมาจากความประทับใจส่วนตัวต่อนักเรียน (McMilland. 2004: 228 – 229; & Miller. 2009: 277 – 278; & Payne. 1997: 223 – 225) นอกจากนี้ สิ่งแวดล้อมขณะการตรวจให้คะแนน ก็อาจมีผลต่อการตรวจให้คะแนนเช่นกัน เช่น บรรยากาศขณะตรวจให้คะแนน เกณฑ์ที่ใช้ในการตรวจให้คะแนน ก้ากวม ไม่มีความเป็นปรนัย ซึ่งวิธีการหนึ่งในแก้ปัญหาดังที่กล่าวมาข้างต้นคือ ต้องมีเกณฑ์การตรวจให้คะแนนที่ชัดเจน มีความเป็นปรนัย มีความเชื่อมั่นและความเที่ยงตรงสูง โดยส่วนใหญ่แล้วผู้เชี่ยวชาญได้แบ่งวิธีการตรวจให้คะแนนออกเป็น 2 รูปแบบ คือ 1) แบบแยกองค์ประกอบ (Analytic Scoring Rubrics) เกณฑ์แบบนี้ใช้เวลาตรวจค่อนข้างนาน สร้างความเมื่อยล้าหากชิ้นงานที่ต้องตรวจมีจำนวนมาก แต่ข้อดีคือ มีสารสนเทศที่ครบถ้วนส่งกลับไปให้นักเรียน เพื่อพัฒนาการเขียนให้ดียิ่งขึ้น และ 2) แบบรวมองค์ประกอบ (Holistic Scoring Rubrics) แม้ว่ารูปแบบนี้ตรวจง่าย ใช้เวลาน้อย และมีความเชื่อมั่นสูง แต่ก็ยากที่จะอธิบายเหตุผลของคะแนนที่นักเรียนได้รับ (ส.วาสนา ประवालพฤกษ์. 2533: 39 – 42; ชัยฤทธิ์ ศิลาเดช. 2544: 29; บุญเชิด ภิญญ์-อนันตพงษ์. 2544: 90; และ สำนักทดสอบทางการศึกษากกรมวิชาการ. 2544: 35 – 38) ส่วนรูปแบบที่ 3) แบบผสมผสาน (Annotated Scoring Rubrics) เป็นการตรวจให้คะแนนที่เพิ่มขึ้นมาใหม่ (Nitko. 1996) เป็นการรวมเกณฑ์การให้คะแนนแบบรวมองค์ประกอบและแบบแยกองค์ประกอบไว้ด้วยกัน เริ่มด้วยการให้คะแนนในภาพรวมด้วยเกณฑ์แบบรวมองค์ประกอบแล้วเลือกให้คะแนนอีกเพียงบางคุณลักษณะของการให้คะแนนแบบแยกองค์ประกอบที่ผู้ตรวจคิดว่าสำคัญ เพื่อเป็นการประหยัดเวลาในการตรวจ และมีสารสนเทศที่จำเป็น ย้อนกลับไปให้นักเรียน จะเห็นได้ว่าวิธีการตรวจให้คะแนนทั้ง 3 รูปแบบ ดังที่กล่าวข้างต้น มีข้อดีและข้อด้อยแตกต่างกัน และนอกจากนี้ ผลการวิจัยของ นิศารัตน์ คงสวัสดิ์ (2544) และ ปวีณา ปิอาทิตย (2545) พบว่า เมื่อใช้รูปแบบวิธีการตรวจให้คะแนนแตกต่างกัน ผู้ตรวจให้คะแนนจะให้คะแนนแตกต่างกัน จะเห็นได้ว่าวิธีการตรวจให้คะแนนส่งผลทำให้คะแนนที่ได้จากการตรวจมีค่าแตกต่างกัน ดังนั้นผู้วิจัยจึงสนใจที่จะสร้างวิธีการตรวจให้คะแนนในการวัดความสามารถในการเขียนของนักเรียนที่แตกต่างกันทั้ง 3 รูปแบบ เพื่อวัดความสามารถในการเขียนเรียงความและการเขียนจดหมายของนักเรียน แล้วศึกษาว่าเมื่อผู้ตรวจใช้วิธีการตรวจให้คะแนนที่แตกต่างกันแล้ว คะแนนของนักเรียนจะมีลักษณะเป็นอย่างไร

นอกจากปัจจัยที่ส่งผลต่อการแตกต่างกันของคะแนน ที่ได้จากการตรวจให้คะแนนการวัดความสามารถในการเขียนของนักเรียนข้างต้นแล้ว การวิเคราะห์ตามโมเดลของราสช์แบบหลายองค์ประกอบ (Many-Facet Rasch Measurement : MFRM) ยังพบว่า การวิเคราะห์ส่วนใหญ่ได้มุ่งเน้นการระบุคุณลักษณะของผู้ตรวจให้คะแนนอีกด้วย (Rater's Characteristics) เช่น ความเข้มงวด/ใจดี (Severity/ Leniency) ผู้ตรวจให้คะแนนที่มีแนวโน้มให้คะแนนที่ต่ำจะถูกระบุว่ามีความเข้มงวด และผู้ตรวจให้คะแนนที่มีแนวโน้มให้คะแนนที่สูง จะถูกระบุว่า มีความใจดี ซึ่งจะพิจารณาจากค่าศูนย์ (0) เครื่องหมายบวก (+) ลบ (-) ของค่า Logit ในโปรแกรม FACETS 3.70.1 ของ Linacre John M. (2012) ดังเช่น งานวิจัยของ นากามูระ (Nakamura. 2002) ที่ศึกษาการใช้งาน MFRM ในการวิเคราะห์การสอบการเขียน นอกจากนี้ ได้มีการต่อยอดการใช้งาน MFRM โดยใช้หาความลำเอียงด้านเพศของผู้ตรวจให้คะแนน (Differential Rater Functioning : DRF) โดยใช้หลักการของการวิเคราะห์ความเข้มงวด/ใจดีของผู้ตรวจให้คะแนนข้ามกลุ่มเพศ ดังเช่นงานวิจัยของเอคเคส (Eckes. 2005) ที่ศึกษาการวิเคราะห์ MFRM ในการตรวจสอบผลกระทบจากผู้ตรวจให้คะแนนในการสอบ TestDaF และในปี 2009 ได้มีการนำ MFRM มาระบุคุณลักษณะผู้ตรวจให้คะแนน คือ การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป (Differential Rater Functioning Over Time : DRIFT) โดยใช้หลักการของการวิเคราะห์ความเข้มงวด/ใจดีของผู้ตรวจให้คะแนนข้ามเวลา (Time Facet) ดังเช่น งานวิจัยของ เมฟอร์ด และ วูฟ (Myford; & Wolfe. 2009) ที่ได้ศึกษาการเฝ้าตรวจสอบการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป ซึ่งการใช้ MFRM ในการระบุคุณลักษณะผู้ตรวจให้คะแนนนั้น จะนำไปสู่ความเข้าใจในความคลาดเคลื่อนของผู้ตรวจให้คะแนนมากขึ้น นอกจากนี้ การศึกษาคุณลักษณะผู้ตรวจให้คะแนนนั้นไม่ปรากฏในงานวิจัยในประเทศ ดังนั้นผู้วิจัยจึงสนใจศึกษาว่า ผู้ตรวจให้คะแนนความสามารถในการเขียนภาษาไทยนั้นมีคุณลักษณะเป็นอย่างไร โดยมีรายละเอียดในแต่ละคุณลักษณะดังนี้ คือ

ความเข้มงวด/ใจดี (Severity/Leniency) ความเข้มงวดและความใจดีของผู้ตรวจให้คะแนนสามารถอธิบายได้ในรูปแบบของโครงสร้างตัวแปรแฝงที่ได้รับการอธิบายเป็นตำแหน่งต่างๆ ของผู้ตรวจให้คะแนนใน Underlying Linear Continuum โดยที่ผู้ตรวจให้คะแนนที่มีแนวโน้มให้คะแนนที่ต่ำจะถูกระบุว่า มีความเข้มงวด และผู้ตรวจให้คะแนนที่มีแนวโน้มให้คะแนนที่สูงจะถูกระบุว่ามีความใจดี การค้นหาคุณลักษณะความเข้มงวด/ใจดีของผู้ตรวจให้คะแนนพบได้จากงานวิจัยของ นากามูระ (Nakamura. 2002) ที่ศึกษาการใช้งานโมเดลของราสช์ แบบหลายองค์ประกอบในการวิเคราะห์การสอบการเขียน และงานวิจัยของ วูฟ (Wolfe. 2004) ที่ศึกษาการระบุผลกระทบจากผู้ตรวจให้คะแนนโดยใช้โมเดลตัวแปรแฝง ซึ่งผลการวิจัยทั้งสองชิ้น พบว่า ระดับความเข้มงวดและใจดีของผู้ตรวจให้คะแนนมีค่าแตกต่างกันเล็กน้อย ดังนั้นผู้วิจัยจึงสนใจที่จะศึกษาว่า ผู้ตรวจให้คะแนนที่เป็นกลุ่มตัวอย่างในการวิจัยในครั้งนี้ จะมีผู้ที่มีคุณลักษณะเข้มงวดและใจดีหรือไม่ และเมื่อใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน 3 รูปแบบคุณลักษณะความเข้มงวด/ใจดีของผู้ตรวจให้คะแนนจะมีการเปลี่ยนแปลงหรือไม่ อย่างไร

ความลำเอียงด้านเพศของผู้ตรวจให้คะแนน (Differential Rater Functioning : DRF) เป็นการต่อยอดการวิเคราะห์ความเข้มงวด/ใจดี โดยความลำเอียงด้านเพศของผู้ตรวจให้คะแนนนั้นมีหลักการคล้ายคลึงกับการทำหน้าที่ต่างกันของข้อสอบ (Differential Item Functioning) ที่ได้รับการสังเกตผ่าน

กลุ่มย่อยที่เกี่ยวข้องกับผู้สอบ โดยใช้หลักการของการวิเคราะห์ความเข้มงวด/ใจดีของผู้ตรวจให้คะแนนข้ามกลุ่มเพศ ดังเช่นงานวิจัยของ เอกเคส (Eckes. 2005) ที่ศึกษาการวิเคราะห์โมเดลของราล์ซ แบบหลายองค์ประกอบในการตรวจสอบผลกระทบจากผู้ตรวจให้คะแนนในการสอบภาษาเยอรมันในฐานะภาษาต่างประเทศ ในงานวิจัยชิ้นนี้ มีผู้ตรวจให้คะแนนเป็นครูผู้เชี่ยวชาญด้านภาษา จำนวน 29 คน และตรวจสอบไม่พบการลำเอียงด้านเพศอย่างมีนัยสำคัญทางสถิติ แต่พบว่า คะแนนในส่วนของการเขียนและการพูด เพศหญิงได้คะแนนสูงกว่าเพศชาย ผู้วิจัยจึงสนใจว่าเมื่อใช้กลุ่มตัวอย่างที่ผู้ตรวจให้คะแนนเป็นนักศึกษาจะตรวจพบคุณลักษณะความลำเอียงด้านเพศของผู้ตรวจให้คะแนนหรือไม่ และเมื่อใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน 3 รูปแบบ คุณลักษณะความลำเอียงด้านเพศของผู้ตรวจให้คะแนนจะมีการเปลี่ยนแปลงหรือไม่ อย่างไร

การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป (Differential Rater Functioning Over Time : DRIFT) ใช้หลักการของการวิเคราะห์ความเข้มงวด/ใจดีของผู้ตรวจให้คะแนนข้ามเวลา (Time Facet) โดยใช้คะแนนจากจุดเวลาที่แตกต่างกัน เช่น เข้า-บ่าย ครั้งที่ 1 – 3 หรือวันที่ 1 – 5 เป็นต้น ดังเช่น งานวิจัยของ เมฟอร์ด และ วูฟ (Myford; & Wolfe. 2009) ที่ได้ศึกษาการเฝ้าตรวจสอบการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไปโดยใช้โมเดลของราล์ซแบบหลายองค์ประกอบงานวิจัยชิ้นนี้เน้นไปที่การตรวจสอบการเขียนในการสอบ AP ELC ใช้ผู้ตรวจให้คะแนน 101 คน โดยแบ่งองค์ประกอบของเวลาออกเป็น 8 จุดเวลา ในเวลา 4 วัน (เข้า-บ่าย) ซึ่งผลการวิจัยพบว่า ผู้ตรวจให้คะแนนไม่ทำหน้าที่ต่างกันเมื่อเวลาผ่านไป ผู้วิจัยจึงสนใจว่าจะตรวจพบคุณลักษณะการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไปหรือไม่ และเมื่อใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน 3 รูปแบบ คุณลักษณะการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไปจะมีการเปลี่ยนแปลงหรือไม่ อย่างไร

จากที่กล่าวมาข้างต้นผู้วิจัย จึงสนใจสร้างและพัฒนาคุณภาพวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 ซึ่งเป็นชั้นที่มีความสำคัญในการตรวจสอบพัฒนาการของการเขียนในระดับประถมศึกษาตอนต้น โดยศึกษาความเที่ยงตรงเชิงโครงสร้างและความเที่ยงตรงตามสภาพของการวัดด้วยการศึกษาพฤติกรรมของผู้ตรวจให้คะแนน ในเงื่อนไขด้านวิธีการตรวจให้คะแนน ด้านคุณลักษณะของผู้ตรวจให้คะแนน และสัดส่วนจำนวนผู้ตรวจให้คะแนน เพื่อขยายขอบเขตความเข้าใจในการตรวจให้คะแนนแบบทดสอบวัดความสามารถในการเขียนให้มีประสิทธิภาพยิ่งขึ้นต่อไป

ความมุ่งหมายของการวิจัย

การวิจัยครั้งนี้มีความมุ่งหมายสำคัญเพื่อ

1. สร้างและพัฒนาคุณภาพของเครื่องมือวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3
2. เพื่อศึกษาคุณลักษณะของผู้ตรวจให้คะแนน คือ ความเข้มงวด/ใจดี ความลำเอียงด้านเพศ และการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป

3. เพื่อตรวจสอบความเที่ยงตรงของการวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3

คำถามของการวิจัย

เพื่อให้การดำเนินงานศึกษาวิจัยครั้งนี้บรรลุตามความมุ่งหมายข้างต้น ผู้วิจัยจึงได้ทำการรวบรวมและวิเคราะห์ข้อมูลเพื่อตอบคำถามการวิจัยต่อไปนี้

1. เครื่องมือวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 คือ วิธีการให้คะแนนการเขียนเรียงความ และข้อสอบการเขียนจดหมายและวิธีการตรวจให้คะแนนการเขียนจดหมายที่ผู้วิจัยสร้างขึ้นมีคุณภาพหรือไม่ อย่างไร

2. การระบุคุณลักษณะของผู้ตรวจให้คะแนน จากการวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 ในครั้งนี้ จะพบผู้ตรวจให้คะแนนที่มีคุณลักษณะต่อไปนี้หรือไม่

2.1 มีผู้ตรวจให้คะแนนที่มีคุณลักษณะเข้มงวด และใจดีหรือไม่ อย่างไร

2.2 มีผู้ตรวจให้คะแนนที่มีคุณลักษณะลำเอียงเข้าข้างเพศชายและลำเอียงเข้าข้างเพศหญิงหรือไม่ อย่างไร

2.3 มีผู้ตรวจให้คะแนนที่มีคุณลักษณะการให้คะแนนไม่คงที่เมื่อเวลาผ่านไปหรือไม่ อย่างไร

2.4 ผู้ตรวจคนเดียวกันจะมีคุณลักษณะแตกต่างกันหรือไม่เมื่อใช้วิธีการตรวจแตกต่างกัน

3. การวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 จากการวิจัยครั้งนี้มีหลักฐานแสดงความเที่ยงตรงในด้านต่างๆ ต่อไปนี้หรือไม่

3.1 มีความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย คือ ความเที่ยงตรงเชิงเหมือนและมีความเที่ยงตรงเชิงจำแนกหรือไม่ อย่างไร

3.2 มีความเที่ยงตรงตามสภาพภายใต้การใช้วิธีการตรวจให้คะแนนแตกต่างกัน ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนนที่มีคุณลักษณะแตกต่างกันหรือไม่ และเงื่อนไขใดที่มีความเที่ยงตรงตามสภาพสูงสุด

สมมติฐานของการวิจัย

เพื่อให้การดำเนินงานศึกษาวิจัยครั้งนี้ บรรลุตามความมุ่งหมายข้างต้น ผู้วิจัยจึงได้ทำการรวบรวมและวิเคราะห์ข้อมูลเพื่อหาคำตอบในสมมติฐาน ดังต่อไปนี้

1. การระบุคุณลักษณะของผู้ตรวจให้คะแนน

1.1 มีผู้ตรวจให้คะแนนที่มีคุณลักษณะเข้มงวด และใจดี โดยมีค่า Logit สูงกว่าเกณฑ์ที่ตั้งไว้

1.2 มีผู้ตรวจให้คะแนนที่มีคุณลักษณะลำเอียงเข้าข้างเพศชาย และลำเอียงเข้าข้างเพศหญิงอย่างมีนัยสำคัญทางสถิติ

1.3 มีผู้ตรวจให้คะแนนที่มีคุณลักษณะการให้คะแนนไม่คงที่ข้ามเวลา อย่างมีนัยสำคัญทางสถิติ

1.4 ผู้ตรวจคนเดียวกันมีคุณลักษณะที่ไม่แตกต่างกันเมื่อใช้วิธีการตรวจแตกต่างกัน อย่างมีนัยสำคัญทางสถิติ

2. ความเที่ยงตรงของการวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3

2.1 มีความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย โดยที่ความเที่ยงตรงเชิงเหมือน คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการวัดคุณลักษณะเดียวกันด้วยวิธีการตรวจให้คะแนนต่างกัน มีค่าสูง อย่างมีนัยสำคัญทางสถิติ และความเที่ยงตรงเชิงจำแนก คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการวัดคุณลักษณะต่างกันด้วยวิธีการตรวจให้คะแนนต่างกัน หรือระหว่างการวัดคุณลักษณะต่างกันด้วยวิธีการตรวจให้คะแนนเดียวกัน มีค่าต่ำ

2.2 มีความเที่ยงตรงตามสภาพภายใต้การใช้วิธีการตรวจให้คะแนนแตกต่างกัน ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนนที่มีคุณลักษณะแตกต่างกัน โดยมีเงื่อนไขคือ 1) เมื่อผู้ตรวจมีคุณลักษณะลำเอียง การให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนน 2 คน ที่มีคุณลักษณะเข้มงวด 1 คน และใจดี 1 คน มีความเที่ยงตรงตามสภาพสูงสุด และ 2) เมื่อผู้ตรวจมีคุณลักษณะเป็นกลาง การให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนน 2 คน ที่มีคุณลักษณะเป็นกลางไม่เข้มงวด/ใจดี มีความเที่ยงตรงตามสภาพสูงสุด

ความสำคัญของการวิจัย

งานวิจัยนี้นอกจากจะมีจุดประสงค์หลักเพื่อขยายขอบเขตความเข้าใจในการทดสอบและวัดผลสัมฤทธิ์ด้วยข้อสอบการเขียนแล้ว ยังมีจุดประสงค์เพื่อส่งเสริมการพัฒนาความสามารถในการตรวจให้คะแนนอีกด้วย โดยผู้วิจัยได้สร้างและพัฒนาคุณภาพของเครื่องมือวัดความสามารถในการเขียนคือวิธีการตรวจให้คะแนนเรียงความ 3 รูปแบบที่แตกต่างกัน และได้สร้างข้อสอบการเขียนจดหมาย และวิธีการตรวจให้คะแนนจดหมาย โดยคำนึงถึงองค์ประกอบที่สามารถนำไปใช้ในการพัฒนาการเขียนของนักเรียนในอนาคต นอกจากนี้การระบุคุณลักษณะของผู้ตรวจให้คะแนน คือ ความเข้มงวด/ใจดีของผู้ตรวจให้คะแนน ความลำเอียงด้านเพศของผู้ตรวจให้คะแนน และการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป ยังเป็นการขยายขอบเขตความเข้าใจความคลาดเคลื่อนที่เกิดจากผู้ตรวจให้คะแนนและยังสามารถนำไปใช้พัฒนาต่อยอดโดยใช้เป็นส่วนหนึ่งของเครื่องมือในการอบรมการเป็นผู้ตรวจให้คะแนนของนิสิต/นักศึกษาที่เรียนคณะศึกษาศาสตร์และครุศาสตร์ เพื่อให้นิสิต/นักศึกษารับรู้จุดบกพร่องในการตรวจให้คะแนนและปรับปรุงตนเอง เป็นการเพิ่มความยุติธรรมในการตรวจให้คะแนนในอนาคตได้เป็นอย่างดี และจากการตรวจสอบความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย และการตรวจสอบความเที่ยงตามสภาพของการวัดความสามารถในการเขียนของนักเรียน ยังเป็นการศึกษาเพื่อความเข้าใจในผลกระทบของความคลาดเคลื่อนที่เกิดจาก วิธีการตรวจให้คะแนน คุณลักษณะของผู้ตรวจให้คะแนน และสัดส่วนจำนวนของผู้ตรวจให้คะแนนอีกด้วย

ขอบเขตของการวิจัย

1. ตัวแปร

1.1 ตัวแปรอิสระ

1.1.1 วิธีการตรวจให้คะแนน

1.1.1.1 แบบแยกองค์ประกอบ

1.1.1.2 แบบรวมองค์ประกอบ

1.1.1.3 แบบผสมผสาน

1.1.2 คุณลักษณะของผู้ตรวจให้คะแนน

1.1.2.1 ความเข้มงวด/ใจดี

1.1.2.2 ความลำเอียงด้านเพศ

1.1.2.3 การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป

1.1.3 สัดส่วนของจำนวนของผู้ตรวจให้คะแนน

1.1.3.1 ในกรณีผู้ตรวจให้คะแนนลำเอียง

1.1.3.1.1 กรณีผู้ตรวจให้คะแนนลำเอียง 2 คน

1.1.3.1.1.1 ความเข้มงวด/ใจดีของผู้ตรวจให้คะแนน

1.1.3.1.1.1.1 เข้มงวด 0 คน : ใจดี 2 คน

1.1.3.1.1.1.2 เข้มงวด 1 คน : ใจดี 1 คน

1.1.3.1.1.1.3 เข้มงวด 2 คน : ใจดี 0 คน

1.1.3.1.1.2 ความลำเอียงด้านเพศของผู้ตรวจให้คะแนน

1.1.3.1.1.2.1 เข้าข้างเพศชาย 0 คน : เข้าข้างเพศหญิง 2 คน

1.1.3.1.1.2.2 เข้าข้างเพศชาย 1 คน : เข้าข้างเพศหญิง 1 คน

1.1.3.1.1.2.3 เข้าข้างเพศชาย 2 คน : เข้าข้างเพศหญิง 0 คน

1.1.3.1.1.3 การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป
คงตัว 0 คน : ไม่คงตัว 2 คน

1.1.3.1.2 กรณีผู้ตรวจให้คะแนนลำเอียง 3 คน

1.1.3.1.2.1 ความเข้มงวด/ใจดีของผู้ตรวจให้คะแนน

1.1.3.1.2.1.1 เข้มงวด 0 คน : ใจดี 3 คน

1.1.3.1.2.1.2 เข้มงวด 1 คน : ใจดี 2 คน

1.1.3.1.2.1.3 เข้มงวด 2 คน : ใจดี 1 คน

1.1.3.1.2.1.4 เข้มงวด 3 คน : ใจดี 0 คน

1.1.3.1.2.3 ความลำเอียงด้านเพศของผู้ตรวจให้คะแนน

1.1.3.1.2.3.1 เข้าข้างเพศชาย 0 คน : เข้าข้างเพศหญิง 3 คน

1.1.3.1.2.3.2 เข้าข้างเพศชาย 1 คน : เข้าข้างเพศหญิง 2 คน

- 1.1.3.1.2.3.3 เข้าข้างเพศชาย 2 คน : เข้าข้างเพศหญิง 1 คน
 1.1.3.1.2.3.4 เข้าข้างเพศชาย 3 คน : เข้าข้างเพศหญิง 0 คน
 1.1.3.1.2.4 การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป
 คงตัว 0 คน : ไม่คงตัว 3 คน

1.1.3.2 ในกรณีที่ผู้ตรวจให้คะแนนเป็นกลาง

- 1.1.3.2.1 กรณีที่ผู้ตรวจให้คะแนนเป็นกลาง 2 คน
 1.1.3.2.1.1 เป็นกลาง ไม่เข้มงวด/ใจดี 2 คน
 1.1.3.2.1.2 เป็นกลาง ไม่เข้าข้างเพศชาย/เข้าข้างเพศหญิง 2 คน
 1.1.3.2.1.3 เป็นกลาง คงตัว 2 คน
 1.1.3.2.2 กรณีที่ผู้ตรวจให้คะแนนเป็นกลาง 3 คน
 1.1.3.2.2.1 เป็นกลาง ไม่เข้มงวด/ใจดี 3 คน
 1.1.3.2.2.2 เป็นกลาง ไม่เข้าข้างเพศชาย/เข้าข้างเพศหญิง 3 คน
 1.1.3.2.2.3 เป็นกลาง คงตัว 3 คน

1.2 ตัวแปรตาม

ความเที่ยงตรงของการวัดความสามารถในการเขียนของนักเรียน

1. ความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย
2. ความเที่ยงตรงตามสภาพ

2. ประชากรและกลุ่มตัวอย่าง

2.1 ประชากร

ประชากรที่ใช้ในการวิจัยมี 2 กลุ่ม คือ

1. ความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 ภาคการเรียนที่ 2 ปีการศึกษา 2553 ของโรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน เพื่อใช้เป็นผู้ทดสอบในการสร้างและพัฒนาคุณภาพของเครื่องมือวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 จำนวน 16,968 คน

2. ผู้ตรวจให้คะแนนจากนักศึกษาคณะศึกษาศาสตร์ที่เรียนวิชาเอกภาษาไทย ชั้นปีที่ 4 ภาคการเรียนที่ 1 ปีการศึกษา 2554 ในสถาบันอุดมศึกษา สังกัดสำนักงานคณะกรรมการอุดมศึกษา เพื่อใช้เป็นผู้ตรวจให้คะแนนความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 จำนวน 402 คน

2.2 กลุ่มตัวอย่าง

กลุ่มตัวอย่างที่ใช้ในการวิจัยมี 2 กลุ่ม คือ

1. ความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 ภาคการเรียนที่ 2 ปีการศึกษา 2553 ของโรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน สำนักงานเขตพื้นที่การศึกษาประถมศึกษาสุรินทร์ เขต 1 ที่ได้มาจากการสุ่มแบบหลายขั้นตอน จากนักเรียนที่เข้าสอบ

การอ่านออกเสียง การเขียน และการคิดคำนวณ ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 ในวันที่ 23 กุมภาพันธ์ 2554 จำนวน 50 คน

2. ผู้ตรวจให้คะแนนจากนักศึกษาครุศึกษาศาสตร์ ที่เรียนวิชาเอกภาษาไทย ชั้นปีที่ 4 ภาคการเรียนที่ 1 ปีการศึกษา 2555 ในมหาวิทยาลัยราชภัฏสุรินทร์ บุรีรัมย์ และร้อยเอ็ด ที่ได้มาจากการสุ่มอย่างง่าย จำนวน 130 คน

ข้อจำกัดสำหรับการวิจัย

เนื่องจากในขั้นตอนแรกของการประมาณค่าพารามิเตอร์ คือต้องทดสอบความเหมาะสมกลมกลืนของข้อมูลเชิงประจักษ์กับโมเดล Partial Credit Model โดยพิจารณาค่า Weighted mean square residual : INFIT และ Unweighted mean square residual : OUTFIT ที่กำหนดขอบเขตของการยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลไว้ และในการเก็บข้อมูลเพื่อการวิจัยในครั้งนี้เป็นการเก็บข้อมูลภายใต้สถานการณ์การตรวจให้คะแนนจริง ดังนั้นการที่ข้อมูลจะเป็นไปตามข้อตกลงเบื้องต้นของโมเดล จึงจำเป็นต้องใช้กลุ่มตัวอย่างขนาดใหญ่ เพื่อให้การประมาณค่าพารามิเตอร์มีความถูกต้อง (Clauser; & Mazor. 1998: 34) แต่เนื่องจากครูที่สอนวิชาภาษาไทยนั้นมีจำนวนน้อยและกระจายอยู่ตามโรงเรียน ยกแก่การจัดอบรมและการเก็บรวบรวมข้อมูล ดังนั้นผู้วิจัยจึงทดแทนโดยใช้นักศึกษาชั้นปีที่ 4 คณะครุศาสตร์ วิชาเอกภาษาไทย ซึ่งผ่านการเรียนการสอนที่เน้นการใช้ภาษาไทยมาแล้วเป็นเวลา 4 ปี จนมีความเชี่ยวชาญด้านการใช้ภาษาไทย และจะต้องฝึกงานเป็นครูภาษาไทยในเวลาอันใกล้ มาเป็นผู้ตรวจให้คะแนนความสามารถในการเขียนของนักเรียน

นิยามศัพท์เฉพาะ

1. ความเที่ยงตรงของการวัดความสามารถในการเขียนของนักเรียน หมายถึง ความสอดคล้องหรือความสัมพันธ์ระหว่างคะแนนที่ได้จากการวัดกับสิ่งที่ต้องการวัด โดยพิจารณาจาก

1.1 มีความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย การตรวจสอบความเที่ยงตรงนั้น ทำได้โดยพิจารณาค่าสหสัมพันธ์เมื่อใช้วิธีการตรวจ และคุณลักษณะที่แตกต่างกัน โดยที่ความเที่ยงตรงเชิงเหมือน ควรมีค่าสูง และความเที่ยงตรงเชิงจำแนก ควรมีค่าต่ำ

1.2 ความเที่ยงตรงตามสภาพภายใต้การใช้วิธีการตรวจให้คะแนนแตกต่างกัน ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนนที่มีคุณลักษณะแตกต่างกัน ซึ่งผู้วิจัยใช้การแสดงหลักฐานความเที่ยงตรงตามสภาพโดยใช้สูตรสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson Product Moment Correlation Coefficient) เพื่อหาค่าสหสัมพันธ์ r จากนั้นแปลงสหสัมพันธ์ r เป็น z' แล้วจึงนำเปรียบเทียบ เพื่อหาว่า ภายใต้เงื่อนไข การใช้วิธีการตรวจให้คะแนนแตกต่างกัน ในสัดส่วนของจำนวนของผู้ให้คะแนนที่มีคุณลักษณะผู้ให้คะแนนแตกต่างกัน เงื่อนไขใดที่จะทำให้สหสัมพันธ์ระหว่างคะแนนการเขียนเรียงความ และคะแนนจากการเขียนจดหมายของนักเรียน มีค่ามากที่สุด แบ่งออกเป็น 2 กรณีคือ ในกรณีที่ผู้ตรวจมีคุณลักษณะลำเอียง และในกรณีที่ผู้ตรวจมีคุณลักษณะเป็นกลาง

2. การเขียนเรียงความ หมายถึง การเขียนร้อยแก้วที่เรียบเรียงข้อความเป็นเรื่องราว โดยอาศัยความรู้ ความเข้าใจ ความคิดเห็น หรือประสบการณ์ของผู้เขียนที่เกี่ยวข้องกับเรื่องนั้นๆ ถ่ายทอดออกมาเพื่อแสดงความสามารถของผู้เขียน วัดโดยใช้ข้อสอบการเขียนเรียงความในส่วนของ การเขียนตามจินตนาการ จากข้อสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 แล้วตรวจให้คะแนนด้วยวิธีการตรวจให้คะแนน 3 รูปแบบที่ผู้วิจัยสร้างขึ้น

3. การเขียนจดหมาย หมายถึง ข้อความที่เขียนขึ้นเพื่อเล่าเรื่องประสบการณ์ของผู้เขียนที่เกี่ยวข้องกับเรื่องนั้นๆ ถ่ายทอดออกมาเพื่อส่งสารโต้ตอบจากบุคคลหนึ่งไปยังบุคคลหนึ่ง วัดโดยใช้แบบทดสอบการเขียนจดหมายที่ผู้วิจัยสร้างขึ้น แล้วตรวจให้คะแนนด้วยวิธีการตรวจให้คะแนน 3 รูปแบบ

4. วิธีการตรวจให้คะแนน หมายถึง เกณฑ์การตรวจให้คะแนนความสามารถในการเขียนของนักเรียน แบ่งออกเป็น 3 รูปแบบ คือ

4.1 แบบแยกองค์ประกอบ (Analytic Scoring Rubrics) คือ การให้คะแนนงานชิ้นใดชิ้นหนึ่ง โดยมีการแยกองค์ประกอบของการให้คะแนน และอธิบายคุณภาพของงานในแต่ละองค์ประกอบเป็นระดับ

4.2 แบบรวมองค์ประกอบ (Holistic Scoring Rubrics) คือ การให้คะแนนงานชิ้นใดชิ้นหนึ่ง โดยดูภาพรวมของชิ้นงานว่ามีความเข้าใจในความคิดรวบยอด การสื่อความหมาย กระบวนการที่ใช้ และผลงาน แล้วเขียนอธิบายคุณภาพของงาน หรือความสำเร็จของงานเป็นชิ้นๆ

4.3 แบบผสมผสาน (Annotated Scoring Rubrics) เป็นการรวมเกณฑ์การตรวจให้คะแนนแบบรวมองค์ประกอบ และแบบแยกองค์ประกอบไว้ด้วยกัน เริ่มด้วยการให้คะแนนในภาพรวมด้วยเกณฑ์การให้คะแนนแบบรวมองค์ประกอบ แล้วเลือกให้คะแนนอีกเพียงบางคุณลักษณะแบบแยกองค์ประกอบ

5. คุณลักษณะของผู้ตรวจให้คะแนน หมายถึง คุณลักษณะของผู้ตรวจให้คะแนนเรียงความที่ได้จากการคำนวณจากโปรแกรม FACETS 3.70.1 ของ Linacre John M. (2012)

5.1 ความเข้มงวด/ใจดี (Severity/Leniency) หมายถึง ระบุความเข้มงวด/ใจดี จากค่าศูนย์ (0) เครื่องหมายบวก (+) ลบ (-) ของค่า Logit โดยผู้วิจัยกำหนดให้ผู้ตรวจให้คะแนนที่เข้มงวดจะมีค่า Logit มากกว่า +1.00 ให้คะแนนเป็นกลาง $-1 < \text{Logit} < +1$ และให้คะแนนใจดีจะมีค่า Logit น้อยกว่า -1.00

5.2 ความลำเอียงด้านเพศ (Differential Rater Functioning: DRF) หมายถึง ระบุความลำเอียงด้านเพศของผู้ตรวจให้คะแนน จากค่า Z-Score ที่มีนัยสำคัญทางสถิติ ถ้า Z-Score ไม่มีนัยสำคัญทางสถิติ แสดงว่าเป็นกลาง สุดทำวิจัยพิจารณาว่าผู้ตรวจให้คะแนนคนนั้นๆ ลำเอียงเข้าข้างเพศชายหรือเพศหญิง คะแนนที่มีแนวโน้มใจดี (Logit ต่ำกว่า) กับเพศใดแสดงว่าลำเอียงเข้าข้างเพศนั้น

5.3 การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป (Differential Rater Functioning Over Time : DRIFT) หมายถึง ระบุการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป จากค่า

Z-Score โดยที่ค่า Z-Score ที่มีนัยสำคัญทางสถิติแสดงว่าผู้ตรวจให้คะแนนให้คะแนนไม่คงที่ ถ้า Z-Score ไม่มีนัยสำคัญทางสถิติแสดงว่าผู้ตรวจให้คะแนนให้คะแนนคงที่

6. สัดส่วนของจำนวนของผู้ตรวจให้คะแนน หมายถึง จำนวนผู้ตรวจให้คะแนนแบ่งสัดส่วนตามคุณลักษณะของผู้ตรวจให้คะแนน แบ่งออกเป็น 2 กรณี คือ กรณีผู้ตรวจให้คะแนนลำเอียง และกรณีผู้ตรวจให้คะแนนเป็นกลาง

6.1 ในกรณีผู้ตรวจให้คะแนนลำเอียง

6.1.1 กรณีผู้ตรวจให้คะแนนลำเอียง 2 คน

6.1.1.1 ความเข้มงวด/ใจดีของผู้ตรวจให้คะแนน

6.1.1.1.1 เข้มงวด 0 คน : ใจดี 2 คน

6.1.1.1.2 เข้มงวด 1 คน : ใจดี 1 คน

6.1.1.1.3 เข้มงวด 2 คน : ใจดี 0 คน

6.1.1.2 ความลำเอียงด้านเพศของผู้ตรวจให้คะแนน

6.1.1.2.1 เข้าข้างเพศชาย 0 คน : เข้าข้างเพศหญิง 2 คน

6.1.1.2.2 เข้าข้างเพศชาย 1 คน : เข้าข้างเพศหญิง 1 คน

6.1.1.2.3 เข้าข้างเพศชาย 2 คน : เข้าข้างเพศหญิง 0 คน

6.1.1.3 การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป

คงตัว 0 คน : ไม่คงตัว 2 คน

6.1.2 กรณีผู้ตรวจให้คะแนนลำเอียง 3 คน

6.1.2.1 ความเข้มงวด/ใจดีของผู้ตรวจให้คะแนน

6.1.2.1.1 เข้มงวด 0 คน : ใจดี 3 คน

6.1.2.1.2 เข้มงวด 1 คน : ใจดี 2 คน

6.1.2.1.3 เข้มงวด 2 คน : ใจดี 1 คน

6.1.2.1.4 เข้มงวด 3 คน : ใจดี 0 คน

6.1.2.2 ความลำเอียงด้านเพศของผู้ตรวจให้คะแนน

6.1.2.2.1 เข้าข้างเพศชาย 0 คน : เข้าข้างเพศหญิง 3 คน

6.1.2.2.2 เข้าข้างเพศชาย 1 คน : เข้าข้างเพศหญิง 2 คน

6.1.2.2.3 เข้าข้างเพศชาย 2 คน : เข้าข้างเพศหญิง 1 คน

6.1.2.2.4 เข้าข้างเพศชาย 3 คน : เข้าข้างเพศหญิง 0 คน

6.1.2.3 การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป

คงตัว 0 คน : ไม่คงตัว 3 คน

6.2 ในกรณีที่ผู้ตรวจให้คะแนนเป็นกลาง

6.2.1 กรณีที่ผู้ตรวจให้คะแนนเป็นกลาง 2 คน

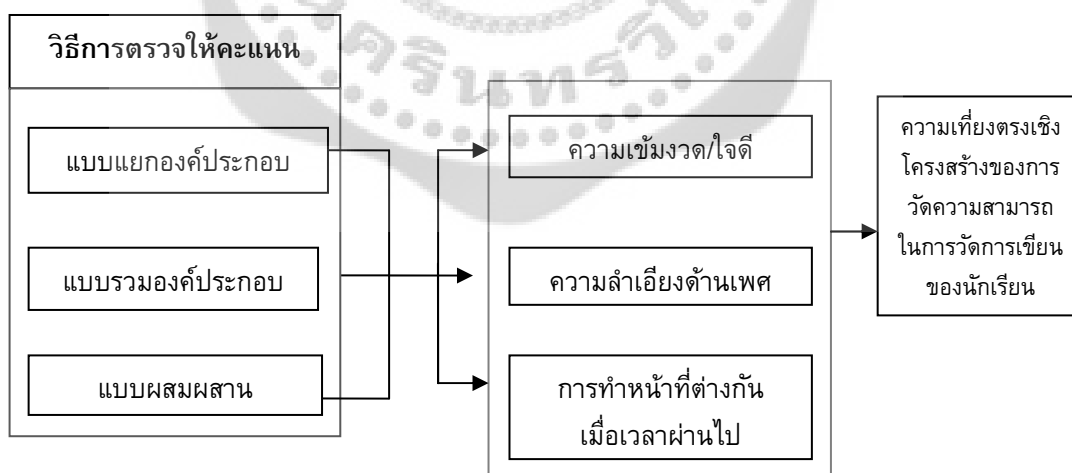
6.2.1.1 เป็นกลาง ไม่เข้มงวด/ใจดี 2 คน

- 6.2.1.2 เป็นกลาง ไม่เข้าข้างเพศชาย/เข้าข้างเพศหญิง 2 คน
- 6.2.1.3 เป็นกลาง คงตัว 2 คน
- 6.2.2 กรณีที่ผู้ตรวจให้คะแนนเป็นกลาง 3 คน
 - 6.2.2.1 เป็นกลาง ไม่เข้มงวด/ใจดี 3 คน
 - 6.2.2.2 เป็นกลาง ไม่เข้าข้างเพศชาย/เข้าข้างเพศหญิง 3 คน
 - 6.2.2.3 เป็นกลาง คงตัว 3 คน

กรอบแนวคิดในการวิจัย

ในงานวิจัยชิ้นนี้ผู้วิจัยต้องการแสดงหลักฐานความเที่ยงตรง 2 ประเภท คือ ความเที่ยงตรงเชิงโครงสร้าง และความเที่ยงตรงตามสภาพ ในที่นี้จะแยกการนำเสนอกรอบแนวคิดออกเป็น 2 กรอบจำแนกตามชนิดของความเที่ยงตรง

จุดประสงค์หลักของการศึกษาความเที่ยงตรงเชิงโครงสร้าง คือ การศึกษาพฤติกรรมของผู้ตรวจให้คะแนนจากภาวะเงื่อนไขที่กำหนด เพื่อความเข้าใจในโครงสร้างการวัดและสามารถนำผลนั้นไปพัฒนาในลำดับต่อไป การวิจัยนี้ได้แบ่งตัวแปรออกเป็น 2 ส่วน ส่วนแรก คือ วิธีการตรวจให้คะแนน ซึ่งมีอยู่ 3 รูปแบบ คือ 1) แบบแยกองค์ประกอบ 2) แบบรวมองค์ประกอบ และ 3) แบบผสมผสาน และส่วนที่ 2 คือ คุณลักษณะของผู้ตรวจให้คะแนน ซึ่งมีอยู่ 3 คุณลักษณะ คือ ความเข้มงวด/ใจดี ความลำเอียงด้านเพศ และการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป ซึ่งตัวแปร 2 ชนิดนี้น่าจะส่งผลต่อความเที่ยงตรงเชิงโครงสร้างของการวัดความสามารถในการเขียนของนักเรียน



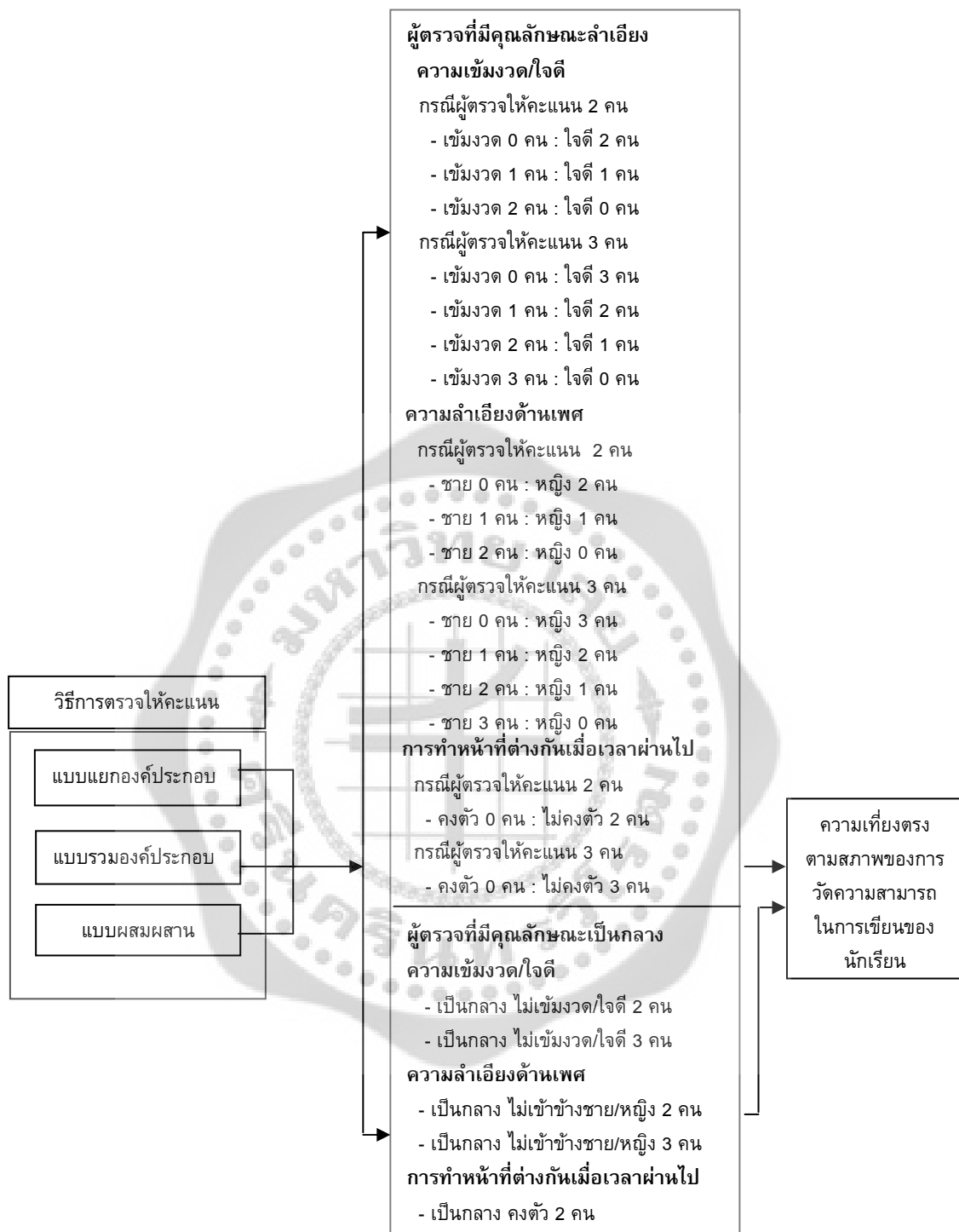
ภาพประกอบ 1 กรอบแนวคิดในการศึกษาความเที่ยงตรงเชิงโครงสร้าง

จากภาพประกอบ 1 ผู้วิจัยต้องการศึกษาความเที่ยงตรงเชิงโครงสร้างของการวัดความสามารถในการเขียนของนักเรียน โดยตรวจสอบว่ามีความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย

หรือไม่ โดยพิจารณาว่ามีความเที่ยงตรงเชิงเหมือน และมีความเที่ยงตรงเชิงจำแนก ผู้วิจัยกำหนดให้ใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน 3 รูปแบบ ซึ่งวิธีการตรวจทั้ง 3 รูปแบบนี้จะวิเคราะห์ด้วยโมเดลเชิงคณิตศาสตร์ที่แตกต่างกันตามคุณลักษณะของผู้ตรวจให้คะแนน ซึ่งจำแนกออกเป็นความเข้มงวด/ใจดี ความลำเอียงด้านเพศ และการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป

จุดประสงค์หลักของการศึกษาความเที่ยงตรงตามสภาพคือ การศึกษาความสัมพันธ์ระหว่างคะแนนคะแนนการเขียนเรียงความ และคะแนนจากการเขียนจดหมายของนักเรียน ซึ่งทดสอบในเวลาใกล้เคียงกัน ภายใต้การใช้วิธีการตรวจให้คะแนนแตกต่างกัน ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนนที่มีคุณลักษณะแตกต่างกัน แบ่งออกเป็น 2 กรณีคือ ในกรณีที่ผู้ตรวจมีคุณลักษณะลำเอียง และในกรณีที่ผู้ตรวจมีคุณลักษณะเป็นกลาง





ภาพประกอบ 2 กรอบแนวคิดในการศึกษาความเที่ยงตรงตามสภาพ

จากภาพประกอบ 2 ผู้วิจัยต้องการศึกษาความเที่ยงตรงตามสภาพของการวัดความสามารถในการเขียนเรียงความของนักเรียน กรณีผู้ตรวจให้คะแนนมีคุณลักษณะลำเอียง และกรณีผู้ตรวจให้คะแนนมีคุณลักษณะเป็นกลาง โดยพิจารณาผลที่ได้จากการตรวจให้คะแนนด้วยวิธีการตรวจให้คะแนนที่แตกต่างกัน 3 รูปแบบ ทั้งนี้ผลที่ได้จากการตรวจให้คะแนนอาจจะยังขึ้นอยู่กับสัดส่วนของจำนวนของผู้ตรวจให้คะแนนที่มีคุณลักษณะผู้ตรวจให้คะแนนแตกต่างกัน โดยในกรณีผู้ตรวจให้คะแนนมีคุณลักษณะลำเอียงนั้น สัดส่วนของจำนวนของผู้ตรวจให้คะแนนที่มีคุณลักษณะผู้ตรวจให้คะแนนแตกต่างกัน มีดังนี้คือ ความเข้มงวด/ใจดี ในกรณีผู้ตรวจให้คะแนน 2 คน (คือ เข้มงวด 0 คน : ใจดี 2 คน, เข้มงวด 1 คน : ใจดี 1 คน, เข้มงวด 2 คน : ใจดี 0 คน) และในกรณีผู้ตรวจให้คะแนน 3 คน (คือ เข้มงวด 0 คน : ใจดี 3 คน, เข้มงวด 1 คน : ใจดี 2 คน, เข้มงวด 2 คน : ใจดี 1 คน, เข้มงวด 3 คน : ใจดี 0 คน) ความลำเอียงด้านเพศ ในกรณีผู้ตรวจให้คะแนน 2 คน (คือ เข้าข้างเพศชาย 0 คน : เข้าข้างเพศหญิง 2 คน, เข้าข้างเพศชาย 1 คน : เข้าข้างเพศหญิง 1 คน, เข้าข้างเพศชาย 2 คน : เข้าข้างเพศหญิง 0 คน) และในกรณีผู้ตรวจให้คะแนน 3 คน (คือ เข้าข้างเพศชาย 0 คน : เข้าข้างเพศหญิง 3 คน, เข้าข้างเพศชาย 1 คน : เข้าข้างเพศหญิง 2 คน, เข้าข้างเพศชาย 2 คน : เข้าข้างเพศหญิง 1 คน, เข้าข้างเพศชาย 3 คน : เข้าข้างเพศหญิง 0 คน) และการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป ในกรณีผู้ตรวจให้คะแนน 2 คน (คือ คงตัว 0 คน : ไม่คงตัว 2 คน) และในกรณีผู้ตรวจให้คะแนน 3 คน (คือ คงตัว 0 คน : ไม่คงตัว 3 คน) และในกรณีผู้ตรวจให้คะแนนมีคุณลักษณะเป็นกลาง สัดส่วนของจำนวนของผู้ตรวจให้คะแนนที่มีคุณลักษณะผู้ตรวจให้คะแนนแตกต่างกัน มีดังนี้คือ ในกรณีผู้ตรวจให้คะแนน 2 คน (ไม่เข้มงวด/ใจดี 2 คน, ไม่เข้าข้างเพศชาย/เข้าข้างเพศหญิง 2 คน, คงตัว 2 คน) และในกรณีผู้ตรวจให้คะแนน 3 คน (ไม่เข้มงวด/ใจดี 3 คน, ไม่เข้าข้างเพศชาย/เข้าข้างเพศหญิง 3 คน, คงตัว 3 คน) ซึ่งในแต่ละเงื่อนไขการใช้วิธีการตรวจให้คะแนนแตกต่างกัน ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนนที่มีคุณลักษณะผู้ตรวจให้คะแนนแตกต่างกันนั้น น่าจะมีค่าความเที่ยงตรงตามสภาพของการวัดความสามารถในการเขียนของนักเรียนที่แตกต่างกัน ตามบริบทของเกณฑ์ในการเปรียบเทียบที่ผู้วิจัยนำมาใช้ คือ คะแนนจากการเขียนเรียงความ และคะแนนจากการเขียนจดหมายที่วัดในเวลาใกล้เคียงกัน

บทที่ 2

เอกสารและงานวิจัยที่เกี่ยวข้อง

การวิจัยครั้งนี้ มีความมุ่งหมายเพื่อสร้างและพัฒนาคุณภาพของเครื่องมือวัดความสามารถในการเขียนของนักเรียน เพื่อค้นหาคุณลักษณะของผู้ตรวจให้คะแนน คือ ความเข้มงวด/ใจดี ความลำเอียงทางเพศ และการทำหน้าที่ต่างกันของผู้ให้คะแนนเมื่อเวลาผ่านไป และเพื่อหาความเที่ยงตรงของการวัดความสามารถในการเขียนของนักเรียน ดังนั้นเพื่อให้เอกสารและงานวิจัยที่เกี่ยวข้องครอบคลุมขอบเขตของการวิจัย ผู้วิจัยจึงแบ่งการนำเสนอออกเป็น 4 ตอน คือ ตอนที่ 1 การเขียน ตอนที่ 2 คุณลักษณะของผู้ตรวจให้คะแนน ตอนที่ 3 ความเที่ยงตรง และตอนที่ 4 งานวิจัยที่เกี่ยวข้อง โดยมีรายละเอียดดังต่อไปนี้

ตอนที่ 1 การเขียน

ในตอนที่ผู้วิจัยได้นำเสนอเอกสารที่เกี่ยวข้องกับการเขียนไว้ดังนี้ คือ การเขียนเรียงความ ความหมาย ความสำคัญของการเขียนเรียงความ ปัญหาที่พบจากการตรวจให้คะแนนการเขียนเรียงความ การเขียนจดหมาย ความหมาย และประเภทของการเขียนจดหมาย การสร้างข้อสอบการวัดความสามารถในการเขียน และโมเดลการวัดความสามารถในการตรวจให้คะแนนการเขียน ดังนี้

1.1 การเขียนเรียงความ

การเขียนเรียงความ หมายถึง การนำย่อหน้า ตั้งแต่ 2 ย่อหน้าขึ้นไปมาเขียนเรียงต่อกัน โดยที่ย่อหน้าต่างๆ ที่นำมาประกอบเรียงเป็นเรียงความนั้นจะต้องมีความเกี่ยวข้องและสัมพันธ์ซึ่งกันและกัน (เปลื้อง ณ นคร. 2512; ชัยนันท์ นันทพันธ์. 2526; ชำนาญ สุภนิตย์; และ สุทิน พูลสวัสดิ์. 2547; นเรศ สุรสิทธิ์. 2550; และ Tragam Kalchayanant. 2552) นอกจากนี้เรียงความ เป็นการเขียนร้อยแก้วที่เรียบเรียงข้อความเป็นเรื่องราวถ่ายทอดออกมาเป็นภาษาเขียน โดยใช้กลวิธีทั้งทางด้านการเขียนและภาษาให้เหมาะสมกับเนื้อหา ตลอดจนการเลือกสรรถ้อยคำ สำนวนโวหาร และการใช้ภาษาให้ถูกต้อง (ถวัลย์ มาตรฐาน. 2545; ประพนธ์ เรื่องณรงค์. 2545; ผจงวาด พูลแก้ว; พัชณี ปี่เพราะ; และ สุวรรณี โสประดิษฐ์. 2551; และ ชูลีพร ดัดงาม. 2552) เพื่อแสดงความรู้ ความคิด ความรู้สึกของผู้เขียนที่เขียนถึงบุคคล สถานที่ เรื่อง หรือสิ่งใดสิ่งหนึ่ง โดยอาศัยความรู้ ความเข้าใจ ความคิดเห็น หรือประสบการณ์ของผู้เขียนที่เกี่ยวข้องกับเรื่องนั้นๆ โดยที่ผู้เขียนต้องการถ่ายทอดความรู้ ความคิด และความรู้สึกมายังผู้อ่าน (เปลื้อง ณ นคร. 2512; ถวัลย์ มาตรฐาน. 2545; ประพนธ์ เรื่องณรงค์. 2545; ผจงวาด พูลแก้ว; พัชณี ปี่เพราะ; และ สุวรรณี โสประดิษฐ์. 2551; กระทรวงศึกษาธิการ. 2552ก; และ ชูลีพร ดัดงาม. 2552)

ดังนั้น การเขียนเรียงความ หมายถึง การเขียนร้อยแก้วที่เรียบเรียงข้อความเป็นเรื่องราวที่เขียนถึงบุคคล สถานที่ เรื่อง หรือสิ่งใดสิ่งหนึ่ง โดยอาศัยความรู้ ความเข้าใจ ความคิดเห็น หรือประสบการณ์ของผู้เขียนที่เกี่ยวข้องกับเรื่องนั้นๆ ถ่ายทอดออกมาเพื่อแสดงความสามารถของผู้เขียน

ความสำคัญของการเขียนเรียงความ การเขียนเรียงความมีความสำคัญอย่างมาก นอกจากเป็นเครื่องวัดความรู้ความสามารถของนักเรียน (เปลื้อง ณ นคร. 2512) แล้ว การเขียนเรียงความ ยังเป็นวิธีการขั้นพื้นฐานที่สำคัญยิ่งของการใช้ภาษาไทย ซึ่งสัญลักษณ์การเขียนของคนไทยก็คืออักษร ไทย ดังนั้นการเขียนเรียงความ จึงเป็นส่วนหนึ่งที่ใช้ในการเรียนในโรงเรียนทุกระดับ (ชัยนันท์ นันทพันธ์. 2526: 1; กรมวิชาการ. 2546: 21 – 22; และ ชำนาญ สุภนิตย์; และ สุทิน พูลสวัสดิ์. 2547) การเขียนเรียงความเป็นเครื่องมือสำคัญในการสื่อสารระหว่างผู้เขียนกับผู้อ่าน โดยใช้สัญลักษณ์ที่กำหนดรู้ร่วมกัน เป็นสื่อกลางของการสร้างความเข้าใจถูกต้องตรงกัน (ถวัลย์ มาศจรัส. 2545; และ กรมวิชาการ. 2546: 21 – 22) การที่จะทำให้ผู้รับสารหรือผู้อ่านเกิดความเข้าใจถึงสิ่งที่ผู้เขียนต้องการสื่อสารอย่างถูกต้องแล้ว ผู้เขียนจะต้องจัดระเบียบความคิดของตนเองออกมาอย่างเป็นระบบ และต้องรู้จักเลือกใช้ถ้อยคำภาษาไทย ในการเขียนที่ถูกต้องชัดเจน เพื่อสื่อความหมายที่ตนเองต้องการอย่างครบถ้วนสมบูรณ์ การเขียนโดยใช้ภาษาที่คลุมเครือมีผลทำให้ผู้อ่านตีความหมายผิดพลาดได้ ซึ่งจะเห็นได้จากภาษากฎหมายหลายๆ ฉบับ ที่จะต้องให้ทฤษฎีหรือศาลรัฐธรรมนูญอ่านแล้วตีความใหม่ เพื่อหาข้อยุติสำหรับใช้เป็นบรรทัดฐาน ในการอธิบายความหมายกันต่อไป (ถวัลย์ มาศจรัส. 2545) และนอกจากนี้ในปัจจุบันการเขียนเรียงความ เป็นสิ่งจำเป็นอย่างยิ่งสำหรับนักเรียนนักศึกษา ในการสอบเข้าศึกษาต่อในวิทยาลัยระดับปริญญาตรี หรือบัณฑิตวิทยาลัยระดับปริญญาโท ตลอดจนการสอบชิงทุนเพื่อไปทำงานหรือศึกษาต่อของแต่ละหน่วยงาน โดยทั่วไปแล้วนักศึกษาต่างชาติที่ต้องการเข้าศึกษาต่อในวิทยาลัย หรือมหาวิทยาลัยในสหรัฐอเมริกา ทั้งระดับปริญญาตรีและปริญญาโท มีความจำเป็นอย่างยิ่งที่จะต้องเขียนเรียงความ ประกอบใบสมัครของวิทยาลัยหรือมหาวิทยาลัยนั้นๆ เพื่อประกอบการพิจารณาในการเข้าเรียน นอกจากนี้เมื่อมีการรับสมัครเข้าทำงานรับราชการหรือเอกชน ก็จะต้องทดสอบการใช้ภาษาไทยและภาษาอังกฤษ เพื่อการสื่อสาร การสอบวิชาชุดคุณระดับต่างๆ ก็วางระเบียบกำหนดให้วิชาเรียงความมีคะแนนมากที่สุดกว่า วิชาอื่นๆ ทั้งนี้งานอาชีพบางอย่างโดยเฉพาะงานราชการจำเป็นต้องอาศัยการเขียนในลักษณะต่างๆ อยู่มาก เพราะถ้ามีพื้นฐานทางภาษาดีแล้ว ก็จะสามารถนำหลักการเขียนเรียงความไปประยุกต์ในงานเขียนแบบต่างๆ ได้ เช่น การร่างหนังสือราชการ งานสารบรรณ การเขียนรายงาน เป็นต้น (ชัยนันท์ นันทพันธ์. 2526: 1; และ ชำนาญ สุภนิตย์; และ สุทิน พูลสวัสดิ์. 2547: 257)

การที่เด็กได้เขียนเรียงความเสมอๆ ย่อมช่วยให้เด็กได้ใช้วิชาให้เกิดประโยชน์แก่การเรียน ในสาขาวิชาอื่นๆ ด้วย (รองรัตน์ อิศรภักดี; และ เทือก กุสุมา ณ อยุธา. 2513: 106; กรมวิชาการ. 2546: 21 – 22; และ กระทรวงศึกษาธิการ. 2552ข) การเรียนการสอนภาษาไทยในปัจจุบันมิได้มุ่งหวังให้นักเรียนอ่านออกเขียนได้เพียงอย่างเดียว หากมุ่งหวังให้นักเรียนนำความรู้ ความสามารถไปใช้ให้เกิดประโยชน์ได้จริง (ถวัลย์ มาศจรัส. 2545; และ กรมวิชาการ. 2546: 21 – 22) ดังนั้นสาระการเรียนรู้ ที่นำมาใช้ในการจัดการเรียนรู้เพื่อให้ผู้เรียนบรรลุตามมาตรฐานที่กำหนด จึงมีองค์ประกอบที่สำคัญ 3 ด้าน คือ ความรู้ ทักษะกระบวนการ และคุณธรรมจริยธรรม ค่านิยม การกำหนดสาระการเรียนรู้นี้ ถือเป็นสาระแกนกลางที่ทุกคนจะต้องเรียน เนื่องจากได้กำหนดจากมาตรฐานการเรียนรู้ช่วงชั้นที่วิเคราะห์มาจากมาตรฐานการเรียนรู้ 12 ปี การกำหนดสาระการเรียนรู้ภาษาไทย ได้กำหนดตามสาระทั้ง 5 หัวข้อ ได้แก่ การอ่าน การเขียน การฟัง การดูและการพูด หลักการใช้ภาษา วรรณคดีและวรรณกรรมเพื่อให้

เห็นภาพรวมทั้งหมดว่ามีอะไรที่ต้องเรียนรู้บ้าง แต่ในการนำไปใช้จัดการเรียนการสอนจริง ครูผู้สอนต้องจัดในลักษณะบูรณาการ ซึ่งเป็นกระบวนการเรียนรู้ตามธรรมชาติของภาษาไทย การบูรณาการอาจทำได้หลายวิธี วิธีที่สะดวกและเหมาะสมมาก คือ การใช้วรรณคดีหรือวรรณกรรมเป็นแกนธรรมชาติภาษาไทยเป็นเรื่องของทักษะ ซึ่งจะแยกเนื้อหาสาระของทักษะแต่ละชั้นปีโดยเด็ดขาดไม่ได้ จำเป็นต้องมีกระบวนการฝึกทักษะต่างๆ การจับใจความ การเลือกใช้คำให้ตรงความหมาย การเขียนเรื่องแสดงความรู้ ความคิด ประสบการณ์ ความรู้สึก ความต้องการจินตนาการ การนำความรู้จากการอ่านไปใช้ในการตัดสินใจ การแก้ปัญหาและการดำเนินชีวิต ฯลฯ จำเป็นต้องสอนทุกชั้นเพราะเป็นเรื่องของทักษะของภาษา แต่ละชั้นจะมีเนื้อหาในการฝึกทักษะที่เพิ่มความซับซ้อนและยากมากขึ้น เช่น จำนวนคำเพิ่มขึ้น ประโยคที่ใช้จะยาวและซับซ้อนขึ้น เรื่องที่นำมาให้อ่านจะยาวมากขึ้น มีเหตุการณ์หลายเหตุการณ์ มีการฝึกคิดในระดับที่สูงขึ้นสำหรับภาษาที่ใช้มีหลากหลายรูปแบบ และใช้ภาษาอย่างมีศิลปะมากขึ้น เป็นต้น เมื่อครบช่วงชั้นแล้วนักเรียนควรได้เรียนครบถ้วนตามสาระที่กำหนด ขณะเดียวกันสถานศึกษาอาจนำสาระการเรียนรู้ในท้องถิ่นมาสอดแทรกหรือเพิ่มเติมได้ เพื่อให้การเรียนรู้มีประสิทธิภาพมากขึ้น กว้างขวางขึ้น และได้เรียนรู้เรื่องราวของท้องถิ่นด้วย (กรมวิชาการ. 2546: 21 – 22) ดังนั้นการเขียนจึงสำคัญอย่างยิ่งสำหรับทุกคนที่ต้องสื่อสารกันด้วยภาษาเขียนเป็นหลัก (ถวัลย์ มาศจรัส. 2545)

การทดสอบทุกรูปแบบมีหน้าที่เหมือนกันคือ มุ่งวัดผลสัมฤทธิ์ในการเรียนของนักเรียน แต่ในกรณีที่ใช้ข้อสอบเรียงความนั้นมีส่วนที่ดีที่เพิ่มจากแบบอื่นๆ หลายประการเช่น 1) ให้อิสระในการตอบ เปิดโอกาสให้ผู้สอบได้เสนอความคิด และรายละเอียดต่างๆ ได้อย่างเสรีภายใต้ขอบเขตที่กำหนดไว้ โดยไม่จำกัดทิศทางที่ผู้สอบกำหนดให้เท่านั้น 2) แก้ไขปัญหาการเดา ซึ่งในกรณีที่ผู้สอบไม่ทราบเรื่องนั้นๆ ย่อมไม่มีโอกาสในการทำข้อสอบถูก 3) สะดวกที่จะใช้กับผู้สอบจำนวนน้อยๆ 4) ประหยัดเวลาในการทำข้อสอบ 5) ส่งเสริมความคิดนอกเนกนัย (Divergent Thinking) โดยผู้สอบสามารถเสนอแนวคิดแปลกใหม่ และริเริ่มในแนวทางใหม่ๆ (ไพศาล หวังพานิช. 2545)

จากความสำคัญดังที่กล่าวมาข้างต้น ผู้วิจัยจึงสนใจที่จะใช้การเขียนเรียงความเป็นเครื่องมือที่ใช้ในการวิจัย เพราะนอกจากจะเป็นการพัฒนาการวัดความสามารถในการวัดความสามารถในการเขียนเรียงความของนักเรียนแล้ว ยังเป็นการพัฒนาความสามารถในการคิดผู้ตรวจให้คะแนนในอนาคตอีกด้วย

ปัญหาของการตรวจให้คะแนนเรียงความ ในส่วนของการจัดการเรียนการสอนที่นอกจากจะมุ่งเน้นคุณภาพโดยรวมของนักเรียนเป็นสำคัญแล้ว การเรียนการสอนเรียงความของครู ยังสามารถนำไปใช้ในการตัดสินผลการเรียนของกลุ่มสาระการเรียนรู้ การอ่าน คิดวิเคราะห์และเขียน คุณลักษณะอันพึงประสงค์ และกิจกรรมพัฒนาผู้เรียนอีกด้วย ผู้สอนต้องคำนึงถึงการพัฒนาผู้เรียนแต่ละคนเป็นหลัก และต้องเก็บข้อมูลของผู้เรียนทุกด้านอย่างสม่ำเสมอและต่อเนื่องในแต่ละภาคเรียน รวมทั้งสอนซ่อมเสริมผู้เรียนให้พัฒนาจนเต็มสมรรถภาพ ในระดับประถมศึกษานั้นผู้เรียนต้องมีเวลาเรียนไม่น้อยกว่าร้อยละ 80 ของเวลาทั้งหมด ต้องได้รับการประเมินทุกตัวชี้วัด และผ่านเกณฑ์ที่สถานศึกษากำหนด ต้องได้รับการตัดสินผลการเรียนทุกรายวิชา และต้องได้รับการประเมิน และมีผลการประเมินผ่านตามเกณฑ์ที่สถานศึกษากำหนด ในการอ่าน คิดวิเคราะห์และเขียน คุณลักษณะอันพึงประสงค์ และกิจกรรมพัฒนาผู้เรียน โดยที่ในเกณฑ์การจบการศึกษาระดับประถมศึกษา ผู้เรียนต้องมีผลการประเมินการอ่าน คิดวิเคราะห์ และเขียน

ในระดับผ่านเกณฑ์การประเมินตามที่สถานศึกษากำหนด และนอกจากนี้จากหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 กำหนดมาตรฐานการเรียนรู้ในกลุ่มสาระการเรียนรู้ภาษาไทยสาระที่ 2 การเขียน มาตรฐาน ท 2.1 ใช้กระบวนการเขียน เขียนสื่อสาร เขียนเรียงความ ย่อความ และเขียนเรื่องราวในรูปแบบต่างๆ เขียนรายงานข้อมูลสารสนเทศ และรายงานการศึกษาค้นคว้าอย่างมีประสิทธิภาพ (กรมวิชาการ, 2552ข) และนอกจากนี้การเลื่อนชั้นทั้งระดับประถมศึกษาและมัธยมศึกษา ถ้าผู้เรียนมีข้อบกพร่องเพียงเล็กน้อย และสถานศึกษาพิจารณาเห็นว่า สามารถพัฒนาและสอนซ่อมเสริมได้ให้อยู่ในดุลยพินิจของสถานศึกษาที่จะผ่อนผันให้เลื่อนชั้นได้ แต่หากผู้เรียนไม่ผ่านรายวิชาจำนวนมาก และมีแนวโน้มว่าจะเป็นปัญหาต่อการเรียนในระดับชั้นที่สูงขึ้น สถานศึกษาอาจตั้งคณะกรรมการพิจารณาให้เรียนซ้ำชั้นได้ ทั้งนี้ให้คำนึงถึงวุฒิภาวะและความรู้ความสามารถของผู้เรียนเป็นสำคัญ

ด้วยความจริงที่ว่า ไม่มีข้อสอบชนิดใดดีพร้อมทุกอย่าง ดังนั้นเป็นปัญหาของครูในฐานะที่เป็นผู้ต้องให้คะแนนความสามารถในการเขียนเรียงความของนักเรียน จึงต้องคำนึงถึงและแก้ไขข้อบกพร่องนั้น ให้มีน้อยที่สุด ข้อจำกัดที่สำคัญในการใช้ข้อสอบเรียงความ (ไพศาล หวังพานิช. 2545) คือ 1) ตรวจยาก การตรวจให้คะแนนถือว่ามีที่ยากกว่าแบบทดสอบอื่นๆ สาเหตุสำคัญของปัญหานี้ มีหลายกรณี เช่น คำตอบของผู้สอบแตกต่างกันไปตามความคิดของแต่ละคน ซึ่งผู้ออกข้อสอบ อาจไม่ได้กำหนดขอบเขตของการตอบไว้อย่างชัดเจนเพียงพอ ผู้ตรวจให้คะแนนไม่มีเกณฑ์การตรวจให้คะแนนที่แน่นอน หรือผู้ตรวจให้คะแนนมีภาวะทางอารมณ์ที่ส่งผลกระทบต่อผลการตรวจให้คะแนนได้ง่าย 2) ถ้ามได้จำกัดเนื่องจากข้อสอบแบบนี้ผู้สอบต้องใช้เวลาในการคิดและเขียน เป็นเหตุให้ไม่สามารถใช้คำถามได้มากข้อ ต้องจำกัดการถามเฉพาะในบางเนื้อหาเท่านั้น ลักษณะเช่นนี้อาจทำให้การวัดไม่ครอบคลุมครบถ้วน 3) ต้องใช้เวลาในการตอบและการตรวจให้คะแนน ผู้สอบต้องใช้เวลามากสำหรับคิดและเขียน และผลจากการตอบเหล่านี้ ครูผู้ตรวจให้คะแนนต้องใช้เวลาในการตรวจมากพอสมควรและผลที่ได้มักไม่มีความเป็นปรนัย 4) ก่อให้เกิดการเขียนที่วกวน แม้ว่าจะขจัดปัญหาในการเอาใจได้ แต่ไม่สามารถป้องกันผู้สอบบางคนที่ยินยอมการตอบววกวนไม่ตรงจุดหรือการตอบแบบครอบจักรวาล ซึ่งเกิดจากที่ผู้สอบคิดว่า ขอให้เขียนอะไรก็ได้บางอย่างลงไปดีกว่าไม่เขียนอะไรเลย 5) คะแนนมักจะขึ้นอยู่กับความสามารถในการเขียน ซึ่งจะพบว่าความสามารถในการเขียนของผู้สอบมักมีความสัมพันธ์กันสูงกับคุณภาพสาระที่เขียน ดังนั้นคะแนนส่วนหนึ่งมักจะได้จากความสามารถทางภาษาของผู้สอบ

นอกจากปัญหาที่เกิดจากครูในการเรียนการสอนในโรงเรียน เพื่อให้เป็นไปตามมาตรฐานและปัญหาจากการใช้ข้อสอบเรียงความดังที่กล่าวในข้างต้นแล้ว ปัญหาอีกส่วนหนึ่งคือการตรวจให้คะแนนเรียงความโดยครูผู้สอนนั่นเอง ซึ่งได้มีผู้กล่าวถึงปัญหาจากความคลาดเคลื่อนของการตรวจให้คะแนนเรียงความไว้มากมายหลายท่าน ผู้วิจัยได้แบ่งออกเป็น 2 ส่วนใหญ่ๆ คือ ความคลาดเคลื่อนที่มาจากภายนอกผู้ตรวจให้คะแนน และความคลาดเคลื่อนที่มาจากภายในผู้ตรวจให้คะแนน

1. ความคลาดเคลื่อนที่มีแหล่งที่มาจากภายนอกผู้ตรวจให้คะแนนซึ่งเป็นสิ่งที่นอกเหนือการควบคุมของผู้ให้คะแนน เช่น เกณฑ์การตรวจให้คะแนน ความยากของเรียงความ ความสามารถของผู้สอบ เป็นต้น อาจเกิดจากตัวนักเรียนเอง เป็นต้นว่า ลายมือ ความถูกต้องของการเขียนตัวสะกด การันต์ การเรียบเรียงถ้อยคำ สำนวนภาษา หลักไวยากรณ์ สิ่งเหล่านี้ มีผลต่อการตรวจให้คะแนน

เป็นอย่างมาก ซึ่งครูจะต้องชี้แจงให้นักเรียนทราบอย่างชัดเจนว่า การตรวจให้คะแนนนั้นคำนึงถึงสิ่งใดบ้าง เพื่อให้นักเรียนได้เพิ่มความระมัดระวัง (พวงรัตน์ ทวีรัตน์. 2530: 119)

2. ความคลาดเคลื่อนที่มาจากภายในผู้ตรวจให้คะแนน จะเป็นความคลาดเคลื่อนที่มาจากคุณลักษณะในตัวผู้ให้คะแนนเป็นหลัก เช่น อารมณ์ของผู้ตรวจให้คะแนนขณะตรวจ ซึ่งผู้ตรวจอาจเกิดจากความเหนื่อยล้า ความเบื่อ หรือความฉุนเฉียวจากตัวกระดาษคำตอบขณะทำการตรวจ ให้คะแนน ผู้ตรวจต่างคนก็จะมีมาตรฐานต่างกัน ทำให้คะแนนผิดเพี้ยนไป ยกตัวอย่างเช่น ครูคนหนึ่งกำลังตรวจข้อสอบนักเรียนอยู่ มีโทรเลขจากกรุงเทพฯ มาบอกว่าสอบเข้าเรียนที่ประสานมิตรได้ ครูคนนี้จะตื่นเต้นดีใจ คะแนนจะมีแนวโน้มสูงขึ้นกว่าปกติ ในทางตรงกันข้าม ถ้าครูคนนี้ได้รับหนังสือจากจังหวัดลงโทษลดขั้นเงินเดือน เพราะไม่แต่งเครื่องแบบข้าราชการในวันที่รัฐมนตรีกว่าการกระทรวงศึกษาเดินทางมาตรวจราชการ อารมณ์ของครูคนนี้จะขุ่นมัวทันที คะแนนจะมีแนวโน้มต่ำลงกว่าปกติ (บุญเชิด ภิญญโณนันทพงษ์. 2521: 125; และ พวงรัตน์ ทวีรัตน์. 2530: 119) การคล้อยตามกันจากพฤติกรรมส่วนตัวของผู้ให้คะแนน (The Halo Effect) (บุญเชิด ภิญญโณนันทพงษ์. 2521: 125; Saal; Downey; & Lahey. 1980; และ พวงรัตน์ ทวีรัตน์. 2530: 119) เป็นความลำเอียงชนิดหนึ่งที่เป็นผลจากผู้ตรวจให้คะแนนได้รับอิทธิพลจากผู้ตอบอันเป็นความประทับใจของผู้ตรวจให้คะแนนที่ได้ประสบจากผู้เขียนตอบนั้น จะทำให้ผู้ตรวจให้คะแนนประเมินค่าคะแนนนั้นออกมาสูงขึ้นหรือต่ำลง เนื่องจากความประทับใจในลักษณะเด่น-ด้อย ของผู้เขียนตอบ ตัวอย่างเช่น เพียงพร เป็นนักเรียนที่ประพฤติดี เคารพครู ขยันทำการบ้าน คุณลักษณะเหล่านี้ จะมีอิทธิพลต่อครูในการตรวจให้คะแนนเพียงพรสูงขึ้น ส่วนสันติ เป็นนักเรียนที่ซน ดื้อกระด้าง เป็นที่ระอาใจของครูอยู่เสมอ การตรวจให้คะแนนของสันติ ซึ่งดีเท่าเพียงพร มีแนวโน้มที่จะได้คะแนนต่ำกว่า เพียงพร หรือการที่ผู้ตรวจให้คะแนนอ่านคำตอบข้อแรกของผู้สอบคนใดคนหนึ่ง ถ้าผู้สอบตอบได้ดีมากก็จะให้คะแนนสูง และเมื่อตรวจข้อต่อไปก็มีแนวโน้มจะให้คะแนนสูงตามไปด้วย แม้ว่าคำตอบข้ออื่นๆ จะไม่ดีเท่าที่ควร และในทางตรงกันข้ามถ้าคำตอบข้อแรกผู้สอบตอบไม่ค่อยดี ก็จะได้คะแนนต่ำ และข้อต่อไป ก็จะมีแนวโน้มต่ำลงไปด้วยเช่นกัน นอกจากนี้ ความรู้จักมักคุ้นก็จะเป็นผลต่อการตรวจให้คะแนน เช่น ถ้าครูทราบว่ากระดาษคำตอบนั้นเป็นของนักเรียนที่เรียนดีและขยัน ก็จะให้คะแนนสูงกว่าที่ควรจะเป็น และในทางตรงกันข้ามถ้าเป็นของนักเรียนที่เรียนอ่อน และเป็นนักเรียนเกเรก็จะให้คะแนนต่ำกว่าเป็นจริง และความเข้มงวดหรือใจดีของผู้ให้คะแนน (Generosity) (Saal; Downey; & Lahey. 1980; และ พวงรัตน์ ทวีรัตน์. 2530: 119) ผู้ตรวจให้คะแนนอาจจะให้คะแนนสูงกับนักเรียนทุกคน และอีกคนอาจจะให้คะแนนต่ำกับนักเรียนทุกคน ครูหลายคนมักจะให้คะแนนผู้สอบหรือนักเรียนสูงกว่าที่ควรจะเป็นเสมอ เช่น ให้เกรด A เกือบก่อนห้องต่างๆ ที่ตามความเป็นจริงแล้วอาจมีคนที่ทำคะแนนได้ถึงเกณฑ์ที่จะให้เกรด A เพียงครึ่งห้องหรือเพียงหนึ่งในสามของนักเรียนทั้งห้องเท่านั้น ฯลฯ ดังนั้นคะแนนของข้อสอบจะได้รับผลกระทบจากคุณลักษณะผู้ให้คะแนน (Rater Characteristics) พอๆ กับได้รับผลกระทบจากความยากของข้อสอบและความสามารถของผู้สอบ

จากความหมาย ความสำคัญ และปัญหาตั้งที่กล่าวมาข้างต้นผู้วิจัยจึงสนใจที่จะใช้การเขียนเรียงความเป็นเครื่องมือที่ใช้ในการวิจัย โดยที่เน้นความสนใจไปที่ความคลาดเคลื่อนที่มาจากภายในผู้ตรวจให้คะแนน และความคลาดเคลื่อนที่มาจากภายนอกผู้ตรวจให้คะแนน เพราะนอกจากจะเป็นการพัฒนา

การวัดความสามารถในการวัดความสามารถในการเขียนเรียงความของนักเรียนแล้ว ยังเป็นการพัฒนาความสามารถในการคิดผู้ตรวจให้คะแนนในอนาคตอีกด้วย

1.2 การเขียนจดหมาย

การเขียนจดหมายเป็นวิธีสื่อสารที่เป็นลายลักษณ์อักษรแทนการพูด ถึงแม้ในปัจจุบันเทคโนโลยีสื่อสารมีให้เลือกสื่อสารได้หลายช่องทาง แต่จดหมายก็ยังคงมีความจำเป็นอยู่มาก (บุญลักษณ์ เอี่ยมสำอางค์ เกื้อกมล พฤกษ์ประมุข; และ โสภิต พิทักษ์. 2551) โดยเฉพาะการติดต่อกิจกรรมระหว่างหน่วยงานต่างๆ ทั้งนี้อาจเนื่องมาจากการเขียนจดหมายสามารถ ทำให้ผู้รับจดหมายได้รับข้อความที่มีความสมบูรณ์ชัดเจน ทำให้เกิดความเข้าใจที่ตรงกัน อีกทั้งจดหมายยังสามารถใช้เป็นหลักฐานของการติดต่อระหว่างกันได้ นอกจากนี้จดหมายในบางรูปแบบยังสามารถสร้างความรู้สึกระหว่างกันได้อีกด้วย

ประเภทของจดหมาย แบ่งออกเป็น 4 ประเภท (บุญลักษณ์ เอี่ยมสำอางค์; เกื้อกมล พฤกษ์ประมุข; และ โสภิต พิทักษ์. 2551) ดังนี้

1. จดหมายส่วนตัว เป็นจดหมายที่เขียนติดต่อกันระหว่างบุคคลที่คุ้นเคยกัน อาจเป็นญาติสนิท ครู อาจารย์ เพื่อไต่ถามทุกข์สุขส่วนตัว เล่าเรื่องราวที่พบเห็นมา แสดงความเสียใจ แสดงความยินดี หรือขอบคุณ หรือแจ้งกิจกรรมบางอย่างที่มีความเกี่ยวข้องกันระหว่างผู้ส่งกับผู้รับ

คำขึ้นต้น คำลงท้าย และสรรพนามที่ใช้ในการเขียนจดหมายส่วนตัวที่นิยมกัน มีดังนี้

ผู้รับจดหมาย บิดา มารดา ญาติผู้ใหญ่ ผู้ที่สนิทกัน เสมอกัน หรือต่ำกว่าบุคคลทั่วไป

คำขึ้นต้น กราบเท้า ... ที่รักและเคารพยิ่ง ... ที่รัก ... ที่คิดถึง
เรียน ... ที่นับถือ

คำลงท้าย ด้วยความเคารพอย่างสูง รักและคิดถึง ด้วยความรักเสมอ
ขอแสดงความนับถือ ด้วยความนับถือ

สรรพนาม บุรุษที่ 1 ลูก ผม ดิฉัน

บุรุษที่ 2 คุณพ่อ คุณแม่ คุณตา คุณย่า

บุรุษที่ 1 ผม ฉัน พี่ ป้า คุณ

บุรุษที่ 2 คุณ เธอ หลาน

บุรุษที่ 1 ผม ดิฉัน ข้าพเจ้า

บุรุษที่ 2 ท่าน

2. จดหมายธุรกิจ เป็นจดหมายที่เขียนติดต่อกันระหว่างบริษัท ห้างร้าน และองค์กรต่างๆ เพื่อติดต่อกันในเรื่องเกี่ยวกับธุรกิจ พาณิชยกรรม และการเงิน ใช้เขียนติดต่อกับบุคคลหรือองค์กรต่างๆ เช่น ห้างร้าน บริษัท สมาคม เพื่อติดต่อกิจกรรมต่างๆ เช่น เขียนจดหมายขอลาหยุดเรียน เมื่อเจ็บป่วยหรือมีกิจกรรมจำเป็น หรือเชิญวิทยากรมาบรรยายขอเข้าชมกิจการของสถานประกอบการต่างๆ เป็นต้น การเขียนจดหมายธุรกิจมีแบบของการเขียนตามแบบของหนังสือราชการ

3. จดหมายกิจธุระ เป็นจดหมายที่บุคคลเขียนติดต่อกับบุคคลอื่น หรือระหว่างบุคคลกับบริษัท ห้างร้าน องค์กร เพื่อแจ้งธุระต่างๆ เช่น นัดหมาย ขอสมัครงาน ขอทราบผลการสอบบรรจุพนักงาน ขอความช่วยเหลือ และขอคำแนะนำเพื่อประโยชน์ในด้านการงานต่างๆ แบ่งตามลักษณะด้านการติดต่อทางธุรกิจเป็น 2 ประเภท คือ

3.1 จดหมายเกี่ยวกับธุรกิจโดยตรง ได้แก่ จดหมายขายสินค้าและบริการ จดหมายสอบถามและตอบแบบสอบถาม จดหมายขอเปิดเครดิตและจดหมายตอบรับการขอเปิดเครดิต

3.2 จดหมายที่เกี่ยวกับการซื้อขายโดยตรง ได้แก่ จดหมายสมัครงาน จดหมายขอบคุณ จดหมายแสดงความยินดี

4. จดหมายราชการ หรือที่เรียกว่าหนังสือราชการ เป็นจดหมายที่เขียนติดต่อกันระหว่างส่วนราชการต่างๆ หรือบุคคลเขียนไปถึงส่วนราชการ หรือส่วนราชการมีไปถึงส่วนราชการด้วยตนเอง ตลอดจนส่วนราชการนั้นเขียนจดหมายไปถึงตัวบุคคล การเขียนจดหมายราชการต้องคำนึงถึงแบบของหนังสือราชการตามระเบียบงานสารบรรณด้วย ข้อความในหนังสือดังกล่าวถือว่าเป็นหลักฐานทางราชการ มีสภาพผูกมัดถาวร ดังนั้นจึงควรเขียนให้กระชับและชัดเจน หนังสือราชการมี 2 ประเภท คือ

4.1 หนังสือภายนอก ได้แก่ จดหมาย หรือหนังสือที่สร้างขึ้นตามแบบวิธีโดยใช้กระดาษตราครุฑอยู่กลางกระดาษส่วนบน ใช้เป็นหนังสือติดต่อระหว่างราชการ หรือส่วนราชการต่อหน่วยงานอื่นๆ และบุคคลภายนอก

4.2 หนังสือภายใน ได้แก่ จดหมาย หรือหนังสือติดต่อราชการที่เป็นแบบพิธีน้อยกว่าหนังสือภายนอก เป็นหนังสือที่มีการติดต่อระหว่างส่วนราชการในหน่วยงานเดียวกัน ตามปกติใช้กระดาษบันทึกข้อความหรือใช้กระดาษตราครุฑ

4.2.1 หนังสือราชการทั่วไปแบ่งออกเป็น 2 ลักษณะ การเลือกใช้ลักษณะใดลักษณะหนึ่งนั้น ผู้ออกหนังสือจะเป็นผู้พิจารณา ได้แก่

4.2.2 ข้อความลับ การกำหนดชั้นความลับ หมายถึง กำหนดความสำคัญของเอกสารหรือข่าวสารว่า เรื่องใดมีความสำคัญมากน้อยเพียงใด และควรมีความลับชั้นไหน เพื่อให้ผู้มีหน้าที่เกี่ยวข้องได้ระมัดระวังให้เหมาะสมกับคำหรือความสำคัญ ชั้นความลับแบ่งออกเป็น 4 ชั้น คือ ลับที่สุด ลับมาก ลับ และปกปิด (การดำเนินการเกี่ยวกับเอกสารลับไม่ว่าระดับชั้นใดก็มีความสำคัญเช่นเดียวกัน จะแตกต่างกันเพียงแต่ระดับผลของความเสียหายที่เกิดจากการรั่วไหลของระดับชั้นที่มีความแตกต่างกันเท่านั้นเอง)

4.2.3 ชั้นความเร็ว เป็นหนังสือที่ต้องจัดส่งและดำเนินการด้วยความเร็วเป็นพิเศษ แบ่งเป็น 3 ประเภท คือ

4.2.3.1 **ด่วนที่สุด** ให้เจ้าหน้าที่ปฏิบัติในทันทีที่ได้รับหนังสือนั้น

4.2.3.2 **ด่วนมาก** ให้เจ้าหน้าที่ปฏิบัติโดยเร็ว

4.2.3.3 **ด่วน** ให้เจ้าหน้าที่ปฏิบัติเร็วกว่าปกติ เท่าที่จะทำได้

การเขียนเรียงความแตกต่างกับการเขียนจดหมายในส่วนที่เป็นองค์ประกอบของรูปแบบของการเขียน โดยที่รูปแบบการเขียนเรียงความประกอบไปด้วย คำนำ เนื้อเรื่อง สรุป แต่รูปแบบการเขียนจดหมายจะประกอบไปด้วย ที่อยู่ วันที่ การขึ้นต้น และการลงท้าย แต่เนื้อเรื่องในจดหมายนั้นมีส่วนที่คล้ายกับการเขียนเรียงความ คือ เมื่อจบการเขียนส่วนที่เป็นการขึ้นต้นแล้ว เนื้อเรื่องที่เขียนจะประกอบไปด้วย คำนำ เนื้อเรื่อง และสรุป แล้วจึงจบการเขียนจดหมายด้วยการเขียนคำลงท้าย เหตุผลที่ผู้วิจัยเลือกการเขียนทั้งสองแบบนี้ เพื่อเป็นเครื่องมือการวิจัยเนื่องจากการวัดความสามารถในการเขียนที่ใช้เสมอกับนักเรียนชั้นประถมศึกษาปีที่ 3 โดยเครื่องมือการวัดความสามารถในการเขียนเรียงความ ผู้วิจัยได้ใช้ข้อสอบที่เป็นมาตรฐานคือ การเขียนเรียงความในส่วนของเขียนตามจินตนาการ จากข้อสอบการอ่านออกเสียง การเขียน และการคิดคำนวณชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 ส่วนข้อสอบการเขียนจดหมายนั้น ผู้วิจัยได้สร้างและพัฒนาขึ้นใหม่

1.3 การวัดความสามารถในการเขียน

เครื่องมือการวิจัยนี้แบ่งออกเป็น 2 ส่วน คือ ส่วนที่เป็นการสร้างเกณฑ์การตรวจให้คะแนนแบบทดสอบวัดความสามารถในการเขียนเรียงความของนักเรียน โดยใช้ข้อสอบมาตรฐานจากสำนักทดสอบทางการศึกษา และการสร้างแบบทดสอบวัดความสามารถในการเขียนจดหมาย พร้อมการสร้างเกณฑ์การตรวจให้คะแนนแบบทดสอบวัดความสามารถในการเขียนจดหมายของนักเรียน ดังนั้นผู้วิจัยจึงแบ่งการนำเสนอการวัดความสามารถในการเขียนออกเป็น 2 ส่วนคือ ส่วนที่เป็นแบบทดสอบวัดความสามารถในการเขียน และส่วนที่เป็นวิธีการตรวจให้คะแนนความสามารถในการเขียน ดังนี้

1.3.1 การสร้างแบบทดสอบวัดความสามารถในการเขียนของนักเรียน

ในการสร้างแบบทดสอบวัดความสามารถในการเขียนนี้มีอยู่ 2 ส่วน คือ การใช้ข้อสอบมาตรฐานจากสำนักทดสอบทางการศึกษา ในส่วนของเขียน ที่ได้คัดสำเนามาจากข้อสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 มาใช้เป็นเครื่องมือในการวิจัยแล้วสร้างเกณฑ์การตรวจให้คะแนนใหม่ และส่วนที่ 2 คือ แบบทดสอบการเขียนจดหมายที่ผู้วิจัยสร้างขึ้น โดยมีรายละเอียดดังนี้

1. รายละเอียดของข้อสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 มีรายละเอียด ดังนี้

ข้อสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 นั้น ได้สร้างขึ้นสืบเนื่องจากสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ได้กำหนดนโยบายเพื่อยกระดับคุณภาพการศึกษาของผู้เรียน โดยกำหนดให้มีการประเมินผลสัมฤทธิ์ทางการเรียนของนักเรียน รูปแบบของการประเมินคุณภาพการศึกษาขั้นพื้นฐานเพื่อการประกันคุณภาพผู้เรียน ทั้งนี้มุ่งนำเสนอผล เพื่อ

การพัฒนาคุณภาพการจัดการเรียนการสอนของสถานศึกษาให้ทัดเทียมกันและให้เกิดผลตามกลยุทธ์ การพัฒนาคุณภาพและมาตรฐานการศึกษาทุกระดับ โดยสำนักทดสอบทางการศึกษาได้ดำเนินการประเมิน ผลสัมฤทธิ์ทางการเรียนในชั้นประถมศึกษาปีที่ 3 กับนักเรียนในทุกโรงเรียนทุกคน การดำเนินการประเมิน คุณภาพการศึกษาในครั้งนี้ นอกจากจะมีการทดสอบผลสัมฤทธิ์ทางการเรียน โดยใช้แบบทดสอบชนิด เลือกตอบแล้ว ยังดำเนินการตรวจสอบความสามารถของนักเรียนในเรื่องของ การอ่านออก เขียนได้ คิดคำนวณ และคิดอย่างเป็นระบบได้ จึงกำหนดให้มีการประเมิน (Performance Test) ในกลุ่มสาระ การเรียนรู้ภาษาไทยและคณิตศาสตร์ เพื่อตรวจสอบความสำเร็จของการดำเนินการพัฒนาการศึกษา ตามกลยุทธ์ของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน ที่เน้นการประเมินเพื่อพัฒนาการเรียนรู้ (Assessment for Learning) โดยมุ่งเน้นให้หน่วยงานทางการศึกษาต่างๆ นำผลการประเมินไปใช้ในการ พัฒนาคุณภาพให้ตรงตามเป้าหมาย

วัตถุประสงค์ของการทดสอบ 1) เพื่อตรวจสอบความสามารถของนักเรียนในเรื่อง ของการอ่านออกเสียง การเขียน และการคิดคำนวณ 2) เพื่อจัดทำข้อมูลพื้นฐานสำหรับไปใช้ในการปรับปรุง และพัฒนาคุณภาพการเรียนรู้ และ 3) เพื่อนำข้อมูลพื้นฐานไปใช้ในการสนับสนุน และพัฒนาการจัด การเรียนรู้

กลุ่มเป้าหมายของการทดสอบ นักเรียนชั้นประถมศึกษาปีที่ 3 ทุกคน ทุกโรงเรียน ในสังกัดสำนักงานคณะกรรมการการศึกษา ขั้นพื้นฐาน ปีการศึกษา 2553

ขอบเขตของการทดสอบ การทดสอบกลุ่มสาระการเรียนรู้ภาษาไทยและคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 3 มีขอบเขต ของการวัดตามกรอบมาตรฐานการเรียนรู้กลุ่มสาระการเรียนรู้ภาษาไทย และคณิตศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ดังนี้

กลุ่มสาระการเรียนรู้ภาษาไทย

สาระที่ 1 การอ่าน

มาตรฐาน ท 1.1 ใช้กระบวนการอ่านสร้างความรู้และความคิดเพื่อนำไปใช้ตัดสินใจ แก้ปัญหาในการดำเนินชีวิต และมีนิสัยรักการอ่าน

สาระที่ 2 การเขียน

มาตรฐาน ท 2.1 ใช้กระบวนการเขียนเขียนสื่อสาร เขียนเรียงความ ย่อความ และเขียนเรื่องราวในรูปแบบต่างๆ เขียนรายงานข้อมูลสารสนเทศและรายงานการศึกษาค้นคว้าอย่างมีประสิทธิภาพ

กลุ่มสาระการเรียนรู้คณิตศาสตร์

สาระที่ 1 จำนวนและการดำเนินการ

มาตรฐาน ค 1.2 เข้าใจถึงผลที่เกิดขึ้นจากการดำเนินการของจำนวน และความสัมพันธ์ระหว่างการดำเนินการต่างๆ และสามารถใช้ในการดำเนินการในการแก้ปัญหา

ลักษณะของแบบทดสอบการเขียนตามจินตนาการเป็นรายบุคคลมีรายละเอียด ดังนี้

1. ให้นักเรียนเขียนเรื่องตามจินตนาการโดยมีเนื้อหาสาระสอดคล้องกับภาพ

2. เขียนเรื่องให้สอดคล้องหรือสัมพันธ์กับภาพที่กำหนดให้ลงในกระดาษ คำตอบ โดยต้องปฏิบัติตามหลักการเขียน ดังนี้

2.1 ตั้งชื่อเรื่องให้สัมพันธ์หรือสอดคล้องกับภาพ

2.2 เนื้อเรื่องที่เป็นสาระสำคัญสัมพันธ์หรือสอดคล้องต่อเนื่องกันตลอดทั้งเรื่องและสอดคล้องกับชื่อเรื่อง

2.3 การเขียนได้ถูกต้องตามอักขรวิธี

2.4 ความยาวของเนื้อเรื่องตามที่กำหนดไม่น้อยกว่า 10 บรรทัด

3. เวลาในการสอบ 40 นาที คะแนนเต็ม 20 คะแนน

4. เขียนตอบลงในแบบทดสอบ

การตรวจให้คะแนน : การตรวจให้คะแนนแบบทดสอบความสามารถในการเขียน

1. สำนักงานเขตพื้นที่การศึกษาจัดหาสถานที่เพื่อตั้งเป็นศูนย์การตรวจแบบทดสอบชนิด เขียนตอบ โดยอาจใช้ห้องประชุมของสำนักงานเขตพื้นที่การศึกษา หรือโรงเรียนที่มีความพร้อมและเหมาะสม

2. แต่งตั้งคณะกรรมการตรวจกระดาษคำตอบจากครูผู้มีประสบการณ์สอนกลุ่มสาระการเรียนรู้ภาษาไทย โดยตั้งกรรมการเป็นชุดๆ ละ 3 คน จำนวนชุดของกรรมการขึ้นอยู่กับปริมาณนักเรียน ที่เข้าสอบในสำนักงานเขตพื้นที่การศึกษานั้นๆ

3. ในการจัดชุดคณะกรรมการตรวจ ให้มีการสลับกรรมการเพื่อมิให้ตรวจนักเรียนของตนเอง

4. จัดประชุมชี้แจงแนวทางการดำเนินการตรวจให้คะแนน เพื่อให้คณะกรรมการเข้าใจเกณฑ์ การตรวจและสามารถดำเนินการตรวจให้คะแนนได้ตามเกณฑ์ที่กำหนด

5. คณะกรรมการดำเนินการตรวจให้คะแนนต้องใช้เกณฑ์การตรวจให้คะแนนประกอบการตรวจให้คะแนน

6. เมื่อกรรมการแต่ละคนตรวจให้คะแนนแล้ว ให้กรอกคะแนนลงในช่องสำหรับผู้ตรวจ กรอกคะแนน รวมคะแนนทุกด้านไว้ในช่องรวมในกระดาษคำตอบ

7. นำคะแนนของคณะกรรมการทั้ง 3 คนรวมกันแล้วนำมาหาค่าเฉลี่ย (เศษตั้งแต่ 0.5 ปัดขึ้น)

8. กรอกคะแนนที่ได้จากการเฉลี่ย กรอกลงในแบบบันทึกคะแนนการทดสอบวัดความสามารถในการเขียน

9. รวบรวมข้อสอบและแบบบันทึกผลส่งสำนักงานเขตพื้นที่การศึกษา

10. สำนักงานเขตพื้นที่การศึกษาดำเนินการตรวจทาน และกรอกคะแนนลงในแบบบันทึกคะแนนรวมของนักเรียนแต่ละคน ส่งผลให้ส่วนกลางทางระบบ EPCC ตามแบบบันทึกคะแนนรวมการทดสอบความสามารถในการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2553

ในงานวิจัยขั้นนี้ผู้วิจัยใช้ประโยชน์จากส่วนหนึ่งของกระดาษคำตอบในส่วนของกรเขียนที่ได้คัดสำเนามาจากข้อสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 มาใช้เป็นเครื่องมือในการวิจัย และสร้างเกณฑ์การตรวจให้คะแนนขึ้นมาใหม่เพื่อนำไปใช้ในการทดลองต่อไป

2. การสร้างข้อสอบการเขียนจดหมาย โดยที่ในการสร้างแบบทดสอบวัดความสามารถในการเขียนจะต้องวางจุดประสงค์ให้แน่ชัดก่อนว่าจะให้เขียนเรื่องนั้นๆ อย่างไร เขียนไปในแนวทางใด เช่น เขียนขยายความจากแนวคิดที่ตั้งไว้ตามโครงเรื่อง เขียนหัวใจให้น่าเชื่อถือ หรือเขียนแสดงความรู้ ความคิดอย่างเป็นระบบ เป็นต้น มีการแบ่งชนิดของการเขียนเรียงความตามจุดมุ่งหมาย (กระทรวงศึกษาธิการ. 2552ก) ไว้ดังนี้

การเขียนจดหมายเรื่องใดเรื่องหนึ่งอาจใช้โวหารหลายแบบประกอบกัน นักเรียนต้องพิจารณาความเหมาะสมให้ตรงกับจุดประสงค์ที่ครูต้องการ ปัจจุบันจะเน้นเขียนในเชิงสร้างสรรค์ เพื่อเน้นคุณค่าในงานเขียนเป็นสำคัญไม่ว่าจะเป็นงานเขียนชนิดใด ดังนั้นผู้วิจัยจึงตั้งใจจะกำหนดให้นักเรียนเขียนจดหมายในหัวข้อที่สร้างสรรค์ และนักเรียนสามารถนำความรู้ที่ได้จากการค้นคว้า เพื่อการเขียนจดหมายนั้น ไปใช้ประโยชน์ในชีวิตประจำวันได้

องค์ประกอบและรูปแบบของจดหมาย จดหมายไม่ว่าประเภทใด จะต้องประกอบด้วย

สิ่งต่างๆ ดังนี้

ที่อยู่เล็กน้อย

2 นิ้ว

1. ที่อยู่ของผู้เขียน เริ่มกึ่งกลางหน้าระหว่างเส้นคั่นหน้ากับริมของขอบกระดาษ
2. วัน เดือน ปี ที่เขียนจดหมาย ให้เอียงมาทางซ้ายของตำแหน่งที่เขียน
3. คำขึ้นต้น ห่างจากขอบกระดาษด้านซ้าย 1 นิ้ว
4. เนื้อหา ขึ้นอยู่กับย่อหน้าตามปกติ อาจอยู่ห่างจากขอบกระดาษด้านซ้าย
5. คำลงท้ายอยู่แนวเดียวกับที่อยู่ของผู้เขียน
6. ชื่อผู้เขียน อยู่ใต้คำลงท้าย ถ้าเข้าไปเล็กน้อย

ตัวอย่าง รูปแบบการเขียนจดหมายทั่วไป

สถานที่เขียนจดหมาย

วันที่ เดือน ปี

ระยะ 1 นิ้ว คำขึ้นต้น

ประมาณ 2 นิ้ว

เนื้อหา

.....

คำลงท้าย

ชื่อผู้เขียน

หลักการทั่วไปในการเขียนจดหมาย

1. หลักการทั่วไปในการเขียนจดหมาย การเขียนจดหมายควรคำนึงถึงสิ่งต่อไปนี้

1.1 การใช้ถ้อยคำ จดหมายที่ต้องใช้ถ้อยคำในการเขียนให้ถูกต้องเหมาะสมกับประเภทของจดหมายและผู้รับจดหมายด้วย หลักการใช้ถ้อยคำสำหรับการเขียนจดหมาย

1.1.1 ใช้คำขึ้นต้นและคำลงท้ายจดหมายส่วนตัวให้เหมาะสม และถูกต้องตามแบบแผน

1.1.1.1 การใช้คำขึ้นต้นและคำลงท้ายจดหมายส่วนตัว การเขียนจดหมาย ส่วนตัวไม่มีข้อกำหนดเกี่ยวกับการใช้คำขึ้นต้นและคำลงท้ายที่ตายตัว เพียงแต่เลือกใช้ให้เหมาะสมเท่านั้น

1.1.1.2 คำขึ้นต้นและคำลงท้ายหนังสือราชการ การใช้คำขึ้นต้นและคำลงท้าย หนังสือราชการ ต้องเป็นไปตามระเบียบงานสารบรรณของทางราชการ ในที่นี้จะนำเสนอ

1.1.2 ใช้ภาษาเขียนให้ถูกกับระดับของจดหมาย โดยทั่วไปแล้ว ถ้าเขียนจดหมายส่วนตัวถึงบุคคลที่สนิทสนมกัน ก็จะใช้คำระดับที่ไม่เป็นทางการ แต่ถ้าเขียนจดหมายธุรกิจต่างๆ ก็ใช้คำระดับที่เป็นทางการ

1.1.3 เขียนโดยใช้ถ้อยคำตรงไปตรงมา เพื่อให้ผู้รับจดหมายได้ทราบเรื่อง อย่างรวดเร็วและเข้าใจได้ทันทีตรงตามที่คุณเขียนต้องการ หากจะต้องปฏิบัติตาม ธรรมเนียมปฏิบัติโดยถูกต้องครบถ้วน

1.1.4 เขียนเชิงสร้างสรรค์ คือ เขียนโดยเรียบเรียงถ้อยคำให้ผู้อ่านเกิดความรู้สึกชื่นชม ชวนอ่าน แสดงน้ำใจที่ดีต่อผู้ได้รับจดหมาย

1.2 มารยาทในการเขียนจดหมาย

1.2.1 เลือกกระดาษ ของ ที่สะอาดเรียบร้อย หากเป็นไปได้ควรใช้กระดาษที่ทำขึ้นเพื่อการเขียนจดหมายโดยตรง แต่ถ้าหาไม่ได้ก็ควรใช้กระดาษที่มีสีสุภาพ กระดาษที่ใช้เขียนควรเป็นกระดาษเต็มแผ่น ไม่ฉีกขาด ไม่ยู่ยิบเยิน ไม่สกปรก ของจดหมายที่ดีที่สุดคือ ของที่การสื่อสารแห่งประเทศไทยจัดทำขึ้น เพราะมีขนาด และคุณภาพได้มาตรฐาน ของประเภทนี้มีจำหน่ายตามที่ทำการไปรษณีย์โทรเลขทุกแห่ง ถ้าหาของจดหมายของการสื่อสารแห่งประเทศไทยไม่ได้ ก็อาจเลือกซื้อจากของที่มีเอกชนทำขึ้นจำหน่าย ซึ่งถ้าเป็นในกรณีหลังนี้ ควรเลือกของที่มีสีสุภาพ ไม่ควรมีลวดลาย ไม่ควรใช้ของที่มีตราครุฑส่งจดหมายที่มีใช้หนังสือราชการ ไม่ควรใช้ซองที่มีขอบซองเป็นลายขวาง น้ำเงินสลับกัน ซึ่งเป็นซองสำหรับส่งจดหมายไปรษณีย์อากาศไปยังต่างประเทศ ในการส่งจดหมายในประเทศ

1.2.2 เขียนหนังสือให้ชัดเจน อ่านง่าย การเขียนตัวอักษรค่อนข้างโต และเว้นช่องไฟค่อนข้างห่าง จะช่วยให้จดหมายนั้นอ่านง่ายไม่ควรเขียนด้วยดินสอดำหรือหมึกสีแดง เพราะถือกันว่าไม่สุภาพ แม้หมึกหรือดินสอดสีต่างๆ ก็ไม่ควรเขียนจดหมาย สีที่เหมาะสมคือสีน้ำเงิน และสีดำ ไม่ควรเขียนให้มีตัวผิด ตัวตก ต้องแก้ต้องเติม มีรอยชูดลบขีดฆ่า หรือมีเส้นโยง ข้อความรุงรัง ทำให้ดูสกปรกไม่งามตา

1.2.3 จะต้องศึกษาให้ถูกต้องก่อนว่า ผู้ที่เราจะเขียนจดหมาย ไปถึงนั้น เป็นใคร มีตำแหน่งหน้าที่อะไร การเขียนข้อความในจดหมายก็ดี การจำหน่ายของก็ดี จะต้องระบุตำแหน่งหน้าที่ ชื่อยศของผู้นั้นให้ถูกต้องและต้องสะกดชื่อ นามสกุล ยศ ตำแหน่ง ของผู้นั้น ให้ถูกต้องด้วย

1.2.4 เมื่อเขียนจดหมายเสร็จแล้ว ต้องพับให้เรียบร้อยแล้วบรรจุซองจำหน่ายของให้ถูกต้องครบถ้วน ปิดดวงตราไปรษณียากรให้ครบถ้วนตามราคาและถูกตำแหน่ง ก่อนที่จะนำไปส่ง

1.2.5 เขียนจำหน่ายของจดหมาย

1.2.5.1 เขียนชื่อ นามสกุลของผู้รับให้ถูกต้อง ชัดเจน อ่านง่าย ถ้าผู้รับเป็นแพทย์ เป็นอาจารย์ หรือมียศตำรวจ ทหาร หรือมีคำนำหน้านามแสดงเกียรติยศหรือฐานันดรศักดิ์ ก็ใช้ถ้อยคำพิเศษเหล่านั้นนำหน้าชื่อ คำนำหน้าชื่อควรเขียนเต็ม ไม่ควรใช้คำย่อ ถ้าทราบตำแหน่งก็ระบุตำแหน่งลงไปด้วยในกรณีที่ ไม่ทราบรายละเอียดดังกล่าว ควรใช้คำว่า คุณ นำหน้าชื่อผู้รับในการจำหน่ายของจดหมายนั้น

1.2.5.2 ระบุสถานที่ของผู้รับให้ถูกต้อง ชัดเจนและมีรายละเอียดพอที่บุรุษไปรษณีย์จะนำจดหมายไปส่งได้ไม่ผิดพลาด ระบุเลขที่บ้าน ห้างร้านหรือสำนักงาน ซอย ตรอก ถนน หมู่บ้าน ตำบล อำเภอ (ในกรณีต่างจังหวัด) หรือแขวง เขต (ในกรณีกรุงเทพมหานคร) ที่สำคัญคือจะต้องระบุรหัสไปรษณีย์ถูกต้องทุกครั้ง จดหมายจะถึงผู้รับเร็วขึ้น

หมายเหตุ การสื่อสารแห่งประเทศไทยได้จัดทำเอกสารแสดงรหัสไปรษณีย์ของอำเภอและจังหวัด ต่างๆ สำหรับแจกจ่ายให้ประชาชน ท่านจะติดต่อขอรับได้จากที่ทำการไปรษณีย์โทรเลขทุกแห่ง

1.2.5.3 การจำหน่ายของ การสื่อสารแห่งประเทศไทย แนะนำให้เขียนนามและที่อยู่ พร้อมด้วยรหัสไปรษณีย์ของผู้ส่งไว้ที่มุมบนด้านซ้ายมือของซองและเขียนชื่อ ผู้รับพร้อมที่อยู่และรหัสไปรษณีย์ไว้ตรงกลาง ดังตัวอย่าง

หมายเหตุ การสื่อสารแห่งประเทศไทย มีบริการพิเศษต่างๆ ที่จะช่วยป้องกันมิให้ จดหมายสูญหาย หรือช่วยให้จดหมายถึงมือผู้รับได้รวดเร็ว ทันเวลา เช่น บริการ EMS เป็นต้น ผู้สนใจจะใช้บริการต่างๆ ดังกล่าว จะต้องไปติดต่อที่ทำการไปรษณีย์โทรเลขโดยตรง เพราะจะต้องกรอกแบบรายการบางอย่าง การเขียนข้อความในทำนองที่ว่า “ขอให้ส่งด่วน” ลงบนซองจดหมาย ไม่ทำให้จดหมายถึงเร็วขึ้นแต่อย่างใด

ประเภทของจดหมาย จดหมายแบ่งออกเป็น 4 ประเภท คือ จดหมายส่วนตัว จดหมายกิจธุระ จดหมายธุรกิจ และจดหมายราชการหรือหนังสือราชการ

1. จดหมายส่วนตัว คือ จดหมายที่บุคคล ซึ่งรู้จักคุ้นเคยกันติดต่อกันด้วย วัตถุประสงค์ที่เป็นการส่วนตัว เช่น เพื่อส่งข่าวคราว ถามทุกข์สุข เล่าเรื่องราว ฯลฯ เป็นการติดต่ออย่างไม่เป็นทางการ เช่น จดหมายเล่าเรื่องราวทุกข์สุข จดหมายแสดงความรู้สึกยินดี เสียใจขอบคุณ หรือขอโทษในกรณีต่างๆ เป็นต้น การเขียนจดหมายส่วนตัวแม้จะยินยอมให้ใช้ถ้อยคำที่แสดงความสนิทสนม เป็นกันเองได้ แต่ก็ควรระมัดระวังอย่าให้ผู้อ่านเข้าใจผิด และควรแสดงความสำรวมมากกว่าการพูดกัน โดยปกติ จดหมายส่วนตัวที่มีเนื้อหาเป็นการขอบคุณ หรือแสดงความยินดีอาจเขียนลงในบัตรที่ออกแบบไว้ อย่างสวยงาม แทนการเขียนในกระดาษก็ได้ การเขียนจดหมายส่วนตัว นิยมให้เขียนด้วยลายมือที่อ่านง่าย แสดงความตั้งใจเขียน ไม่นิยมใช้การพิมพ์ดีดจดหมายหรือจำหน่ายของจดหมายส่วนตัว

2. จดหมายกิจธุระ คือ จดหมายติดต่อบetweenบุคคลกับบุคคลหรือบุคคล กับหน่วยงานด้วยเรื่องที่มีใช้เรื่องส่วนตัว แต่เป็นเรื่องที่เกี่ยวกับงาน เช่น การสมัครงาน การติดต่อสอบถาม การขอความร่วมมือ ฯลฯ ภาษาที่ใช้จึงต้องสุภาพและกล่าวถึงแต่ธุระเท่านั้น ไม่มีข้อความที่แสดงความสัมพันธ์ เป็นการส่วนตัวต่อกัน

3. จดหมายธุรกิจ คือ จดหมายติดต่อในเรื่องที่เกี่ยวกับธุรกิจ เช่น การเสนอ ขายสินค้า การขอทราบรายละเอียดเพิ่มเติมเกี่ยวกับสินค้าหรือบริการ การสั่งซื้อสินค้า การติดตาม ทวงหนี้ ฯลฯ จดหมายกิจธุระและจดหมายธุรกิจ มีลักษณะการติดต่อที่เป็นทางการมากกว่า จดหมาย ส่วนตัวจึงต้องใช้คำที่สุภาพ ง่าย และมีเนื้อหาความกะทัดรัด เข้าใจได้ตรงกัน

4. จดหมายราชการหรือหนังสือราชการ คือ เอกสารที่ถือเป็นหลักฐานใน ราชการ ได้แก่ หนังสือที่มีที่ไปที่มาระหว่างส่วนราชการ หรือหนังสือที่ส่วนราชการมีไปถึงหน่วยงานอื่น ซึ่งมีใช้ส่วนราชการ หรือมีไปถึงบุคคลภายนอก หรือหนังสือที่หน่วยงานอื่น ซึ่งมีใช้ส่วนราชการ หรือ บุคคลภายนอกเขียนมาถึงส่วนราชการ จดหมายราชการต้องใช้ถ้อยคำและรูปแบบการเขียนให้ถูกต้อง ตามระเบียบที่ทางราชการกำหนดไว้ ระเบียบดังกล่าวเรียกว่า ระเบียบงานสารบรรณ

กล่าวโดยสรุปในงานวิจัยชิ้นนี้ได้มีการใช้แบบทดสอบวัดความสามารถในการเขียนมีอยู่ 2 ส่วน คือ การใช้ข้อสอบมาตรฐานจากสำนักทดสอบทางการศึกษา และข้อสอบการเขียนจดหมายที่ผู้วิจัยสร้างขึ้นเอง โดยข้อสอบการเขียนจดหมายที่ผู้วิจัยสร้างขึ้นเอง จะกำหนดให้นักเรียนเขียนตอบจดหมายส่วนตัว ในเชิงสร้างสรรค์

1.3.2 การสร้างเกณฑ์การตรวจให้คะแนนแบบทดสอบวัดความสามารถในการเขียน ของนักเรียน

เพื่อความสะดวกในการอธิบายการสร้างเกณฑ์การตรวจให้คะแนนแบบทดสอบวัด ความสามารถในการเขียนของนักเรียน ผู้วิจัยได้แบ่งการอธิบายออกเป็น 3 ส่วน คือ ความหมายของ วิธีการตรวจให้คะแนน ประเภทของวิธีการตรวจให้คะแนน และการสร้างเกณฑ์การตรวจให้คะแนน

1.3.2.1 ความหมายของวิธีการตรวจให้คะแนน วิธีการตรวจให้คะแนน หรือเกณฑ์การตรวจให้คะแนน (Rubrics) คือ ชุดของแนวทางในการตรวจให้คะแนนผลการปฏิบัติเรื่องใดเรื่องหนึ่ง สำหรับใช้ประเมินคุณภาพการปฏิบัติงานของผู้เรียน แนวทางในการตรวจให้คะแนนนั้น อาจทำในรูปของมาตรฐานประเมินค่า หรือแบบตรวจสอบรายการ ประกอบด้วยประเด็นต่างๆ (Criteria) ที่จะใช้พิจารณางานหนึ่งๆ และคำอธิบายระดับคุณภาพของแต่ละประเด็นประเมิน ซึ่งอาจเรียงลำดับตั้งแต่ดีเลิศไปจนถึงต้องปรับปรุง หรือให้เป็นระดับตัวเลขตั้งแต่มากที่สุด (เช่น 4) ไปจนถึงน้อยที่สุด (เช่น 0) ประเด็นประเมินอาจกำหนดเพิ่มเติมได้หลายข้อ คำอธิบายระดับคุณภาพควรอธิบายให้ชัดเจนที่สุด กระชับที่สุด เป็นคำอธิบายที่สามารถบอกได้ว่า ทำไมจึงต้องดีเลิศ ดี ต้องปรับปรุง (เสาวนีย์ เกรียรติ. 2540: 159; และ บุญเชิด ภิญญอนันตพงษ์. 2544: 90) วิธีการตรวจให้คะแนนแบ่งออกเป็น วิธีการตรวจให้คะแนนเฉพาะ (Specific Rubric) เป็นแนวทางในการตรวจให้คะแนนเฉพาะของคำถามแบบปลายเปิดในแต่ละข้อ สำหรับวิธีการตรวจให้คะแนนทั่วไป (General Rubric) จะเป็นตัวแบบของวิธีการตรวจให้คะแนน เพื่อใช้พิจารณาเกณฑ์เฉพาะ รายละเอียดจะเปลี่ยนแปลงสำหรับระดับคะแนนที่แตกต่างกัน ซึ่งวิธีการตรวจให้คะแนนจะเน้นการแสดงความสามารถของนักเรียนในด้านความเข้าใจในความคิดรวบยอด ความรู้เกี่ยวกับกระบวนการหรือขั้นตอน ทักษะในการแก้ปัญหา การใช้เทคโนโลยีใหม่ๆ ในการประมวลผลข้อมูล และการสื่อความหมายกับผู้อื่น (ชัยฤทธิ์ ศีลาเดช. 2544: 29)

กล่าวโดยสรุป วิธีการตรวจให้คะแนน หรือเกณฑ์การตรวจให้คะแนน (Rubrics) คือ ชุดของแนวทางในการตรวจให้คะแนนผลการปฏิบัติเรื่องใดเรื่องหนึ่ง สำหรับใช้ประเมินคุณภาพการปฏิบัติงานของผู้เรียน ประกอบด้วยประเด็นต่างๆ (Criteria) ที่จะใช้พิจารณางานหนึ่งๆ และคำอธิบายระดับคุณภาพของแต่ละประเด็นประเมิน ซึ่งเกณฑ์การตรวจให้คะแนนจะเน้นการแสดงความสามารถของนักเรียนในด้านความเข้าใจในความคิดรวบยอด ความรู้เกี่ยวกับกระบวนการหรือขั้นตอน ทักษะในการแก้ปัญหา การใช้เทคโนโลยีใหม่ๆ ในการประมวลผลข้อมูล และการสื่อความหมายกับผู้อื่น

1.3.2.2 ประเภทของวิธีการตรวจให้คะแนน สำหรับงานวิจัยชิ้นนี้ผู้วิจัยแบ่งประเภทของวิธีการตรวจให้คะแนนไว้ 3 แบบ แต่ผู้เชี่ยวชาญส่วนใหญ่ได้แบ่งวิธีการตรวจให้คะแนนเป็น 2 แบบ (ส.วาสนา ประवालพฤษ์. 2533: 39 – 42; ชัยฤทธิ์ ศีลาเดช. 2544: 29; บุญเชิด ภิญญอนันตพงษ์. 2544: 90; และ กรมวิชาการ. 2544: 35 – 38) คือ

1.3.2.2.1 การตรวจให้คะแนนแบบแยกองค์ประกอบ (Analytic Scoring Rubrics) เพื่อให้การมองคุณภาพของงานหรือความสามารถของนักเรียนได้อย่างชัดเจน จึงได้มีการแยกองค์ประกอบของการตรวจให้คะแนน และอธิบายคุณภาพของงานในแต่ละองค์ประกอบเป็นระดับ

1.3.2.2.2 การตรวจให้คะแนนแบบรวมองค์ประกอบ (Holistic Scoring Rubrics) คือการตรวจให้คะแนนงานชิ้นใดชิ้นหนึ่ง โดยดูภาพรวมของชิ้นงานว่ามีความเข้าใจในความคิดรวบยอด การสื่อความหมาย กระบวนการที่ใช้ และผลงาน แล้วเขียนอธิบายคุณภาพของงาน หรือความสำเร็จของงานเป็นชั้นๆ

1.3.2.2.3 การตรวจให้คะแนนแบบผสมผสาน (Annotated Scoring Rubrics) เป็นการตรวจให้คะแนนที่เพิ่มขึ้นมาใหม่ (Nitko, 1996) จะรวมแบบรวมองค์ประกอบและแยกองค์ประกอบไว้ด้วยกัน เริ่มด้วยการตรวจให้คะแนนในภาพรวมของการปฏิบัติงานด้วยแบบรวมองค์ประกอบ แล้วเลือกให้คะแนนอีกเพียงบางคุณลักษณะของงานแบบแยกองค์ประกอบ ซึ่งการตรวจให้คะแนนเพียงบางคุณลักษณะนี้ จะไม่มีผลต่อการเปลี่ยนแปลงคะแนนที่ประเมินแบบรวมองค์ประกอบ ประโยชน์ก็คือจะมีความรวดเร็วในการตรวจให้คะแนน และเป็นการให้ผู้ตรวจให้คะแนนได้เลือกเฉพาะบางคุณลักษณะที่โดดเด่นเพียงไม่กี่องค์ประกอบเพื่อเป็นผลสะท้อน (Feedback) ให้แก่ผู้เรียน แต่ไม่มีประโยชน์ในการวินิจฉัยผู้เรียนว่าบกพร่องในคุณลักษณะใด เพราะหลายๆ คุณลักษณะไม่ได้ถูกประเมิน

1.3.2.3 การสร้างเกณฑ์การตรวจให้คะแนน ได้มีผู้เสนอวิธีการสร้างเกณฑ์การตรวจให้คะแนนไว้หลายวิธี แต่ละวิธีมีจำนวนขั้นตอนแตกต่างกัน เช่น บัญเชิด ภิญโญนนตพงษ์ (2545) มี 3 ขั้นตอน โครงการพัฒนาทรัพยากรมนุษย์ (ม.ป.ป.) และ เสาวนีย์ เกียรติ (2540: 159) มี 7 ขั้นตอน และ ชัยฤทธิ์ ศิลาดเดช (2540: 68) มี 8 ขั้นตอน ดังนั้นผู้วิจัยจึงเลือกขั้นตอนการสร้างที่ละเอียดที่สุดจากโครงการพัฒนาทรัพยากรมนุษย์ ซึ่งได้อธิบายการสร้างเกณฑ์การตรวจให้คะแนนมีขั้นตอนในการดำเนินการ 7 ขั้นตอน ดังนี้

ขั้นที่ 1 วิเคราะห์ผลการเรียนรู้ของนักเรียน วิเคราะห์ผลการเรียนรู้ของนักเรียนในแต่ละจุดประสงค์ หรือแต่ละหัวเรื่อง (Theme) เมื่อนักเรียนได้ปฏิบัติกิจกรรมการเรียนรู้แล้วจะเกิดผลการเรียนรู้อะไรบ้าง ผลการเรียนรู้ของนักเรียนไม่จำเป็นจะต้องมีครบทุกประเภท ทุกครั้งที่นักเรียนได้ปฏิบัติกิจกรรมการเรียนรู้ บางครั้งในการวางแผนการสอน ครูคาดหวังในผลการเรียนรู้ของนักเรียนเพียงกระบวนการอย่างเดียว บางครั้งครูคาดหวังทั้งกระบวนการและผลงาน

ขั้นที่ 2 กำหนดประเด็นที่จะต้องประเมิน การกำหนดประเด็นที่ต้องการประเมินอาจให้นักเรียน ร่วมกันเสนอความคิด ในการกำหนดประเด็นประเมินผลการเรียนรู้แต่ละอย่างของเขา แต่อย่างไรก็ตามเนื่องจากการวางแผนการสอน ครูผู้สอนได้มีการคาดหวังความสำเร็จจากการเรียนรู้ของนักเรียนเอาไว้ล่วงหน้าแล้ว จึงควรกำหนดรายการประเมินที่สำคัญๆ ของความสำเร็จจากการเรียนแต่ละด้านเอาไว้ล่วงหน้าด้วย เพื่อจะได้แจ้งให้นักเรียนทราบและเพื่อการตรวจสอบผลงานของนักเรียนเอง

ขั้นที่ 3 คัดเลือกประเด็นประเมินผลที่สำคัญ การกำหนดประเด็นที่จะต้องประเมินเราจะพบว่า ในผลการเรียนรู้หนึ่งอย่าง จะมีประเด็นที่ต้องการประเมินมาก โดยเฉพาะถ้าเปิดโอกาสให้นักเรียนมีส่วนร่วมในการเสนอประเด็นการประเมินด้วยแล้วจะมีมุมมองในการประเมินที่หลากหลาย หรือรวมประเด็นที่สำคัญๆ

ขั้นที่ 4 เลือกรูปแบบในการสร้างเกณฑ์ การสร้างเกณฑ์การประเมินผลการเรียนรู้ของนักเรียน เพื่อพิจารณาตัดสินให้คะแนน (Scoring Rubrics) มีรูปแบบในการสร้าง 3 รูปแบบ

รูปแบบที่ 1 การตรวจให้คะแนนแบบรวมองค์ประกอบ (Holistic Scoring Rubrics) การสร้างเกณฑ์การประเมินแบบรวมองค์ประกอบ หมายถึงการตรวจให้คะแนนผลการเรียนรู้โดยรวมทุกประเด็นที่กำหนดเพื่อการประเมิน แล้วเขียนอธิบายคุณภาพของผลการเรียนรู้แต่ละระดับ

ตาราง 1 เกณฑ์การตรวจให้คะแนนแบบรวมองค์ประกอบ

ระดับคุณภาพ	คำอธิบาย
ระดับ 3	หมายถึง ตอบคำถามจากเรื่องที้อ่านถูกทุกข้อ บอกเนื้อหาสาระถูกต้องได้ใจความ ต่อเนื่อง บอกข้อคิดได้ตรงประเด็นสมบูรณ์ เสนอความคิดเห็นด้วยเหตุผลและ ประโยชน์
ระดับ 2	หมายถึง ตอบคำถามจากเรื่องที้อ่านผิดไม่เกิน 3 ข้อ จาก 5 ข้อ บอกเนื้อหา สาระได้ถูกต้องแต่วกวน บอกข้อคิดได้ตรงประเด็นแต่ไม่ต่อเนื่อง เสนอ ความคิดเห็นด้วยเหตุผล
ระดับ 1	หมายถึง ตอบคำถามจากเรื่องที้อ่านผิดมากกว่า 3 ข้อ ใน 5 ข้อ บอกเนื้อหา สาระได้บ้าง บอกข้อคิดได้บ้าง แต่วกวน เสนอความคิดเห็นแต่ไม่แสดงเหตุผล

รูปแบบที่ 2 การตรวจให้คะแนนแบบแยกองค์ประกอบ (Analytic Scoring Rubrics) การตรวจให้คะแนนแบบแยกองค์ประกอบ หมายถึง การตรวจให้คะแนนโดยการแยกองค์ประกอบ ของสิ่งที่จะประเมิน เพื่อให้มองเห็นคุณภาพของงาน หรือความสามารถของนักเรียนได้อย่างชัดเจน ผลการประเมินจะบ่งบอกถึงจุดเด่น จุดด้อยของแต่ละประเด็นประเมินได้ชัดเจน การสร้างเกณฑ์การประเมิน ในรูปแบบนี้จะต้องเขียนคำอธิบายคุณภาพของงานในแต่ละองค์ประกอบ และแต่ละระดับขององค์ประกอบ ให้ชัดเจน

ตาราง 2 เกณฑ์การตรวจให้คะแนนแบบแยกองค์ประกอบ

ระดับคุณภาพ	การตอบคำถามจากเรื่องที้อ่าน
ระดับ 3 (ดี)	ตอบคำถามถูกต้องทุกข้อ
ระดับ 2 (พอใช้)	ตอบคำถามผิดไม่เกิน 3 ข้อ จาก 5 ข้อ
ระดับ 1 (ควรปรับปรุง)	ตอบคำถามผิดมากกว่า 3 ข้อ จาก 5 ข้อ

รูปแบบที่ 3 การตรวจให้คะแนนแบบผสมผสาน (Annotated Scoring Rubrics) จะรวมแบบรวมองค์ประกอบและแบบแยกองค์ประกอบไว้ด้วยกัน เริ่มด้วยการตรวจให้คะแนนในภาพรวม ของการปฏิบัติงานด้วยแบบรวมองค์ประกอบ แล้วเลือกให้คะแนนอีกเพียงบางคุณลักษณะของงานแบบ แยกองค์ประกอบ

ขั้นที่ 5 กำหนดค่าระดับคุณภาพในการประเมิน ค่าระดับคุณภาพ คือ ตัวเลขที่บ่งบอกถึงคะแนนการประเมินผลการเรียนรู้ของนักเรียนซึ่งผู้สอนเป็นผู้กำหนด หรืออาจจะให้นักเรียนมีส่วนร่วมในการกำหนดค่าระดับคุณภาพก็ได้ ตัวเลขค่าระดับคุณภาพอาจจะมีระดับ 0 – 1 – 2 – 3 หรือ 0 – 3 – 4 – 5 – 6 หรือ 0 – 10 – 15 – 20 ในกรณีที่กำหนดค่าระดับต่ำสุดที่เลข 0 นั้น หมายถึงนักเรียนไม่มีผลการเรียนรู้หรือไม่มีผลงานเข้ารับการประเมิน หรือไม่ได้ดำเนินการตามรายการประเมินนั้น และการกำหนดค่าระดับจาก 0 – 10 แสดงว่า ผู้กำหนดค่าระดับคุณภาพ พิจารณาแล้วเห็นว่า คุณภาพของงานหรือสิ่งที่ประเมินควรมีคะแนนสูงสุดคือ 10

ขั้นที่ 6 บรรยายคุณภาพการประเมินแต่ละระดับ เมื่อได้เลือกรูปแบบในการกำหนดเกณฑ์การตรวจให้คะแนน และกำหนดตัวเลขค่าระดับแล้วจะต้องเขียนคำอธิบายขอบข่ายการพิจารณาตัดสินให้คะแนนแต่ละระดับคุณภาพให้ชัดเจน ผลงานหรือกระบวนการในลักษณะไหน อย่างไร จึงจะมีคุณภาพระดับ 4 ลักษณะไหนอย่างไรจึงมีคุณภาพระดับ 3 ระดับ 2 และระดับ 1 คำบรรยายคุณภาพแต่ละระดับจะเกิดประโยชน์สูงสุด

ตาราง 3 คำอธิบายคุณภาพ “การพูดอภิปราย”

ระดับคุณภาพ	คำอธิบาย
4 ดีมาก	หมายถึง มีความคิดริเริ่มสร้างสรรค์ มีการนำเสนอข้อเท็จจริงและแสดงเหตุผลอย่างเหมาะสม มีการยกตัวอย่างเพื่อสนับสนุนประเด็นอภิปราย การออกเสียงถูกต้องชัดเจน มีมารยาทในการพูด และรักษาเวลาได้
3 ดี	หมายถึง มีความคิดริเริ่มสร้างสรรค์ มีการนำเสนอข้อเท็จจริงและแสดงเหตุผลอย่างเหมาะสม แต่ขาดตัวอย่างเพื่อสนับสนุนประเด็นอภิปราย ส่วนการออกเสียงถูกต้องชัดเจน มีมารยาทในการพูด และรักษาเวลาได้ดี
2 ใช้ได้	หมายถึง ไม่มีความคิดริเริ่มสร้างสรรค์ มีการนำเสนอข้อเท็จจริงและข้อคิดเห็น แต่ขาดตัวอย่างเพื่อสนับสนุนประเด็นอภิปราย มีการออกเสียงถูกต้องชัดเจน แต่ไม่รักษาเวลาและมารยาทในการพูด
1 ใช้ไม่ได้	หมายถึง มีการนำเสนอข้อเท็จจริงและข้อคิดเห็น แต่ขาดตัวอย่างเพื่อสนับสนุนประเด็นอภิปราย การออกเสียงไม่ถูกต้องชัดเจน ขาดการรักษาเวลาและมารยาทในการพูด
0	หมายถึง ไม่มีการอภิปราย

ขั้นที่ 7 กำหนดคะแนนการตัดสินระดับคุณภาพ การกำหนดคะแนนการตัดสินระดับคุณภาพเป็นการกำหนดช่วงคะแนนจากการประเมินผลการเรียนรู้

ตาราง 4 การกำหนดคะแนนการตัดสินระดับคุณภาพ

คะแนน	ระดับคุณภาพ
9 – 12	ดี
5 – 8	พอใช้
1 – 4	ควรปรับปรุง

จากความหมาย การแบ่งประเภท และวิธีการสร้างเกณฑ์ ที่ได้กล่าวไว้ข้างต้นจะพบว่าประเภทของเกณฑ์การตรวจให้คะแนนส่งผลให้เกิดความคลาดเคลื่อนต่อคะแนนการเขียนเรียงความ ดังนั้นผู้วิจัยจึงนำวิธีการตรวจให้คะแนน ทั้ง 3 รูปแบบ คือ การตรวจให้คะแนนแบบแยกองค์ประกอบ (Analytic Scoring Rubrics) การตรวจให้คะแนนแบบรวมองค์ประกอบ (Holistic Scoring Rubrics) และการตรวจให้คะแนนแบบผสมผสาน (Annotated Scoring Rubrics) มาสร้างเป็นวิธีการตรวจให้คะแนนเรียงความ และวิธีการตรวจให้คะแนนจดหมาย

1.4 โมเดลการวัดความสามารถในการตรวจให้คะแนนการเขียน

ในส่วนนี้ ผู้วิจัยจะกล่าวถึงโมเดลที่เหมาะสมที่ใช้ในการวัดความสามารถในการเขียนของนักเรียน ตลอดจนการใช้สถิติต่างๆ ที่ใช้วัดความคลาดเคลื่อนจากการตรวจให้คะแนนของผู้ตรวจให้คะแนน ดังนี้

การวัดทางการศึกษาและจิตวิทยาเป็นการวัดคุณลักษณะภายในของมนุษย์ ซึ่งไม่สามารถสังเกตได้โดยตรง แต่มีความสำคัญและจำเป็นต้องศึกษาเพราะการวัดคุณลักษณะภายในจะทำให้เข้าใจการเกิดพฤติกรรมภายนอกของมนุษย์ที่สามารถสังเกตได้โดยตรง อันจะนำไปสู่การทำนาย ควบคุม และพัฒนาพฤติกรรมมนุษย์ การวัดคุณลักษณะภายในของมนุษย์จำเป็นต้องอาศัยทฤษฎีการทดสอบ เพื่อทำความเข้าใจคุณลักษณะของสิ่งที่มุ่งวัดโครงสร้างของการวัดและการพัฒนาเครื่องมือสำหรับทดสอบ ทฤษฎีการทดสอบเป็นองค์ความรู้ที่มีนัยทั่วไปเกี่ยวกับการทดสอบ วิธีการแก้ปัญหา การทดสอบและพัฒนาเครื่องมือการสอบ ซึ่งจะช่วยให้นักวัดผลสามารถทำการสร้าง และพัฒนาแบบสอบให้มีคุณภาพ สามารถแปลความหมายผลการวัดได้อย่างถูกต้องและสามารถนำเสนอสารสนเทศไปใช้สำหรับการตัดสินใจได้อย่างเหมาะสม

ทฤษฎีการทดสอบแบบมาตรฐานเดิม (Classical Test Theory) ถูกใช้เป็นแนวคิดพื้นฐานสำหรับการพัฒนาแบบทดสอบทางการศึกษาและจิตวิทยาเป็นส่วนใหญ่ตลอดช่วงศตวรรษที่ 20 ที่ผ่านมา ในปัจจุบันก็ยังมีการใช้กันอย่างต่อเนื่องกว้างขวาง แต่ด้วยข้อจำกัดหลายประการของการวัดตามทฤษฎีการทดสอบแบบมาตรฐานเดิม ทำให้มีนักทฤษฎีทางการทดสอบหลายท่านได้ก่อกระแสของการวัดคุณลักษณะภายในของบุคคลแนวใหม่ เพื่อแก้ไขจุดอ่อนเกี่ยวกับข้อตกลงเบื้องต้นของทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory) (ศิริชัย กาญจนวาสี. 2550: 5) จนมีพัฒนาการของการทดสอบแนวใหม่ ซึ่ง

สามารถจำแนกออกเป็น 2 แนวทางที่สำคัญ ได้แก่ ทฤษฎีการสรุปอ้างอิง (Generalizability Theory) หรือเขียนย่อว่า G-Theory และทฤษฎีการตอบสนองข้อสอบ (Item Response Theory) หรือเขียนย่อว่า IRT แต่ละแนวทางมีแนวคิด ดังนี้

ทฤษฎีการทดสอบแบบมาตรฐานเดิม ถือว่า คะแนนความคลาดเคลื่อนของการวัดมีลักษณะเป็นหนึ่งเดียวที่แบ่งแยกไม่ได้ (Unique error) จึงมีข้อจำกัดสำหรับการศึกษาค่าความเชื่อมั่นของแบบวัดหรือแบบทดสอบที่จะต้องทำการวิเคราะห์ภายใต้แหล่งความคลาดเคลื่อนครั้งละ 1 แหล่ง เช่น การศึกษาความคลาดเคลื่อนจากช่วงเวลาของการทดสอบ สามารถวิเคราะห์ได้จากความเชื่อมั่นของการสอบซ้ำ (Test-Retest Reliability) การศึกษาความคลาดเคลื่อนจากความยาวของข้อสอบ สามารถวิเคราะห์ได้จากความเชื่อมั่นของ สเปียร์แมน-บราวน์ (Spearman-Brown Formula) เป็นต้น

สำหรับ ทฤษฎีการสรุปอ้างอิง ถือว่า คะแนนความคลาดเคลื่อนของการวัดสามารถเกิดจากหลายแหล่ง (Multiple Sources of Error) ที่สามารถประมาณค่าแยกกันภายใต้การวิเคราะห์ครั้งเดียวกันได้ ทฤษฎีนี้ต้องการสรุปอ้างอิงความน่าเชื่อถือของคะแนนที่สังเกตได้ จากแบบสอบที่ทำการวัดในเงื่อนไขหรือสถานการณ์ต่างๆ ของการทดสอบไปยังคะแนนจริง เงื่อนไขหรือสถานการณ์ของการทดสอบที่จะใช้มีได้หลากหลายขึ้นอยู่กับกรอบแบบ (Design) เช่น อาจกำหนดเงื่อนไขเป็นความยาวของแบบสอบ รูปแบบของข้อสอบ จำนวนครั้งของการสอบ จำนวนผู้ตรวจ เป็นต้น G-Theory เสนอโมเดลวิเคราะห์ค่าสัมประสิทธิ์ความน่าเชื่อถือของผลการวัด (G-Coefficient) ซึ่งเปรียบเสมือนค่าความเชื่อมั่นของแบบวัดหรือแบบสอบโดยสามารถแปลความหมายได้ 2 ลักษณะ ได้แก่ ความน่าเชื่อถือของผลการวัดสำหรับใช้ตัดสินเชิงเปรียบเทียบ (Relative Decision) และความน่าเชื่อถือของผลการวัดสำหรับใช้ตัดสินเชิงสัมบูรณ์ (Absolute Decision) โดยมีข้อตกลงเบื้องต้นว่า คุณลักษณะที่มุ่งวัดของบุคคลอยู่ในสภาวะที่คงที่ (Steady State) ความผันแปรของผลการวัดเกิดจากแหล่งความคลาดเคลื่อนของการวัดอย่างน้อย 1 แหล่ง และวุฒิภาวะหรือการเรียนรู้ระหว่างการวัดไม่มีผลต่อความผันแปรของคะแนนที่ได้และความแปรปรวนของคะแนนที่สังเกตได้ประกอบด้วยความแปรปรวนของคะแนนจริง ความแปรปรวนของคะแนนความคลาดเคลื่อนที่เป็นระบบ และความแปรปรวนของคะแนนความคลาดเคลื่อนสุ่ม (ศิริชัย กาญจนวาสี, 2550: 6)

ทฤษฎีการทดสอบแบบมาตรฐานเดิม ตั้งอยู่บนฐานของข้อตกลงเบื้องต้นที่สำคัญคือ ค่าพารามิเตอร์ของข้อสอบเป็นค่าเฉพาะของกลุ่มผู้สอบและเป็นอิสระจากความสามารถของผู้สอบความคลาดเคลื่อนมาตรฐานของการวัด (Standard Error of Measurement) เป็นค่าเฉพาะประชากร และเป็นค่าเดียวคงที่สำหรับทุกคนที่เป็นสมาชิกของประชากรนั้น รวมทั้งการเปรียบเทียบคะแนนหรือคุณภาพของแบบสอบสามารถกระทำได้อีกต่อเมื่อแบบสอบมีคุณสมบัติเป็นแบบสอบคู่ขนาน (Parallel Forms)

สำหรับ ทฤษฎีการตอบสนองข้อสอบ ได้เสนอแนวคิดที่ว่า ความน่าจะเป็นของการตอบข้อสอบได้ถูกต้องขึ้นอยู่กับความสามารถจริงของผู้สอบ และคุณลักษณะของข้อสอบอันประกอบด้วยพารามิเตอร์ความยาก อำนาจจำแนกและโอกาสการเดาข้อสอบได้ถูก ระบบความสัมพันธ์ดังกล่าว สามารถแสดงด้วยโมเดลการตอบสนองข้อสอบ ซึ่งอาจเป็นโมเดล 1 พารามิเตอร์ โมเดล 2 พารามิเตอร์ หรือโมเดล 3 พารามิเตอร์ ทฤษฎีนี้ถือว่าค่าพารามิเตอร์ของข้อสอบและความสามารถจริงของผู้สอบมีความสัมพันธ์กัน การประมาณค่าพารามิเตอร์ของข้อสอบ ได้แก่ ค่าความยาก ค่าอำนาจจำแนก และค่าความน่าจะเป็น

ในการเดาข้อสอบได้ถูก จึงต้องพิจารณาร่วมกับความสามารถจริงของผู้ตอบ ดังนั้นเมื่อได้กลุ่มผู้ตอบขนาดใหญ่ที่เป็นตัวแทนประชากรการประมาณค่าพารามิเตอร์ของข้อสอบจะต้องกระทำพร้อมๆ ไปด้วยกับการประมาณค่าความสามารถจริงของผู้สอบ จึงจะทำให้ได้ค่าพารามิเตอร์ที่มีนัยทั่วไป มีความน่าเชื่อถือและไม่แปรผันไปตามความสามารถของกลุ่มผู้สอบในด้านความคลาดเคลื่อนของการวัด ทฤษฎีการตอบสนองข้อสอบสามารถวิเคราะห์ถึงความคลาดเคลื่อนในการวัดของข้อสอบแต่ละข้อและแบบสอบทั้งฉบับ จำแนกตามระดับความสามารถจริงของผู้ตอบ นั่นคือ ข้อสอบแต่ละข้อ รวมทั้งแบบสอบแต่ละฉบับ สามารถมีคุณภาพในด้านการให้สารสนเทศ (ความเชื่อมั่นและความเที่ยงตรง) แตกต่างกันได้สำหรับผู้สอบที่มีระดับความสามารถจริงแตกต่างกัน สำหรับการเปรียบเทียบคะแนนระหว่างแบบสอบ ทฤษฎีนี้ไม่มีข้อจำกัดว่าแบบสอบจะต้องเป็นแบบสอบคู่ขนานกัน และมีเทคนิคที่สามารถเปรียบเทียบคะแนนของแบบสอบต่างฉบับที่มีความยากต่างกันได้อย่างมีประสิทธิภาพ โดยมีข้อตกลงว่า แบบสอบมุ่งวัดคุณลักษณะเดียว (Unidimensionality) มีความเป็นอิสระระหว่างข้อสอบ (Independence) โมเดลการตอบสนองข้อสอบมีรูปแบบเป็นฟังก์ชันโลจิสติก (Logistic Function) และแบบสอบที่ใช้ต้องไม่เป็นแบบสอบประเภทความเร็ว (Speed Test) (ศิริชัย กาญจนวาสี. 2550: 7 – 8)

จากที่ได้กล่าวมาข้างต้น และงานวิจัยที่ผ่านมา จะพบว่า โมเดลการวัดในทฤษฎีการตอบสนองข้อสอบที่เหมาะสมในการใช้ในการตรวจให้คะแนนเรียงความคือ พาเชียลเครดิตโมเดล (Partial Credit Model) เนื่องจากมาสเตอร์ (Masters. 1982; อ้างอิงจาก ศิริชัย กาญจนวาสี. 2550) ได้พัฒนาโมเดลนี้ให้สามารถใช้กับแบบสอบที่มีหลายตัวเลือก และการตรวจให้คะแนนตัวเลือกแต่ละตัวแตกต่างกัน โมเดลนี้ได้รับการพัฒนาให้ดีขึ้นโดย ไรท์ และ มาสเตอร์ (ศิริชัย กาญจนวาสี. 2550) นักวัดผลการศึกษาทั้งสองยังได้ประมวลโมเดลโลจิสติกหนึ่งพารามิเตอร์ หรือ Rasch model แบบต่างๆ ที่มีการพัฒนาขึ้น รวมเรียกว่า โมเดลการวัดขั้นพื้นฐาน (Fundamental Measurement Model) ซึ่งมีคุณสมบัติสำคัญคือ การแยกพารามิเตอร์ผู้สอบ และพารามิเตอร์ข้อคำถามและมีการใช้สถิติที่เพียงพอในการประมาณค่าพารามิเตอร์ของโมเดล ซึ่งรายละเอียดในโมเดลนี้ มีดังนี้

มาสเตอร์ (ศิริชัย กาญจนวาสี. 2550: 96; อ้างอิงจาก Masters. 1982) ได้พัฒนา พาเชียลเครดิตโมเดล Partial Credit Model (PCM) โดยการขยายแนวคิดมาจาก โมเดลของราล์ซ (Rasch Model) ซึ่งเป็นโมเดลที่ใช้ในการประมาณค่าพารามิเตอร์สำหรับข้อสอบที่ให้คะแนนแบบสองค่า (0 – 1) สำหรับใช้ในการวิเคราะห์ข้อสอบหรือข้อคำถามที่มีกระบวนการตอบหลายลำดับขั้น ซึ่งจำเป็นต้องมีการตรวจให้คะแนนการตอบถูกต้องหรือตอบถูกบางส่วนในแต่ละลำดับขั้นของกระบวนการตอบ เช่น ข้อสอบวัดผลสัมฤทธิ์ทางคณิตศาสตร์ ซึ่งมีการตรวจให้คะแนนแบบคำตอบถูกบางส่วน แบบวัดเจตคติ บุคลิกภาพ เชาวน์ปัญญา ที่มีการตรวจให้คะแนนคำตอบเป็นลำดับขั้น

พาเชียลเครดิตโมเดล มีลักษณะเป็นโมเดลที่พัฒนาขยายต่อจากโมเดลการตอบสนองข้อสอบที่มี 1 พารามิเตอร์ (1-Parameter Model) จึงมีลักษณะพารามิเตอร์มาตรฐานคล้ายโมเดลของราล์ซ และใช้หลักการคำนวณความน่าจะเป็นของการตอบแต่ละระดับขั้นการตอบโดยตรงแบบขั้นตอนเดียว (Direct IRT Method) ลักษณะข้อคำถามแต่ละข้อ (i) อธิบายด้วยค่าพารามิเตอร์ระดับความยากระหว่าง

รายการคำตอบที่อยู่ถัดไป (Item Step Difficulty, δ_{ij}) โดยทุกข้อมีค่าพารามิเตอร์ความชันเท่ากัน (Equal Slope) สมมุติว่าข้อคำถาม i มีคะแนน $x = 0, 1, \dots, m_i$ โดยมีจำนวนรายการคำตอบ $K_i = m_i + 1$ เช่น รายการข้อคำถามที่มี 4 รายการคำตอบ ไม่เคย (0) น้อย (1) ปานกลาง (2) มาก (3) เป็นต้น ในการตอบคำถามข้อนี้ ผู้ตอบจะเลือกรายการสูงสุดต้องคิด 3 ขั้นตอนโดยทำการตัดสินใจ ขั้นที่ 1 เลือกระหว่างไม่เคย (0) กับน้อย (1) ขั้นที่ 2 เลือกระหว่างน้อย (1) กับปานกลาง(2) และขั้นที่ 3 เลือกระหว่างปานกลาง (2) กับมาก (3) ค่า δ_{ij} ไม่ได้บอกถึง θ ของผู้ตอบที่มีโอกาส 0.50 ในการเลือกรายการคำตอบที่อยู่เหนือถัดไปอย่างโมเดล GRM แต่ δ_{ij} แสดงถึงระดับความยากสัมพัทธ์ของแต่ละขั้นการตอบ ภายในคำถามข้อเดียวกัน แต่ละขั้นตอนอาจง่ายกว่าหรือยากกว่าบางขั้นตอน δ_{ij} ยังแสดงถึงตำแหน่ง θ ที่ต้องการเลือกรายการคำตอบตัดกัน จึงบ่งบอกถึงผู้ตอบที่มีคุณลักษณะ θ จะมีโอกาสเลือกรายการคำตอบใดมากกว่ากัน

สำหรับ $x = j$ รายการคำตอบโค้งฟังก์ชันการเลือกรายการคำตอบของโมเดล พาเซียลเครดิตโมเดลเขียนได้ดังนี้ (ศิริชัย กาญจนวาสี. 2550: 97)

$$P_{ix}(\theta) = \frac{\exp\left[\sum_{j=0}^x (\theta - \delta_{ij})\right]}{\sum_{r=0}^{m_i} \left[\exp\left[\sum_{j=0}^r (\theta - \delta_{ij})\right]\right]}$$

$$\text{เมื่อ } \sum (\theta - \delta_{ij}) \equiv 0$$

$P_{ix}(\theta)$ แทน ความน่าจะเป็นผู้ตอบซึ่งมีคุณลักษณะ θ จะตอบข้อ i ด้วยการเลือกหรือสามารถทำรายการคำตอบขั้นที่ x จากจำนวน m_i ขั้น (Step)

δ_{ij} แทน ค่าพารามิเตอร์ระดับความยากของขั้นตอนที่ j ในข้อ i (Item step Difficulty) เมื่อ $j = 1, 2, \dots, m_i$ ค่าที่สูงแสดงถึงขั้นการตอบนั้นมีความยากสูงกว่าขั้นอื่น

ความสัมพันธ์ระหว่างคุณลักษณะ θ กับระดับขั้นของรายการคำตอบแต่ละข้อคำถามตามโมเดล พาเซียลเครดิตโมเดลสามารถนำมาคำนวณคะแนนที่คาดหวัง (Expected score) หรือคะแนนจริง (True score) ของแต่ละข้อคำถามได้ตามสูตร ดังนี้

$$E(X) = \sum_{x=0}^{m_i} xP_x(\theta)$$

โค้งของฟังก์ชัน ดังกล่าวแสดงถึงการแจกแจงคะแนนรายชื่อที่คาดหวังของผู้ตอบ ที่มีคุณลักษณะ θ เมื่อรวมโค้งของทุกข้อเข้าด้วยกันทั้งฉบับ จะทำให้ได้โค้งการแจกแจงคะแนนดิบรวม ที่คาดหวังของผู้ตอบที่มีคุณลักษณะ θ ข้อดีของพาเซี่ยลเครดิตโมเดล ซึ่งคล้ายกับโมเดลของ ราล์ช คือ คะแนนดิบที่คาดหวังเป็นค่าสถิติที่เพียงพอสำหรับการคำนวณค่าของคุณลักษณะ θ ดังนั้นคะแนนดิบของข้อคำถามชุดเดียวกันที่เป็นไปตามโมเดล พาเซี่ยลเครดิตโมเดล จะสอดคล้องหรือสมมูลกับตำแหน่งคุณลักษณะของผู้ตอบบนสเกลของ θ แต่อย่างไรก็ตามคงต้องอยู่บนพื้นฐานว่าข้อคำถามทุกข้อจะต้องมีความสัมพันธ์ที่เท่าเทียมกันกับคุณลักษณะ θ ดังนั้นโมเดลนี้จึงเหมาะสมสำหรับการวัดความสามารถในการเขียนเรียงความ

เกณฑ์ที่ใช้ในการตัดสินพารามิเตอร์ต่าง ๆ ของข้อสอบ

ในขั้นแรกของการประมาณค่าพารามิเตอร์คือต้องเปรียบเทียบความเหมาะสมกลมกลืนของข้อมูลเชิงประจักษ์กับโมเดล Partial Credit Model เนื่องจากเป็นดัชนีที่บ่งชี้ว่า ข้อมูลเชิงประจักษ์ที่เก็บข้อมูลมาได้ นั้น สอดคล้องกลมกลืนกับโมเดลหรือไม่ โดยพิจารณาค่า Weighted mean square residual : INFIT และ Unweighted mean square residual : OUTFIT ที่กำหนดขอบเขตของการยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลอยู่ระหว่าง 0.5 ถึง 1.5 (Wright, BD; & Linacre, JM. 1994: 370)

1. พารามิเตอร์ของนักเรียน (Student Ability : θ) ระดับความสามารถเฉลี่ยของนักเรียนอยู่ในวิธียอมรับได้คือ -3 ถึง +3 (ศิริชัย กาญจนวาสี. 2550: 55) และค่าความเชื่อมั่นแยกส่วนของผู้สอบ (Person Separation Reliability; Person Reliability) ซึ่งคล้ายคลึงกับสัมประสิทธิ์แอลฟา (Englehard. 2002; Myford; & Wolfe. 2003) เป็นอัตราส่วนของความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนสังเกตในการประมาณค่าความสามารถของผู้สอบ ถ้าค่านี้มีค่าสูงแสดงว่ามีความแตกต่างระหว่างความสามารถของผู้สอบ แสดงถึงความสามารถในการจำแนกความสามารถของผู้สอบแต่ละคนออกจากกัน เป็นดัชนีที่แสดงถึงความเชื่อถือได้ของคะแนนความสามารถของผู้สอบที่ได้จากข้อสอบและผู้ตรวจ

2. พารามิเตอร์ของข้อสอบ (Step Difficulty : δ) หรือจุดโค้งการตอบที่ตัดกัน โดยที่ค่าความยากในการเปลี่ยนขั้น ควรเรียงลำดับจากน้อยไปมาก และการทดสอบความสอดคล้องของโมเดลของการตรวจให้คะแนนเรียงความตามโปรแกรม FACETS จะให้พิจารณาค่า Quality Control ในตาราง (Linacer. 2012: 194) ที่ต้องมีค่าเรียงลำดับจากน้อยไปมากโดยไม่ข้ามลำดับและไม่ติด " * " ใดๆ จากการคำนวณของโปรแกรม ถ้าพบว่า ค่าความยากในการเปลี่ยนขั้นของวิธีการตรวจให้คะแนนเรียงความเป็นไปตามเงื่อนไข แสดงว่ามีความเที่ยงตรงในการแบ่งขั้น (Validity of the Categorization) และข้อมูลเชิงประจักษ์สอดคล้องกับโมเดล Partial Credit Model

3. พารามิเตอร์ของผู้ตรวจให้คะแนน (Rater Severity) ให้พิจารณาค่าความเชื่อมั่นแยกส่วนของผู้ตรวจ (Rater Separation Reliability) ซึ่งแสดงถึงความแปรปรวนของการให้คะแนนที่ไม่ต้องการให้เกิดขึ้นระหว่างผู้ตรวจให้คะแนนในระดับของความเข้มงวด/ใจดี ซึ่งดัชนีนี้ไม่ได้บ่งชี้ว่ามีความสอดคล้องกันระหว่างผู้ตรวจให้คะแนน แต่บ่งชี้ว่าผู้ตรวจให้คะแนนมีความเข้มงวดในระดับเดียวกัน

หรือไม่ กล่าวอีกนัยหนึ่งความเชื่อมั่นแยกส่วนของผู้ตรวจให้คะแนนซึ่งถึงความเชื่อมั่นของความแตกต่าง (Reliably Different) มากกว่าระดับความเชื่อมั่นที่คล้ายคลึงกัน (Reliably Similar)

ทฤษฎีการสรุปอ้างอิง (Generalizability Theory : G-Theory) เป็นทฤษฎีทางสถิติของการวิเคราะห์ความน่าเชื่อถือของผลการวัดในสถานการณ์ของการวัดผลลักษณะต่างๆ ที่เป็นเป้าหมายของการนำเครื่องมือไปใช้ ความน่าเชื่อถือของผลการวัด หมายถึง ความถูกต้องของการสรุปอ้างอิง (Generalization) จากคะแนนที่สังเกตได้ไปยังคะแนนจริงของบุคคล โดยคะแนนจริงเป็นคะแนนที่พึงได้ของผู้สอบแต่ละคน การสอบภายใต้สถานการณ์หรือเงื่อนไขของการวัดที่ยอมรับได้ทั้งหมด ตามแนวคิดของ G-Theory คะแนนที่ได้จากการวัดภายใต้สถานการณ์หรือเงื่อนไขเดียวด้วยแบบทดสอบชุดเดียว และทำการสอบครั้งเดียว คะแนนที่ได้จะไม่สามารถเชื่อถือได้เต็มที่เพราะคะแนนที่ได้จะไม่เป็นตัวแทนที่ดีของคะแนนจริง ซึ่งเป็นคะแนนเฉลี่ยที่ได้จากการทดสอบในหลายๆ สถานการณ์ด้วยแบบทดสอบหลายๆ ชุด และทำการทดสอบหลายๆ ครั้ง

ผู้สอบคนเดียวกันเมื่อทำการทดสอบหลายๆ ครั้งในหลายสถานการณ์มักจะจะได้คะแนนในแต่ละครั้งแตกต่างกัน ความไม่คงเส้นคงวาของคะแนนเกิดจากความคลาดเคลื่อนของการวัด ซึ่งทฤษฎีการทดสอบแบบมาตรฐานเดิมให้คำตอบด้วยการแยกวิเคราะห์สำหรับการทดสอบแต่ละครั้ง โดยพิจารณาว่าความคลาดเคลื่อนของการวัดมาจากแหล่งเดียว (Single Source of Error) เช่น ความคลาดเคลื่อนมาจากการทิ้งช่วงระยะเวลาของการสอบซ้ำก็จะคำนวณค่าความเชื่อมั่นแบบ Test-Retest Method ถ้าความคลาดเคลื่อนมาจากชุดของแบบทดสอบคู่ขนานก็จะคำนวณค่าความเชื่อมั่นแบบ Equivalent Forms Method ถ้าความคลาดเคลื่อนมาจากความยาวของแบบทดสอบก็จะคำนวณค่าความเชื่อมั่นด้วยสูตรของ Spearman-Brown Formula เป็นต้น ในขณะที่ G-Theory ได้พิจารณาถึงแหล่งความคลาดเคลื่อนพหุ (Multiple Sources of Error) ที่สามารถวิเคราะห์ความคลาดเคลื่อนจากหลายแหล่งพร้อมกัน ผู้บริหารการทดสอบจึงสามารถตัดสินใจได้ว่าควรใช้สถานการณ์หรือเงื่อนไขการวัดแบบใด จึงจะได้คะแนนที่เชื่อถือถึงระดับที่ต้องการ ผลการวิเคราะห์ด้วย G-Theory จะให้ค่าสัมประสิทธิ์เชิงสรุปที่แสดงถึงระดับความน่าเชื่อถือของคะแนนที่ได้จากการวัด (Level of Dependability) เรียกว่า สัมประสิทธิ์การสรุปอ้างอิง ความน่าเชื่อถือของผลการวัด (Generalizability Coefficient) ซึ่งคล้ายกับสัมประสิทธิ์ความเชื่อมั่น (Reliability Coefficient) ในทฤษฎีการทดสอบแบบมาตรฐานเดิม (ศิริชัย กาญจนวาสี. 2550: 13 – 14)

การตรวจสอบความน่าเชื่อถือของผลการวัดนั้นทำให้ผู้บริหารการสอบตัดสินใจความน่าเชื่อถือของผลการวัดได้ 2 ลักษณะ

1. ความน่าเชื่อถือของคะแนนในการนำผลไปใช้ในเชิงสัมพัทธ์ (Relative) เพื่อแปลผลเปรียบเทียบระหว่างบุคคล เช่น นาย ก มีความสามารถสูงเพียงใด เมื่อเทียบกับกลุ่มผู้สอบด้วยกัน หรือ ก มีความสามารถสูงหรือต่ำกว่าบุคคลในกลุ่มเพียงใด หรือ ก ได้คะแนนคิดเป็นคะแนนมาตรฐานหรือเปอร์เซ็นต์ไทล์เท่าไร เป็นต้น (ศิริชัย กาญจนวาสี. 2550: 15)

สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมบูรณ์ (P_{Abs}) เมื่อคะแนนความคลาดเคลื่อนเป็นความแปรปรวนของคะแนนความคลาดเคลื่อนสัมบูรณ์ สัมประสิทธิ์นี้บ่งบอก

ความเชื่อมั่นของแบบสอบ ในสถานการณ์ของการตัดสินใจที่ขึ้นกับคะแนนของผู้สอบตามลำพัง ไม่มี การเปรียบเทียบภายในกลุ่มหรือระหว่างกลุ่ม เช่น ความเชื่อมั่นของแบบสอบอิงเกณฑ์ ซึ่งใช้แสดง ความน่าเชื่อถือสำหรับตรวจสอบความสามารถของผู้สอบเทียบกับเกณฑ์ที่กำหนดไว้ล่วงหน้า เป็นต้น (ศิริชัย กาญจนวาสี. 2550: 18)

สำหรับการตัดสินใจเชิงสัมบูรณ์ (Absolute Decision) องค์กรประกอบความแปรปรวน รวมทั้งหมด ยกเว้นองค์ประกอบความแปรปรวนของบุคคล ($\sigma^2_{P_i}$) ถือเป็นความแปรปรวนของคะแนน ความคลาดเคลื่อนแบบสัมบูรณ์ (σ^2_{Abs}) (ศิริชัย กาญจนวาสี. 2550: 25)

2. ความน่าเชื่อถือของคะแนนในการนำผลไปใช้ในเชิงสัมบูรณ์ (Absolute) เพื่อ แปลผลเปรียบเทียบกับเกณฑ์มาตรฐาน เช่น นาย ก มีความสามารถอยู่ระดับใด สูงหรือต่ำเพียงใด ตามเกณฑ์มาตรฐาน หรือ ก มีความสามารถผ่านระดับมาตรฐานหรือไม่ โดยไม่ได้สนใจเปรียบเทียบกับ ความสามารถของเพื่อนในกลุ่ม เป็นต้น (ศิริชัย กาญจนวาสี. 2550: 15)

สัมประสิทธิ์การสรุปอ้างอิงสำหรับการตัดสินใจเชิงสัมพัทธ์ (ρ_{Rel}^2) เมื่อคะแนน ความคลาดเคลื่อนเป็นความแปรปรวนของคะแนนความคลาดเคลื่อนสัมพัทธ์ สัมประสิทธิ์นี้บ่งบอกความเชื่อมั่น ของแบบสอบ ในสถานการณ์ของการตัดสินใจที่มีการเปรียบเทียบคะแนนระหว่างผู้สอบ เช่น ความเชื่อมั่น ของแบบสอบอิงกลุ่ม ซึ่งใช้แสดงความน่าเชื่อถือสำหรับเปรียบเทียบคะแนนระหว่างผู้สอบด้วยการเข้าโค้ง การแจกแจงคะแนนและการตัดเกรด หรือการสอบแข่งขันหาผู้สอบที่ได้คะแนนสูงเมื่อเทียบกับกลุ่ม เพื่อ เข้าศึกษาต่อในสถานศึกษาต่างๆ เป็นต้น (ศิริชัย กาญจนวาสี. 2550: 19)

การตัดสินใจเชิงสัมพัทธ์ (Relative Decision) องค์กรประกอบความแปรปรวน ที่เกี่ยวข้องกับการเปรียบเทียบตำแหน่งสัมพัทธ์กับบุคคล ซึ่งได้แก่ องค์กรประกอบความแปรปรวนที่แสดง ปฏิสัมพันธ์ระหว่างเงื่อนไขของการวัดกับบุคคล เช่น π_i , ρ_o , ρ_{io} เมื่อ i คือ ความยาวของข้อสอบ และ o คือ จำนวนครั้งของการสอบ เป็นต้น ผลรวมขององค์ประกอบความแปรปรวนดังกล่าว ถือเป็น ความแปรปรวนของคะแนนความคลาดเคลื่อนแบบสัมพัทธ์ (ศิริชัย กาญจนวาสี. 2550: 25)

เงื่อนไขการวัดที่ผู้วิจัยต้องการวัดประกอบไปด้วย วิธีการตรวจให้คะแนนและ ผู้ให้คะแนน ดังนั้นผู้วิจัยจึงเลือกการประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงสำหรับการออกแบบการวัด ที่มีสองฟาเซต (G-Coefficient for Two Facet Design) แบบ Two-Facet Crossed Design มาใช้ในการ ประมาณค่าความเชื่อมั่น

การออกแบบเอกภพสำหรับสององค์ประกอบ (Two-Facet Design) แบบ Two-Facet Crossed Design

1. Population : ผู้สอบหรือบุคคล (P)
2. Universe : ความเป็นไปได้ทุกอย่างภายใต้เงื่อนไขการรวมกันของ ฟาเซต
3. Facet 1 : วิธีการตรวจให้คะแนน (I)
4. Facet 2 : ผู้ให้คะแนน (R)

5. Design : p X i X r design

G-Study คือ กลุ่มตัวอย่าง 50 คน ถูกให้คะแนนโดยใช้วิธีการตรวจให้คะแนน 3 วิธี โดยมีจำนวนผู้ตรวจให้คะแนน 3 คน

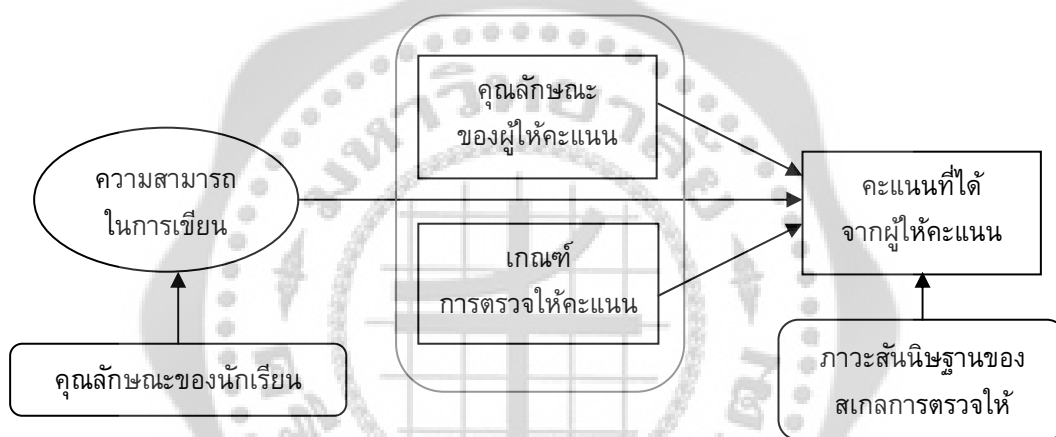
$$\sigma_{X_{pir}}^2 = \sigma_p^2 + \sigma_r^2 + \sigma_{pi}^2 + \sigma_{pr}^2 + \sigma_{ir}^2 + \sigma_{pir,e}^2$$

เมื่อ	$\sigma_{X_{pio}}^2$	แทน	ความแปรปรวนของคะแนนที่สังเกตได้
	σ_p^2	แทน	ความแปรปรวนของคะแนนเอกภพของผู้สอบ (μ_p)
	σ_i^2	แทน	ความแปรปรวนของคะแนนเฉลี่ยวิธีการตรวจให้คะแนน (μ_i)
	σ_r^2	แทน	ความแปรปรวนของคะแนนเฉลี่ยของผู้ตรวจให้คะแนน (μ_r)
	σ_{pi}^2	แทน	ความแปรปรวนของคะแนนปฏิสัมพันธ์ระหว่างผู้สอบกับวิธีการตรวจให้คะแนน
	σ_{pr}^2	แทน	ความแปรปรวนของคะแนนปฏิสัมพันธ์ระหว่างผู้สอบกับผู้ตรวจให้คะแนน
	σ_{ir}^2	แทน	ความแปรปรวนของคะแนนปฏิสัมพันธ์ระหว่างวิธีการตรวจให้คะแนนกับผู้ตรวจให้คะแนน
	$\sigma_{pir,e}^2$	แทน	ความแปรปรวนของคะแนนคลาดเคลื่อนหรือส่วนที่เหลือ (e_{pir})

ในการวิจัยครั้งนี้ผู้วิจัยใช้ค่าความเชื่อมั่น และดัชนีความสอดคล้องของผู้ตรวจให้คะแนนเพื่อเป็นหลักฐานแสดงคุณภาพของเครื่องมือที่ใช้ในการวิจัยตามทฤษฎีมาตรฐานเดิม หาค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ และค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ เพื่อแสดงคุณภาพของเครื่องมือตามทฤษฎีการสรุปอ้างอิง จากโปรแกรม EDUG 6.1-e ของสถาบัน IRDP (2010) ผู้วิจัยใช้การออกแบบเอกภพสำหรับสององค์ประกอบ แบบ Two-Facet Crossed Design โดยกำหนดให้ P คือ นักเรียน I คือ วิธีการตรวจให้คะแนน และ R คือ ผู้ตรวจให้คะแนน Design ที่ใช้สำหรับงานวิจัยนี้ คือ p X i X r design และหาค่าพารามิเตอร์ต่างๆ ตามทฤษฎีการตอบสนองข้อสอบ โดยในที่นี้จะวิเคราะห์โดยใช้ Partial Credit Model ซึ่งจะได้พารามิเตอร์ของนักเรียน พารามิเตอร์ของเรียงความ และพารามิเตอร์ของผู้ตรวจให้คะแนน จากโปรแกรม FACETS (Linacre. 2012)

ตอนที่ 2 คุณลักษณะของผู้ตรวจให้คะแนน

การตรวจสอบผู้ให้คะแนนที่ทำหน้าที่วัดประสิทธิภาพของการทำงานให้แก่คนทำงาน ได้มีการพิจารณาถึงการอบรมผู้ให้คะแนนที่ได้ใช้ประโยชน์เป็นการสนับสนุนวิธีการลดจำนวนของความไม่ลงรอยกันจากการตรวจให้คะแนนต่างๆ เนื่องมาจากธรรมชาติของการตรวจให้คะแนนเป็นการให้จากความคิดของบุคคลแต่ละบุคคล (Smith. 1986; & Woehr. 1994; citing Mulqueen; et al. 2000) การศึกษาส่วนมากได้ตรวจสอบแนวโน้มการตรวจให้คะแนนของผู้ให้คะแนน (Rater Tendencies) เช่น ความใจดี ความคล้อยตามกัน และความแม่นยำของผู้ให้คะแนน (Rater Accuracy) (Woehr; & Huffcutt. 1994; citing Mulqueen; et al. 2000) นอกจากนี้ยังมีผู้นำเสนอรูปแบบการวัดความสามารถในการเขียนเรียงความไว้ดังนี้



ภาพประกอบ 3 รูปแบบการวัดในการสอบเรียงความของ Community English Program (CEP)

รูปแบบนี้ใช้ครั้งแรกโดย แองเจิลฮาร์ท (Engelhard. 1992; citing Eckes. 2005) ในความพยายามที่จะระบุปัจจัยต่างๆ ที่มีอิทธิพลต่อการตรวจให้คะแนนแบบสังเกตในการตรวจสอบความสามารถด้านการเขียนโดยใช้เกณฑ์การตรวจให้คะแนนเชิงวิเคราะห์

นอกจากโมเดลที่ได้นำเสนอแล้วผู้ให้คะแนนยังได้ถูกพบว่า เป็นผลกระทบเชิงระบบของคะแนนจากการสอบ (Lunz; et al. 1991; Engelhard. 1994; & Fitzpatrick; Ercikan; Yen; & Ferrara. 1998; citing Yen, M.; & Fitzpatrick. 2006: 122) และการประเมินผลกระทบจากผู้ให้คะแนน (Rater Effect) สามารถนำไปใช้ประโยชน์สำหรับ 1) วัดผู้ให้คะแนนอย่างละเอียด 2) ระบุตัวผู้ให้คะแนนที่ให้คะแนนไม่ดี และ 3) แยกข้อสอบที่ยากออกจากผลกระทบจากผู้ให้คะแนน นอกจากนี้จากบันทึกของ แองเจิลฮาร์ท ปี 1994 ได้กล่าวไว้ว่า การพัฒนาความเข้าใจในพฤติกรรมของผู้ให้คะแนนจะช่วยให้พวกเราพัฒนาวิธีการที่มีทฤษฎีเป็นฐาน ไปสู่การประเมินผลกระทบจากผู้ให้คะแนนและความสำคัญของผลที่

ตามมาได้ โปรแกรม FACETS สร้างขึ้นมาโดย Linacre ในปี 1991 เป็นการสาธิตว่า จะสามารถขยายโมเดลของราสส์ ให้สามารถนำไปใช้ประเมินความเข้มงวดของผู้ให้คะแนนได้อย่างไร โดยใช้ Random Coefficient Multidimensional Logit Model มาใช้ โดยที่คะแนนของข้อสอบจากผู้ให้คะแนนได้ถูกจัดการให้เป็นเสมือน Multiple-Rater-Items และความไม่เป็นอิสระจากกันระหว่างการตรวจให้คะแนนไม่ได้ถูกนำมาพิจารณา

การวิเคราะห์ผู้ให้คะแนนโดยใช้โมเดลของราสส์หลายองค์ประกอบ (Many-Facet Rasch Model : MFRM) ถูกใช้เป็นทางเลือกต่อทฤษฎีการสรุปอ้างอิง (Generalizability Theory: G-Theory) การวิเคราะห์ MFRM และ G-theory ให้ข้อมูลเกี่ยวกับองค์ประกอบพาเซทและปฏิสัมพันธ์ซึ่งกันและกัน อย่างไรก็ตาม G-Theory แบ่งส่วนความผันแปรที่บ่งชี้ลักษณะต่อแต่ละองค์ประกอบพาเซท โดยการวิเคราะห์โครงสร้างความแปรปรวน และมุ่งเน้นกลุ่มที่เป็นหน่วยของการวิเคราะห์ เทคนิคการวิเคราะห์แบบราสส์ หลายองค์ประกอบมาจากทฤษฎีการตอบสนองข้อสอบที่วัดโมเดลตัวแปรแฝง และมุ่งเน้นองค์ประกอบแต่ละบุคคลของแต่ละความแปรปรวน ประโยชน์ของรูปแบบนี้ เป็นประโยชน์เนื่องจากให้ข้อมูลที่เป็นระดับบุคคลที่สามารถถูกใช้สำหรับจุดประสงค์ต่างๆ ในผลของการตอบสนองกลับโดยตรงต่อผู้ให้คะแนน แต่ละบุคคลที่เกี่ยวข้องกับประสิทธิภาพการตรวจให้คะแนนของบุคคลนั้น เช่นเดียวกับข้อมูลเฉพาะที่เกี่ยวข้องกับความยาก ประสิทธิภาพของข้อสอบ และสเกลในรูปแบบการตรวจให้คะแนน (Stahl; & Lunz. 1992; citing Mulqueen; et al. 2000)

งานวิจัยส่วนมากที่เกี่ยวข้องกับผลกระทบของผู้ให้คะแนนมีการตรวจสอบผลกระทบโดยใช้วิธีการแตกต่างกัน งานวิจัยต่างๆ ได้มุ่งเน้นในส่วนของความรู้อของผู้ให้คะแนน และความสัมพันธ์ระหว่างกระบวนการ และเน้นที่จุดรวมเนื้อหาความรู้ของผู้ให้คะแนนและความมีประสิทธิภาพของผู้ให้คะแนน เช่นงานวิจัยของ ชาญวิทย์ จรัสสุทธิอิศวร (2550) ที่ได้ศึกษาการแสดงผลหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนประเมินทักษะกระบวนการทางวิทยาศาสตร์โดยใช้ทฤษฎีมาตรฐานเดิมและทฤษฎีการตอบข้อคำถามตามโมเดลของราสส์แบบหลายองค์ประกอบ เป็นต้น การศึกษาผลกระทบนี้ ได้มุ่งเน้นในคุณลักษณะต่างๆ ของผู้ให้คะแนนขึ้นงานการตรวจให้คะแนน และสภาพแวดล้อมการตรวจให้คะแนนที่เกี่ยวข้องกับการมีอยู่ของผลกระทบของผู้ให้คะแนนในการตรวจให้คะแนน ในบางการศึกษานั้น ได้แนะนำว่า พฤติกรรมในการนำเสนอในส่วนของข้อมูลที่ถูกแสดงต่อผู้ให้คะแนนและวิธีการที่ผู้ให้คะแนนได้ดำเนินการตรวจให้คะแนนนั้น อาจจะแนะนำชนิดที่แน่นอนของ ผลกระทบของผู้ให้คะแนน เช่น Proximity errors หรือ Halo effects (Wolfe. 2004: 37)

ในปี 1989 ลินาเคอร์ (Linacre. 1994) ได้ปรับขยายโมเดลของราสส์สำหรับการตอบข้อสอบแบบลำดับขั้น (Order Categories) ข้อสอบที่มีการตรวจให้คะแนนแบบบางส่วน (Partial Credit Test Items) ซึ่งจากเดิมโมเดลของราสส์จะมุ่งเน้นไปที่สององค์ประกอบ คือ ข้อสอบและผู้สอบ แต่ในทางปฏิบัติของการตรวจให้คะแนนแบบหลายค่านี้มีอีกองค์ประกอบหนึ่งที่มีผลต่อการให้คะแนน คือ ผู้ตัดสิน (Judge) หรือผู้ตรวจให้คะแนน (Rater) โดยลินาเคอร์ได้ปรับขยายโมเดลไปสู่สถานการณ์ที่มากกว่าสององค์ประกอบ (Facets) โดยเพิ่มองค์ประกอบที่สนใจเข้าไปในโมเดลทำให้สามารถประมาณค่าพารามิเตอร์ของผู้ตรวจให้คะแนน (Rater Parameter) ได้ โมเดลที่พัฒนาขึ้นนี้เรียกว่า โมเดลของราสส์แบบหลายองค์ประกอบ

(Many-Facets Rasch Model : MFRM) ลิณาเคอร์ มีแนวคิดว่าจะแนมาจาก 4 องค์ประกอบ คือ ความสามารถของผู้สอบ (Ability) ความยากของข้อสอบ (Difficulty) ความเข้มงวดของผู้ตรวจให้คะแนน (Rater Severity) และโครงสร้างของมาตราประมาณค่า (Rating Scale) ฟังก์ชันความสัมพันธ์ขององค์ประกอบทั้งหมด จะแสดงด้วยสมการคณิตศาสตร์ สามารถใช้ตรวจสอบคุณภาพของการตรวจให้คะแนน แบบบางส่วน (Partial Credit) โดยสามารถบ่งชี้พฤติกรรมกรรมการตรวจให้คะแนนของผู้ตรวจว่า มีความเข้มงวดหรือใจดีมากเกินไปใด ผู้วิจัยสามารถใช้โครงสร้างทางสถิติในการวิเคราะห์เพื่อสรุปรูปแบบ (Pattern) ของผู้ตรวจให้คะแนน (Rater) และผู้สอบ (Rater) และคุณลักษณะ (Trait) หรือข้อสอบ (Item) และตัวแปรอื่นๆ ที่สนใจ ทั้งนี้การวิเคราะห์ราส์ชแบบหลายองค์ประกอบ สามารถแยกคำนวณในแต่ละองค์ประกอบ (Facet) อย่างเป็นอิสระจากองค์ประกอบอื่น (Myford; & Wolfe. 2003: 402)

โมเดลของราส์ชแบบหลายองค์ประกอบมีลักษณะที่สำคัญ คือ เป็นโมเดลแบบเส้นตรงเชิงบวก (Additive Linear Model) ที่แปลงคะแนนสังเกตของผู้ตอบแต่ละคน แต่ละรายการของกลุ่มตัวอย่างให้อยู่ในรูปของคะแนน Logit (Logit Score – Log Odds Score) และใช้วิธีการวิเคราะห์ข้อมูลตามแนวทางการถดถอย (Regression) โดยมีอัตราส่วนของชั้นความน่าจะเป็นที่ต่อเนื่องกัน (Successive Category Probability) เป็นตัวแปรตามและองค์ประกอบต่างๆ (Facets) เป็นตัวแปรอิสระ นอกจากนี้ยังสามารถปรับขยายโมเดลให้ใช้กับข้อมูลที่เป็นมาตราประเมินค่า และการตรวจให้คะแนนความรู้บางส่วน ซึ่งเป็นการให้คะแนนแบบหลายค่า (Polytomous) ที่มีหลายองค์ประกอบ ในการวิเคราะห์องค์ประกอบต่างๆ จะถูกวิเคราะห์พร้อมกันอย่างเป็นอิสระ และเมื่อปรับให้อยู่บนมาตรา Logit เดียวกัน (Calibrate) ทำให้สามารถวัดความเข้มงวดของผู้ตรวจให้คะแนนอยู่บนมาตราเดียวกันกับความสามารถของผู้สอบ และความยากของข้อสอบได้ โมเดลของราส์ชแบบหลายองค์ประกอบเป็นโมเดลที่เป็นเอกมิติ (Unidimension) ที่สามารถรวมค่าพารามิเตอร์ความสามารถของนักเรียนเข้ากับองค์ประกอบอื่นๆ ได้ เช่น รวมเอาองค์ประกอบความยากของข้อคำถาม และองค์ประกอบของผู้ตรวจให้คะแนน เป็นต้น

โมเดลของราส์ชแบบหลายองค์ประกอบ มีความเหมาะสมหากผู้วิจัยมีจุดมุ่งหมาย เพื่อรวบรวมผลการให้คะแนนที่ได้จากการวัดองค์ประกอบต่างๆ แยกกันเพื่อจะรวมเป็นคะแนนรวม (Total Score) โดยมีข้อตกลงเบื้องต้นว่า ชุดของประชากรที่ได้รับการวัดและชุดของข้อคำถามนั้นจะต้องสามารถจัดเป็นลำดับเฉพาะ

สถิติทดสอบความสอดคล้องของโมเดล (Fit Statistics) เป็นการแสดงถึงระดับความสอดคล้องของข้อมูลเชิงประจักษ์กับค่าความคาดหวังจากโมเดลของราส์ชแบบหลายองค์ประกอบ การตรวจสอบความสอดคล้องนี้มีหลายแนวทาง แนวทางหนึ่งที่ยอมรับใช้คือ การใช้ค่ามาตรฐานของส่วนที่เหลือ (Standardized Residual) เป็นดัชนีบ่งชี้คุณภาพของการให้คะแนน นอกจากนี้การตรวจสอบความสอดคล้องของข้อมูลกับโมเดลยังมีประโยชน์เกี่ยวกับความเที่ยงตรงและความเชื่อมั่นของแบบทดสอบด้วย (Myford; & Wolfe. 2003: 407) ค่ามาตรฐานของส่วนที่เหลือ จะให้รายละเอียดเกี่ยวกับการตรวจสอบความสอดคล้องของคะแนนสังเกตกับรูปแบบการให้คะแนนที่คาดหวังของโมเดล (Expecting Rating Patterns) ค่านี้สามารถสรุปความแตกต่างขององค์ประกอบ (Facets) และความแตกต่างของสมาชิกในองค์ประกอบ (Elements)

ค่ามาตรฐานของส่วนที่เหลือ จึงเป็นดัชนีแสดงถึงความสอดคล้องของข้อมูลเชิงประจักษ์กับโมเดลเชิงทฤษฎี (Englehard. 1994: 97) เมื่อนำค่ามาตรฐานของส่วนที่เหลือยกกำลังสองแล้วหาค่าเฉลี่ยทั้งผู้สอบและข้อสอบเพื่อสรุปเป็นค่าสถิติของผู้ตรวจแต่ละคน แสดงถึง ค่าความไม่สอดคล้อง (Misfit) ทั้งหมดในการตรวจของผู้ตรวจให้คะแนน (Linacre. 1994) ซึ่งค่าเฉลี่ยนี้จะให้ค่าดัชนีใน 2 ลักษณะ ดังนี้

1. สถิติ Outfit เป็นดัชนีของค่าเฉลี่ยยกกำลังสองของส่วนที่เหลือแบบไม่ถ่วงน้ำหนัก (Unweighted Mean Square Residual) ที่แสดงถึงความแตกต่างระหว่างคะแนนสังเกตกับค่าความคาดหวัง สถิติ Outfit, u_j สำหรับองค์ประกอบผู้ตรวจแสดงถึงค่าเฉลี่ยยกกำลังสองของค่ามาตรฐานส่วนที่เหลือจากผู้สอบและข้อสอบทั้งหมดของผู้ตรวจ สามารถคำนวณได้จากสมการ (Myford; & Wolf. 2003: 402)

$$u_j = \frac{\sum_{n=1}^N \sum_{l=1}^I Z_{nlj}^2}{NI}$$

เมื่อ N แทน จำนวนผู้สอบ
I แทน จำนวนข้อสอบ

เนื่องจากสถิติ Outfit เป็นสถิติที่ค่อนข้างไว (Sensitivity) ดังนั้นจึงมีค่าสถิติอีกค่าหนึ่งที่มีความละเอียดขึ้นโดยมีตัวถ่วงน้ำหนัก

2. สถิติ Infit เป็นดัชนีของค่าเฉลี่ยยกกำลังสองแบบถ่วงน้ำหนัก (Weighted Mean Squares) สามารถคำนวณแต่ละสมาชิกขององค์ประกอบได้ สถิตินี้มีความไวต่อค่าที่คาดหวังต่ำ (Outlying Unexpected) สามารถคำนวณได้จากสมการ (Englehard. 1994: 97)

$$v_j = \frac{\sum_{n=1}^N \sum_{l=1}^I W_{nlj} Z_{nlj}^2}{\sum_{n=1}^N \sum_{l=1}^I W_{nlj}}$$

เมื่อ W_{nlj} เป็นตัวถ่วงน้ำหนัก

ค่า Infit และ Outfit มีค่าตั้งแต่ 0 ถึง ∞ การที่จะตัดสินว่าจะใช้ค่า Infit และ Outfit ในช่วงใดขึ้นอยู่กับจุดประสงค์ที่ใช้ ซึ่งในงานวิจัยขั้นนี้ต้องการแปลความหมายโดยใช้ผู้ตรวจให้คะแนนเป็นจำนวนมาก จึงใช้เกณฑ์การตัดสินจาก (Linacre. 2012: 257) ดังนี้

- > 2.0 แปลว่า มีระบบการวัดที่บิดเบือนผิดปกติ
- 1.5 – 2.0 แปลว่า ไม่สามารถสร้างโครงสร้างของการวัดได้
- 0.5 – 1.5 แปลว่า สามารถวัดได้ตามโครงสร้าง
- < 0.5 แปลว่า วัดตามโครงสร้างได้ต่ำ

ผู้วิจัยใช้เกณฑ์นี้ในการแปลความหมายความสอดคล้องกลมกลืนของข้อมูลเชิงประจักษ์ที่ได้จากการวิจัยกับโมเดล Partial Credit Model

ความเชื่อมั่นแยกส่วน (Reliability of Separation Index) เป็นดัชนีที่มีประโยชน์มากในการตรวจสอบตัวแปรแฝง ในงานวิจัยชิ้นนี้ใช้ค่าความเชื่อมั่นแยกส่วนในการอธิบายคุณภาพของผู้สอบและผู้ตรวจให้คะแนน ดังนี้

1. ค่าความเชื่อมั่นแยกส่วนของผู้สอบ (Person Separation Reliability; Person Reliability) คล้ายคลึงกับสัมประสิทธิ์แอลฟา (Englehard. 2002; Myford; & Wolfe. 2003) เป็นอัตราส่วนของความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนสังเกตในการประมาณค่าความสามารถของผู้สอบ ถ้าค่านี้มีค่าสูงแสดงว่า มีความแตกต่างระหว่างความสามารถของผู้สอบ แสดงถึงความสามารถในการจำแนกความสามารถของผู้สอบแต่ละคนออกจากกัน เป็นดัชนีที่แสดงถึงความเชื่อถือได้ของคะแนนความสามารถของผู้สอบที่ได้จากข้อสอบและผู้ตรวจ

2. ความเชื่อมั่นแยกส่วนของผู้ตรวจ (Rater Separation Reliability) แสดงถึง ความแปรปรวนของการให้คะแนนที่ไม่ต้องการให้เกิดขึ้นระหว่างผู้ตรวจให้คะแนนในระดับของความเข้มงวด/ใจดี ซึ่งดัชนีนี้ไม่ได้บ่งชี้ว่า มีความสอดคล้องกันระหว่างผู้ตรวจให้คะแนน แต่บ่งชี้ว่าผู้ตรวจให้คะแนนมีความเข้มงวดในระดับเดียวกันหรือไม่ กล่าวอีกนัยหนึ่ง ความเชื่อมั่นแยกส่วนของผู้ตรวจให้คะแนนบ่งชี้ถึงความเชื่อมั่นของความแตกต่าง (Reliably Different) มากกว่าระดับความเชื่อมั่นที่คล้ายคลึงกัน (Reliably Similar)

จากการศึกษาเอกสารเพิ่มเติมของผู้วิจัยในเรื่องการวิเคราะห์แบบ MFRM จะพบว่า การวิเคราะห์ส่วนใหญ่มุ่งเน้นการระบุคุณลักษณะผู้ให้คะแนน คือ ความเข้มงวด/ใจดี เช่น งานวิจัยการใช้ MFRM ในโปรแกรมการอบรมผู้สอนการบินของ มัลควีน และคณะ (Mulqueen; et al. 2000) งานวิจัยของ นากามูระ (Nakamura. 2002) ที่ศึกษาการใช้งาน MFRM ในการวิเคราะห์การสอบการเขียนงานวิจัยของ วูฟ (Wolfe. 2004) ที่ศึกษาการระบุผลกระทบจากผู้ให้คะแนนโดยใช้โมเดลตัวแปรแฝง และงานวิจัยของ อีรมณีรัตน์ และ โยดโควสกี (Iramaneerat; & Yudkowsky. 2006) ได้ศึกษาความคลาดเคลื่อนของผู้ให้คะแนนในการประเมินทักษะทางการแพทย์ นอกจากนี้ได้มีการต่อยอดการใช้งาน MFRM โดยใช้หาความลำเอียงด้านเพศของผู้ให้คะแนน (Differential Rater Functioning) โดยใช้หลักการของการวิเคราะห์ความเข้มงวด/ใจดีของผู้ให้คะแนนข้ามกลุ่มเพศ เช่น งานวิจัยของ เอกเคส (Eckes. 2005) ที่ศึกษาการวิเคราะห์ MFRM ในการตรวจสอบผลกระทบจากผู้ให้คะแนนในการสอบ TestDaF และงานวิจัยของ แองเจิลฮาร์ด และ เมฟอร์ด (Engelhard; & Myford. 2003) ที่ได้ศึกษาความสามารถในการให้คำปรึกษาโดยใช้ MFRM และในปี 2009 ได้มีการนำ MFRM มาระบุคุณลักษณะผู้ให้คะแนนโดยการนำองค์ประกอบของเวลา (Time Facet) มาใช้ในการวิเคราะห์การทำหน้าที่ต่างกันของผู้ให้คะแนน เมื่อเวลาผ่านไป (Differential Rater Functioning Over Time) โดย เมฟอร์ด และวูฟ (Myford; & Wolfe. 2009) ซึ่งมีรายละเอียด ดังนี้

1. ความเข้มงวด/ใจดีของผู้ให้คะแนน (Severity/Leniency)

ผลกระทบของผู้ให้คะแนนที่ได้รับการวิจัยโดยทั่วไป เกิดขึ้นเมื่อผู้ให้คะแนนใช้ประโยชน์

จากเกณฑ์การตรวจให้คะแนนเดิม ในระดับความเข้มงวดที่แตกต่างกันที่ได้รับการปรับปรุงภายในโครงสร้างงานการตรวจให้คะแนนนั้น ในกรณีการจัดลำดับของคะแนนต่างๆ ที่ถูกให้คะแนนโดยผู้ให้คะแนนที่แตกต่างกัน มีความคงที่มาก แต่คะแนนเฉลี่ยแตกต่างกันของผู้ให้คะแนนที่มีแนวโน้มให้มอบหมายคะแนนที่ต่ำ ถูกเรียกว่า ความเข้มงวด หรือขณะที่ผู้ให้คะแนนที่มีแนวโน้มให้มอบหมายคะแนนที่สูงถูกเรียกว่า ความใจดี ในกรณีที่ความโปร่งใสของผู้ให้คะแนนที่เป็นความเข้มงวดและความใจดี มอบหมายคะแนนที่มีความสัมพันธ์อย่างสมบูรณ์กับการตรวจให้คะแนนที่คาดหวังไว้ ด้วยเหตุผลนี้ความเบี่ยงเบนมาตรฐานของส่วนที่เหลือที่ได้รับการสังเกตเกือบใกล้ 0 และส่วนที่เหลือ และ $E(X)$ จะไม่ขึ้นอยู่กับส่วนใด ๆ ความเข้มงวด และความใจดีของผู้ให้คะแนนสามารถถูกทำเป็นรูปแบบที่อธิบายได้ในโครงสร้างตัวแปรแฝงที่ได้รับการอธิบายเป็นตำแหน่งต่างๆ ของผู้ให้คะแนนในการ Underlying Linear Continuum และส่วนที่เหลือจากรูปแบบนี้ นำเอาตำแหน่งต่างๆ ไปพิจารณา

ในโครงสร้างการวัดต่างๆ ที่มีการอ้างอิงแบบ Norm ผู้ให้คะแนนความเข้มงวด และความใจดี ไม่ส่งอิทธิพลต่อตำแหน่งที่ข้องเกี่ยวของสิ่งต่างๆ ของการวัดเมื่อการเปรียบเทียบถูกกระทำระหว่างผู้ให้คะแนน เมื่อผู้ให้คะแนนทั้งหมดให้คะแนนสิ่งต่างๆ ของการวัดทั้งหมด ค่าเฉลี่ยอย่างง่าย หรือผลรวมของคะแนนต่างๆ ได้รับความหมายโดยผู้ให้คะแนนทั้งหมด สร้างการกระจายของคะแนนที่คาดหวังกับการกระโดดอย่างคงที่ของการกระจายนั้น ด้วยเหตุนี้การเปรียบเทียบแบบที่มีการอ้างอิงแบบ Norm จะส่งผลในการตัดสินใจที่เชื่อถือได้อย่างสมบูรณ์เกี่ยวกับสิ่งต่างๆ ของการวัดในกรณีของความเข้มงวด และความใจดีที่เป็นแบบบริสุทธ์กับการออกแบบการตรวจให้คะแนนอย่างสมบูรณ์ อย่างไรก็ตาม เมื่อแบบการตรวจให้คะแนนต่างๆ ไม่สมบูรณ์ (เช่น เมื่อมีสัดส่วนของผู้ให้คะแนนที่มอบหมายการตรวจให้คะแนนต่อสิ่งใด ๆ ของการวัดที่กำหนดไว้คุณลักษณะที่อยู่ในส่วนธรรมดาในโครงการ การตรวจให้คะแนนที่มีสเกลกว้างสำหรับค่าที่ต่ำสุด) การมีอยู่ของผู้ให้คะแนนความเข้มงวด และความใจดี อาจมีอิทธิพลต่อตำแหน่งที่ข้องเกี่ยวของสิ่งต่างๆ แต่ละสิ่งของการวัดขึ้นอยู่กับระดับของความเข้มงวด และความใจดีของกลุ่มย่อยของผู้ให้คะแนน ซึ่งมอบหมายคะแนนต่อชิ้นงานของการวัดนั้น อิทธิพลของผู้ให้คะแนนความเข้มงวดและความใจดี สามารถถูกทำให้ลดน้อยลง โดยการเพิ่มจำนวนของผู้ให้คะแนนที่มอบหมายคะแนนต่อสิ่งที่มีกำหนดไว้ของการวัด และทำให้แน่ใจว่าการมอบหมายของผู้ให้คะแนนต่อสิ่งต่างๆ ของการวัดนั้นเป็นแบบสุ่ม

รูปแบบราศีชหลายองค์ประกอบขยายกลุ่มของรูปแบบราศีชต่อกรณีต่างๆ ซึ่งฟาเซทหลายองค์ประกอบของราศีชของการวัดถูกกำหนดเป็นรูปแบบ (เช่น องค์ประกอบต่างๆ ของบริบทการวัดที่อาจให้ข้อผิดพลาดทางการวัด) ในกรณีของผลกระทบของผู้ให้คะแนนรูปแบบราศีชแบบ หลายองค์ประกอบ อาจถูกใช้เพื่อควบคุมข้อผิดพลาดที่ถูกให้โดยความแปรผันเชิงระบบระหว่างทั้งข้อคำถามและผู้ให้คะแนน การทำนายการให้เพิ่มเติมขององค์ประกอบของบริบทการวัดต่อ Logit ของการสังเกตชนิดสเกลการตรวจให้คะแนนกับชนิดสเกลการตรวจให้คะแนนต่ำต่อไป โดยใช้ตัววัดที่แสดงถึงสิ่งของการวัดและฟาเซทของบริบทการวัดเช่นผู้ให้ คะแนนและข้อคำถามต่างๆ เราสามารถแสดงความสัมพันธ์เชิงคณิตศาสตร์ดั้งสมการ

$$\ln\left(\frac{P_x}{P_{x-1}}\right) = \theta_n - \lambda_r - \delta_{ik}$$

ขณะที่ θ_n เป็นจุดที่สิ่งของการวัดตามเส้นตรงที่ต่อเนื่อง, λ_r , เป็นจุดของผู้ให้คะแนน r และ δ_{ik} เป็นจุดของ Threshold (k) ระหว่างประเภทต่างๆ ของ x และ $x-1$ บนสเกลการตรวจให้คะแนนสำหรับข้อ l จุดของ Threshold แต่ละอันถูกจัดให้คงที่ผ่านข้อทั้งหมด ซอฟแวร์ของราล์ชใช้วิธีการต่างๆ ของความสูงสุดเพื่อประเมินค่าต่างๆ ของตัววัดที่มีพื้นฐานบนการตรวจให้คะแนนต่างๆ ที่ได้รับการสังเกต

ตามที่เป็นหลักฐานในสมการ จุดของการตามผู้ให้คะแนน มีอิทธิพลโดยตรงกับความเป็นไปได้ของการสังเกตชนิดเฉพาะสำหรับสิ่งของการวัดในคำถาม ดังนั้นมันกระจางที่ว่า λ_r แสดงความเข้มงวด และความใจดี ของผู้ให้คะแนน r โดยการตรวจสอบความสำคัญที่เกี่ยวข้องของการประเมิน λ_r สำหรับชุดเฉพาะการตรวจให้คะแนน เราสามารถระบุผู้ให้คะแนนที่เป็นแบบความเข้มงวด หรือความใจดีเกี่ยวข้องสัมพันธ์เพื่อทุกคนของผู้ให้คะแนนข้อผิดพลาดมาตรฐานต่างๆ สามารถรับการประเมินสำหรับแต่ละการประเมินตัววัด และสถิติ Wald สามารถถูกคำนวณเพื่อระบุผู้ให้คะแนนที่ได้รับมาจากกลุ่มค่าเฉลี่ย (เช่น ค่า สมมติฐานกลาง) ของ 0 ด้วยระดับสำคัญทางสถิติ

$$\chi^2_{\text{Wald}} = \left(\frac{\lambda_r}{\text{SE}_{\lambda_r}}\right)^2$$

ค่าต่างๆ ที่คาดไว้ที่มีพื้นฐานตามรูปแบบสามารถถูกคำนวณสำหรับการรวมกันของผู้ให้คะแนนด้วยการวัดและ สิ่งที่วัดกับข้อคำถาม และการ x เป็นการนับของจำนวนของสเกลการตรวจให้คะแนน Thresholds เกินโดยสิ่งของการวัด (จาก 0 ถึง m)

$$E_{\text{nir}} = \sum_{n=1}^N \sum_{i=1}^I k \pi_{\text{nirk}}$$

$$\pi_{\text{nirk}} = \frac{\exp \sum_{j=0}^x \theta_n - \lambda_r - \delta_{ik}}{\sum_{k=0}^m \exp \sum_{j=0}^k \theta_n - \lambda_r - \delta_{ik}}$$

$$\delta_{i0} \equiv 0 \quad \text{ดังนั้น} \quad \sum_{j=0}^0 \theta_n - \lambda_r - \delta_{ik} = 0$$

ส่วนที่เหลือของการตรวจให้คะแนนที่ได้รับการสังเกตจากความคาดหวังที่มีพื้นฐาน ส่วน โมเดล, $X_{nir} - E_{nir}$, สามารถถูกใช้เพื่อระบุผู้ให้คะแนนความเที่ยงตรง/ความไม่เที่ยงตรง และผู้ให้คะแนนความเป็นกลาง/ความสุดโต่ง โดยเฉพาะอย่างยิ่งส่วนที่เหลือความเบี่ยงเบนมาตรฐานที่น้อยกว่า มีความสัมพันธ์กับผู้ให้คะแนนความเที่ยงตรง และความเป็นกลาง ขณะที่ส่วนที่เหลือความเบี่ยงเบนมาตรฐานที่มากกว่ามีความสัมพันธ์กับผู้ให้คะแนนความเป็นกลาง และความสุดโต่งผู้ให้คะแนนความเที่ยงตรง และความเป็นกลาง สามารถถูกทำให้แตกต่างโดยการตรวจสอบความสัมพันธ์ระหว่างส่วนที่เหลือและการตรวจให้คะแนนแบบที่คาดหวังไว้สัมพันธ์กับส่วนที่เหลือนั้น ($r_{residual, expected}$) หากความสัมพันธ์ระหว่างส่วนที่เหลือและการตรวจให้คะแนนแบบที่คาดหวังไว้ใกล้ 0 แล้วผู้ให้คะแนนความเที่ยงตรง นั้นถูกระบุในทางกลับกัน หากความสัมพันธ์ที่คาดหวังไว้-ส่วนที่เหลือ เป็นลบ แล้วผู้ให้คะแนนความเป็นกลาง ถูกระบุผู้ให้คะแนนความไม่เที่ยงตรง และความสุดโต่ง สามารถถูกทำให้แตกต่างในทางคล้ายคลึงกัน ความสัมพันธ์ระหว่างส่วนที่เหลือและการตรวจให้คะแนนแบบที่คาดหวังไว้ควรจะเป็น 0 ในกรณีของผู้ให้คะแนนความไม่เที่ยงตรง และควรจะเป็นบวกในกรณีของความสุดโต่ง

การวิเคราะห์รหัสหลายองค์ประกอบ ได้ให้มาตรวัดความเข้มงวดในหน่วยที่เป็น Logit ค่าเฉลี่ยความเข้มงวดของผู้ให้คะแนนอยู่ Logit ที่ 0 ผู้ให้คะแนนที่มีความใจดีมากกว่าค่าเฉลี่ยมีมาตรการวัด Logit ติดลบ ในขณะที่ผู้ให้คะแนนที่มีความเข้มงวดมากกว่าค่าเฉลี่ยมีมาตรการวัด Logit เป็นค่าบวก การตัดสินใจระดับความเข้มงวดของผู้ให้คะแนนมีความแตกต่างกันอย่างมีนัยสำคัญหรือไม่ ด้วยค่าสถิติไคสแควร์ที่ทดสอบสมมติฐานกลางว่า ไม่มีความแตกต่างกันอย่างมีนัยสำคัญในการวัดความเข้มงวดของผู้ให้คะแนน ระดับการมีนัยสำคัญของไคสแควร์ ($p < 0.05$) บ่งชี้ว่า ผู้ให้คะแนนอย่างน้อย 2 คน มีระดับความเข้มงวดแตกต่างกันอย่างมีนัยสำคัญ

ดังนั้นกล่าวโดยสรุป การระบุความเข้มงวด/ใจดี หมายถึง ผู้ตรวจให้คะแนนที่มีแนวโน้มให้คะแนนที่ต่ำจะถูกระบุว่า มีความเข้มงวด และผู้ตรวจให้คะแนนที่มีแนวโน้มให้คะแนนที่สูงจะถูกระบุว่า มีความใจดี โดยในขั้นแรกนั้นต้องพิจารณาว่าผู้ตรวจให้คะแนนแต่ละคนให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล Partial Credit Model หรือไม่ โดยพิจารณาค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนที่มีการถ่วงน้ำหนัก Weighted mean square residual : INFIT และค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนที่ไม่มีการถ่วงน้ำหนัก Unweighted mean square residual : OUTFIT ที่กำหนดขอบเขตของการยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลอยู่ระหว่าง 0.5 ถึง 1.5 จากนั้นจึงระบุความเข้มงวด/ใจดี จากค่าศูนย์ (0) เครื่องหมายบวก (+) ลบ (-) ของค่า Logit ผู้ตรวจให้คะแนนที่เข้มงวดจะมีค่า Logit มากกว่า +1.00 ให้คะแนนเป็นกลาง $-1 < \text{Logit} < +1$ และให้คะแนนใจดีจะมีค่า Logit น้อยกว่า -1.00

2. ความลำเอียงด้านเพศของผู้ให้คะแนน (Differential Rater Functioning : DRF)

ผลกระทบของผู้ให้คะแนน เช่น ความเข้มงวด หรือความใจดี ความคล้อยตามกัน หรือแนวโน้มที่เป็นกลาง ได้รับการมองเป็นแหล่งของความแปรปรวนทางวิธีการ (Method Variance) นั่นก็คือ เป็นแหล่งของความแปรปรวนอย่างเป็นระบบในการตรวจให้คะแนนแบบมีการสังเกต ซึ่งสัมพันธ์กับ

ผู้ให้คะแนนและไม่สัมพันธ์กับผู้ถูกให้คะแนน (Cronbach. 1995; Hoyt. 2000; Myford; & Wolfe. 2003; citing Eckes. 2005)

การศึกษาต่างๆ มุ่งเน้นการตรวจสอบประสิทธิภาพทางภาษาที่ถูกระบุคล้ายคลึงกัน ในระดับที่สำคัญของผลกระทบหลักของผู้ให้คะแนน โดยเฉพาะอย่างยิ่งในเรื่องความเข้มงวด หรือความใจดี ของผู้ให้คะแนน ความแตกต่างเช่นนี้ถูกยืนยันว่ายังคงมีอยู่หลังจากการอบรมผู้ให้คะแนน และยังคงไว้ในกลุ่มของผู้ให้คะแนนตลอดช่วงเวลาที่ผ่านมา นอกจากนี้ยังพบผลกระทบที่สำคัญสำหรับการมีปฏิสัมพันธ์ระหว่างผู้ให้คะแนนและผู้ถูกให้คะแนน การมีปฏิสัมพันธ์รูปแบบผู้ให้คะแนนและชิ้นงาน และการมีปฏิสัมพันธ์รูปแบบผู้ให้คะแนนและเกณฑ์ (Wigglesworth. 1993; citing Eckes. 2005)

การทบทวนความหมายต่างๆ สำหรับการอบรมผู้ให้คะแนนที่แนะนำโดย McNamara (1996: 127; citing Eckes. 2005) เพื่อยอมรับว่า จุดมุ่งหมายที่เหมาะสมมากที่สุดของการอบรมผู้ให้คะแนนคือ ต้องทำให้ผู้ให้คะแนนมีความคงที่ภายในเพื่อที่จะทำการสร้างรูปแบบเชิงสถิติของคุณลักษณะต่างๆ ที่เป็นไปได้ แต่นอกเหนือจากสิ่งนี้ เพื่อยอมรับความแปรปรวนในคุณลักษณะต่างๆ ของผู้ให้คะแนนที่คงที่เป็นข้อเท็จจริงของชีวิต ซึ่งต้องถูกชดเชยสำหรับบางวิธีการ นอกจากนี้ในส่วนของแนวโน้มของคนในการจัดเก็บเอกสารวิจัยทางด้านจิตวิทยาสังคม เพื่อดูพฤติกรรมของคนอื่น ตลอดจนมุมมองด้านเพศอีกเรื่องหนึ่งของความสำคัญอ้างอิงถึงขอบเขตที่ผู้ให้คะแนนต้องมีมุมมองและการประเมิน โดยมีพื้นฐานด้านเพศเมื่อให้คะแนนประสิทธิภาพของผู้สอบ นี่คือการเรื่องของ Differential Facet Functioning ที่เกี่ยวข้องกับผู้ให้คะแนนหรือ Differential Rater Functioning (DRF) ซึ่งคล้ายคลึงกับ Differential Item Functioning ที่ได้รับการสังเกตผ่านกลุ่มย่อยที่เกี่ยวข้องของผู้สอบ (Engelhard; & Myford. 2003; & Wang. 2000; citing Eckes. 2005) แองเจลิฮาร์ท และ เมฟอร์ด (Engelhard; & Myford. 2003) ที่ได้ศึกษาความสามารถในการให้คำปรึกษาโดยใช้ MFRM ในบริบทของการเขียนเรียงความสำหรับ Advanced Placement English Literature และ Composition Exam แม้ว่าการศึกษาที่เป็นกลุ่มผู้ให้คะแนนไม่ได้แสดงผลกระทบ แบบความเข้มงวด หรือความใจดี ที่เกี่ยวข้องกับเพศของผู้เรียน นักวิจัยต่างสามารถที่จะระบุผู้ให้คะแนนรายบุคคลที่มีแนวโน้มในการมอบหมายคะแนนที่สูงหรือต่ำอย่างคงที่ และไม่คาดหวังไว้กับเรียงความของผู้เรียนเพศชายที่กำหนดให้ ระดับของผู้ให้คะแนนเฉพาะของความเข้มงวดหรือความใจดี และค่าของประสิทธิภาพของผู้เรียนเพศชาย

ในการวิจัยได้กำหนดรูปแบบสเกลการตรวจให้คะแนนสำหรับแต่ละเกณฑ์ (หรือ ชิ้นงาน) เพื่อมีโครงสร้างประเภทของตัวเอง นั่นก็คือ โครงสร้างของสเกลการตรวจให้คะแนนที่สามารถมีความหลากหลายจากเกณฑ์หนึ่ง (หรือ ชิ้นงานหนึ่ง) ไปสู่เกณฑ์หนึ่ง ดังนั้นรูปแบบเฉพาะที่ถูกสร้างขึ้นในการวิเคราะห์ต่างๆ นี้ คือ Three-Facet Partial Credit Model (Linacre; & Wright. 2002; citing Eckes. 2005) เส้นแบ่งที่เป็นมาตรฐานของฟาเซทองค์ประกอบของผู้สอบ ผู้ให้คะแนนเกณฑ์ (หรือชิ้นงาน) และสเกลการตรวจให้คะแนนบนสเกลที่มีช่วงเท่ากันเดียวกัน การสร้างโครงสร้างของการอ้างอิงสำหรับการแปลผลลัพธ์ต่างๆ ของการวิเคราะห์ เมื่อตัววัดต่างๆ ของรูปแบบได้รับการประเมินผลกระทบของการมีปฏิสัมพันธ์ เช่น การมีปฏิสัมพันธ์ระหว่างผู้ให้คะแนนและผู้สอบ หรือระหว่างผู้ให้คะแนนและเกณฑ์ สามารถถูกตรวจสอบได้โดยการตรวจสอบส่วนที่เหลือที่มีมาตรฐาน (เช่น ความแตกต่างที่มีมาตรฐาน

ระหว่างการตรวจให้คะแนนที่ได้รับการสังเกตและคาดหวังไว้) การวิเคราะห์การมีปฏิสัมพันธ์ (หรือการวิเคราะห์ความลำเอียง) ช่วยให้มีรูปแบบการมีปฏิสัมพันธ์ท่ามกลางองค์ประกอบฟาเซท โดยเฉพาะอย่างยิ่งรูปแบบที่บ่งชี้ไปที่ส่วนเบี่ยงเบนที่คงที่จากสิ่งที่ถูกคาดหวังบนพื้นฐานของรูปแบบ

นอกจากนี้การวิเคราะห์ความลำเอียงด้านเพศ โดยการประเมินค่าวัดประสิทธิภาพ การเขียน สำหรับแต่ละกลุ่ม ผู้สอบ หญิงและชาย ตามด้วยค่าวัดต่างๆ ของประสิทธิภาพที่แยกกัน สำหรับแต่ละการรวมกัน หรือทุกๆ การรวมกันของผู้ให้คะแนนแต่ละคนกับกลุ่มเพศตามลำดับ องค์ประกอบฟาเซทที่มีการคำนวณพิเศษมากขึ้นในเรื่องสถิติความลำเอียงด้านเพศ เพื่อทดสอบสมมติฐานกลางว่าไม่มีความลำเอียงด้านเพศในข้อมูล สถิติ Z ที่สำคัญทางสถิติจะชี้ว่า การรวมกันของกลุ่มระหว่าง ผู้ให้คะแนน-เพศ มีผลในการตรวจให้คะแนนสูง-ต่ำอย่างไม่คาดหวัง โดยกำหนดให้ระดับของความเข้มงวดของผู้ให้คะแนนและระดับ ของกลุ่มเพศของความเชี่ยวชาญ ดังนั้นการวิเคราะห์ความลำเอียงทางด้านเพศระหว่างผู้ให้คะแนน-ผู้สอบ ได้ช่วยให้พบว่าผู้ให้คะแนนใดๆ ที่มีระดับการปฏิบัติที่มีความความเข้มงวดหรือความใจดี ที่แตกต่างกันหรือไม่ การตรวจให้คะแนนเรียงความของชาย (หรือ เรียงความของหญิง) มีความเป็นความเข้มงวด หรือความใจดี มากกว่าที่คาดหวังหรือไม่ หรือระดับของความเข้มงวด หรือความใจดีของผู้ให้คะแนนแต่ละคน ไม่มีความแปรปรวนผ่านกลุ่มเพศหรือไม่ โดยใช้สูตรการคำนวณ (Engelhard; & Myford. 2003) ดังนี้

$$\ln[P_{nijk}/P_{nijk-1}] = \theta_n - \xi_i - \alpha_j - \tau_k$$

P_{nijk}	แทน	ความน่าจะเป็นที่นักเรียนคนที่ n จะได้คะแนน k บนข้อสอบข้อที่ i จากผู้ให้คะแนนคนที่ j
P_{nijk-1}	แทน	ความน่าจะเป็นที่นักเรียนคนที่ n จะได้คะแนน k-1 บนข้อสอบข้อที่ i จากผู้ให้คะแนนคนที่ j
θ_n	แทน	ความสามารถของนักเรียนคนที่ n
ξ_i	แทน	ความยากของข้อสอบข้อที่ i
α_j	แทน	ความเข้มงวดของผู้ให้คะแนนคนที่ j
τ_k	แทน	ความยากในการเปลี่ยนระดับจากคะแนน k ไปสู่คะแนน k-1

สูตรที่โปรแกรม FACETS ใช้ในการคำนวณ Z-Score คือ t value ของ Welch-Satterthwaite (Linacre. 2012: 313)

$$t = \frac{(M_1 - M_2)}{\sqrt{SE_1^2 - SE_2^2}}$$

เมื่อ $M_1 - M_2$ แทน ความแตกต่างระหว่างค่าเฉลี่ย

$$\sqrt{SE^2 - SE^2}$$

แทน ความเบี่ยงเบนมาตรฐานร่วมของค่าเฉลี่ย
(Joint standard error of means)

ดังนั้นกล่าวโดยสรุป การระบุความลำเอียงด้านเพศของผู้ตรวจให้คะแนน การวิเคราะห์เปรียบเทียบความเข้มงวด/ใจดีของผู้ตรวจให้คะแนนว่ามีความแตกต่างกันหรือไม่ ข้ามกลุ่ม นักเรียนเพศชาย และกลุ่มนักเรียนเพศหญิง ในขั้นแรกนั้นต้องพิจารณาว่า ผู้ตรวจให้คะแนนแต่ละคนให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล Partial Credit Model หรือไม่โดยพิจารณาค่า INFIT และ OUTFIT ต่อมาจึงระบุความลำเอียงด้านเพศของผู้ตรวจให้คะแนน จากค่า Z-Score ที่มีนัยสำคัญทางสถิติ สุดท้ายจึงพิจารณาว่าผู้ตรวจให้คะแนนคนนั้นๆ ลำเอียงเข้าข้างเพศชายหรือเพศหญิง คะแนนที่มีแนวโน้มใจดี (Logit ต่ำกว่า) กับเพศใดแสดงว่าลำเอียงเข้าข้างเพศนั้น

3. การทำหน้าที่ต่างกันของผู้ให้คะแนนเมื่อเวลาผ่านไป (Differential Rater Functioning Over Time : DRIFT)

ในการสอบที่นักเรียนต้องแสดงความสามารถที่มีความซับซ้อน เช่น การสอบการเขียนเรียงความนั้น ผู้ให้คะแนนได้ให้คะแนนที่คำนึงถึงคุณภาพของการตอบของนักเรียนเป็นสำคัญ แต่การใช้มนุษย์เป็นผู้ให้คะแนนผลกระทบต่อความแม่นยำในการตรวจให้คะแนนได้จากการศึกษาที่ผ่านมาพบว่า ระดับความเข้มงวดของผู้ให้คะแนนนั้นเปลี่ยนแปลงอย่างเป็นระบบเมื่อเวลาผ่านไป (Braun, 1988; & Hoskens; & Wilson, 2002; citing Myford; & Wolfe, 2009) นั่นคือ ผู้ให้คะแนนแสดงหลักฐานความแตกต่างกันของความเข้มงวดเมื่อเวลาผ่านไปนั่นเอง โมเดลราสซ์แบบหลายองค์ประกอบเป็นวิธีการสำหรับตรวจหา และวัดการทำหน้าที่ต่างกันของผู้ให้คะแนนเมื่อเวลาผ่านไปในการวิเคราะห์ราสซ์แบบหลายองค์ประกอบ อาจจะมีรูปแบบโมเดลเชิงคณิตศาสตร์ที่หลากหลายที่ใช้สำหรับตรวจหาความเคลื่อนที่ของผลกระทบจากผู้ให้คะแนน (Dynamic Rater Effects) หนึ่งในนั้นคือโมเดลราสซ์แบบหลายองค์ประกอบที่ประกอบด้วยองค์ประกอบของเวลา (Time Facet) ซึ่งต่อจากนี้ไปจะเรียกว่า โมเดลองค์ประกอบของเวลา (Time Facet Model)

$$\ln\left(\frac{\pi_{nrx}}{\pi_{nrx-1}}\right) = A_n - S_r - R_t - C_x,$$

เมื่อ	π_{nrx}	แทน	ความน่าจะเป็นที่ความสามารถของนักเรียนคนที่ n จะได้คะแนน x จากผู้ให้คะแนนคนที่ r ในเวลาที่ t
	π_{nrx-1}	แทน	ความน่าจะเป็นที่ความสามารถของนักเรียนคนที่ n จะได้คะแนน x-1 จากผู้ให้คะแนนคนที่ r ในเวลาที่ t
	A_n	แทน	ระดับของผลความสำเร็จสำหรับนักเรียนคนที่ n

S_r	แทน	ความเข้มงวดของผู้ให้คะแนนคนที่ r
R_t	แทน	ความเบี่ยงเบนของการตรวจให้คะแนนในเวลาที t จากค่าเฉลี่ยของการตรวจให้คะแนนทั้งหมด
C_x	แทน	ความยากของการข้ามระดับระหว่างระดับสเกล x ไปเป็น $x - 1$

องค์ประกอบของเวลาได้แสดงภาพความเข้มงวดของผู้ให้คะแนนแต่ละคนเสมือนเป็นคุณสมบัติที่คงที่ข้ามเวลา แต่ว่าเหมาะสมสำหรับการเปลี่ยนแปลงค่าเฉลี่ยของการตรวจให้คะแนนข้ามเวลา (ผ่าน R_t) โมเดลองค์ประกอบของเวลาที่มีประโยชน์สำหรับตรวจหาการเปลี่ยนแปลงจำนวนไม่มาก ในค่าเฉลี่ยการตรวจให้คะแนนข้ามผู้ให้คะแนน เสมือนกับความก้าวหน้าของเวลา แม้ว่าการเปลี่ยนแปลงใน R_t เป็นการแสดงภาพผลการเปลี่ยนแปลงเฉลี่ยโดยรวมในการตรวจให้คะแนนระหว่างเวลา แต่ อาจจะเป็นการเปลี่ยนแปลงระหว่างความแตกต่างในความเข้มงวดของผู้ให้คะแนนข้ามเวลา หรือไม่ก็เป็นความแตกต่างในความสำเร็จโดยรวมของนักเรียนผู้ที่ซึ่งถูกให้คะแนนในระหว่างแต่ละช่วงเวลา ซึ่งอาจจะสมมติได้ว่าไม่มีความสำคัญเมื่อความสามารถเป็นการถูกกำหนดอย่างสุ่มไปยังช่วงเวลา โมเดลองค์ประกอบของเวลาเหมาะสมสำหรับการจัดการในการประเมินของผู้ให้คะแนนการตอบของนักเรียน ไปสู่ชิ้นงานการแสดงความสามารถหนึ่งชิ้นงาน แม้ว่าสามารถขยายโมเดลไปสู่การรวมพารามิเตอร์เฉพาะถิ่นของข้อสอบ(หรือชิ้นงาน) สำหรับในบริบทที่การตอบของนักเรียนมีมากกว่าการประเมินข้อสอบข้อเดียว (หรือชิ้นงานชิ้นเดียว) โมเดลองค์ประกอบของเวลานี้สามารถแสดงเสมือนเป็นฟังก์ชันความน่าจะเป็นได้อีกด้วย

$$\pi_{nrkx} = \frac{\exp \sum_{k=0}^x (A_n - S_r - R_t - C_k)}{\sum_{j=0}^m \exp \sum_{k=0}^j (A_n - S_r - R_t - C_k)}$$

เมื่อ	j, k	แทน	จำนวนที่นับได้ของดัชนีที่แสดงพิสัย
	m	แทน	ค่าสูงสุดของระดับมาตรฐานวัด เริ่มต้นด้วยค่า 0

ปัญหาของโมเดลองค์ประกอบของเวลา เมื่อถูกนำไปประยุกต์สู่การวิเคราะห์การทำหน้าที่ต่างกันของผู้ให้คะแนนเมื่อเวลาผ่านไป โมเดลนี้ไม่ได้ประมาณค่าความสามารถของผู้ให้คะแนนแต่ละคนโดยตรงในแต่ละช่วงเวลา ซึ่งมีความเป็นไปได้ที่จะมีการสร้างสูตรโมเดลองค์ประกอบของเวลาขึ้นใหม่ เพื่อที่จะคำนวณการประมาณค่าแยก (Separate Estimate) ความเข้มงวดของผู้ให้คะแนนแต่ละคนในแต่ละจุดเวลา

$$\ln\left(\frac{\pi_{nrtx}}{\pi_{nrtx-1}}\right) = A_n - S_{rt} - C_x$$

เมื่อ	A_n	แทน	ระดับของความสำเร็จของนักเรียนคนที่ n
	C_x	แทน	ความยากของการเปลี่ยนระดับจากระดับ x ไปสู่ $x - 1$
	S_{rt}	แทน	ความเข้มงวดของผู้ให้คะแนนคนที่ r ณ เวลา t

ในโมเดลแยก (Separate Model) ลักษณะพิเศษเฉพาะของพารามิเตอร์ความเข้มงวด ได้ถูกประมาณค่าสำหรับผู้ให้คะแนนแต่ละคน ณ แต่ละจุดเวลา มีการยินยอมให้เปรียบเทียบค่าประมาณความเข้มงวดของผู้ให้คะแนน ณ จุดเวลาที่กำหนดไว้ เพื่อนำไปสู่การประมาณค่าความเข้มงวดของผู้ให้คะแนน ณ จุดเวลาอื่นๆ สิ่งที่สำคัญคือการวิเคราะห์ สามารถสร้างสเกลร่วม (Common Scale) สำหรับทุกจุดเวลาในโมเดลแยก โดยการสร้างเงื่อนไขให้ผู้ให้คะแนนแต่ละคน ณ แต่ละจุดเวลา เป็นเสมือนลักษณะพิเศษเฉพาะบุคคลเมื่อปรับให้เข้ากับข้อมูล

การตรวจสอบความคงที่ของผลกระทบจากผู้ให้คะแนน (Detecting Static Rater Effects) งานวิจัยในอดีต วิธีการรหัสแบบหลายองค์ประกอบได้ถูกใช้ในการตรวจสอบผลกระทบจากผู้ให้คะแนนโดยที่มุ่งตรวจสอบความคงที่ของผลกระทบ (Static effect) การสำรวจโดยส่วนใหญ่ ในความเคลื่อนไหวของผลกระทบ (Dynamic effect) มีจุดสนใจบนความแตกต่างของความเข้มงวด และการศึกษาต่างๆ ได้อธิบายว่า การวิเคราะห์สามารถตรวจสอบรูปแบบการตรวจให้คะแนน (Rating pattern) ได้อย่างไร โดยใช้ดัชนีผลกระทบจากผู้ให้คะแนนเฉพาะ แต่มีการศึกษาในเรื่องดัชนีที่คงที่ของผลกระทบจากผู้ให้คะแนนในเรื่องของกำลังทางสถิติ (Statistical power) หรือระดับความคลาดเคลื่อน (Error rate) เพียงไม่กี่เรื่องเท่านั้นที่ถูกประเมินอย่างเป็นระบบ การวิจัยเรื่องความคงที่ของผลกระทบจากผู้ให้คะแนนก่อนหน้านี้ ยินยอมให้มีการคาดเดาเกี่ยวกับผลกระทบการทำหน้าที่ต่างกันของผู้ให้คะแนน เมื่อเวลาผ่านไปอื่นๆ มากกว่า จะให้มีการคาดเดาความแตกต่างกันของความเข้มงวด ซึ่งอาจเป็นเพราะว่าความแตกต่างกันของความเข้มงวดนั้น ปรากฏชัดเจนด้วยตัวมันเองในการวิเคราะห์แล้ว ในส่วนต่อไปจะแสดงดัชนี 2 ดัชนี ที่ใช้ในการระบุความแตกต่างของความแม่นยำและระดับสเกลที่ใช้

Single rater-rest of the raters correlation (SR-ROR : r_{SR-ROR}) เป็นลักษณะทั่วไปของสหสัมพันธ์แบบเพียร์สัน ที่ทำนายความแม่นยำเสมือนเป็นคุณลักษณะที่คงที่ของความสามารถของผู้ให้คะแนนภายใต้กรอบที่ถูกนำมาใช้ในการอ้างอิง r_{SR-ROR} นี้ถูกคำนวณโดยใช้คะแนนที่ได้มาจากผู้ให้คะแนนเพียงคนเดียว และส่วนที่เหลือจากผู้ให้คะแนนที่ถูกกำหนดไว้ ค่าสหสัมพันธ์ของ r_{SR-ROR} ถ้าผู้ให้คะแนนกำหนดตำแหน่งของระดับความสามารถของนักเรียนโดยสุ่มในการเปรียบเทียบการตรวจให้คะแนนของผู้ให้คะแนนคนอื่นๆ แล้ว ค่าของ r_{SR-ROR} จะเข้าใกล้ 0 แสดงว่าผู้ให้คะแนนไม่มีความแม่นยำ (Inaccuracy) ในทางตรงกันข้าม ถ้าผู้ให้คะแนนมีพฤติกรรมตรวจให้คะแนนสอดคล้องกันกับผู้ให้คะแนนคนอื่นๆ แล้ว ค่า r_{SR-ROR} จะมีค่าเข้าใกล้ 1 แสดงว่า ผู้ให้คะแนนมีความแม่นยำ หมายถึงเหตุที่

สำคัญคือค่าของ r_{SR-ROR} ไม่ควรจะได้รับอิทธิพลจากการตรวจให้คะแนนที่มีแนวโน้มเป็นกลาง เพราะว่า แนวโน้มเป็นกลาง มีผลต่อผลลัพธ์ในความแม่นยำของลำดับตำแหน่งของความสามารถของนักเรียน เนื่องจากการตรวจให้คะแนนที่มีแนวโน้มเป็นกลางมักจะถูกบีบโดยค่าเฉลี่ย

ดัชนี $r_{res,exp}$ สามารถนำไปใช้ประโยชน์ในการตรวจหาแนวโน้มเป็นกลางของการตรวจให้คะแนน เป็นสหสัมพันธ์ระหว่างส่วนที่เหลือและโมเดลที่มีพื้นฐานมาจากคะแนนที่คาดหวังไว้สำหรับ ผู้ให้คะแนนเฉพาะ ในระดับที่ไม่ซับซ้อน ดัชนีนี้สามารถตรวจสอบส่วนที่เหลือของการตรวจให้คะแนน ของผู้ให้คะแนน (r) แต่ละคน ที่ให้คะแนนความสามารถของนักเรียน (n) แต่ละคน และเชื่อมโยงโมเดล ที่มีพื้นฐานจากคะแนนที่คาดหวังสำหรับการตรวจให้คะแนนแต่ละคู่

$$R_{nr} = X_{nr} - E_{nr}$$

เมื่อ R_{nr} แทน ส่วนที่เหลือ
 X_{nr} แทน ความสามารถที่ผู้ให้คะแนนคนที่ r ให้กับนักเรียนคนที่ n
 E_{nr} แทน คะแนนจากโมเดลที่มีพื้นฐานจากการคาดหวังของผู้ให้คะแนนคนที่ r ให้กับนักเรียนคนที่ n และ

$$E_{nr} = \sum_{k=0}^m k\pi_{nrk}$$

เมื่อ π_{nrk} มีความคล้ายคลึงกับโมเดลองค์ประกอบของเวลาแสดงเป็นฟังก์ชัน ความน่าจะเป็น

สามารถคำนวณคะแนนส่วนที่เหลือที่คาดหวังไว้ (Residual-Expected Rating) หรือ R-ER ค่าสหสัมพันธ์ ($r_{res,exp}$) สำหรับผู้ให้คะแนนแต่ละคนเป็นเสมือนค่าเฉลี่ยของแนวโน้มของผู้ให้คะแนน ไปสู่การใช้สเกลกลางๆ มากเกินไปโดย เมื่อผู้ให้คะแนนแสดงหลักฐานของการมีแนวโน้มการใช้สเกล กลางๆ แสดงว่า ผู้ให้คะแนนคนนั้นให้คะแนนความสามารถของนักเรียนที่สูง ในขณะที่การทำนายของ โมเดลราส์ช หลายองค์ประกอบจะให้คะแนนที่ต่ำกว่า ในทางตรงกันข้ามถ้าผู้ให้คะแนนให้คะแนนความสามารถ ของนักเรียนต่ำ แต่การทำนายของโมเดลราส์ชแบบหลายองค์ประกอบจะให้ค่าคะแนนที่สูงกว่า ดังผล จากการวิจัยแบบแผนการกระจายของส่วนที่เหลือ (แกน y) และคะแนนที่คาดหวัง (แกน x) ควรจะ แสดงความชันที่เป็นลบสำหรับผู้ให้คะแนนที่มีแนวโน้มการใช้สเกลกลางๆ (นั่นคือ ค่าที่เป็นลบของ $r_{res,exp}$) ในกรณีของความไม่แม่นยำที่คงที่ (Static Inaccuracy) มันจะไม่มีประโยชน์อะไรเลยถ้าค่าสหสัมพันธ์ มีค่าเข้าใกล้ 0 ดังนั้นค่าสหสัมพันธ์ R-ER ควรจะอ่อนไหวต่อผลกระทบของการมีแนวโน้มการใช้สเกล กลางๆ แต่ควรจะไม่เป็นอิทธิพลโดยผลกระทบจากความไม่แม่นยำที่คงที่ (Static Inaccuracy Effect)

การตรวจสอบความเคลื่อนไหวของผลกระทบจากผู้ให้คะแนน (Detecting Dynamic Rater Effects) มีความเป็นไปได้ที่จะเปลี่ยนแปลงดัชนีความคงที่ของผลกระทบไปสู่สถิติที่อ่อนไหว ต่อผลกระทบการทำหน้าที่ต่างกันของผู้ให้คะแนนเมื่อเวลาผ่านไป โดยการเปรียบเทียบค่าที่ได้มาจาก

ผู้ให้คะแนนเฉพาะในสองช่วงจุดเวลา (เช่น หนึ่งสามารถนำไปสู่สถิติการทดสอบเพื่อเปรียบเทียบ สอง สหสัมพันธ์ SR-ROR) ระวังไว้เสมอว่าเมื่อผู้ให้คะแนนมีความแม่นยำสูง SR-ROR จะสูงในการเปรียบเทียบ ว่าผู้ให้คะแนนคนใดที่ไม่แม่นยำ ดังนั้นผู้ให้คะแนนที่กลายมาเป็นผู้ที่ให้คะแนนแม่นยำมากกว่า เมื่อเวลาผ่านไปจะสร้างการตรวจให้คะแนนที่ผลลัพธ์จะมีค่า SR-ROR สูง และผู้ที่ให้คะแนนแม่นยำน้อยกว่า เมื่อเวลาผ่านไปจะสร้างการตรวจให้คะแนนที่ผลลัพธ์จะมีค่า SR-ROR ต่ำ เราสามารถทดสอบสมมติฐานกลาง ที่ค่าสัมประสิทธิ์สหสัมพันธ์ 2 ค่ามีค่าเท่ากัน โดยทำสหสัมพันธ์ Fisher's เปลี่ยนรูปให้เป็นมาตรฐาน

$$Z_{SR-ROR} = \frac{\ln(1 + r_{SR-ROR}) - \ln(1 - r_{SR-ROR})}{2}$$

การเปลี่ยนรูปนี้ได้สร้างค่าที่เป็นบวกของ Z_{SR-ROR} เมื่อ r_{SR-ROR} มีค่าเข้าใกล้ 1.00, สร้างค่าที่เป็นลบของ Z_{SR-ROR} เมื่อ r_{SR-ROR} มีค่าเข้าใกล้ -1.00 และสร้างค่าที่เป็น .00 ของ Z_{SR-ROR} เมื่อ r_{SR-ROR} มีค่าเข้าใกล้ .00 ต่อมาสามารถเปลี่ยนรูปสหสัมพันธ์ไปสู่การทดสอบสมมติฐาน เพื่อสรุป โดยไม่คำนึงถึงว่าเป็นความแตกต่างอย่างมีนัยสำคัญเชิงสถิติ

$$Z_{SR-ROR_c, SR-ROR_b} = \frac{Z_{r_{SR-ROR_c}} - Z_{r_{SR-ROR_b}}}{\sqrt{\frac{1}{N_c - 3} + \frac{1}{N_b - 3}}}$$

เมื่อ $Z_{r_{SR-ROR_c}}$ แทน ค่าที่ได้จากการแปลงรูปด้วยวิธี Fisher ณ เวลา c
 $Z_{r_{SR-ROR_b}}$ แทน ค่าที่ได้จากการแปลงรูปด้วยวิธี Fisher ณ เวลา b
 N_c แทน จำนวนของการตรวจให้คะแนนในเวลา $Z_{r_{SR-ROR_c}}$ เป็นฐาน
 N_b แทน จำนวนของการตรวจให้คะแนนในเวลา $Z_{r_{SR-ROR_b}}$ เป็นฐาน

ค่าของสถิติสามารถเปรียบเทียบกับ Standard Normal Distribution มีข้อตกลงเบื้องต้นว่าเวลา B เกิดก่อนเวลา C แล้ว ค่าที่มากกว่า .196 บ่งชี้ว่า SR-ROR ที่เปรียบเทียบกับเวลาที่ C สูงกว่า สหสัมพันธ์ ณ เวลา B อย่างมีนัยสำคัญทางสถิติ (ในระดับ .05) เป็นสัญญาณว่ามีการเพิ่มขึ้นของความแม่นยำเมื่อเวลาผ่านไป ค่าที่ต่ำกว่า -1.96 บ่งชี้ว่า SR-ROR ที่เปรียบเทียบกับเวลาที่ C ต่ำกว่า สหสัมพันธ์ ณ เวลา B อย่างมีนัยสำคัญทางสถิติ (ในระดับ .05) เป็นสัญญาณว่ามีการลดลงของความแม่นยำเมื่อเวลาผ่านไป ค่าที่ต่ำกว่า .00 บ่งชี้ว่ามีการเปลี่ยนแปลงเล็กน้อยของความแม่นยำเมื่อเวลาผ่านไป

สามารถอธิบายด้วยภาพในความเปลี่ยนแปลงของสหสัมพันธ์ R-ER เมื่อเวลาผ่านไป เพื่อวัดความแตกต่างของสเกลที่ใช้ ระวังไว้ว่าสหสัมพันธ์ระหว่างส่วนที่เหลือและการตรวจให้คะแนน ที่คาดหวังมีค่าเข้าใกล้ .00 ในกรณีของทั้งความแม่นยำและไม่แม่นยำ ในทางตรงกันข้ามเมื่อผลกระทบจากการใช้สเกลกลางๆ ถูกค้นพบ สหสัมพันธ์ R-ER จะมีค่าติดลบ ดังนั้นจากนี้ไป การแปลงค่าด้วย Fisher ของ $r_{res,exp}$ ได้สร้างค่าที่เป็นลบสำหรับ $Z_{res,exp}$ เมื่อผู้ให้คะแนนให้คะแนนแสดงออกถึงการใช้

สเกลกลางๆ ค่า $Z_{res,exp}$ เข้าใกล้ .00 เมื่อมีหลักฐานเพียงเล็กน้อยว่า มีการใช้สเกลกลางๆ และค่า $Z_{res,exp}$ เป็นบวก เมื่อมีการกระจายตัวของคะแนนที่ได้จากผู้ให้คะแนนมีค่อนข้างมาก (ตรงข้ามกับการใช้สเกลกลางๆ)

การประยุกต์ข้อมูลไปสู่การทดสอบเชิงสถิติเหมือนว่าจะแสดงโมเดลเริ่มต้น ดังนี้

$$Z_{res,exp_c - res,exp_b} = \frac{Z_{res,exp_c} - Z_{res,exp_b}}{\sqrt{\frac{1}{N_c - 3} + \frac{1}{N_b - 3}}}$$

เมื่อ Z_{res,exp_c} แทน ค่าที่ได้จากการแปลงรูปด้วยวิธี Fisher ของ R-ER ณ เวลา c

$Z_{r_{SR-RO}b}$ แทน ค่าที่ได้จากการแปลงรูปด้วยวิธี Fisher ของ R-ER ณ เวลา b

N_c แทน จำนวนของการตรวจให้คะแนนในเวลาที่ Z_{res,exp_c} เป็นฐาน

N_b แทน จำนวนของการตรวจให้คะแนนในเวลาที่ $Z_{r_{SR-RO}b}$ เป็นฐาน

การเปรียบเทียบสถิตินี้ไปสู่ Standard Normal Distribution และมีข้อตกลงว่า เวลา B เกิดก่อนเวลา C ค่าที่น้อยกว่า -1.96 มีสัญญาณว่ามีการลดลงในการกระจายออกของการใช้สเกลกลางๆ เมื่อเวลาผ่านไป ในทางกลับกัน ค่าที่มากกว่า 1.96 บ่งชี้ว่า มีการเพิ่มขึ้นของการกระจายออกของการใช้สเกลกลางๆ ของการตรวจให้คะแนนเมื่อเวลาผ่านไปและพร้อมกันกับการลดลงของการใช้สเกลกลางๆ ของผู้ให้คะแนนเมื่อเวลาผ่านไป ค่าที่เข้าใกล้ .00 บ่งชี้ว่า มีการเปลี่ยนแปลงเพียงเล็กน้อยของการตรวจให้คะแนนเมื่อเวลาผ่านไป

ดังนั้นกล่าวโดยสรุป การระบุการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป การวิเคราะห์เปรียบเทียบความเข้มงวด/ใจดีของผู้ตรวจให้คะแนนว่า มีความแตกต่างกันหรือไม่เฉพาะในสองช่วงจุดเวลา คือวันที่ 1 และวันที่ 2 ในขั้นแรกนั้น ต้องพิจารณาว่าผู้ตรวจให้คะแนนแต่ละคนให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล Partial Credit Model หรือไม่โดยพิจารณาค่า INFIT และ OUTFIT ต่อมาจึงระบุการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป จากค่า Z-Score โดยที่ค่า Z-Score ที่มีนัยสำคัญทางสถิติแสดงว่าผู้ตรวจให้คะแนนให้คะแนนไม่คงที่ข้ามเวลา ถ้า Z-Score ไม่มีนัยสำคัญทางสถิติแสดงว่าผู้ตรวจให้คะแนนให้คะแนนคงที่ข้ามเวลา

กล่าวโดยสรุป จุดประสงค์หลักของการวัดผลการศึกษาคือสามารถวัดสมองให้ตรงกับความสามารถที่แท้จริงของนักเรียนเพื่อนำผลนั้นไปพัฒนาในลำดับต่อไป แต่การประเมินผู้ให้คะแนนอาจจะเปลี่ยนแปลงข้ามเวลาหรือข้ามข้อสอบ ข้อสอบบางข้ออาจให้คะแนนยากกว่าข้ออื่นๆ และความแปรปรวนในการอบรมผู้ให้คะแนนอาจจะได้รับผลกระทบจากพฤติกรรมของผู้ให้คะแนน แม้ว่าทฤษฎีการสรุปอ้างอิงจะเสนอแนะวิธีโดยตรงในการประเมินความคงที่ของผู้ให้คะแนน แต่ที่ไม่มีวิธีใดที่จะประเมินความคงที่ของผู้ให้คะแนนจนเมื่อโมเดล IRT ถูกนำมาใช้ โมเดล IRT ส่วนใหญ่สมมติฐานว่าการตอบของผู้สอบในแต่ละข้อ

เป็นอิสระจากการตอบข้อสอบข้ออื่นๆ เป็นเงื่อนไขบนความสามารถของผู้สอบ เมื่อใช้ผู้ให้คะแนนหลายคน (Multiple Raters) ในการประเมินการตอบข้อสอบของผู้สอบ และคะแนนที่หลากหลาย (Multiple rating) จากที่ได้กล่าวไปข้างต้น ที่การวิจัยได้แบ่งออกเป็น 2 ส่วนใหญ่ๆ คือ ความคลาดเคลื่อนที่มีแหล่งที่มาจากภายในผู้ตรวจให้คะแนน และความคลาดเคลื่อนที่มีแหล่งที่มาจากภายนอกผู้ตรวจให้คะแนน ดังนั้น ผู้วิจัยจึงได้คัดเลือกความคลาดเคลื่อนที่ส่งผลต่อการตรวจให้คะแนนออกเป็นความคลาดเคลื่อนที่มีแหล่งที่มาจากภายในผู้ตรวจให้คะแนน คือ ความเข้มงวด/ใจดี (Severity/Leniency) ความลำเอียงทางเพศของผู้ให้คะแนน (Differential Rater Functioning : DRF) และการทำหน้าที่ต่างกันของผู้ให้คะแนนเมื่อเวลาผ่านไป (Differential Rater Functioning Over Time : DRIFT) และความคลาดเคลื่อนที่มีแหล่งที่มาจากภายนอกผู้ตรวจให้คะแนน คือ วิธีการตรวจให้คะแนนการเขียนเรียงความ ซึ่งมีอยู่ 3 แบบ คือ การตรวจให้คะแนนแบบรวมองค์ประกอบ (Holistic Scoring Rubrics) การตรวจให้คะแนนแบบแยกองค์ประกอบ (Analytic Scoring Rubrics) และการตรวจให้คะแนนแบบผสมผสาน (Annotated Scoring Rubrics)

ตอนที่ 3 ความเที่ยงตรง

ในตอนนี้นำผู้วิจัยได้แบ่งเอกสารที่เกี่ยวข้องกับความเที่ยงตรงไว้ดังนี้ คือ ความหมาย ความสำคัญ และประเภทของความเที่ยงตรง ความหมายและการแสดงหลักฐานความเที่ยงตรง ซึ่งจะกล่าวถึงที่มา และรายละเอียดของตัวแปรตาม ดังนี้

1. ความหมายของความเที่ยงตรง

การหาความเที่ยงตรง เป็นการตรวจสอบความสมเหตุสมผลของการแปลความหมายของเครื่องมือวัด โดยมีเป้าหมายเพื่อแปลความหมายของข้อมูลที่ได้จากการวัดว่าข้อมูลที่ได้มานั้นเป็นตัวแทนที่ดีในการบรรยายคุณลักษณะ และพยากรณ์คุณลักษณะอื่นได้ดีเพียงไร ความเที่ยงตรงถือเป็นระดับความแม่นยำที่เครื่องมือวัดจะวัดในสิ่งที่ต้องการวัด ถ้ามีความถูกต้องแม่นยำสูงก็แสดงว่าเครื่องมือวัดนั้นมีความเที่ยงตรงสูง ความเที่ยงตรงเป็นคุณสมบัติที่สำคัญของเครื่องมือทุกชนิด ทั้งแบบทดสอบผลสัมฤทธิ์ทางการศึกษา แบบทดสอบวัดความถนัด แบบทดสอบวัดเจตคติ แบบวัดบุคลิกภาพ ฯลฯ ซึ่งทุกฉบับจะต้องมีคุณภาพด้านความเที่ยงตรง จึงจะเชื่อได้ว่าเป็นเครื่องมือวัดที่ดีและผลที่ได้จากการวัดจะถูกต้องตรงตามความต้องการ (บุญเชิด ภิญญอนันตพงษ์. 2545: 171 – 172)

ความเที่ยงตรง (Validity) หมายถึง ความสามารถของเครื่องมือในการวัดในสิ่งที่ต้องการวัดได้ตามจุดมุ่งหมายที่กำหนด (ล้วน สายยศ; และ อังคณา สายยศ. 2543; มหาวิทยาลัยสุโขทัยธรรมาธิราช. 2549; เตือนใจ เกตุษา. 2549; เยาวดี วิบูลย์ศรี. 2552; Linderman; & Merenda. 1979; Carnines; & Zeller. 1987; Lyman. 1991; & Gruijter; & Van, der Kamp. 2008) นอกจากนี้ยังเป็นระดับความสอดคล้องหรือความสัมพันธ์ระหว่างคะแนนที่ได้จากการวัดกับสิ่งที่ต้องการวัดนั่นเอง (บุญเชิด ภิญญอนันตพงษ์. 2547; & Gruijter; & Van, der Kamp. 2008) ความเที่ยงตรงเป็นคุณสมบัติที่สำคัญมากของแบบทดสอบที่ผู้ใช้แบบทดสอบจะต้องคำนึงถึง (ล้วน สายยศ; และ อังคณา สายยศ. 2543; บุญเชิด ภิญญอนันตพงษ์.

2547; มหาวิทยาลัยสุโขทัยธรรมาธิราช. 2549; เตือนใจ เกตุษา. 2549; & Linderman; & Merenda. 1979) ความเที่ยงตรงของแบบทดสอบมีประโยชน์ทั้งในทางทฤษฎีการทดสอบและในทางปฏิบัติ โดยทั่วไปแบบทดสอบมีจุดมุ่งหมายเพื่อทำความเข้าใจ เพื่ออธิบาย และเพื่อทำนายพฤติกรรมต่างๆ ของผู้สอบ จุดมุ่งหมายของแบบทดสอบจะบรรลุผลย่อมขึ้นอยู่กับความสามารถในการอ้างอิงคะแนนนั้น ไปยังผู้สอบได้อย่างแท้จริง แบบทดสอบที่ใช้ในการวิจัยนี้เป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน แบบทดสอบที่มีความเที่ยงตรงย่อมทำให้ผลที่ได้สามารถระบุว่านักเรียนมีผลสัมฤทธิ์ตรงตามจุดมุ่งหมายของการวัด

ถึงแม้ว่า ความเที่ยงตรงจะมีความสำคัญในการวัดผลการศึกษามากก็ตาม แต่บางครั้งจะพบว่า ได้เกิดความเข้าใจผิดในกลุ่มนักการศึกษาในเรื่องการอ้างถึงความเที่ยงตรงของแบบทดสอบ เพราะแบบทดสอบฉบับหนึ่งนั้นจะเป็นเพียงเครื่องมือที่ใช้เพื่ออ้างอิงคะแนนไปสู่สิ่งที่วัดเท่านั้น ถ้าหากใช้เครื่องมือผิดโดยการนำไปใช้ในสถานการณ์ที่ไม่เหมาะสม ก็มิได้หมายความว่าเครื่องมือนั้นมีข้อบกพร่อง หากแต่เป็นข้อบกพร่องที่มีต้นเหตุมาจากผู้ใช้ การวัดความเที่ยงตรง นอกจากจะมีความสำคัญในทางการศึกษาแล้วยังมีความสำคัญต่อกิจกรรมอื่นๆ ที่ใช้ระเบียบวิธีการทางวิทยาศาสตร์อีกด้วย

ดังนั้น ความเที่ยงตรง หมายถึง ความสามารถของเครื่องมือในการวัดในสิ่งที่ต้องการวัดได้ตามจุดมุ่งหมายที่กำหนด

2. ประเภทของความเที่ยงตรง

การแบ่งประเภทของความเที่ยงตรง มักแบ่งออกตามเกณฑ์ที่ใช้ในการอ้างอิง ซึ่งจากการศึกษาเอกสารพบว่าการแบ่งประเภทความเที่ยงตรงออกเป็น 2 – 4 ประเภท โดยส่วนใหญ่แล้วแบ่งออกเป็น 3 ประเภท (ล้วน สายยศ; และ อังคณา สายยศ. 2543; บุญเชิด ภิญโญอนันตพงษ์. 2547; เตือนใจ เกตุษา. 2549; มหาวิทยาลัยสุโขทัยธรรมาธิราช. 2549; เยาวดี วิบูลย์ศรี. 2552; Carnines; & Zeller. 1987; & Gruijter; & Van, der Kamp. 2008) คือ

1. ความเที่ยงตรงตามเนื้อหา (Content Validity) เป็นความสามารถของเครื่องมือที่วัดได้ตามเนื้อหาที่ต้องการจะวัด ซึ่งการพิจารณาความเที่ยงตรงชนิดนี้จะใช้การวิเคราะห์อย่างมีเหตุผล (Rational Analysis) แบ่งออกเป็น 2 ประเภทย่อย คือ ความเที่ยงตรงเชิงเหตุผล (Logical Validity) และความเที่ยงตรงเชิงประจักษ์ (Face Validity)

2. ความเที่ยงตรงเชิงโครงสร้าง (Construct Validity) เป็นความสามารถของเครื่องมือที่วัดได้ตรงตามลักษณะหรือตามทฤษฎีต่างๆ ของโครงสร้างนั้น

3. ความเที่ยงตรงตามเกณฑ์สัมพัทธ์ (Criterion-Related Validity) เป็นความสามารถของเครื่องมือที่เอาผลการวัดของแบบทดสอบไปหาความสัมพันธ์กับเกณฑ์ที่ต้องการ แบ่งได้เป็น 2 ประเภทย่อย คือ ความเที่ยงตรงตามสภาพ และความเที่ยงตรงเชิงพยากรณ์

ในงานวิจัยขั้นนี้ผู้วิจัยสนใจจะแสดงความเที่ยงตรงตามเนื้อหา ในประเภทย่อยความเที่ยงตรงเชิงประจักษ์ ความเที่ยงตรงเชิงโครงสร้าง ตามวิธีลักษณะหลากหลาย-วิธีหลาย และความเที่ยงตรงตามเกณฑ์สัมพัทธ์ในประเภทย่อย ความเที่ยงตรงตามสภาพ โดยมีรายละเอียด ดังนี้

3. การแสดงหลักฐานความเที่ยงตรง

3.1 การแสดงหลักฐานความเที่ยงตรงตามเนื้อหา ความเที่ยงตรงตามเนื้อหา คือการลงความคิดเห็นเกี่ยวกับความสอดคล้องระหว่างเนื้อหาของเครื่องมือวัดกับเนื้อหาของเกณฑ์ ถ้ามีความสอดคล้องกันมากแสดงว่า มีหลักฐานแสดงความเที่ยงตรงตามเนื้อหาสูง ซึ่งเครื่องมือที่มีความเที่ยงตรงตามเนื้อหาสูง เครื่องมือนั้น จะต้องมียุทธศาสตร์คำถามที่นำมาสร้างเป็นข้อคำถามแต่ละข้อครอบคลุมเนื้อหาอย่างครบถ้วน คำว่า "เนื้อหา" ซึ่งเป็นเกณฑ์ของความเที่ยงตรงตามเนื้อหา ไม่ได้หมายถึงเฉพาะเนื้อหาวิชาเพียงอย่างเดียว แต่ยังหมายถึงพฤติกรรมแสดงออกตามเนื้อหาด้วย วิธีการแสดงหลักฐานความเที่ยงตรงตามเนื้อหาจะใช้ดุลยพินิจของผู้เชี่ยวชาญ โดยอาศัยหลักฐานผลการตัดสินของผู้เชี่ยวชาญในคุณลักษณะที่ต้องการจะวัด ถือได้ว่าเป็นการแสดงหลักฐานที่มีความเป็นอัตนัยมาก (จินตนา ธนวิบูลย์ชัย, 2535: 71) ดังนั้นสิ่งที่พึงระลึกไว้เสมอก็คือ ข้อสรุปที่ได้จากการตรวจสอบความเที่ยงตรงนี้ จะมีลักษณะเป็นเพียงว่าแบบทดสอบที่สร้างขึ้นนั้นดูเหมือนว่า จะวัดคุณลักษณะที่ต้องการวัดเท่านั้นยังสรุปไม่ได้ว่าสามารถวัดคุณลักษณะนั้นได้อย่างแท้จริง แต่การหาความเที่ยงตรงตามเนื้อหานี้นับได้ว่าควรเป็นขั้นตอนที่สำคัญแรกของการหาความเที่ยงตรงชนิดอื่นๆ

เครื่องมือในการวิจัยในครั้งนี้ ต้องผ่านการหาความเที่ยงตรงเชิงประจักษ์ ซึ่งเป็นประเภทย่อยของความเที่ยงตรงตามเนื้อหา เพื่อให้แน่ใจว่าเครื่องมือที่ใช้ในการวิจัย จะสามารถวัดได้ตรงกับจุดประสงค์ของการวิจัยมากที่สุด

3.2 การแสดงหลักฐานเที่ยงตรงเชิงโครงสร้าง หลักฐานการแสดงความเที่ยงตรงตามโครงสร้างเป็นการแสดงว่า เครื่องมือวัดนั้นสามารถวัดภายใต้ขอบเขต ความหมาย หรือคุณลักษณะประจำตามโครงสร้างทางทฤษฎีที่สมมติขึ้นนั้น ได้เพียงใด คำว่า "โครงสร้าง" (Construct) ในที่นี้หมายถึงตัวประกอบหรือองค์ประกอบ (Factor) ดังนั้นการหาความเที่ยงตรงเชิงโครงสร้างจึงเป็นการหาความสอดคล้องระหว่างองค์ประกอบที่วัดในเครื่องมือวัดกับองค์ประกอบที่ต้องการวัดว่าวัดในองค์ประกอบเดียวกันหรือไม่ (บุญเชิด ภิญญอนันตพงษ์, 2545: 177) และพึงระลึกอยู่เสมอว่าตัวแบบทดสอบไม่ใช่สิ่งที่เราสนใจ แต่ความสนใจจะอยู่ที่แบบทดสอบได้สะท้อนปรากฏการณ์ในโลกของความเป็นจริง หรือพฤติกรรม หรือคุณสมบัติบางประการหรือไม่ เพราะลักษณะที่เป็นแบบแผนและนามธรรมไม่สามารถสังเกตได้โดยตรงด้วยสัมผัสทั้ง 5 แต่ต้องอาศัยการอ้างอิงจากปรากฏการณ์ที่สังเกตได้ (จินตนา ธนวิบูลย์ชัย, 2535: 80)

การแสดงหลักฐานความเที่ยงตรงเชิงโครงสร้างจะต้องให้นิยามปฏิบัติการ (Operational Definition) โดยอาศัยทฤษฎีใดทฤษฎีหนึ่งเป็นหลัก แล้วจึงสร้างเครื่องมือวัดสิ่งนั้นตามคำนิยามปฏิบัติการ ซึ่งนิยามปฏิบัติการจะประกอบไปด้วย 3 ส่วน คือ คุณลักษณะ (Trait) ที่ต้องการวัด สิ่งเร้าหรือสถานการณ์ที่จะกระตุ้นให้คุณลักษณะนั้นแสดงออกมา และการตอบสนองแสดงออกที่สามารถสังเกตเห็นได้ ความเที่ยงตรงเชิงโครงสร้างจะมีลักษณะคล้ายกันกับความเที่ยงตรงตามเนื้อหาและความเที่ยงตรงตามเกณฑ์ แต่ความแตกต่างอยู่ที่หลักฐานความเที่ยงตรงเชิงโครงสร้างต้องสะสมรายละเอียดจากหลายแหล่ง การกำหนดเกณฑ์เพียง

เกณฑ์เดียวจึงไม่เพียงพอที่จะอธิบายความเที่ยงตรงเชิงโครงสร้างได้ ต้องอาศัยรายละเอียดที่มากกว่า หรือต้องอาศัยเกณฑ์หลายเกณฑ์ (บุญเชิด ภิญโญนันตพงษ์. 2545: 178)

วิธีการแสดงหลักฐานความเที่ยงตรงตามโครงสร้างนั้น เป็นการแสดงหลักฐานว่า เครื่องมือนั้นสามารถวัดขอบเขต ความหมาย หรือคุณลักษณะประจำตามโครงสร้างที่สมมติขึ้นนั้นได้ เพียงใด การแสดงหลักฐานความเที่ยงตรงตามโครงสร้างมีหลายวิธี เช่น วิธีพิจารณาเปรียบเทียบกับเกณฑ์ที่กำหนด วิธีเปรียบเทียบจากกลุ่มที่ต่างกัน วิธีเปรียบเทียบกับเครื่องมือมาตรฐานที่วัดคุณลักษณะเดียวกัน วิธีการหาความสอดคล้องภายในเครื่องมือวัด วิธีการวิเคราะห์องค์ประกอบ และวิธีการวิเคราะห์ลักษณะหลากหลาย ซึ่งงานวิจัยครั้งนี้สนใจการหาความเที่ยงตรงเชิงโครงสร้างโดยใช้วิธีการวิเคราะห์ลักษณะหลากหลาย-วิธีหลาย

การวิเคราะห์ลักษณะหลากหลาย-วิธีหลาย (Multitrait-Multimethod Analysis) หมายถึง การวิเคราะห์ความสัมพันธ์ระหว่างผลการวัดคุณลักษณะต่างๆ ของมนุษย์ด้วยวิธีการต่างๆ โดยถือหลักการว่า ถ้าวัดลักษณะเดียวในตัวมนุษย์ด้วยวิธีวัดหลายๆ วิธีแล้ว ผลการวัดด้วยวิธีเหล่านั้น ย่อมมีความสัมพันธ์กันสูง แต่ถ้าวัดลักษณะต่างๆ ของมนุษย์ด้วยวิธีการที่ต่างกัน ผลจากการวัดเหล่านั้นย่อมไม่มีความสัมพันธ์กัน หรือมีความสัมพันธ์กันต่ำมาก

แนวคิดและหลักการของการวิเคราะห์ลักษณะหลากหลาย-วิธีหลาย แคมป์เบล และ ฟิสค์ (สำเร็จ บุญเรืองรัตน์. 2547: 120; อ้างอิงจาก Campbell; & Fiske. 1995) เป็นผู้เสนอความคิดเมตริกซ์ ลักษณะหลากหลาย-วิธีหลาย (Multitrait-Multimethod Matrix : MTMM) คำว่า ลักษณะหลากหลาย ตรงกับภาษาอังกฤษว่า Multitrait ส่วนคำว่าวิธีหลาย ตรงกับคำภาษาอังกฤษว่า Multimethod โดยเมตริกซ์จะแสดงค่าสหสัมพันธ์ระหว่างผลการวัดลักษณะต่างๆ ของมนุษย์ด้วยวิธีต่างๆ กัน มนุษย์แต่ละคนมีลักษณะ (Trait) ต่างๆ อยู่ในตัวมากมายหลายชนิด เช่น มีความรู้ มีความถนัดด้านภาษา มีความถนัดด้านคณิตศาสตร์ มีความถนัดด้านมิติสัมพันธ์ มีความซื่อสัตย์ มีความรับผิดชอบ มีมนุษยสัมพันธ์ มีวินัย มีความขยันอดทน มีความเมตตา กรุณา เป็นต้น

ถ้าให้	T_1	แทน	ลักษณะที่ 1
	T_2	แทน	ลักษณะที่ 2
	T_3	แทน	ลักษณะที่ 3
	T_n	แทน	ลักษณะที่ n

จะเห็นว่า มีลักษณะต่างๆ กัน ตั้งแต่ $T_1, T_2, T_3, \dots, T_n$ ในแต่ละบุคคลมีลักษณะเหล่านี้ มากน้อยแตกต่างกัน และแต่ละลักษณะจะมีวิธีวัด (Method) ได้หลายวิธี เช่น วัดความรู้ อาจใช้แบบทดสอบ เลือกตอบเป็นวิธีที่ 1 การสัมภาษณ์เพื่อประเมินความรู้เป็นวิธีที่ 2 หรือวิธีอื่นๆ เพื่อวัดความรู้ การวัดว่า บุคคลใดมีความรับผิดชอบสูง ก็อาจวัดได้หลายวิธี เช่น มอบหมายงานให้ทำแล้วสังเกต ประเมินความรับผิดชอบหรือสร้างแบบทดสอบเพื่อวัดความรับผิดชอบ เป็นต้น

ถ้าให้	M_1	แทน	วิธีที่ 1 สำหรับวัดลักษณะที่ 1
	M_2	แทน	วิธีที่ 2 สำหรับวัดลักษณะที่ 1

M_3 แทน วิธีที่ 3 สำหรับวัดลักษณะที่ 1

M_1 แทน วิธีที่ 1 สำหรับวัดลักษณะที่ 1

ลักษณะที่ 1 สามารถวัดได้ด้วยวิธี $M_1, M_2, M_3, \dots, M_1$ นักวัดผลพยายามสร้างเครื่องมือหรือพัฒนาเทคนิควิธีการวัดต่างๆ อยู่เสมอ เพื่อจะได้รู้สภาพลักษณะต่างๆ ของมนุษย์ เมื่อสร้างวิธีวัดแล้วก็ต้องตรวจสอบความเที่ยงตรงของเครื่องมือเหล่านั้นว่า มีความเที่ยงตรงมากน้อยเพียงใด แคมป์เบล และ ฟิงค์ กำหนดหลักการนี้ไว้ ดังนี้

1. ถ้าแบบทดสอบต่างชนิด ต่างก็วัดลักษณะเดียวกันแล้ว สหสัมพันธ์จากผลการทดสอบด้วยแบบทดสอบเหล่านั้นจะต้องสูง เรียกลักษณะของแบบทดสอบเช่นนี้ว่ามีความเที่ยงตรงเชิงเหมือน (Convergent Validity)

2. ถ้าแบบทดสอบต่างชนิดต่างก็วัดลักษณะซึ่งต่างกันแล้ว ผลการทดสอบด้วยแบบทดสอบเหล่านั้น จะไม่มีสหสัมพันธ์กันหรือมีสหสัมพันธ์กันต่ำ เรียกลักษณะของแบบทดสอบเช่นนี้ว่าความเที่ยงตรงเชิงจำแนก (Discriminant Validity)

จากค่าสหสัมพันธ์ลักษณะหลากหลายวิธีเหล่านั้นสามารถสรุปเป็นเกณฑ์ในการวิเคราะห์ความเที่ยงตรงของแบบทดสอบได้ดังนี้ (สำเร็จ บุญเรืองรัตน์. 2547: 120 – 123)

1. ค่าความเชื่อมั่นของแบบทดสอบจะสูงกว่าค่าความเที่ยงตรงเชิงเหมือน
2. ค่าความเที่ยงตรงเชิงเหมือนจะมีค่ามากกว่าค่าความเที่ยงตรงเชิงจำแนก
3. ค่าความเที่ยงตรงเชิงเหมือนจะมีค่าสูง และมีนัยสำคัญทางสถิติ
4. ค่าความเที่ยงตรงเชิงจำแนกจะมีค่าต่ำ และค่าที่ใกล้เคียงกับ 0 หรือเป็น 0

3.3 การแสดงหลักฐานความเที่ยงตรงตามสภาพ การแสดงหลักฐานความเที่ยงตรงตามสภาพสามารถหาได้โดยการนำคะแนนที่ได้จากแบบทดสอบและคะแนนที่เป็นตัวเกณฑ์มาคำนวณหาความสัมพันธ์ โดยหาสัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) โดยใช้สูตรสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson Product Moment Correlation Coefficient) (ล้วน สายยศ; และ อังคณา สายยศ. 2543; บุญเชิด ภิญญอนันตพงษ์. 2547; มหาวิทยาลัยสุโขทัยธรรมาธิราช. 2549; และ เยาวดี วิบูลย์ศรี. 2552) ดังนี้

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

เมื่อ X แทน คะแนนที่ได้จากแบบทดสอบ

Y แทน คะแนนเกณฑ์

N แทน จำนวนผู้สอบ

การแสดงหลักฐานความเที่ยงตรงตามสภาพของแบบทดสอบอิงเกณฑ์ หมายถึง แบบทดสอบอิงเกณฑ์ที่สามารถวัดความรอบรู้ (Master) และไม่รอบรู้ (Nonmaster) ของนักเรียนในการทดสอบ (Test status) ในแต่ละจุดประสงค์ว่าตรงกับสภาพความจริงหรือไม่ ถ้าตรงกับสภาพความจริง ก็แสดงว่า

มีความเที่ยงตรงสูงจากค่าของ “ ϕ ” มีค่าระหว่าง -1 ถึง +1 ถ้าค่า “ ϕ ” มีค่าใกล้ +1 แสดงว่ามีความเที่ยงตรงต่ำหรือไม่เที่ยงตรง

แต่ในกรณีที่ต้องการเปรียบเทียบค่าสหสัมพันธ์ของกลุ่มตัวอย่างกับประชากรนั้น โดยปกติแล้วการแจกแจงของค่าสหสัมพันธ์นั้นมีลักษณะการแจกแจงแบบเบ้ ดังนั้นจึงแปลงสหสัมพันธ์ r เป็น z' ซึ่งจะมีการแจกแจงใกล้เคียงกับโค้งปกติ (ฉัตรศิริ ปิยะพิมลสิทธิ์. ม.ป.ป: ออนไลน์) โดยใช้สูตร

$$z' = \frac{1}{2} [\ln(1+r) - \ln(1-r)]$$

เมื่อ \ln แทน ลอการิทึมฐานธรรมชาติ

สำหรับงานวิจัยขั้นนี้ผู้วิจัยใช้การแสดงผลฐานความเที่ยงตรงตามสภาพของแบบทดสอบอิงกลุ่ม โดยใช้สูตรสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson Product Moment Correlation Coefficient) เพื่อหาค่าสหสัมพันธ์ r แต่ในกรณีที่ต้องการเปรียบเทียบค่าสหสัมพันธ์ของกลุ่มตัวอย่างกับประชากรนั้น โดยปกติแล้วการแจกแจงของค่าสหสัมพันธ์นั้นมีลักษณะการแจกแจงแบบเบ้ ดังนั้นจึงแปลงสหสัมพันธ์ r เป็น z' ซึ่งจะมีการแจกแจงใกล้เคียงกับโค้งปกติ จากนั้นจึงนำค่าที่ได้มาเปรียบเทียบความแตกต่างระหว่างค่าสหสัมพันธ์พิชเซอร์ z' เพื่อหาว่าภายใต้เงื่อนไข การใช้วิธีการตรวจให้คะแนนแตกต่างกัน ในสัดส่วนของจำนวนของผู้ให้คะแนนที่มีคุณลักษณะผู้ให้คะแนนแตกต่างกัน เงื่อนไขใดจะสัมพันธ์กับคะแนนการเขียนเรียงความและคะแนนจากการสอบการเขียนจดหมายของนักเรียนมากที่สุด

ในงานวิจัยขั้นนี้หาหลักฐานความเที่ยงตรงของความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 คือ ความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย และความเที่ยงตรงตามสภาพภายใต้การใช้วิธีการตรวจให้คะแนนแตกต่างกัน ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนนที่มีคุณลักษณะแตกต่างกัน เงื่อนไขใดที่มีความเที่ยงตรงตามสภาพสูงสุด

ตอนที่ 4 งานวิจัยที่เกี่ยวข้อง

ในตอนนีผู้วิจัยได้รวบรวมเอกสารที่เกี่ยวข้องกับการวัดความสามารถในการเขียนของนักเรียนคุณลักษณะของผู้ตรวจให้คะแนน และความเที่ยงตรงไว้ ดังนี้

1. การตรวจให้คะแนนการเขียน

นิคาร์ตัน คงสวัสดิ์ (2544) ได้ศึกษาผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียงที่มีจำนวนผู้ตรวจ และวิธีการตรวจต่างกัน โดยมีจุดมุ่งหมายสำคัญเพื่อศึกษาและเปรียบเทียบผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียงที่มีผู้ตรวจให้คะแนนจำนวนแตกต่างกัน คือ 2, 3 และ 4 คน เมื่อใช้วิธีการตรวจวิธีเดียวกัน และใช้วิธีการตรวจต่างวิธีกัน คือ การตรวจวิธีวิเคราะห์ย่อย (Analytic Method) และการตรวจวิธีประเมินรวม (Holistic Method) เมื่อมีผู้ตรวจจำนวนเท่ากัน กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 5 ของโรงเรียนสังกัดคณะกรรมการการศึกษาเอกชน จำนวน 60 คน และผู้ตรวจแบบทดสอบความเรียงเป็นครูผู้สอนวิชาสร้างเสริมประสบการณ์ชีวิต ชั้นประถมศึกษาปีที่ 5

ของโรงเรียนสังกัดสำนักงานคณะกรรมการการศึกษาเอกชน จำนวน 4 คน ตรวจสอบทดสอบความเรียง จำนวน 5 ข้อ ผลการวิจัยพบว่า เมื่อมีผู้ตรวจจำนวนแตกต่างกัน โดยใช้วิธีการตรวจเดียวกัน ค่าเฉลี่ยของความเชื่อมั่นของผู้ตรวจมีความแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ และเมื่อผู้ตรวจจำนวนเท่ากัน โดยใช้วิธีการตรวจต่างกัน ค่าเฉลี่ยของความเชื่อมั่นของผู้ตรวจมีความแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ

อัญญารัตน์ เจริญพฤษนิภา (2546) ได้พัฒนาแบบประเมินทักษะการอ่าน คิทธิวิเคราะห์ เขียนของนักเรียนชั้นประถมศึกษา ซึ่งสร้างและพัฒนาแบบประเมินทักษะการอ่าน คิทธิวิเคราะห์ เขียนของนักเรียนชั้นประถมศึกษา โดยใช้แนวคิดในการประเมินทักษะการอ่าน คิทธิวิเคราะห์ เขียนมาเป็นกรอบแนวคิด แล้วสนทนากลุ่ม กลุ่มครู นักเรียน ผู้ปกครอง เพื่อสร้างตัวบ่งชี้และเกณฑ์การตรวจให้คะแนนแบบแยกองค์ประกอบ กลุ่มตัวอย่างคือ นักเรียนระดับชั้นประถมศึกษาปีที่ 4 – 6 จำนวน 3,065 คน ซึ่งได้มาจากการสุ่มแบบหลายขั้นตอน ผลการวิจัยพบว่า 1) แบบประเมินทักษะการอ่าน คิทธิวิเคราะห์ เขียน มี 10 ตัวบ่งชี้ ได้แก่ การประเมินทักษะการอ่าน ตัวบ่งชี้ คือ ความคล่องในการอ่าน ความเข้าใจในการอ่าน วิธีการอ่านสาร การประเมินทักษะการคิทธิวิเคราะห์ ตัวบ่งชี้ คือ วิเคราะห์เนื้อหา วิเคราะห์ความสัมพันธ์ วิเคราะห์หลักการ การประเมินทักษะการเขียน ตัวบ่งชี้ คือ เนื้อเรื่อง ลำดับเรื่อง ไวยากรณ์ กลไกการเขียน 2) คุณภาพของแบบประเมินทักษะการอ่าน คิทธิวิเคราะห์ เขียน พบว่า แบบประเมินทักษะการอ่าน คิทธิวิเคราะห์ เขียน มีความตรงตามเนื้อหา ค่าดัชนีความสอดคล้องระหว่างตัวบ่งชี้กับทักษะที่ประเมิน และระหว่างเกณฑ์การตรวจให้คะแนนกับตัวบ่งชี้มีค่าเท่ากับ 0.83 – 1.00 มีความตรงเชิงเกณฑ์สัมพันธ์ ที่ได้จากการวิเคราะห์ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนที่ได้จากการประเมินทั้ง 3 ทักษะกับผลสัมฤทธิ์ทางการเรียน มีความสัมพันธ์อย่างมีนัยสำคัญที่ระดับ .01 มีความตรงเชิงจำแนกจากการทดสอบความแตกต่างระหว่างค่าเฉลี่ยของนักเรียน ที่มีทักษะทางการเรียนสูงและต่ำ พบว่า มีความแตกต่างกันอย่างมีนัยสำคัญที่ระดับ .01 และมีความตรงเชิงโครงสร้างที่ได้จากการวิเคราะห์องค์ประกอบเชิงยืนยัน

ปวีณา ปีอาทิตย์ (2545) ได้ศึกษาจำนวนผู้ประเมินและจำนวนงานเขียนที่เหมาะสมเมื่อใช้เกณฑ์การตรวจให้คะแนนที่ต่างกัน มีวัตถุประสงค์เพื่อ 1) เปรียบเทียบความเชื่อมั่นซึ่งแทนด้วยค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนการประเมินงานเขียนจากผู้ประเมินจำนวน 4 คน เมื่อใช้เกณฑ์การตรวจให้คะแนนแบบภาพรวม (Holistic Scoring Rubric) และผู้ประเมินจำนวน 4 คนที่ใช้เกณฑ์การตรวจให้คะแนนแบบแยกองค์ประกอบ (Analytic Scoring Rubric) และ 2) เพื่อหาจำนวนผู้ประเมินและจำนวนงานเขียนที่เหมาะสมเมื่อใช้เกณฑ์การตรวจให้คะแนนแบบภาพรวม (Holistic Scoring Rubric) หรือเกณฑ์การตรวจให้คะแนนแบบแยกองค์ประกอบ (Analytic Scoring Rubric) กลุ่มตัวอย่างประกอบด้วย ผลงานการเขียนวิชาภาษาไทยของนักเรียนชั้นประถมศึกษาปีที่ 5 จำนวน 117 ผลงาน และผู้ประเมินจำนวน 8 คน ที่ใช้เกณฑ์การตรวจให้คะแนนแบบภาพรวม และเกณฑ์การตรวจให้คะแนนแบบแยกองค์ประกอบ ผลการวิจัยพบว่า ภายใต้ชิ้นงานจำนวน 117 ชิ้น ผู้ประเมินจำนวน 4 คน เมื่อใช้เกณฑ์การตรวจให้คะแนนแบบภาพรวม ค่าสัมประสิทธิ์การสรุปอ้างอิงเท่ากับ 0.71484 และเมื่อใช้เกณฑ์การตรวจให้คะแนนแบบแยกองค์ประกอบ ค่าสัมประสิทธิ์การสรุปอ้างอิงเท่ากับ 0.81410 ถ้าต้องการ

ใช้ผู้ประเมิน 1 คน และกำหนดให้ค่าสัมประสิทธิ์การสุรุปอ้างอิงของเกณฑ์การตรวจให้คะแนนแบบภาพรวมเป็น 0.2, 0.4, 0.6 และ 0.8 ใช้จำนวนชิ้นงานอย่างน้อย 4, 8, 16 และ 28 ชิ้น ตามลำดับ ส่วนเกณฑ์การตรวจให้คะแนนแบบแยกองค์ประกอบ เมื่อกำหนดค่าสัมประสิทธิ์การสุรุปอ้างอิงเป็น 0.2 และ 0.4 ใช้จำนวนชิ้นงานน้อยกว่า 4 ชิ้น เมื่อกำหนดค่าสัมประสิทธิ์การสุรุปอ้างอิงเป็น 0.6 ใช้จำนวนชิ้นงานอย่างน้อย 8 ชิ้น และเมื่อกำหนดค่าสัมประสิทธิ์การสุรุปอ้างอิงเป็น 0.8 ใช้จำนวนชิ้นงานอย่างน้อย 12 ชิ้น ทั้งนี้ขึ้นอยู่กับผู้ปฏิบัติว่า จะเลือกใช้เกณฑ์การตรวจให้คะแนนให้เหมาะสมกับการเรียนการสอน

พรณี เจียมสุบุตร (2543) ได้ศึกษาเปรียบเทียบความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่มีจำนวนผู้ตรวจและวิธีการตรวจต่างกัน โดยมีจุดมุ่งหมาย เพื่อเปรียบเทียบค่าความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่มีจำนวนผู้สอบ 2 คน และ 3 คน โดยใช้วิธีการตรวจวิธีประเมินรวม และวิธีวิเคราะห์ย่อย โดยศึกษากับกลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2542 ของโรงเรียนในสังกัดกรมสามัญศึกษา กรุงเทพมหานคร กลุ่ม 7 จำนวน 70 คน ซึ่งเลือกมาโดยการสุ่มแบบหลายขั้นตอน เครื่องมือที่ใช้ในการศึกษาเป็นแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ ชั้นมัธยมศึกษาปีที่ 2 เรื่องสมการและอสมการ ผลการวิจัยพบว่า ค่าความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ด้วยวิธีวิเคราะห์ย่อย ที่มีจำนวนผู้ตรวจ 2 คน และ 3 คน มีค่าเท่ากับ 0.92 และ 0.96 ตามลำดับ ซึ่งแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนการตรวจให้คะแนนด้วยวิธีรวมที่มีจำนวนผู้ตรวจ 2 คน และ 3 คน มีค่าเท่ากับ 0.97 และ 0.98 ตามลำดับ ซึ่งแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ค่าความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่ใช้วิธีการตรวจให้คะแนนแบบวิธีรวม และวิธีวิเคราะห์ของผู้ตรวจให้คะแนน 2 คน มีค่าเท่ากับ 0.97 และ 0.92 ตามลำดับ ซึ่งแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ส่วนแบบทดสอบที่ใช้วิธีการตรวจให้คะแนนวิธีประเมินรวม และวิธีการวิเคราะห์ของผู้ตรวจ 3 คน มีค่าเท่ากับ 0.98 และ 0.96 ตามลำดับ ซึ่งแตกต่างอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

อุษณีย์ บัวศิริพันธ์ (2543) ได้ศึกษาเปรียบเทียบค่าสัมประสิทธิ์การสุรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจ จำนวนผู้ตรวจ และประสบการณ์ของผู้ตรวจ แตกต่างกัน โดยมีจุดมุ่งหมาย เพื่อเปรียบเทียบค่าสัมประสิทธิ์การสุรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนแตกต่างกัน มีจำนวนผู้ตรวจให้คะแนนต่างกัน และผู้ตรวจมีประสบการณ์แตกต่างกัน กลุ่มตัวอย่างเป็นนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2540 ของโรงเรียนในสังกัดกรมสามัญศึกษา จังหวัดจันทบุรี จำนวน 60 คน ได้มาจากการสุ่มแบบหลายขั้นตอน ผลการวิจัยพบว่า ค่าสัมประสิทธิ์การสุรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจแตกต่างกัน จำนวนผู้ตรวจเท่ากัน มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ค่าสัมประสิทธิ์การสุรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจให้คะแนนแตกต่างกัน ผู้ตรวจมีประสบการณ์เหมือนกัน มีค่าแตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ค่าสัมประสิทธิ์การสุรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีจำนวนผู้ตรวจแตกต่างกัน วิธีการตรวจให้คะแนนเหมือนกัน มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01 และค่าสัมประสิทธิ์

การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีผู้ตรวจให้คะแนนมีประสบการณ์ต่างกัน วิธีการตรวจเหมือนกัน มีค่าแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .01

น้ำผึ้ง อินทะเนตร (2554) ได้ศึกษาคุณลักษณะของคะแนนแบบทดสอบปลายเปิดวิชาคณิตศาสตร์ เมื่อจำนวนผู้ตรวจและรูปแบบการตรวจให้คะแนนต่างกัน โดยใช้โมเดลการสรุปอ้างอิง และโมเดลหลายองค์ประกอบของราล์ซ โดยมุ่งหมายเพื่อศึกษาคุณลักษณะของแบบทดสอบปลายเปิดวิชาคณิตศาสตร์ โดยใช้โมเดลการสรุปอ้างอิงและโมเดลหลายองค์ประกอบของราล์ซ ภายใต้เงื่อนไขจำนวนผู้ตรวจต่างกัน และรูปแบบการตรวจให้คะแนนแตกต่างกัน 3 ลักษณะ เครื่องมือที่ใช้คือ แบบทดสอบปลายเปิดวิชาคณิตศาสตร์ ระดับมัธยมศึกษาตอนต้น ตามหลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 จำนวน 12 ข้อ กลุ่มตัวอย่างคือ นักเรียนชั้นมัธยมศึกษาที่ 4 จังหวัดน่าน จำนวน 180 คน ซึ่งได้มาจากการสุ่มแบบสองชั้นตอน ผลการวิจัยพบว่า เมื่อวิเคราะห์ด้วยโมเดลการสรุปอ้างอิง เมื่อใช้รูปแบบการตรวจให้คะแนนเดียวกันในทุกเงื่อนไขจำนวนผู้ตรวจ ความแปรปรวนขององค์ประกอบเดียวกัน มีค่าใกล้เคียงกัน คะแนนในทุกเงื่อนไขที่ต่างกันมีความเที่ยงตรงตามสภาพสูงและแตกต่างกัน อย่างไม่มีนัยสำคัญทางสถิติ และเมื่อวิเคราะห์ด้วยโมเดลหลายองค์ประกอบของราล์ซ ในทุกเงื่อนไขจำนวนผู้ตรวจและรูปแบบการตรวจให้คะแนน ความแปรปรวนของผู้สอบมีค่าสูงสุด รองลงมาคือ ความแปรปรวนของข้อสอบ และความแปรปรวนของผู้ตรวจมีค่าต่ำสุด

2. คุณลักษณะของผู้ตรวจให้คะแนน

มัลควีน เบเกอร์ และ ดีสม์คเคส (Mulqueen; Baker; & Dismukes. 2002) ได้ศึกษาเรื่องการอบรมการตรวจให้คะแนนผู้สอนการบิน ประโยชน์ของโมเดลทฤษฎีการตอบข้อสอบหลายองค์ประกอบ โดยใช้เทคนิคการวิเคราะห์แบบราล์ซ หลายองค์ประกอบเพื่อวิเคราะห์ผลของโปรแกรมการอบรม IRR-Training เพื่อพัฒนาการตรวจให้คะแนนของผู้สอนการบิน กลุ่มตัวอย่างคือ ผู้สอนการบิน 33 คน ในสายการบินพาณิชย์ของอเมริกา โดยตรวจให้คะแนนประสิทธิภาพลูกเรือจากเทปวิดีโอการทำงานของลูกเรือที่มีระดับการปฏิบัติงานแตกต่างกัน 3 เทป ผลการวิจัยพบว่า ผู้สอนการบินมีระดับความเข้มงวดแตกต่างกันอย่างมีนัยสำคัญทางสถิติ วิดีโอเทปของลูกเรือสามารถแยกความแตกต่างของความสามารถได้อย่างดีเยี่ยม ในการให้เกรดพบว่าบางทักษะยากกว่าอย่างมีนัยสำคัญทางสถิติ การวิเคราะห์ปฏิสัมพันธ์พบว่าคะแนนผลรวมของความสามารถของผู้สอนและลูกเรือได้ถูกคาดหวังไว้ถูกต้องแล้ว

นากามูระ (Nakamura. 2002) ได้ศึกษาการใช้งานราล์ซแบบหลายองค์ประกอบในการวิเคราะห์การสอบการเขียน มีจุดประสงค์เพื่อปรับปรุงประสิทธิภาพในการใช้ผู้ตรวจหลายคนให้คะแนนในการวัดผลความสามารถในการเขียนของนักเรียน และลดภาระงานในการตรวจให้คะแนนด้วยการใช้แนวทางเชิงสถิติ กลุ่มตัวอย่างคือนักเรียนในมหาวิทยาลัยประเทศญี่ปุ่นจำนวน 32 คน ผู้ตรวจให้คะแนน ซึ่งครูผู้สอนจำนวน 4 คน เครื่องมือคือ ข้อสอบการเขียนเรียงความในหัวข้ออิสระ ตรวจให้คะแนนโดยใช้เกณฑ์ 7 ประการ ในมาตราประมาณค่า 4 ระดับ ผลการวิจัยพบว่า ระดับความเข้มงวด/ใจดีของผู้ให้คะแนน 4 คน มีความแตกต่างกัน แต่มีดัชนีสอดคล้องระหว่างผู้ให้คะแนนสูง และภาระงานของผู้ตรวจให้คะแนนสามารถลดลงได้ ด้วยการจับคู่ผู้ตรวจให้คะแนนอย่างเป็นระบบ

แองเจิลฮาร์ท และ เมฟอร์ด (Engelhard; & Myford. 2003) ได้ศึกษาเรื่องการเฝ้าตรวจสอบความสามารถในการให้คำปรึกษาในโปรแกรม Advance Placement English Literature ประวัติศาสตร์ ภาษาอังกฤษและการประพันธ์ ด้วยโมเดลของราล์ชหลายองค์ประกอบ กลุ่มตัวอย่างประกอบด้วย นักเรียน 174,857 คน และผู้ให้คำปรึกษา 612 คน ผลการวิจัยพบว่า มีความเข้มงวด/ใจดีของผู้ให้คำปรึกษา มีความแตกต่างกันอย่างมีนัยสำคัญทางสถิติ ไม่พบความลำเอียงด้านกลุ่มผู้เรียน และไม่พบความลำเอียงด้านเพศ

วูล์ฟ (Wolfe. 2004) ได้ศึกษาเรื่องการระบุผลกระทบของผู้ให้คะแนนโดยใช้โมเดลคุณลักษณะแฝง เพื่อระบุผลกระทบจากผู้ให้คะแนนจากการสอบ Advanced Placement Examination ใน 3 ผลกระทบ คือ ความเที่ยงตรง/ความไม่เที่ยงตรง ความเข้มงวด/ใจดี ความเป็นกลาง/ความสุดโต่ง เครื่องมือที่ใช้ คือ เรียงความ 28 เรื่องจากการสอบในส่วนของ English Literature and Composition ตรวจโดยผู้ตรวจ 101 คน ที่ได้รับการอบรมการตรวจให้คะแนนมาเป็นอย่างดี ผลการวิจัยพบว่า ความเที่ยงตรง/ความไม่เที่ยงตรง ผู้ให้คะแนนที่มีความเที่ยงตรงให้คะแนนเรียงความ 86% พบผู้ให้คะแนน 2 คน มีความไม่เที่ยงตรงสูงอย่างเห็นได้ชัด ความเข้มงวด/ใจดี มีนักเรียนที่ผ่านการทดสอบ 6% ที่ถูกให้คะแนนโดยผู้ให้คะแนนที่ใจดี และ 5% ที่ถูกให้คะแนนโดยผู้ให้คะแนนที่เข้มงวด ความเป็นกลาง/ความสุดโต่ง ผู้ให้คะแนนแบบสุดโต่งให้คะแนนกระจายมากกว่าผู้ให้คะแนนปกติ 79% ส่วนผู้ให้คะแนนเป็นกลางให้คะแนนใกล้เคียงกับค่าเฉลี่ย

ปาร์ค (Park. 2004) ได้วิจัยข้อสอบ English as a Second Language (ESL) Placement Test โดยการใช้การวิเคราะห์ของราล์ชแบบหลายองค์ประกอบ มีจุดประสงค์เพื่อวิเคราะห์ความสามารถในการแยกผู้สอบที่มีระดับความสามารถต่างกันออกจากกันของเกณฑ์การตรวจให้คะแนน ตรวจสอบความผิดปกติของความสามารถของผู้สอบ ตรวจสอบระดับของความเข้มงวดของผู้ตรวจให้คะแนน และความเหมาะสมของสเกลที่ใช้ กลุ่มตัวอย่างประกอบด้วยนักเรียนที่เรียนภาษาอังกฤษในฐานะภาษาต่างประเทศ จำนวน 99 คน ที่ลงทะเบียนเรียนวิชา CEP : Community English Program มหาวิทยาลัยโคลัมเบีย และครูผู้ให้คะแนน 17 คน ที่ผ่านการอบรมการตรวจให้คะแนนมาแล้ว โดยเครื่องมือที่ใช้คือ เรียงความจากนักเรียน 99 ชิ้น เกณฑ์การให้คะแนน 4 ด้าน คือ ความสมบูรณ์โดยรวมของชิ้นงาน การจัดการโครงสร้างการเขียน การควบคุมขอบเขตของเนื้อหา และการใช้ภาษา เป็นมาตรฐานประมาณค่า 4 ระดับ เรียงความแต่ละชิ้นตรวจให้คะแนนโดยครู 2 คน ผลการวิจัยพบว่า มีผู้สอบที่ได้คะแนนความสามารถในการเขียนสูงมากกว่าที่ระดับของความเข้มงวดของผู้ให้คะแนนมีความแตกต่างกัน ขอบเขตของเกณฑ์การตรวจให้คะแนนมีความยากแตกต่างกัน และมีความเหมาะสมของสเกลที่ใช้

เอกเคส (Eckes. 2005) ได้ทดสอบผลกระทบของผู้ให้คะแนนของการประเมินประสิทธิภาพในการเขียนและการพูดของแบบทดสอบภาษาเยอรมันในฐานะภาษาต่างประเทศ (Test of German as a Foreign Language : TestDaF) โดยการใช้การวิเคราะห์ราล์ชแบบหลายองค์ประกอบ ซึ่งงานวิจัยนี้ได้ศึกษาผลกระทบของผู้ให้คะแนน โดยเน้นในเรื่องคะแนนผลกระทบหลัก การมีปฏิสัมพันธ์ระหว่างผู้ให้คะแนนและองค์ประกอบที่เกี่ยวข้อง คือ ผู้สอบและเกณฑ์การตรวจให้คะแนน และการค้นหาการทำหน้าที่แตกต่างกันด้านเพศของผู้ให้คะแนน ซึ่งเครื่องมือ TestDaF ประกอบด้วย 4 ส่วน คือ การอ่าน

ทำความเข้าใจ การฟังทำความเข้าใจ การแสดงความคิดและความรู้สึกด้วยการเขียน และการแสดงความคิดและความรู้สึกด้วยการพูด กลุ่มตัวอย่าง ผู้สอบ ในส่วนการเขียนมีผู้เข้าร่วม 1,359 คน (หญิง 747 คน และชาย 612 คน) ในส่วนการพูดผู้เข้าร่วม 1,348 คน (หญิง 741 คน และชาย 607 คน) ผู้ให้คะแนนตรวจให้คะแนนการเขียน 29 คน (หญิง 23 คน และชาย 6 คน) และให้คะแนนการพูด 31 คน (หญิง 26 คน และชาย 5 คน) ผลการวิจัยพบว่า 1) ความเข้มงวดของผู้ให้คะแนนมีความแตกต่างกันสูง 2) มีความคงที่อย่างยุติธรรมในการตรวจให้คะแนนในภาพรวม 3) มีความคงที่ต่ำในความสัมพันธ์ระหว่างผู้ให้คะแนนกับเกณฑ์การตรวจให้คะแนน ซึ่งมีค่าต่ำกว่าในความสัมพันธ์กับผู้สอบ 4) มีความลำเอียงเชิงกลุ่ม ซึ่งไม่ก่อให้เกิดความลำเอียงทางเพศ

อิรมณีรัตน์ และ โยโดโควสกี (Iramaneerat; & Yudkowsky. 2006) ได้ศึกษาความคลาดเคลื่อนของผู้ให้คะแนนในการประเมินทักษะทางการแพทย์ มีจุดประสงค์เพื่อประเมินทักษะทางการแพทย์ 6 ทักษะของนักเรียนชั้นปีที่ 4 ในโรงเรียนการแพทย์มิถุเวสเทิร์น และค้นหาความคลาดเคลื่อนที่มีแหล่งมาจากผู้ให้คะแนน 4 ชนิด คือ ความเข้มงวด/ใจดี ความไม่ลงรอยกัน การคล้อยตามกัน และระดับของความเข้มงวด กลุ่มตัวอย่างประกอบด้วย นักเรียนชั้นปีที่ 4 จำนวน 173 คน ผู้ให้คะแนน 17 คนที่มาจากผู้ที่แสดงเป็นคนไข้ ผลการวิจัยพบว่า ทักษะทางการแพทย์ของนักเรียนมีการกระจายเป็นปกติ ความเข้มงวด/ใจดีของผู้ให้คะแนนมีระดับความเข้มงวดแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 พบความไม่ลงรอยกันของผู้ให้คะแนนคนที่ 4 กับผู้ให้คะแนนคนที่ 3, 4, 7 และ 8 มีอิสระจากการคล้อยตามกัน และระดับของความเข้มงวดของผู้ให้คะแนนเป็นอิสระจากกัน

เมฟฟอร์ด และ วูฟ (Myford; & Wolfe. 2009) ได้ศึกษาเรื่องการเฝ้าตรวจสอบความสามารถของผู้ให้คะแนนเมื่อเวลาผ่านไป กรอบการทำงานสำหรับการตรวจสอบความแตกต่างในความแม่นยำและลำดับมาตรฐานค่าที่ใช้ มีจุดประสงค์เพื่อแสดงดัชนีทางสถิติต่างๆ เพื่อระบุการทำหน้าที่ต่างกันของผู้ให้คะแนนเมื่อเวลาผ่านไป กลุ่มตัวอย่างคือนักเรียน 51,233 คน ผู้ตรวจให้คะแนน 101 คน เครื่องมือคือ ข้อสอบข้อที่ 1 ของ AP ELC โดยแบ่งฟาเซของค์ประกอบของเวลาออกเป็น 8 จุดเวลา ในเวลา 4 วัน (เช้า - บ่าย) ผลการวิจัยพบว่า เมื่อเวลาผ่านไปผู้ให้คะแนนมีแนวโน้มใช้มาตรฐานค่าระดับกลางๆ เพิ่มขึ้น คะแนนของนักเรียนเพิ่มขึ้นเล็กน้อยเมื่อเวลาผ่านไป และความแม่นยำของผู้ให้คะแนนเพิ่มขึ้นเมื่อเวลาผ่านไป

3. ความเที่ยงตรง

กนกวรณ รัตนธน (2544) ได้ศึกษาเรื่องการเปรียบเทียบความตรงตามสภาพและความคงที่ในการประมาณค่าความสามารถของผู้สอบในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เมื่อระดับความสามารถของผู้สอบและอัตราการใช้ข้อสอบซ้ำแตกต่างกัน มีจุดประสงค์เพื่อเปรียบเทียบความตรงตามสภาพและความคงที่ในการประมาณค่าความสามารถ ของผู้สอบในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์ เมื่อระดับความสามารถของผู้สอบ และอัตราการใช้ข้อสอบซ้ำแตกต่างกัน กลุ่มตัวอย่างคือ นักเรียนชั้นมัธยมศึกษาปีที่ 3 ภาคเรียนที่ 2 ปีการศึกษา 2543 จังหวัดสุพรรณบุรี จำนวน 319 คน เครื่องมือที่ใช้ในการวิจัยคือ แบบสอบวัดความสามารถทาง

ภาษาอังกฤษ คลังข้อสอบวิชาภาษาอังกฤษ และโปรแกรมคอมพิวเตอร์สำหรับการทดสอบแบบปรับ
 เหมาะกับความสามารถของผู้สอบ (MicroCAT) ผลการวิจัยพบว่า 1) ระดับความสามารถของผู้สอบ
 และอัตราการใช้ข้อสอบซ้ำที่แตกต่างกัน ทำให้ความตรงตามสภาพในการประมาณค่าความสามารถ
 ของผู้ทดสอบ ในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์แตกต่างกัน
 โดยผู้สอบที่มีระดับความสามารถสูง และมีอัตราการใช้ข้อสอบซ้ำที่ถูกต้องขึ้น ส่งผลให้ความตรงตาม
 สภาพของการทดสอบลดลง 2) ระดับความสามารถของผู้สอบและอัตราการใช้ข้อสอบซ้ำที่แตกต่างกัน
 ทำให้ความคงที่ในการประมาณค่าความสามารถของผู้สอบ ในการทดสอบแบบปรับเหมาะกับความสามารถ
 ของผู้สอบด้วยคอมพิวเตอร์แตกต่างกัน โดยผู้สอบที่มีระดับความสามารถสูงและมีอัตราการใช้ข้อสอบ
 ซ้ำที่สูงขึ้น ส่งผลให้ความคงที่ในการประมาณค่าความสามารถของผู้สอบลดลง 3) ปฏิสัมพันธ์ระหว่าง
 ระดับความสามารถของผู้สอบและอัตราการใช้ข้อสอบซ้ำร่วมกัน ส่งผลต่อความคงที่ในการประมาณค่า
 ความสามารถของผู้สอบ โดยผู้สอบที่มีระดับความสามารถสูงและมีอัตราการใช้ข้อสอบซ้ำที่สูงขึ้น จะ
 ทำให้แนวโน้มของการประมาณค่าความสามารถของผู้สอบสูงขึ้นด้วย 4) ปฏิสัมพันธ์ระหว่างจำนวนครั้ง
 ที่ทดสอบและอัตราการใช้ข้อสอบซ้ำร่วมกัน ส่งผลต่อความคงที่ในการประมาณค่าความสามารถของ
 ผู้สอบ เมื่อจำนวนครั้งที่ทดสอบเพิ่มขึ้นและมีอัตราการใช้ข้อสอบซ้ำที่สูงขึ้น จะทำให้แนวโน้มของการประมาณ
 ค่าความสามารถของผู้สอบสูงขึ้นด้วย

เอมอร์ จังศิริพรภรณ์ (2545) ได้ศึกษาเรื่องการเปรียบเทียบคุณภาพของแบบสอบเลือกตอบ
 เมื่อตรวจด้วยวิธีการตรวจให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม การวิจัยครั้งนี้ มีวัตถุประสงค์
 เพื่อเปรียบเทียบคุณภาพของแบบสอบในด้านความตรงตามสภาพอำนาจจำแนก ความยาก ฟังก์ชัน
 สารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศเฉลี่ยโดยใช้ทฤษฎี
 การตอบสนองข้อสอบเมื่อตรวจสอบให้คะแนนแบบประเพณีนิยมกับวิธีให้คะแนนความรู้บางส่วน ซึ่ง
 ประกอบด้วยวิธีประยุกต์การตรวจให้คะแนนของคูมบ์ และวิธีประยุกต์การตรวจให้คะแนนของเดรสเซล
 และ สมิท กลุ่มตัวอย่างที่ใช้ในการวิจัย ได้แก่ นิสิต คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย ชั้นปีที่ 3
 ที่ลงทะเบียนเรียน วิชาการประเมินผลการเรียนการสอน ในภาคการศึกษาต้น ปีการศึกษา 2544 จำนวน
 297 คน เครื่องมือที่ใช้ในการวิจัยครั้งนี้ เป็นแบบสอบที่ใช้เป็นแบบวัดผลสัมฤทธิ์ทางการเรียน วิชาการ
 ประเมินผลการเรียนการสอนที่มีวิธีการตรวจให้คะแนน 3 วิธี ดังกล่าวจากแบบสอบชุดเดียวกัน เก็บข้อมูล
 โดยให้ทำข้อสอบวัดผลสัมฤทธิ์ทางการเรียน วิชาการประเมินผลการเรียนการสอนที่สร้างขึ้น 3 ครั้ง ด้วย
 แบบสอบชุดเดียวกัน แต่วิธีการตอบแตกต่างกัน โดยการเก็บข้อมูลแต่ละครั้งห่างกัน 2 สัปดาห์ วิเคราะห์
 ข้อมูลโดยใช้โปรแกรม BILOG สำหรับการตรวจให้คะแนนแบบประเพณีนิยม และใช้โปรแกรม PARSCALE
 สำหรับการตรวจให้คะแนนด้วยวิธีประยุกต์การตรวจให้คะแนนของคูมบ์และวิธีประยุกต์การตรวจให้คะแนน
 ของ เดรสเซล และ สมิท ผลการวิจัยสรุปได้ดังนี้ 1) การเปรียบเทียบความตรงตามสภาพ ระหว่าง
 วิธีการตรวจให้คะแนนที่แตกต่างกัน พบว่าทั้ง 3 วิธี มีค่าความตรงตามสภาพไม่มีความแตกต่างกัน
 อย่างมีนัยสำคัญทางสถิติ 2) การเปรียบเทียบอำนาจจำแนกและความยากระหว่างวิธีการตรวจให้คะแนน
 ที่แตกต่างกัน พบว่า วิธีประยุกต์การตรวจให้คะแนนของคูมบ์มีค่าอำนาจจำแนกเฉลี่ยสูงสุด และมีค่า
 ความยากเฉลี่ยต่ำที่สุด 3) การเปรียบเทียบฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ

และอัตราส่วนสารสนเทศระหว่างวิธีการตรวจให้คะแนนที่แตกต่างกัน พบว่า วิธีประยุกต์การตรวจให้คะแนนของ คูมภ์ มีค่าฟังก์ชันสารสนเทศของข้อสอบ ฟังก์ชันสารสนเทศของแบบสอบ และอัตราส่วนสารสนเทศเฉลี่ยสูงสุด รองลงมาคือ วิธีประยุกต์การตรวจให้คะแนนของเดรสเซล และ สมิท ส่วนวิธีประเพณีนิยมมีค่าต่ำที่สุด

ปิยะรัตน์ ประมวลทรัพย์ (2546) ได้ศึกษาการพัฒนาแบบสอบอัตนัยประยุกต์วัดสมรรถนะทางวิชาชีพพยาบาล มีจุดประสงค์เพื่อพัฒนาแบบสอบอัตนัยประยุกต์วัดสมรรถนะทางวิชาชีพพยาบาล และศึกษาคุณภาพแบบสอบในด้านความเที่ยงตรงตามเนื้อหา ความเที่ยงตรงเชิงโครงสร้าง ความเที่ยงตรงตามสภาพ ความเชื่อมั่นและความเป็นปรนัยในการตรวจให้คะแนน กลุ่มตัวอย่างคือ นักศึกษาพยาบาล ชั้นปีที่ 4 ปีการศึกษา 2545 วิทยาลัยพยาบาลบรมราชชนนี เครือข่ายภาคกลาง จำนวน 137 คน ได้มาโดยการสุ่มแบบหลายขั้นตอน แบบสอบอัตนัยประยุกต์ที่ผู้วิจัยพัฒนาขึ้น ประกอบด้วย 7 กรณีศึกษา เป็นสถานการณ์จริงเกี่ยวกับสมรรถนะทางวิชาชีพพยาบาล มีจำนวน 30 ข้อกระทง คะแนนเต็ม 100 คะแนน ผลการวิจัยพบว่า แบบสอบอัตนัยประยุกต์วัดสมรรถนะทางวิชาชีพพยาบาลฉบับนี้ มีความเที่ยงตรงตามเนื้อหา สามารถวัดสมรรถนะทางวิชาชีพพยาบาล ด้วยเกณฑ์ความสอดคล้องของผู้ทรงคุณวุฒิตั้งแต่ร้อยละ 80 ขึ้นไป มีความเที่ยงตรงตามสภาพ โดยค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนจากการทำแบบสอบอัตนัยประยุกต์กับคะแนนจากการประเมินผลการปฏิบัติงานพยาบาลวิชาชีพ มีค่า .750 มีนัยสำคัญทางสถิติที่ระดับ .05 และการทดสอบโดยเทคนิคกลุ่มที่รู้จัก พบว่า คะแนนเฉลี่ยของพยาบาลวิชาชีพที่มีสมรรถนะทางวิชาชีพพยาบาลสูง และคะแนนเฉลี่ยของพยาบาลวิชาชีพที่มีสมรรถนะทางวิชาชีพพยาบาลต่ำ มีความแตกต่างอย่างมีนัยสำคัญที่ระดับ .05 ส่วนความเชื่อมั่นของแบบสอบมีค่าสัมประสิทธิ์แอลฟาของครอนบาค .746 และมีความเป็นปรนัยในการตรวจให้คะแนน มีค่าสัมประสิทธิ์สหสัมพันธ์ของผู้ตรวจให้คะแนน 3 ท่าน คือ $r_{12} = .977$, $r_{13} = .977$, $r_{23} = .987$ และทุกค่ามีนัยสำคัญทางสถิติที่ระดับ .01

ธิดารัตน์ ธนะคำดี (2552) ได้ศึกษาพัฒนาแบบวัดความมีวินัยในตนเองสำหรับนักเรียน ช่วงชั้นที่ 2 โรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษา กรุงเทพมหานคร เขต 1 โดยมีจุดมุ่งหมายเพื่อแสดงหลักฐานความเที่ยงตรงและความเชื่อมั่นของแบบวัดความมีวินัยในตนเอง และเปรียบเทียบความมีวินัยในตนเองของนักเรียน โดยจำแนกตามเพศและระดับชั้นของนักเรียน กลุ่มตัวอย่างที่ใช้ในการวิจัยในครั้งนี้เป็นนักเรียนช่วงชั้นที่ 2 ภาคเรียนที่ 2 ปีการศึกษา 2551 จำนวน 506 คนที่ได้มาจากการสุ่มแบบหลายขั้นตอน (Multi-stage Random Sampling) เครื่องมือที่ใช้ในการวิจัยเป็นแบบวัดความมีวินัยในตนเอง ที่มีรูปแบบการวัดต่างกัน 2 ฉบับ คือ ฉบับข้อความและฉบับสถานการณ์ แต่ละฉบับประกอบด้วย 6 ด้าน คือ ด้านความรับผิดชอบ ด้านความเชื่อมั่นในตนเอง ด้านความอดทน ด้านความซื่อสัตย์ ด้านความเป็นผู้นำและด้านการปฏิบัติตามกฎระเบียบของสังคม ทำการวิเคราะห์ความเที่ยงตรงเชิงโครงสร้าง ด้วยวิธีการวิเคราะห์องค์ประกอบเชิงยืนยัน และการวิเคราะห์ด้วยวิธีหลายลักษณะหลายวิธี (Multitrait-Multimethod) วิเคราะห์ความเชื่อมั่นของข้อคำถามรายข้อ (Item-Reliability Index) ความเชื่อมั่นรายฉบับ (Cronbach Coefficient) และการวิเคราะห์ความแปรปรวนสองทาง (Two-Way MANOVA) ผลการวิจัยพบว่า 1) ความเที่ยงตรงเชิงโครงสร้างของแบบวัดความมีวินัยในตนเองจากการวิเคราะห์องค์ประกอบ

พบว่า แบบวัดทั้งฉบับข้อความและฉบับสถานการณ์ มี 6 องค์ประกอบ คือด้านความรับผิดชอบ ด้านความเชื่อมั่น ในตนเอง ด้านความอดทน ด้านความซื่อสัตย์ ด้านความเป็น ผู้นำ และด้านการปฏิบัติ ตามกฎระเบียบของสังคม เป็นไปตามโครงสร้างที่สังเคราะห์ไว้ แบบวัดฉบับข้อความมีค่าน้ำหนักองค์ประกอบ ตั้งแต่ 0.319 – 0.959 และแบบวัดสถานการณ์ มีค่าน้ำหนักองค์ประกอบตั้งแต่ 0.318 – 0.924 เมื่อพิจารณาความเที่ยงตรงเชิงโครงสร้างด้วยวิธีวิเคราะห์หลายลักษณะหลักวิธีของแบบวัดความมีวินัย ในตนเองทั้ง 2 ฉบับ พบว่า มีค่าสัมประสิทธิ์สหสัมพันธ์ของความเที่ยงตรงเชิงเหมือน ตั้งแต่ 0.782 – 0.931 และมีค่าสัมประสิทธิ์สหสัมพันธ์ของความเที่ยงตรงเชิงจำแนกตั้งแต่ 0.055 – 0.455 ซึ่งต่ำกว่าค่าความเที่ยงตรง เชิงเหมือน 2) ความเชื่อมั่นของแบบวัดความมีวินัยในตนเองฉบับข้อความแต่ละด้านมีค่าความเชื่อมั่น ตั้งแต่ 0.682 – 0.798 ความเชื่อมั่นรวมทั้งฉบับมีค่า 0.879 แบบวัดความมีวินัยในตนเองฉบับสถานการณ์ แต่ละด้านมีค่าความเชื่อมั่นตั้งแต่ 0.609 – 0.738 ความเชื่อมั่นรวมทั้งฉบับมีค่า 0.858 และ 3) นักเรียน ที่มีเพศต่างกัน และระดับชั้นต่างกัน มีวินัยในตนเองแตกต่างกัน โดยนักเรียนเพศหญิงมีวินัยในตนเองสูงกว่า นักเรียนเพศชาย นักเรียนชั้นประถมศึกษาปีที่ 5 มีวินัยในตนเองสูงกว่านักเรียนชั้นประถมศึกษาปีที่ 6 และนักเรียนชั้นประถมศึกษาปีที่ 4 ไม่พบว่า มีผลปฏิสัมพันธ์ที่เกิดจากความแตกต่างระหว่างระดับชั้น และเพศส่งผลร่วมกันต่อความมีวินัยในตนเองของนักเรียน

ตาราง 5 การสังเคราะห์งานวิจัยที่เกี่ยวข้อง

ชื่อผู้วิจัย (ปีที่วิจัย)	ตัวแปรอิสระ									ประเด็นที่นำมาพัฒนาต่อ	
	วิธีการตรวจ			คุณลักษณะ			จำนวนผู้ตรวจ				
	รวมองค์ประกอบ	แยกองค์ประกอบ	เชิงวัด/ใจดี	ลำเอียงด้านเพศ	ทำหน้าที่ต่างกัน	เมื่อเวลาผ่านไป	2 คน	3 คน	อื่นๆ		
พรรณี เจียมสุขบุตร (2543)	✓	✓					✓	✓		จำนวนผู้ให้คะแนนและวิธีการตรวจให้ คะแนนส่งผลต่อค่าความเชื่อมั่นของ แบบทดสอบ	
อุษณีย์ บัวศิริพันธ์ (2543)	✓								6	การเปรียบเทียบคะแนนที่มีวิธีการ ตรวจให้คะแนนแตกต่างกัน มีจำนวน ผู้ตรวจให้คะแนนต่างกัน และผู้ตรวจ มีประสบการณ์แตกต่างกัน	
กนกวรรณ รัตนชน (2544)										✓	การใช้ความเที่ยงตรงตามสภาพเป็น ตัวแปรตามในการวิเคราะห์แบบ TRT
นิศารัตน์ คงสวัสดิ์ (2544)	✓	✓					✓	✓	4	จำนวนผู้ให้คะแนนและวิธีการตรวจให้ คะแนนส่งผลต่อค่าความเชื่อมั่นของ ผู้ให้คะแนน	

ตาราง 5 (ต่อ)

ชื่อผู้วิจัย (ปีที่วิจัย)	ตัวแปรอิสระ							ความเที่ยงตรง	ประเด็นที่นำมาพัฒนาต่อ	
	วิธีการตรวจ		คุณลักษณะ			จำนวนผู้ตรวจ				
	รวมองค์ประกอบ	แยกองค์ประกอบ	เข้มงวด/ใจดี	ลำเอียงด้านเพศ	ทำหน้าที่ต่างกัน	เสียเวลาผ่านไป	2 คน			3 คน
ปวีณา ปีอาทิตย์ (2545)	✓	✓							8	จำนวนผู้ให้คะแนนและวิธีการตรวจให้คะแนนที่แตกต่างกันส่งผลให้ค่าสัมประสิทธิ์การสรุปอ้างอิงแตกต่างกัน
เอมอร จังศิริพรปกรณ์ (2545)								✓		การใช้ความเที่ยงตรงตามสภาพเป็นตัวแปรตามในการวิเคราะห์แบบ TRT ในเงื่อนไขวิธีการตรวจให้คะแนนที่แตกต่างกัน
ปิยะรัตน์ ประมวล ทรัพย์ (2546)								✓	3	การใช้ความเที่ยงตรงตามสภาพเป็นตัวแปรตามในการศึกษาคุณภาพแบบทดสอบ
อัญญารัตน์ เจริญพถฉิมินถ (2546)	✓									วิธีการดำเนินการสร้างตัวบ่งชี้และเกณฑ์การตรวจให้คะแนนแบบแยกองค์ประกอบ
ธิดารัตน์ ธนะคำดี (2552)								✓		การใช้การวิเคราะห์ความเที่ยงตรงเชิงโครงสร้างด้วยวิธีลักษณะหลากหลายวิธีหลาย(Multitrait-Multimethod)
น้ำผึ้ง อินทะเนตร (2554)		✓					✓	✓	4	การแปลความหมายจากการใช้ MFRM มาใช้ในการหาคุณภาพเครื่องมือที่ใช้ในการวิจัย
มลลิวีน เบเคอร์ และ ดีสมัคเคส (2002)			✓						33	การใช้ MFRM มาใช้ในการหาประสิทธิภาพของเครื่องมือ โดยใช้คะแนนที่ได้จากการอบรมผู้ให้คะแนนและหาระดับความเข้มงวดของผู้ให้คะแนน
นาคามูระ (2002)	✓		✓						4	การใช้ MFRM มาใช้ในการหาระดับความเข้มงวด/ใจดีของผู้ให้คะแนนเรียงความ

ตาราง 5 (ต่อ)

ชื่อผู้วิจัย (ปีที่วิจัย)	ตัวแปรอิสระ			ความเที่ยงตรง	ประเด็นที่นำมาพัฒนาต่อ					
	วิธีการตรวจ		จำนวนผู้ตรวจ							
	รวมองค์ประกอบ	แยกองค์ประกอบ								
			เชิงบวก/ใจดี	ลำเอียงด้านเพศ	ทำหน้าที่ต่างกัน	เมื่อเวลาผ่านไป	2 คน	3 คน	อื่นๆ	
ปาร์ค (2004)	✓	✓							17	การใช้ MFRM มาใช้ในการหาระดับความเข้มงวด/ใจดีของผู้ให้คะแนน และหาระดับความเหมาะสมของสเกลการตรวจให้คะแนน
เอกเคส (2005)			✓	✓					29	การใช้ MFRM มาใช้ในการหาระดับความเข้มงวด/ใจดี และความลำเอียงด้านเพศของผู้ให้คะแนน
อิรมณีรัตน์ และ โยโดควาสกี (2006)	✓	✓							17	การใช้ MFRM มาใช้ในการหาระดับความเข้มงวด/ใจดีของผู้ให้คะแนน
เมฟอร์ด และ วูฟ (2009)	✓				✓				101	การใช้ MFRM มาใช้ในการหาการทำหน้าที่ต่างกันของผู้ให้คะแนนเมื่อเวลาผ่านไป

จากตาราง 5 จะเห็นได้ว่า วิธีการตรวจให้คะแนนที่แตกต่างกันคือการตรวจให้คะแนนแบบรวมองค์ประกอบและแบบแยกองค์ประกอบจะส่งผลต่อการตรวจให้คะแนนการเขียน ทั้งในด้านความเชื่อมั่นจากผู้ตรวจให้คะแนนและค่าสัมประสิทธิ์ต่างๆ ดังนั้นเมื่อผู้วิจัยใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน คะแนนที่ได้จากผู้ตรวจให้คะแนนคนเดียวกันคะแนนที่ได้อาจจะแตกต่างกัน ในส่วนของคุณลักษณะของผู้ตรวจให้คะแนนในปัจจุบันมีการพัฒนาการค้นหาคุณลักษณะที่ลำเอียงเพื่อที่จะพัฒนาผู้ตรวจให้คะแนนในหลายวิธีโดยเฉพาะอย่างยิ่งการหาคุณลักษณะที่ลำเอียงด้วยการใช้ทฤษฎีการตอบสนองข้อสอบ โดยการแบ่งองค์ประกอบในการคำนวณแล้วนำมาเปรียบเทียบว่า ระดับความเข้มงวด/ใจดีของผู้ตรวจให้คะแนนมีความลำเอียงหรือไม่ โดยผู้วิจัยสนใจความเข้มงวด/ใจดี ความลำเอียงด้านเพศ และการทำหน้าที่ต่างกันเมื่อเวลาผ่านไปเป็นพิเศษ ซึ่งงานวิจัยในต่างประเทศและในประเทศยังมีจำนวนน้อย ดังนั้นผู้วิจัยจึงสนใจว่าเมื่อนำมาใช้กับผู้ตรวจให้คะแนนที่เป็นกลุ่มตัวอย่างผลจะเป็นอย่างไร และจำนวนผู้ให้คะแนนที่แตกต่างกันมีผลที่แตกต่างกัน เมื่อมีจำนวนผู้ตรวจให้คะแนนมากขึ้น ความเชื่อมั่นมักจะ

เพิ่มขึ้นตาม แต่ในสถานการณ์จริงการตรวจให้คะแนนเรียงความเป็นการตรวจให้คะแนนที่ค่อนข้างใช้เวลานาน ดังนั้นจึงเป็นไปได้หน่อยที่จะมีผู้ตรวจจำนวนมาก ดังนั้นผู้วิจัยจึงมีความสนใจที่จะศึกษาความเที่ยงตรงเมื่อผู้ให้คะแนนมีวิธีการตรวจให้คะแนน 3 วิธีแตกต่างกัน คุณลักษณะลำเอียงแตกต่างกันในสัดส่วนจำนวนผู้ให้คะแนน 2 และ 3 คน



บทที่ 3

วิธีดำเนินการวิจัย

1. การกำหนดประชากรและกลุ่มตัวอย่าง
2. การสร้างเครื่องมือที่ใช้ในการวิจัย
3. การเก็บรวบรวมข้อมูล
4. การจัดกระทำและวิเคราะห์ข้อมูล

การกำหนดประชากรและกลุ่มตัวอย่าง

ประชากร ประชากรที่ใช้ในการวิจัยมี 2 กลุ่ม คือ

1. ความสามารถในการเขียนของนักเรียน ชั้นประถมศึกษาปีที่ 3 ภาคการเรียนที่ 2 ปีการศึกษา 2553 ของโรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน เพื่อใช้เป็นผู้ทดสอบในการสร้างและพัฒนาคุณภาพของเครื่องมือวัดความสามารถในการเขียนของนักเรียนชั้น

ประถมศึกษาปีที่ 3 จำนวน 16,968 คน

2. ผู้ตรวจให้คะแนนจากนักศึกษาคณะศึกษาศาสตร์ ที่เรียนวิชาเอกภาษาไทย ชั้นปีที่ 4 ภาคการเรียนที่ 1 ปีการศึกษา 2554 ในสถาบันอุดมศึกษา สังกัดสำนักงานคณะกรรมการอุดมศึกษา เพื่อใช้เป็นผู้ตรวจให้คะแนนความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 จำนวน 402 คน

กลุ่มตัวอย่าง กลุ่มตัวอย่างที่ใช้ในการวิจัยมี 2 กลุ่ม คือ

1. ความสามารถในการเขียนของนักเรียน ชั้นประถมศึกษาปีที่ 3 ภาคการเรียนที่ 2 ปีการศึกษา 2553 ของโรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน สำนักงานเขตพื้นที่การศึกษาประถมศึกษาสุรินทร์ เขต 1 ที่ได้มาจากการสุ่มแบบหลายขั้นตอน จากนักเรียนที่เข้าสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 ในวันที่ 23 กุมภาพันธ์ 2554 จำนวน 50 คน โดยมีขั้นตอนการสุ่มดังนี้

ขั้นที่ 1 จำแนกนักเรียนจากคะแนนที่ได้จากครูที่เป็นกรรมการในการสอบออกเป็น 4 กลุ่ม คือ ดี (16 – 20 คะแนน) พอใช้ (8 – 15 คะแนน) ปรับปรุง (4 – 7 คะแนน) และ เขียนไม่ได้ (0 – 3 คะแนน)

ขั้นที่ 2 จำแนกนักเรียนในแต่ละกลุ่มจากขั้นที่ 1 ออกเป็นนักเรียนที่เป็นเพศชาย และเพศหญิง

ขั้นที่ 3 สุ่มนักเรียนในแต่ละกลุ่มย่อยโดยให้เพศของนักเรียนในแต่ละกลุ่มย่อย มีจำนวนใกล้เคียงกันจำนวน 50 คน ได้นักเรียนชาย จำนวน 26 คน คิดเป็นร้อยละ 52 และนักเรียนหญิง จำนวน 24 คน คิดเป็นร้อยละ 48

2. ผู้ตรวจให้คะแนนจากนักศึกษาครุศึกษาศาสตร์ ที่เรียนวิชาเอกภาษาไทย ชั้นปีที่ 4 ภาคการเรียนที่ 1 ปีการศึกษา 2555 ในมหาวิทยาลัยราชภัฏสุรินทร์ บุรีรัมย์ และ ร้อยเอ็ด ที่ได้มาจากการสุ่มอย่างง่าย จำนวน 130 คน จำแนกเป็นนักศึกษาชาย จำนวน 34 คน คิดเป็นร้อยละ 26.15 และ นักศึกษาหญิง จำนวน 96 คน คิดเป็นร้อยละ 73.85

ตาราง 6 แสดงจำนวนประชากรและกลุ่มตัวอย่าง

มหาวิทยาลัยราชภัฏกลุ่มภาคตะวันออกเฉียงเหนือ	ประชากร	กลุ่มตัวอย่าง
มหาวิทยาลัยราชภัฏเลย	14	-
มหาวิทยาลัยราชภัฏอุดรธานี	71	-
มหาวิทยาลัยราชภัฏมหาสารคาม	95	-
มหาวิทยาลัยราชภัฏร้อยเอ็ด	40	22
มหาวิทยาลัยราชภัฏบุรีรัมย์	104	73
มหาวิทยาลัยราชภัฏสุรินทร์	78	35
รวม	402	130

การสร้างเครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยมีอยู่ 2 ส่วน 1) วิธีการตรวจให้คะแนนการเขียนเรียงความในส่วนของ การเขียน จากข้อสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 และ 2) ข้อสอบการเขียนจดหมายและวิธีการตรวจให้คะแนนการเขียนจดหมายที่ผู้วิจัยสร้าง

1. การสร้างวิธีการตรวจให้คะแนนการเขียนเรียงความ จำแนกเป็น 3 รูปแบบ คือ วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ แบบรวมองค์ประกอบ และแบบผสมผสาน ซึ่งการสร้างวิธีการตรวจให้คะแนนการเขียนเรียงความมีขั้นตอน ดังนี้

1.1 ศึกษากระดาษคำตอบในส่วนของ การเขียน จากข้อสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 และวิธีการตรวจให้คะแนนจากสำนักทดสอบทางการศึกษา

1.2 ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการสร้างวิธีการตรวจให้คะแนนการเขียนเรียงความ

1.3 สร้างวิธีการตรวจให้คะแนนเรียงความจำแนกตามรูปแบบการตรวจให้คะแนน

1.3.1 วิธีการใช้เกณฑ์การตรวจให้คะแนนเรียงความแบบแยกองค์ประกอบ

1.3.1.1 อ่านเรียงความให้จบ 1 รอบ

1.3.1.2 พิจารณาให้คะแนนในประเด็นที่ 1 – 4 ทีละประเด็นจนครบ

1.3.1.3 ถ้าไม่แน่ใจในประเด็นใดให้วางพักแยกไว้ก่อนแล้วจึงนำมาพิจารณาอีกครั้ง

ภายหลัง

ตาราง 7 ตัวอย่างคะแนนแบบแยกองค์ประกอบ

ประเด็น การตรวจ ระดับคุณภาพ	1. ประเด็นสำคัญของเรียงความ
5	1. ชื่อเรื่องแปลกใหม่ น่าสนใจ ใช้ถ้อยคำกะทัดรัด สอดคล้องกับประเด็นสำคัญของภาพ 2. เนื้อเรื่องเขียนได้สอดคล้องกับชื่อเรื่องและครอบคลุมประเด็นสำคัญของภาพอย่างชัดเจน
4	1. ชื่อเรื่องน่าสนใจ ใช้ถ้อยคำเข้าใจง่าย สอดคล้องกับประเด็นสำคัญของภาพ 2. เนื้อเรื่องเขียนได้สอดคล้องกับชื่อเรื่องและครอบคลุมประเด็นสำคัญของภาพเป็นส่วนใหญ่
3	1. ชื่อเรื่องใช้ถ้อยคำไม่กระชับรัดกุม แต่สอดคล้องกับประเด็นสำคัญของภาพ 2. เนื้อเรื่องเขียนได้สอดคล้องกับชื่อเรื่องและครอบคลุมประเด็นสำคัญของภาพบางส่วน
2	1. ชื่อเรื่องใช้ถ้อยคำไม่กระชับรัดกุม แต่ชื่อเรื่องบางส่วนสอดคล้องกับประเด็นสำคัญของภาพ 2. เนื้อเรื่องเขียนได้สอดคล้องกับชื่อเรื่องแต่ไม่ครอบคลุมประเด็นสำคัญของภาพ
1	1. ตั้งชื่อเรื่องไม่สอดคล้องกับประเด็นสำคัญของภาพ 2. เนื้อเรื่องเขียนไม่สอดคล้องกับชื่อเรื่องและไม่ครอบคลุมประเด็นสำคัญของภาพ

ตาราง 8 ตัวอย่างวิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ

ชื่อ-นามสกุล	คะแนน				รวม	หมายเหตุ
	1. ประเด็นสำคัญ ของเรียงความ	2. รูปแบบการ เขียนเรียงความ	3. การนำเสนอ	4. การใช้ภาษา		
1. ด.ญ.กมลชนก	2	1	2	4	10	
2. ด.ญ.ครองขวัญ	5	5	5	5	20	
3. ด.ช.ปฎิภาณ	2	3	2	2	9	
4. ด.ญ.ภาวิณี	1	1	2	4	8	
5. ด.ช.ฤทธิไกร	3	3	4	3	13	

1.3.2 วิธีการใช้เกณฑ์การตรวจให้คะแนนเรียงความแบบรวมองค์ประกอบ

1.3.2.1 พิจารณาภาพรวมและแบ่งกระดาษออกเป็น 5 กอง คือ กอง 5 คะแนน กอง 4 คะแนน กอง 3 คะแนน กอง 2 คะแนน กอง 1 คะแนน

1.3.2.2 เรียงความชิ้นใดถ้าผู้ตรวจไม่แน่ใจว่าจะอยู่กลุ่มใดให้ผู้ตรวจทำเครื่องหมายไว้และวางไว้กองที่น่าจะอยู่กับกองนั้นๆ หลังจากนั้นจึงนำมาพิจารณาอีกครั้ง

ตาราง 9 ตัวอย่างคะแนนแบบรวมองค์ประกอบ

คำอธิบาย

ระดับคุณภาพ 5 คะแนน

- ชื่อเรื่องแปลกใหม่ น่าสนใจ ใช้ถ้อยคำกระทัดรัด สอดคล้องกับประเด็นสำคัญของภาพ และเนื้อเรื่องเขียนได้สอดคล้องกับชื่อเรื่องและครอบคลุมประเด็นสำคัญอย่างชัดเจน
- เขียนได้ครบถ้วนทั้ง 3 ส่วน ได้แก่ คำนำ เนื้อเรื่อง สรุป โดยที่ความยาวของแต่ละส่วนเป็นสัดส่วนอย่างเหมาะสม และมีจำนวนบรรทัดทั้งหมดไม่ต่ำกว่าที่กำหนด
- เรียบเรียงเนื้อเรื่องเป็นลำดับต่อเนื่องสัมพันธ์กันชัดเจน เชื่อมโยงทุกตอนของเรื่องให้สัมพันธ์กันโดยตลอด สอดแทรกแนวคิดแสดงเหตุผลที่เป็นประโยชน์ต่อสังคม
- สะกดคำถูกต้องทุกคำ เลือกใช้ถ้อยคำได้ตรงกับความหมาย ใช้คำสุภาพ ไม่ใช่ภาษาพูดในภาษาเขียน ไม่มีการเขียนฉีกคำหรือแยกคำเลยตลอดทั้งเรื่อง ลายมือเป็นระเบียบสวยงาม อ่านง่าย สะอาด ไม่มีรอยลบขีดฆ่า

 ตาราง 9 (ต่อ)

 คำอธิบาย

ระดับคุณภาพ 4 คะแนน

1. ชื่อเรื่องน่าสนใจ ใช้ถ้อยคำเข้าใจง่าย สอดคล้องกับประเด็นสำคัญของภาพ และเนื้อเรื่องเขียน ได้สอดคล้องกับชื่อเรื่องและครอบคลุมประเด็นสำคัญเป็นส่วนมาก
2. เขียนได้ครบถ้วนทั้ง 3 ส่วน ได้แก่ คำนำ เนื้อเรื่อง สรุป โดยที่ความยาวของแต่ละส่วนไม่เป็นสัดส่วนเหมาะสม และมีจำนวนบรรทัดทั้งหมดไม่ต่ำกว่าที่กำหนด
3. เรียบเรียงเนื้อเรื่องเป็นลำดับต่อเนื่องสัมพันธ์กันค่อนข้างชัดเจน เชื่อมโยงทุกตอนของเรื่องให้สัมพันธ์กันโดยตลอด สอดแทรกแนวคิดและแสดงเหตุผลที่เป็นประโยชน์ต่อสังคม
4. สะกดคำผิด 1 – 3 คำ เลือกใช้ถ้อยคำได้ตรงความหมาย ใช้คำสุภาพ เขียนฉีกคำหรือแยกคำ 1 – 3 แห่ง ลายมืออ่านง่าย มีรอยลบขีดฆ่า 1 – 3 แห่ง

ระดับคุณภาพ 3 คะแนน

1. ชื่อเรื่องใช้ถ้อยคำไม่กระชับรัดกุม แต่สอดคล้องกับประเด็นสำคัญของภาพ และเนื้อเรื่องเขียน ได้สอดคล้องกับชื่อเรื่องและครอบคลุมประเด็นสำคัญบางส่วน
2. เขียนมีเพียง 2 ส่วน เช่น มีการเขียนเนื้อเรื่องและสรุปแต่ไม่มีการเขียนคำนำ หรือเขียนคำนำและเนื้อเรื่องแต่ไม่มีการเขียนสรุป โดยที่ความยาวของแต่ละส่วนไม่เป็นสัดส่วนเหมาะสม มีจำนวนบรรทัดทั้งหมดไม่ต่ำกว่าที่กำหนด
3. เรียบเรียงเนื้อเรื่องสับสน วกวน แต่ยังอยู่ในกรอบ เสนอความคิดในบางตอนไม่เชื่อมโยงสัมพันธ์กัน สอดแทรกแนวคิดและแสดงเหตุผลที่เป็นประโยชน์เล็กน้อย
4. สะกดคำผิด 4 – 5 คำ ใช้ถ้อยคำไม่เหมาะสมบางแห่ง เขียนฉีกคำหรือแยกคำ 4 – 5 แห่ง ลายมืออ่านง่าย มีรอยลบขีดฆ่า 4 – 5 แห่ง

ระดับคุณภาพ 2 คะแนน

1. ชื่อเรื่องใช้ถ้อยคำไม่กระชับรัดกุม แต่ชื่อเรื่องบางส่วนสอดคล้องกับประเด็นสำคัญของภาพ และเนื้อเรื่องเขียนได้ สอดคล้องกับชื่อเรื่องแต่ไม่ครอบคลุมประเด็นสำคัญ
2. เขียนได้เพียงส่วนเดียว คือ คำนำ เนื้อเรื่อง หรือสรุป และมีจำนวนบรรทัดทั้งหมดต่ำกว่าที่กำหนด
3. เรียบเรียงเนื้อเรื่องสับสน วกวน เสนอความคิดไม่เชื่อมโยงสัมพันธ์กัน สอดแทรกแนวคิดที่เป็นประโยชน์น้อย และขาดการแสดงผลสนับสนุนความคิดเห็น
4. เขียนสะกดคำผิด 6 – 9 คำ ใช้ถ้อยคำไม่เหมาะสมบางแห่ง เขียนฉีกคำหรือแยกคำ 6 – 9 แห่ง ลายมือพออ่านออก มีรอยลบขีดฆ่า 6 – 9 แห่ง

ระดับคุณภาพ 1 คะแนน

1. ตั้งชื่อเรื่องไม่สอดคล้องกับประเด็นสำคัญของภาพ และเนื้อเรื่องเขียนไม่สอดคล้องกับชื่อเรื่องและไม่ครอบคลุมประเด็นสำคัญ
 2. เขียนแล้วแยกส่วนไม่ได้ และมีจำนวนบรรทัดทั้งหมดต่ำกว่าที่กำหนด
 3. เรียบเรียงเนื้อเรื่องสับสน ทำให้ผู้อ่านเข้าใจผิด เสนอความคิดไม่เชื่อมโยงสัมพันธ์กัน ไม่ได้สอดแทรกแนวคิดที่เป็นประโยชน์ และขาดการแสดงผลสนับสนุนความคิดเห็น
 4. เขียนสะกดคำผิดตั้งแต่ 10 คำขึ้นไป ใช้ถ้อยคำไม่เหมาะสมเป็นส่วนใหญ่ เขียนฉีกคำหรือแยกคำตั้งแต่ 10 แห่งขึ้นไป ลายมืออ่านไม่ออก มีรอยลบขีดฆ่าตั้งแต่ 10 แห่งขึ้นไป
-

ตาราง 10 ตัวอย่างวิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ

ชื่อ-นามสกุล	ระดับคุณภาพ					หมายเหตุ
	1 คะแนน	2 คะแนน	3 คะแนน	4 คะแนน	5 คะแนน	
1. ด.ญ.กมลชนก			✓			
2. ด.ญ.ครองขวัญ					✓	
3. ด.ช.ปฎิภาณ		✓				
4. ด.ญ.ภาวิณี	✓					
5. ด.ช.ฤทธิไกร				✓		

1.3.3 วิธีการใช้เกณฑ์การตรวจให้คะแนนเรียงความแบบผสมผสาน

ขั้นตอนที่ 1 ตรวจให้คะแนนโดยใช้วิธีการให้คะแนนแบบรวมองค์ประกอบ

- พิจารณาภาพรวมและแบ่งกระดาษออกเป็น 5 กอง คือ กอง 5 คะแนน กอง 4 คะแนน กอง 3 คะแนน กอง 2 คะแนน กอง 1 คะแนน
- เรียงความชิ้นใดถ้าผู้ตรวจไม่แน่ใจว่าจะอยู่กลุ่มใดให้ผู้ตรวจทำเครื่องหมายไว้และวางไว้กองที่น่าจะอยู่กับกองนั้นๆ หลังจากนั้นจึงนำมาพิจารณาอีกครั้ง

ขั้นตอนที่ 2 คัดเลือกประเด็นที่มีปัญหาในการตรวจ เพื่อนำมาตรวจอีกครั้งโดยใช้วิธีการให้คะแนนแบบแยกองค์ประกอบ

- เลือกเฉพาะบางคุณลักษณะที่โดดเด่น เพื่อเป็นผลสะท้อน (Feedback) ให้แก่ผู้เรียน โดยเลือกประเด็นที่มีประโยชน์ต่อการพัฒนาการเขียนของนักเรียน
- พิจารณาให้คะแนนในประเด็นที่เลือกไว้ที่ละประเด็นจนครบ

ตาราง 11 ตัวอย่างวิธีการตรวจให้คะแนนแบบผสมผสาน

ชื่อ-นามสกุล	คะแนน					หมายเหตุ
	คะแนนที่ได้จาก Holistic	1. ประเด็นสำคัญของ เรียงความ	2. รูปแบบ การเขียนเรียงความ	3. การนำเสนอ	4. การใช้ภาษา	
1. ด.ญ.กมลชนก	3	2	/	2	/	
2. ด.ญ.ครองขวัญ	5	5	/	5	/	
3. ด.ช.ปฎิภาณ	2	2	/	2	/	
4. ด.ญ.ภาวิณี	1	1	/	2	/	
5. ด.ช.ฤทธิไกร	4	3	/	4	/	

3. นำวิธีการตรวจให้คะแนนที่ได้ไปหาความเที่ยงตรงเชิงประจักษ์ โดยผู้เชี่ยวชาญด้านการวัดผลการศึกษาและผู้เชี่ยวชาญด้านภาษาไทย จำนวน 5 ท่าน ได้แก่

- 3.1 รศ.ชูศรี วงศ์รัตน์
- 3.2 ผศ.ดร.สุนันท์ ศลโกสุม
- 3.3 ผศ.จินตนา พุทฺธเมตะ
- 3.4 ผศ.อุทัยวรรณ ปิ่นประชาสรร
- 3.5 อาจารย์จิริฐา ชิตประสงค์

ผลการคำนวณค่า IOC พบว่า ได้ค่าเฉลี่ยจากผลการตัดสินของผู้เชี่ยวชาญแต่ละข้ออยู่ระหว่าง 0.6 – 1.0 นั่นคือ วิธีการตรวจให้คะแนนที่ผู้วิจัยสร้างขึ้นสามารถวัดได้ตามที่นิยามศัพท์กำหนด

4. ปรับปรุงแก้ไขวิธีการตรวจให้คะแนนจากข้อเสนอแนะที่ได้มาจากผู้เชี่ยวชาญ

5. นำวิธีการตรวจให้คะแนนที่ได้ปรับปรุงแล้วมาจัดทำเป็นคู่มือการตรวจให้คะแนนการเขียนเรียงความ จากนั้นนำไปประเมินระดับความเหมาะสม โดยผู้เชี่ยวชาญด้านการวัดผลการศึกษาและผู้เชี่ยวชาญด้านภาษาไทย จำนวน 5 ท่าน ได้แก่

- 5.1 ผศ.ฐาปนีย์ พงษ์ตวันกุล
- 5.2 อาจารย์ ดร.ศุภวรรณ สัจจพิบูล
- 5.3 อาจารย์ ดร.วิไลลักษณ์ ลังกา
- 5.4 อาจารย์ ดร.อรอุมา เจริญสุข

5.5 อาจารย์หทัยรัตน์ ทองมั่น

ผลการประเมินระดับความเหมาะสมพบว่า มีค่าเฉลี่ยแต่ละข้ออยู่ระหว่าง 4.00 – 5.00 แสดงว่า กลุ่มการตรวจให้คะแนนที่ผู้วิจัยสร้างขึ้นมีคุณภาพ สามารถนำไปใช้กับนักศึกษาที่เป็นกลุ่มตัวอย่างได้ ยกเว้นข้อที่ 10 ที่ว่า เวลาที่ใช้ในการจัดกิจกรรมมีความเหมาะสม เนื่องจากผู้เชี่ยวชาญไม่แน่ใจว่าเวลาที่ผู้วิจัยกำหนดไว้ 3 ชั่วโมง จะเพียงพอต่อกระบวนการทดลองหรือไม่ และไม่แน่ใจว่าเวลาที่ใช้ที่เหมาะสมจะใช้เวลาเท่าใด ผู้เชี่ยวชาญส่วนใหญ่จึงไม่ลงคะแนนในส่วนนี้ และให้แก้ไขเวลาตามกระบวนการที่ใช้จริงหลังการทดลองใช้กับประชากรที่ไม่ใช่กลุ่มตัวอย่าง รายละเอียด ดูภาคผนวก ก

6. นำกระดาษคำตอบและคู่มือการตรวจให้คะแนนที่ปรับปรุงแล้วมาคัดสำเนา จัดชุดตามจำนวนของนักศึกษา วิชาเอกภาษาไทย ชั้นปีที่ 4 มหาวิทยาลัยราชภัฏเลย จำนวน 24 ชุด แล้วนำไปทดสอบเพื่อหาคุณภาพและปรับปรุงวิธีการตรวจให้คะแนนในเบื้องต้น และตรวจสอบความเหมาะสมของเวลาในการจัดกระบวนการทดลองโดยใช้รูปแบบการหมุนเวียนสมดุล (Counterbalanced Designs) (องอาจ นัยวัฒน์. 2548) ในวันที่ 27 กุมภาพันธ์ 2555 ผลการทดลองพบว่า 1) แบบแยกองค์ประกอบ มีดัชนีความสอดคล้องของผู้ตรวจให้คะแนน เท่ากับ .825 และค่าความเชื่อมั่นเท่ากับ .926 2) แบบรวมองค์ประกอบ มีดัชนีความสอดคล้องของผู้ตรวจให้คะแนน เท่ากับ .833 และค่าความเชื่อมั่นเท่ากับ .921 และ 3) แบบผสมผสาน มีดัชนีความสอดคล้องของผู้ตรวจให้คะแนน เท่ากับ .830 และค่าความเชื่อมั่นเท่ากับ .983

7. ปรับปรุงแก้ไขวิธีการตรวจให้คะแนนและกระบวนการทดลอง

8. จัดพิมพ์เพื่อใช้เป็นเครื่องมือในการวิจัย

2. การสร้างข้อสอบการเขียนจดหมายและวิธีการตรวจให้คะแนนการเขียนจดหมาย มีขั้นตอน ดังนี้

2.1 กำหนดจุดมุ่งหมายในการวัด โดยผู้วิจัยกำหนดให้นักเรียนเขียนจดหมายโต้ตอบ ในหัวข้อการออกกำลังกาย

2.2 ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการสร้างวิธีการตรวจให้คะแนนการเขียนจดหมาย

2.3 สร้างวิธีการตรวจให้คะแนนจดหมายจำแนกตามวิธีการตรวจให้คะแนน

2.4 นำวิธีที่ได้ไปหาความเที่ยงตรงเชิงประจักษ์โดยผู้เชี่ยวชาญด้านการวัดผลการศึกษา และผู้เชี่ยวชาญด้านภาษาไทย จำนวน 5 ท่าน ได้แก่

2.4.1 รศ.ชูศรี วงศ์รัตน์

2.4.2 ผศ.ดร.สุนันท์ ศลโกสม

2.4.3 ผศ.จินตนา พุททเมตตะ

2.4.4 ผศ.อุทัยวรรณ ปิ่นประชาสรร

2.4.5 อาจารย์จิรฐา ชิตประสงค์

ผลการคำนวณค่า IOC พบว่า ได้ค่าเฉลี่ยจากผลการตัดสินของผู้เชี่ยวชาญแต่ละข้อ อยู่ระหว่าง 0.6 – 1.0 นั่นคือ วิธีการตรวจให้คะแนนที่ผู้วิจัยสร้างขึ้นสามารถวัดได้ตามที่นิยามศัพท์ กำหนด

2.5 ปรับปรุงแก้ไขวิธีการตรวจให้คะแนนจากข้อแนะนำที่ได้มาจากผู้เชี่ยวชาญ ดังนี้

2.5.1 เด็กนักเรียนชั้นประถมศึกษาปีที่ 3 อาจจะไม่ชอบการออกกำลังกาย ดังนั้น ถ้าต้องการกระตุ้นให้นักเรียนตอบสนองความสามารถในการเขียนของนักเรียนในชั้นนี้ ควรจะให้นักเรียน เขียนในหัวข้อเรื่อง สารทุกข์สุขดิบ การเลี้ยงสัตว์ การเล่นเกมส์ หรือกิจกรรมที่ชอบทำในวันหยุด

2.5.2 ควรเตรียมกระดาษข้อสอบการเขียนจดหมายให้อยู่ในรูปแบบการเขียนจดหมาย โดยกำหนดบรรทัดให้ส่วนต่างๆ ของจดหมายให้ชัดเจน คือ ที่อยู่ วันที่ คำขึ้นต้น คำลงท้าย รวมไปถึง การกำหนดริมกระดาษและแนวขึ้นต้น

2.5.3 ควรใช้สรรพนามให้เหมาะสมกับระดับบุคคลที่เขียนถึง (คุณ เธอ ชื่อเพื่อน ผม มัน ชื่อผู้เขียน)

2.5.4 ควรใช้ชื่อเพื่อนที่เป็นเพศกลางๆ หรือเป็นชื่อเล่น

2.5.5 คำสั่งควรแก้ไขเป็น “ จงบรรยาย.....”

2.5.6 ควรลดจำนวนบรรทัดลงเหลือ 10 บรรทัด

2.6 นำกระดาษคำตอบและคู่มือการตรวจให้คะแนนที่ปรับปรุงแล้วมาคัดสำเนา จัดชุดตาม จำนวนของนักศึกษา วิชาเอกภาษาไทย ชั้นปีที่ 4 มหาวิทยาลัยราชภัฏเลย จำนวน 24 ชุด แล้วนำไป ทดสอบเพื่อหาคุณภาพและปรับปรุงวิธีการตรวจให้คะแนนในเบื้องต้น และตรวจสอบความเหมาะสม ของเวลาในการจัดกระบวนการทดลองโดยใช้รูปแบบหมุนเวียนสมดุลงในวันที่ 27 กุมภาพันธ์ 2555 ผลการทดลอง พบว่า 1) แบบแยกองค์ประกอบ มีดัชนีความสอดคล้องของผู้ตรวจให้คะแนน เท่ากับ .818 และค่า ความเชื่อมั่นเท่ากับ .952 2) แบบรวมองค์ประกอบ มีดัชนีความสอดคล้องของผู้ตรวจให้คะแนน เท่ากับ .833 และค่าความเชื่อมั่นเท่ากับ .952 และ 3) แบบผสมผสาน มีดัชนีความสอดคล้องของ ผู้ตรวจให้คะแนน เท่ากับ .819 และค่าความเชื่อมั่นเท่ากับ .981

2.7 ปรับปรุงแก้ไขวิธีการตรวจให้คะแนนและกระบวนการทดลอง

2.8 จัดพิมพ์เพื่อใช้เป็นเครื่องมือในการวิจัย

การเก็บรวบรวมข้อมูล

ผู้วิจัยได้แบ่งการดำเนินการเก็บรวบรวมข้อมูลในการวิจัยออกเป็น 2 ระยะดังนี้

ระยะที่ 1 การใช้ข้อมูลจากกลุ่มตัวอย่างที่เป็นนักเรียน

1. ติดตามประสานงาน ขออนุญาตผู้บริหารและครูผู้สอนของนักเรียนชั้นประถมศึกษาปีที่ 4 ปีการศึกษา 2554 ซึ่งเป็นนักเรียนที่เลื่อนชั้นจากนักเรียนชั้นประถมศึกษาปีที่ 3 กลุ่มเดียวกันกับกลุ่มตัวอย่าง ที่ใช้ในการวิจัยที่เข้าสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ใน โครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 เพื่อกำหนด

วัน เวลา สถานที่ ในการดำเนินการสอบการเขียนจดหมาย จากข้อสอบการเขียนจดหมายที่ผู้วิจัยสร้างขึ้น

2. ดำเนินการทดลองกับกลุ่มตัวอย่างตามที่ได้วางแผนไว้ในวันที่ 29 สิงหาคม 2554 เวลา 08.30 น. – 10.38 น. ณ ห้องสมุดโรงเรียนเมืองสุรินทร์

3. ให้ครูประจำชั้นตรวจให้คะแนนการเขียนเรียงความ

ระยะที่ 2 การใช้ข้อมูลจากกลุ่มตัวอย่างที่เป็นนักศึกษา วิชาเอกภาษาไทยชั้นปีที่ 4

1. ติดต่อประสานงาน ขออนุญาตผู้บริหารและคณาจารย์ผู้สอนของนักศึกษา วิชาเอกภาษาไทยชั้นปีที่ 4 ที่เป็นกลุ่มตัวอย่าง เพื่อกำหนด วัน เวลา สถานที่ในการดำเนินการวิจัย

2. จัดประชุมชี้แจงเพื่อทำความเข้าใจกับนักศึกษาในการตรวจให้คะแนน

3. ดำเนินการวิจัยกับกลุ่มตัวอย่างตามรูปแบบที่ได้วางแผนไว้

การดำเนินการวิจัย

1. นำข้อสอบการเขียนจดหมายที่ผู้วิจัยสร้างขึ้นไปทดสอบกับนักเรียนชั้นประถมศึกษาปีที่ 4 ปีการศึกษา 2554 ซึ่งเป็นนักเรียนที่เลื่อนชั้นจากนักเรียนชั้นประถมศึกษาปีที่ 3 กลุ่มเดียวกันกับกลุ่มตัวอย่างที่ใช้ในการวิจัยที่เข้าสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 แล้วตรวจให้คะแนนโดยครูประจำชั้น จากนั้นนำคะแนนที่ได้มาใช้เป็นส่วนหนึ่งในการหาค่าความเที่ยงตรง ตามสภาพต่อไป

2. นำกระดาษคำตอบในส่วนของ การเขียนที่ได้คัดสำเนา มาจากข้อสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 และคู่มือการตรวจให้คะแนนที่ผ่านการหาคุณภาพแล้ว มาคัดสำเนา จัดชุดตามจำนวนนักศึกษามหาวิทยาลัยราชภัฏ วิชาเอกภาษาไทย ชั้นปีที่ 4 ที่เป็นกลุ่มตัวอย่าง จำนวน 185 ชุด เพื่อนำไปทดลองครั้งที่ 1 ระหว่างวันที่ 8 มิถุนายน 2555 ถึง 15 มิถุนายน 2555 ในครั้งนี้ มีนักศึกษามาเข้าร่วมการวิจัย จำนวน 140 คน ไม่มาเข้าร่วมการวิจัย จำนวน 45 คน

3. จัดอบรมการตรวจให้คะแนนแก่นักศึกษาที่เป็นกลุ่มตัวอย่าง ตามวันและเวลาที่ได้นัดหมาย เริ่มจากชี้แจงให้นักศึกษาได้ทราบถึงความมุ่งหมายและความสำคัญของการวิจัย จากนั้นฝึกสอนนักศึกษา ให้มีความรู้ความเข้าใจในการใช้วิธีการตรวจให้คะแนนเรียงความที่ถูกต้องทั้ง 3 รูปแบบ

4. ทำการทดลองครั้งที่ 1 โดยแบ่งนักศึกษาที่เป็นกลุ่มทดลองออกเป็น 3 กลุ่ม ด้วยวิธีการสุ่มอย่างง่าย แล้วดำเนินการตามรูปแบบการทดลองแบบหมุนเวียนสมดุลงโดยมีรายละเอียด ดังนี้

ตาราง 12 แบบแผนการทดลองแบบ Counterbalanced Designs ที่ใช้ในการวิจัย

กลุ่มที่	ระยะเวลาการให้สิ่งทดลอง (X) และการวัดค่าตัวแปรตาม (O)		
	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3
1	X ₁ O	X ₂ O	X ₃ O
2	X ₃ O	X ₁ O	X ₂ O
3	X ₂ O	X ₃ O	X ₁ O

X₁ = analytic X₂ = Holistic X₃ = Annotated

4.1 แจกเครื่องมือที่ใช้ในการวิจัยโดยให้นักศึกษากลุ่มที่ 1 ตรวจสอบให้คะแนนเรียงความด้วยวิธีการตรวจให้คะแนนแบบรวมองค์ประกอบ กลุ่มที่ 2 ตรวจสอบให้คะแนนเรียงความด้วยวิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ และกลุ่มที่ 3 ตรวจสอบให้คะแนนเรียงความด้วยวิธีการตรวจให้คะแนนแบบผสมผสาน รจนกระทั่งนักศึกษาตรวจให้คะแนนเสร็จสิ้นจึงเก็บเครื่องมือที่ใช้ในการวิจัยกลับคืน

4.2 แจกเครื่องมือที่ใช้ในการวิจัยโดยให้นักศึกษากลุ่มที่ 2 ตรวจสอบให้คะแนนเรียงความด้วยวิธีการตรวจให้คะแนนแบบรวมองค์ประกอบ กลุ่มที่ 3 ตรวจสอบให้คะแนนเรียงความด้วยวิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ และกลุ่มที่ 1 ตรวจสอบให้คะแนนเรียงความด้วยวิธีการตรวจให้คะแนนแบบผสมผสาน รจนกระทั่งนักศึกษาตรวจให้คะแนนเสร็จสิ้นจึงเก็บเครื่องมือที่ใช้ในการวิจัยกลับคืน

4.3 แจกเครื่องมือที่ใช้ในการวิจัยโดยให้นักศึกษากลุ่มที่ 3 ตรวจสอบให้คะแนนเรียงความด้วยวิธีการตรวจให้คะแนนแบบรวมองค์ประกอบ กลุ่มที่ 1 ตรวจสอบให้คะแนนเรียงความด้วยวิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ และกลุ่มที่ 2 ตรวจสอบให้คะแนนเรียงความด้วยวิธีการตรวจให้คะแนนแบบผสมผสาน รจนกระทั่งนักศึกษาตรวจให้คะแนนเสร็จสิ้นจึงเก็บเครื่องมือที่ใช้ในการวิจัยกลับคืน

4.4 นัดหมายวันและเวลาเพื่อทำการทดลองครั้งต่อไปในอีก 5 สัปดาห์

5. นำข้อมูลที่ได้จากการวิจัยครั้งที่ 1 มาหาค่าสถิติพื้นฐาน คือ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ความแปร ความโค้ง สัมประสิทธิ์แอลฟาของครอนบาค และดัชนีความสอดคล้องของผู้ตรวจให้คะแนน ตามทฤษฎีมาตรฐานเดิม หาค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ และค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ เพื่อแสดงคุณภาพของเครื่องมือตามทฤษฎีการสรุปอ้างอิง จากโปรแกรม EDUG 6.1-e ผู้วิจัยใช้การออกแบบเอกภพสำหรับสององค์ประกอบ แบบ Two-Facet Crossed Design โดยกำหนดให้ P คือ นักเรียน I คือ วิธีการตรวจให้คะแนน และ R คือ ผู้ตรวจให้คะแนน Design ที่ใช้สำหรับงานวิจัยนี้ คือ $p \times i \times X \times r$ design ประกอบไปด้วย นักเรียน 50 คน วิธีการตรวจให้คะแนน 3 รูปแบบ และผู้ตรวจให้คะแนน 3 คน (ข้อสอบเรียงความใช้วิธีสุ่มอย่างง่ายในการคัดเลือกผู้ตรวจให้คะแนน ข้อสอบการเขียนจดหมายใช้ครู 3 คน ที่เป็นครูประจำชั้น) และหาค่าพารามิเตอร์ต่างๆ ตามทฤษฎีการตอบสนองข้อสอบ โดยในที่นี้จะวิเคราะห์โดยใช้ Partial Credit Model ซึ่งจะได้พารามิเตอร์ของนักเรียน พารามิเตอร์

ของเรียงความ และพารามิเตอร์ของผู้ตรวจให้คะแนน จากโปรแกรม FACETS เพื่อใช้เป็นหลักฐานในการแสดงคุณภาพของเครื่องมือสำหรับการวิจัยตามความมุ่งหมายของการวิจัยข้อที่ 1

6. วิเคราะห์คุณลักษณะผู้ตรวจให้คะแนนแต่ละคนโดยใช้โปรแกรม FACETS ซึ่งจะได้คุณลักษณะของผู้ตรวจให้คะแนนในครั้งนี้จำนวน 2 คุณลักษณะ คือ ความเข้มงวด/ใจดีของผู้ตรวจให้คะแนน และความลำเอียงด้านเพศ ตามความมุ่งหมายของการวิจัยข้อที่ 2

7. นำกระดาษคำตอบและวิธีการตรวจให้คะแนนมาจัดชุดตามจำนวนของนักศึกษา วิชาเอกภาษาไทย ชั้นปีที่ 4 ที่เป็นกลุ่มตัวอย่าง จำนวน 140 ชุด เพื่อนำไปทดลองครั้งที่ 2 ระหว่างวันที่ 6 – 17 สิงหาคม 2555 ผู้วิจัยต้องการให้นักศึกษาสมัครให้คะแนนครั้งที่ 1 จึงทิ้งระยะในการเก็บข้อมูลครั้งที่ 2 ไปได้ประมาณ 5 สัปดาห์ เนื่องจากการศึกษาของนักจิตวิทยาเกี่ยวกับการจำและการลืมของมนุษย์ พบว่าเราจะสามารถจำเรื่องราวที่ได้อ่านไปประมาณครึ่งหนึ่ง และจะลดลงไปอีกครึ่งหนึ่งของที่เหลือในทุกๆ 7 วัน จนในที่สุดจะนึกไม่ออกเลย (จตุรงค์ ลังกาพินธุ์. ออนไลน์: 1 – 2) ในครั้งนี้มีนักศึกษามาเข้าร่วมการวิจัยจำนวน 130 คน ไม่มาเข้าร่วมการวิจัย จำนวน 10 คน

8. ทำการทดลองครั้งที่ 2 โดยกระทำซ้ำขั้นตอนดังข้อที่ 3

9. นำข้อมูลที่ได้จากการทดลองครั้งที่ 2 ที่มีลักษณะผู้ตรวจให้คะแนนตรวจให้คะแนนสมบูรณ์ ทั้ง 3 รูปแบบจำนวน 130 คน มาวิเคราะห์คุณลักษณะผู้ตรวจให้คะแนนแต่ละคน ซึ่งจะได้คุณลักษณะการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไปของผู้ตรวจให้คะแนน

10. นำการระบุคุณลักษณะมาศึกษาความสอดคล้องของการระบุความเข้มงวด/ใจดี ความลำเอียงด้านเพศ และการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป เมื่อใช้วิธีการตรวจให้คะแนนเรียงความแตกต่างกัน 3 รูปแบบ โดยใช้ค่า χ^2 จาก Friedman Test

11. หาความเที่ยงตรงของความสามารถในการวัดการเขียนเรียงความของนักเรียน ตามความมุ่งหมายของการวิจัยข้อที่ 3 โดยมีขั้นตอน ดังนี้

11.1 หาความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย โดยผู้วิจัยนำค่า Logit และ Z-Score ที่ได้จากการตรวจให้คะแนนโดยใช้วิธีการตรวจให้คะแนนเรียงความ 3 รูปแบบ (Methods) คือ แบบแยกองค์ประกอบ แบบรวมองค์ประกอบ และแบบผสมผสาน และคุณลักษณะผู้ตรวจให้คะแนน 3 ชนิด (Traits) คือ ความเข้มงวด/ใจดี ความลำเอียงด้านเพศ และการทำหน้าที่ต่างกันเมื่อเวลาผ่านไป มาหาความสัมพันธ์โดยใช้สูตรสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

11.2 หาความเที่ยงตรงตามสภาพ โดยผู้วิจัยนำคุณลักษณะของผู้ตรวจให้คะแนนที่ได้จากข้อ 6 และข้อ 9 มาจัดชุดตามสัดส่วนโดยใช้วิธีการสุ่มอย่างง่าย ดังตาราง 11 และ 12 นำค่าที่ได้มาเปรียบเทียบความเที่ยงตรงตามสภาพ โดยการหาความสัมพันธ์โดยใช้สูตรสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันจากนั้นแปลงค่า r ที่ได้เป็นพิชเชอร์ z'

12. สรุปผลการวิจัย

ตาราง 13 แสดงรูปแบบการดำเนินการหาความเที่ยงตรงเชิงโครงสร้างโดยใช้ผู้ตรวจที่มีคุณลักษณะ
ลำเอียง

วิธีการตรวจให้	คุณลักษณะผู้ตรวจให้	สัดส่วนจำนวนผู้ตรวจให้คะแนน (หมายเลขของผู้ตรวจให้คะแนนที่ได้จากการสุ่ม)
คะแนน	คะแนน	
		กรณีผู้ตรวจให้คะแนน 2 คน เข้มงวด 0 คน : ใจดี 2 คน (เข้มงวด - : ใจดี 9, 88) เข้มงวด 1 คน : ใจดี 1 คน (เข้มงวด 1 : ใจดี 69) เข้มงวด 2 คน : ใจดี 0 คน (เข้มงวด 3, 15 : ใจดี -)
	ความเข้มงวด/ใจดี	กรณีผู้ตรวจให้คะแนน 3 คน เข้มงวด 0 คน : ใจดี 3 คน (เข้มงวด - : ใจดี 9, 28, 69) เข้มงวด 1 คน : ใจดี 2 คน (เข้มงวด 15 : ใจดี 58, 116) เข้มงวด 2 คน : ใจดี 1 คน (เข้มงวด 25, 83 : ใจดี 58) เข้มงวด 3 คน : ใจดี 0 คน (เข้มงวด 11, 62, 76 : ใจดี -)
แบบแยก องค์ประกอบ		กรณีผู้ตรวจให้คะแนน 2 คน ชาย 0 คน : หญิง 2 คน (ชาย - : หญิง 44, 54) ชาย 1 คน : หญิง 1 คน (ชาย 28 : หญิง 76) ชาย 2 คน : หญิง 0 คน (ชาย 2, 42 : หญิง -)
	ความลำเอียงด้านเพศ	กรณีผู้ตรวจให้คะแนน 3 คน ชาย 0 คน : หญิง 3 คน (ชาย - : หญิง 52, 101, 123) ชาย 1 คน : หญิง 2 คน (ชาย 65 : หญิง 66, 123) ชาย 2 คน : หญิง 1 คน (ชาย 26, 42 : หญิง 44) ชาย 3 คน : หญิง 0 คน (ชาย 120, 122, 28 : หญิง -)
		กรณีผู้ตรวจให้คะแนน 2 คน
	การทำหน้าที่ต่างกัน เมื่อเวลาผ่านไป	ครั้งที่ 0 คน : ไม่ครั้งที่ 2 คน (ครั้งที่ - : ไม่ครั้งที่ 41, 62) กรณีผู้ตรวจให้คะแนน 3 คน ครั้งที่ 0 คน : ไม่ครั้งที่ 3 คน (ครั้งที่ - : ไม่ครั้งที่ 23, 26, 60)

ตาราง 13 (ต่อ)

วิธีการตรวจให้		สัดส่วนจำนวนผู้ตรวจให้คะแนน
คะแนน	คุณลักษณะผู้ตรวจให้คะแนน	(หมายเลขของผู้ตรวจให้คะแนนที่ได้จากการสุ่ม)
		กรณีผู้ตรวจให้คะแนน 2 คน
		เข้มงวด 0 คน : ใจดี 2 คน (เข้มงวด - : ใจดี 45, 114)
		เข้มงวด 1 คน : ใจดี 1 คน (เข้มงวด 42 : ใจดี 47)
		เข้มงวด 2 คน : ใจดี 0 คน (เข้มงวด 14, 93 : ใจดี -)
	ความเข้มงวด/ใจดี	กรณีผู้ตรวจให้คะแนน 3 คน
		เข้มงวด 0 คน : ใจดี 3 คน (เข้มงวด - : ใจดี 28, 36, 114)
		เข้มงวด 1 คน : ใจดี 2 คน (เข้มงวด 25 : ใจดี 56, 88)
		เข้มงวด 2 คน : ใจดี 1 คน (เข้มงวด 14, 17 : ใจดี 88)
		เข้มงวด 3 คน : ใจดี 0 คน (เข้มงวด 93, 101, 107 : ใจดี -)
		กรณีผู้ตรวจให้คะแนน 2 คน
แบบรวม		ชาย 0 คน : หญิง 2 คน (ชาย - : หญิง 69, 115)
องค์ประกอบ		ชาย 1 คน : หญิง 1 คน (ชาย 99 : หญิง 66)
		ชาย 2 คน : หญิง 0 คน (ชาย 5, 89 : หญิง -)
	ความลำเอียงด้านเพศ	กรณีผู้ตรวจให้คะแนน 3 คน
		ชาย 0 คน : หญิง 3 คน (ชาย - : หญิง 50, 54, 66)
		ชาย 1 คน : หญิง 2 คน (ชาย 99 : หญิง 25, 59)
		ชาย 2 คน : หญิง 1 คน (ชาย 5, 7 : หญิง 53)
		ชาย 3 คน : หญิง 0 คน (ชาย 89, 99, 120 : หญิง -)
		กรณีผู้ตรวจให้คะแนน 2 คน
	การทำหน้าที่ต่างกัน	คงที่ 0 คน : ไม่คงที่ 2 คน (คงที่ - : ไม่คงที่ 87, 102)
	เมื่อเวลาผ่านไป	กรณีผู้ตรวจให้คะแนน 3 คน
		คงที่ 0 คน : ไม่คงที่ 3 คน (คงที่ - : ไม่คงที่ 33, 98, 108)

ตาราง 13 (ต่อ)

วิธีการตรวจให้ คะแนน		คุณลักษณะผู้ตรวจให้คะแนน	สัดส่วนจำนวนผู้ตรวจให้คะแนน (หมายเลขของผู้ตรวจให้คะแนนที่ได้จากการสุ่ม)
แบบ ผสมผสาน	ความเข้มงวด/ใจดี	กรณีผู้ตรวจให้คะแนน 2 คน	เข้มงวด 0 คน : ใจดี 2 คน (เข้มงวด - : ใจดี 88, 102)
		กรณีผู้ตรวจให้คะแนน 3 คน	เข้มงวด 1 คน : ใจดี 1 คน (เข้มงวด 40 : ใจดี 13) เข้มงวด 2 คน : ใจดี 0 คน (เข้มงวด 3, 17 : ใจดี -)
แบบ ผสมผสาน	ความลำเอียงด้าน เพศ	กรณีผู้ตรวจให้คะแนน 2 คน	ชาย 0 คน : หญิง 2 คน (ชาย - : หญิง 76, 115)
		กรณีผู้ตรวจให้คะแนน 3 คน	ชาย 1 คน : หญิง 1 คน (ชาย 92 : หญิง 2) ชาย 2 คน : หญิง 0 คน (ชาย 7, 58 : หญิง -)
แบบ ผสมผสาน	การทำหน้าที่ต่างกัน เมื่อเวลาผ่านไป	กรณีผู้ตรวจให้คะแนน 2 คน	ชาย 0 คน : หญิง 3 คน (ชาย - : หญิง 22, 54, 69)
		กรณีผู้ตรวจให้คะแนน 3 คน	ชาย 1 คน : หญิง 2 คน (ชาย 92 : หญิง 54, 103) ชาย 2 คน : หญิง 1 คน (ชาย 4, 82 : หญิง 69) ชาย 3 คน : หญิง 0 คน (ชาย 7, 58, 120 : หญิง -)
แบบ ผสมผสาน	การทำหน้าที่ต่างกัน เมื่อเวลาผ่านไป	กรณีผู้ตรวจให้คะแนน 2 คน	คงที่ 0 คน : ไม่คงที่ 2 คน (คงที่ - : ไม่คงที่ 84, 116)
		กรณีผู้ตรวจให้คะแนน 3 คน	คงที่ 0 คน : ไม่คงที่ 3 คน (คงที่ - : ไม่คงที่ 13, 50, 106)

ตาราง 14 แสดงรูปแบบการดำเนินการหาความเที่ยงตรงเชิงโครงสร้างโดยใช้ผู้ตรวจที่มีคุณลักษณะเป็นกลาง

วิธีการตรวจให้คะแนน	คุณลักษณะผู้ตรวจให้คะแนน	สัดส่วนจำนวนผู้ตรวจให้คะแนน (หมายเลขของผู้ตรวจให้คะแนนที่ได้จากการสุ่ม)
แบบแยกองค์ประกอบ	ความเข้มงวด/ใจดี	กรณีผู้ตรวจให้คะแนน 2 คน (9, 33) กรณีผู้ตรวจให้คะแนน 3 คน (18,75,103)
	ความลำเอียงด้านเพศ	กรณีผู้ตรวจให้คะแนน 2 คน (13, 92) กรณีผู้ตรวจให้คะแนน 3 คน (48, 76, 113)
	การทำหน้าที่ต่างกันเมื่อเวลาผ่านไป	กรณีผู้ตรวจให้คะแนน 2 คน (31, 102) กรณีผู้ตรวจให้คะแนน 3 คน (88, 95, 125)
	ความเข้มงวด/ใจดี	กรณีผู้ตรวจให้คะแนน 2 คน (63, 110) กรณีผู้ตรวจให้คะแนน 3 คน (48, 76, 113)
แบบรวมองค์ประกอบ	ความลำเอียงด้านเพศ	กรณีผู้ตรวจให้คะแนน 2 คน (46, 100) กรณีผู้ตรวจให้คะแนน 3 คน (97, 114, 118)
	การทำหน้าที่ต่างกันเมื่อเวลาผ่านไป	กรณีผู้ตรวจให้คะแนน 2 คน (47, 89) กรณีผู้ตรวจให้คะแนน 3 คน (49, 104, 129)
	ความเข้มงวด/ใจดี	กรณีผู้ตรวจให้คะแนน 2 คน (73, 86) กรณีผู้ตรวจให้คะแนน 3 คน (33, 63, 87)
	ความลำเอียงด้านเพศ	กรณีผู้ตรวจให้คะแนน 2 คน (68, 109) กรณีผู้ตรวจให้คะแนน 3 คน (57, 65,114)
แบบผสมผสาน	การทำหน้าที่ต่างกันเมื่อเวลาผ่านไป	กรณีผู้ตรวจให้คะแนน 2 คน (72, 123) กรณีผู้ตรวจให้คะแนน 3 คน (62, 110, 126)

สถิติที่ใช้ในการวิจัย

1. สถิติพื้นฐาน คือ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ความเบ้ ความโด่ง สัมประสิทธิ์แอลฟาของครอนบาค และดัชนีความสอดคล้องของผู้ตรวจให้คะแนน

2. โมเดลการวัดในทฤษฎีการตอบสนองข้อสอบที่เหมาะสมในการใช้ในการตรวจให้คะแนน การเขียนคือ พาเชียลเครดิตโมเดล (Partial Credit Model) ดังนี้ (ศิริชัย กาญจนวาสี. 2550)

$$P_{ix}(\theta) = \frac{\exp\left[\sum_{j=0}^x (\theta - \delta_{ij})\right]}{\sum_{r=0}^{m_i} \left[\exp\sum_{j=0}^r (\theta - \delta_{ij})\right]}$$

เมื่อ $\sum (\theta - \delta_{ij}) \equiv 0$

$P_{ix}(\theta)$ แทน ความน่าจะเป็นที่ผู้ตอบซึ่งมีคุณลักษณะ θ จะตอบข้อ i ด้วยการเลือกหรือสามารถทำรายการคำตอบขั้นที่ x จากจำนวน m_i ขั้น (Step)

δ_{ij} แทน ค่าพารามิเตอร์ระดับความยากของขั้นตอนที่ j ในข้อ i (Item step Difficulty) เมื่อ $j = 1, 2, \dots, m_i$ ค่าที่สูงแสดงถึงขั้นตอนการตอบนั้นมีความยากสัมพัทธ์สูงกว่าขั้นอื่น

3. คุณลักษณะผู้ตรวจให้คะแนน ใช้โมเดลราส์ชแบบหลายองค์ประกอบ

3.1 ความเข้มงวด/ใจดี ของผู้ตรวจให้คะแนน ใช้โมเดลราส์ชแบบหลายองค์ประกอบ โดยใช้สูตรการคำนวณ ดังนี้

$$\text{Log} (P_{nijk}/P_{nijk-1}) = B_n - C_j - D_i - F_{ik}$$

เมื่อ P_{nijk} แทน ความน่าจะเป็นของผู้สอบคนที่ n ถูกให้คะแนน k โดยผู้ตรวจให้คะแนนคนที่ j ในข้อที่ i

P_{nijk-1} แทน ความน่าจะเป็นของผู้สอบคนที่ n ถูกให้คะแนน $k - 1$ โดยผู้ตรวจให้คะแนนคนที่ j ในข้อที่ i

B_n แทน ความสามารถของผู้สอบคนที่ n

C_j แทน ความเข้มงวดของผู้ตรวจให้คะแนนคนที่ j

D_i แทน ความยากของข้อสอบข้อที่ i

F_{ik} แทน ความยากของการเปลี่ยนขั้นจาก k เป็น $k-1$ ในข้อที่

i

3.2 ความลำเอียงด้านเพศ ใช้โมเดลราส์ชแบบหลายองค์ประกอบ โดยใช้สูตรการคำนวณ (Engelhard; & Myford. 2003) ดังนี้

$$\ln[P_{nijk}/P_{nijk-1}] = \theta_n - \zeta_i - \alpha_j - \tau_k$$

เมื่อ	P_{nijk}	แทน	ความน่าจะเป็นที่นักเรียนคนที่ n จะได้คะแนน k บนข้อสอบข้อที่ i จากผู้ตรวจให้คะแนนคนที่ j
	P_{nijk-1}	แทน	ความน่าจะเป็นที่นักเรียนคนที่ n จะได้คะแนน $k - 1$ บนข้อสอบข้อที่ i จากผู้ตรวจให้คะแนนคนที่ j
	θ_n	แทน	ความสามารถของนักเรียนคนที่ n
	ζ_i	แทน	ความยากของข้อสอบข้อที่ i
	α_j	แทน	ความเข้มงวดของผู้ตรวจให้คะแนนคนที่ j
	τ_k	แทน	ความยากในการเปลี่ยนระดับจากคะแนน k ไปสู่คะแนน $k - 1$

สูตรที่โปรแกรม FACETS ใช้ในการคำนวณ Z-Score คือ t value ของ Welch-Satterthwaite (Linacre. 2012: 313)

$$t = \frac{(M_1 - M_2)}{\sqrt{SE_1^2 - SE_2^2}}$$

เมื่อ	$M_1 - M_2$	แทน	ความแตกต่างระหว่างค่าเฉลี่ย
	$\sqrt{SE_1^2 - SE_2^2}$	แทน	ความเบี่ยงเบนมาตรฐานร่วมของค่าเฉลี่ย

3.3 การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป (Differential Rater Functioning Over Time : DRIFT) ใช้โมเดลราสส์ แบบหลายองค์ประกอบที่ประกอบด้วยองค์ประกอบของเวลา (Time Facet) (Myford; & Wolfe. 2009) โดยใช้สูตรการคำนวณ ดังนี้

$$\ln\left(\frac{\pi_{njtk}}{\pi_{njtk-1}}\right) = A_n - S_j - R_t - C_k,$$

เมื่อ	π_{njtk}	แทน	ความน่าจะเป็นที่ความสามารถของนักเรียนคนที่ n จะได้คะแนน k จากผู้ตรวจให้คะแนนคนที่ j ในเวลาที่ t
	π_{njtk-1}	แทน	ความน่าจะเป็นที่ความสามารถของนักเรียนคนที่ n จะได้คะแนน $k-1$ จากผู้ตรวจให้คะแนนคนที่ j ในเวลาที่ t
	A_n	แทน	ระดับของผลความสำเร็จสำหรับนักเรียนคนที่ n

S_j	แทน	ความเข้มงวดของผู้ตรวจให้คะแนนคนที่ j
R_t	แทน	ความเบี่ยงเบนของการตรวจให้คะแนนในเวลา t จากค่าเฉลี่ยของการตรวจให้คะแนนทั้งหมด
C_k	แทน	ความยากของการข้ามระดับระหว่างระดับสเกล k ไปเป็น $k-1$

$$Z_{res,exp_c-res,exp_b} = \frac{Z_{res,exp_c} - Z_{res,exp_b}}{\sqrt{\frac{1}{N_c - 3} + \frac{1}{N_b - 3}}}$$

เมื่อ	Z_{res,exp_c}	แทน	ค่าที่ได้จากการแปลงรูปด้วยวิธี Fisher ณ เวลา c
	Z_{res,exp_b}	แทน	ค่าที่ได้จากการแปลงรูปด้วยวิธี Fisher ณ เวลา b
	N_c	แทน	จำนวนของการให้คะแนนในเวลา Z_{res,exp_c} เป็นฐาน
	N_b	แทน	จำนวนของการให้คะแนนในเวลา Z_{res,exp_b} เป็นฐาน

4. การประมาณค่าสัมประสิทธิ์ การสรุปอ้างอิงสำหรับการออกแบบการวัดที่มีสองฟาเซต (G-Coefficient for Two Facet Design) แบบ Two-Facet Crossed Design (ศิริชัย กาญจนวาสี. 2550)

$$\sigma_{X_{pir}}^2 = \sigma_p^2 + \sigma_r^2 + \sigma_{pi}^2 + \sigma_{pr}^2 + \sigma_{ir}^2 + \sigma_{pir,e}^2$$

เมื่อ	$\sigma_{X_{pio}}^2$	แทน	ความแปรปรวนของคะแนนที่สังเกตได้
	σ_p^2	แทน	ความแปรปรวนของคะแนนเอกภพของผู้สอบ (μ_p)
	σ_i^2	แทน	ความแปรปรวนของคะแนนเฉลี่ยวิธีการตรวจให้คะแนน (μ_i)
	σ_r^2	แทน	ความแปรปรวนของคะแนนเฉลี่ยของผู้ตรวจให้คะแนน (μ_r)
	σ_{pi}^2	แทน	ความแปรปรวนของคะแนนปฏิสัมพันธ์ระหว่างผู้สอบกับวิธีการตรวจให้คะแนน
	σ_{pr}^2	แทน	ความแปรปรวนของคะแนนปฏิสัมพันธ์ระหว่างผู้สอบกับผู้ตรวจให้คะแนน
	σ_{ir}^2	แทน	ความแปรปรวนของคะแนนปฏิสัมพันธ์ระหว่างวิธีการตรวจให้คะแนนกับผู้ตรวจให้คะแนน
	$\sigma_{pir,e}^2$	แทน	ความแปรปรวนของคะแนนคลาดเคลื่อนหรือส่วนที่เหลือ (e_{pir})

5. การหาค่าความเที่ยงตรง โดยใช้สัมประสิทธิ์สหสัมพันธ์ (Correlation Coefficient) โดยใช้สูตรสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน (Pearson Product Moment Correlation Coefficient) (ล้วน

สายยศ; และ อังคณา สายยศ. 2543; บุญเชิด ภิญโญอนันตพงษ์. 2547; มหาวิทยาลัยสุโขทัยธรรมาธิราช. 2549; และ ยาวดี วิบูลย์ศรี. 2552) โดยใช้สูตรการคำนวณ ดังนี้

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

เมื่อ X แทน คะแนนที่ได้จากการตรวจให้คะแนนเรียงความ
 Y แทน คะแนนที่ได้จากการตรวจให้คะแนนจดหมาย
 N แทน จำนวนผู้สอบ

จากนั้นแปลงสหสัมพันธ์ r เป็น z' โดยวิธีการแปลงแบบฟิชเชอร์ (Fisher's Transformer) (ฉัตรศิริ ปิยะพิมลสิทธิ์. ออนไลน์) โดยใช้สูตร

$$z' = \frac{1}{2} [\ln(1+r) - \ln(1-r)]$$

เมื่อ \ln แทน ลอการิทึมฐานธรรมชาติ ซึ่งในตาราง Fisher's Z จะแสดงการแปลงค่า r เป็น z' โดยไม่จำเป็นต้องคำนวณเอง

เกณฑ์ที่ใช้ในการพิจารณา

1. พารามิเตอร์ของนักเรียน (Student Ability : θ) ระดับความสามารถเฉลี่ยของนักเรียนอยู่ในวิธียอมรับได้คือ -3 ถึง +3 (ศิริชัย กาญจนวาสี. 2550: 55) และค่าความเชื่อมั่นแยกส่วนของผู้สอบ (Person Separation Reliability; Person Reliability) ซึ่งคล้ายคลึงกับสัมประสิทธิ์แอลฟา (Englehard. 2002; Myford; & Wolfe. 2003) เป็นอัตราส่วนของความแปรปรวนของคะแนนจริงกับความแปรปรวนของคะแนนสังเกตในการประมาณค่าความสามารถของผู้สอบ ถ้าค่านี้มีค่าสูงแสดงว่ามีความแตกต่างระหว่างความสามารถของผู้สอบ แสดงถึงความสามารถในการจำแนกความสามารถของผู้สอบแต่ละคนออกจากกัน เป็นดัชนีที่แสดงถึงความเชื่อถือได้ของคะแนนความสามารถของผู้สอบที่ได้จากข้อสอบและผู้ตรวจ

2. พารามิเตอร์ของข้อสอบ (Step Difficulty : δ) หรือจุดโค้งการตอบที่ตัดกัน โดยที่ค่าความยากในการเปลี่ยนขั้นควรเรียงลำดับจากน้อยไปมาก และการทดสอบความสอดคล้องของโมเดลของการตรวจให้คะแนนเรียงความตามโปรแกรม FACETS จะให้พิจารณาค่า Quality Control ในตาราง (Linacer. 2012: 194) ที่ต้องมีค่าเรียงลำดับจากน้อยไปมากโดยไม่ข้ามลำดับและไม่ติด " * " ใดๆ จากการคำนวณของโปรแกรม ถ้าพบว่า ค่าความยากในการเปลี่ยนขั้นของวิธีการตรวจให้คะแนน

เรียงความเป็นไปตามเงื่อนไข แสดงว่า มีความเที่ยงตรงในการแบ่งชั้น (Validity of the Categorization) และข้อมูลเชิงประจักษ์สอดคล้องกับโมเดล Partial Credit Model

3. พารามิเตอร์ของผู้ตรวจให้คะแนน (Rater Severity) ให้พิจารณาค่าความเชื่อมั่นแยกส่วนของผู้ตรวจ (Rater Separation Reliability) ซึ่งแสดงถึงความแปรปรวนของการให้คะแนนที่ไม่ต้องการให้เกิดขึ้นระหว่างผู้ตรวจให้คะแนนในระดับของความเข้มงวด/ใจดี ซึ่งดัชนีนี้ไม่ได้บ่งชี้ว่ามีความสอดคล้องกันระหว่างผู้ตรวจให้คะแนน แต่บ่งชี้ว่าผู้ตรวจให้คะแนนมีความเข้มงวดในระดับเดียวกันหรือไม่ กล่าวอีกนัยหนึ่ง ความเชื่อมั่นแยกส่วนของผู้ตรวจให้คะแนนบ่งชี้ถึงความเชื่อมั่นของความแตกต่าง (Reliably Different) มากกว่าระดับความเชื่อมั่นที่คล้ายคลึงกัน (Reliably Similar)

4. การระบุความเข้มงวด/ใจดี ในขั้นแรกนั้นต้องพิจารณาว่าผู้ตรวจให้คะแนนแต่ละคนให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล Partial Credit Model หรือไม่ โดยพิจารณาค่า Weighted mean square residual:INFIT และ Unweighted mean square residual:OUTFIT ที่กำหนดขอบเขตของการยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลอยู่ระหว่าง 0.5 ถึง 1.5 จากนั้นจึงระบุความเข้มงวด/ใจดี จากค่าศูนย์ (0) เครื่องหมายบวก (+) ลบ (-) ของค่า Logit โดยผู้วิจัยกำหนดให้ผู้ตรวจให้คะแนนที่เข้มงวดจะมีค่า Logit มากกว่า +1.00 ให้คะแนนเป็นกลาง $-1 < \text{Logit} < +1$ และให้คะแนนใจดีจะมีค่า Logit น้อยกว่า -1.00

5. การระบุความลำเอียงด้านเพศ ในขั้นแรกนั้นต้องพิจารณาว่าผู้ตรวจให้คะแนนแต่ละคนให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล Partial Credit Model หรือไม่โดยพิจารณาค่า INFIT และ OUTFIT ต่อมาจึงระบุความลำเอียงด้านเพศของผู้ตรวจให้คะแนน จากค่า Z-Score ที่มีนัยสำคัญทางสถิติ ถ้า Z-Score ไม่มีนัยสำคัญทางสถิติ แสดงว่า เป็นกลาง สุดท้ายจึงพิจารณาว่าผู้ตรวจให้คะแนนคนนั้นๆ ลำเอียงเข้าข้างเพศชายหรือเพศหญิง คะแนนที่มีแนวโน้มใจดี (Logit ต่ำกว่า) กับเพศใดแสดงว่าลำเอียงเข้าข้างเพศนั้น

6. การระบุการทำหน้าที่ต่างกันเมื่อเวลาผ่านไป ในขั้นแรกนั้นต้องพิจารณาว่า ผู้ตรวจให้คะแนนแต่ละคนให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล Partial Credit Model หรือไม่ โดยพิจารณาค่า INFIT และ OUTFIT ต่อมาจึงระบุการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป จากค่า Z-Score โดยที่ค่า Z-Score ที่มีนัยสำคัญทางสถิติแสดงว่าผู้ตรวจให้คะแนนให้คะแนนไม่คงที่ ถ้า Z-Score ไม่มีนัยสำคัญทางสถิติแสดงว่าผู้ตรวจให้คะแนนให้คะแนนคงที่

บทที่ 4

ผลการวิเคราะห์ข้อมูล

การศึกษาวิจัยในครั้งนี้มีจุดมุ่งหมายเพื่อสร้างและพัฒนาคุณภาพของเครื่องมือวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 เพื่อศึกษาคุณลักษณะของผู้ตรวจให้คะแนน คือ ความเข้มงวด/ใจดี ความลำเอียงด้านเพศ และการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป และเพื่อตรวจสอบความเที่ยงตรงของความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 ผู้วิจัยขอเสนอผลการวิจัยออกเป็น 3 ตอน ดังนี้

ตอนที่ 1 การแสดงคุณภาพของเครื่องมือวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3

ตอนที่ 2 การศึกษาคุณลักษณะของผู้ตรวจให้คะแนน คือ ความเข้มงวด/ใจดี ความลำเอียงด้านเพศ และการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป

ตอนที่ 3 การตรวจสอบความเที่ยงตรงของความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3

เพื่อให้เกิดความเข้าใจในผลของการวิจัยตรงกันผู้วิจัยได้กำหนดสัญลักษณ์และอักษรย่อใช้แทนค่าสถิติและตัวแปรต่างๆ ดังต่อไปนี้

Ana	แทน	วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ
Hol	แทน	วิธีการตรวจให้คะแนนแบบรวมองค์ประกอบ
Ann	แทน	วิธีการตรวจให้คะแนนแบบผสมผสาน
\bar{X}	แทน	ค่าเฉลี่ยเลขคณิตของคะแนน
SD	แทน	ส่วนเบี่ยงเบนมาตรฐาน
SE	แทน	ความคลาดเคลื่อนมาตรฐาน
S_k	แทน	ความเบ้
K_u	แทน	ความโด่ง
α	แทน	ค่าความเชื่อมั่นจากสูตรของครอนบาคแอลฟา
RAI	แทน	ค่าดัชนีสอดคล้องระหว่างผู้ตรวจให้คะแนนจากสูตร RAI
P	แทน	นักเรียน
I	แทน	วิธีการตรวจให้คะแนน
R	แทน	ผู้ตรวจให้คะแนน
PI	แทน	ปฏิสัมพันธ์ระหว่างนักเรียนและวิธีการตรวจให้คะแนน
PR	แทน	ปฏิสัมพันธ์ระหว่างนักเรียนและผู้ตรวจให้คะแนน
IR	แทน	ปฏิสัมพันธ์ระหว่างวิธีการตรวจให้คะแนนและผู้ตรวจให้คะแนน

PIR	แทน	ปฏิสัมพันธ์ระหว่างนักเรียน	วิธีการตรวจให้คะแนน และผู้ตรวจให้คะแนน
SL	แทน	ความเข้มงวด/ใจดี	
DRF	แทน	ความลำเอียงด้านเพศ	
DRIFT	แทน	การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป	
Max.	แทน	ค่าสูงสุด	
Min.	แทน	ค่าต่ำสุด	
Sep.R	แทน	ค่าความเชื่อมั่นแยกส่วน (Separation Reliability)	
Infit	แทน	ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนที่มีการถ่วงน้ำหนัก	
Outfit	แทน	ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนที่ไม่มีการถ่วงน้ำหนัก	
r	แทน	ค่าสหสัมพันธ์จากสูตรสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน	
z'	แทน	ค่าจากการแปลง r ด้วยวิธีการแปลงแบบฟิชเชอร์	

ตอนที่ 1 การแสดงคุณภาพของเครื่องมือวัดความสามารถในการเขียนของนักเรียน ชั้นประถมศึกษาปีที่ 3

การวิเคราะห์ในตอนที่ 1 ผู้วิจัยต้องการตอบคำถามการวิจัยข้อที่ 1 ที่ว่าเครื่องมือวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 ที่ผู้วิจัยสร้างขึ้นมีคุณภาพหรือไม่ อย่างไร จำแนกตามวิธีการตรวจให้คะแนน คือ แบบแยกองค์ประกอบ แบบรวมองค์ประกอบ และแบบผสมผสาน และแบ่งการนำเสนอออกเป็น 2 ส่วน ตามชนิดของเครื่องมือสำหรับวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 ซึ่งมี 2 ชนิด คือ วิธีการตรวจให้คะแนนการเขียนเรียงความในส่วนของเขียนจากข้อสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 และวิธีการตรวจให้คะแนนการเขียนจดหมาย โดยแบ่งการนำเสนอออกเป็น 3 ส่วนย่อยคือ 1.1 การวิเคราะห์คุณภาพข้อสอบตามทฤษฎีมาตรฐานเดิม 1.2 การวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการสรุปอ้างอิง และ 1.3 การวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบ โดยมีรายละเอียด ดังนี้

1.1 การวิเคราะห์คุณภาพข้อสอบตามทฤษฎีมาตรฐานเดิม ผู้วิจัยนำเสนอผลการวิเคราะห์ข้อมูล ได้แก่ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ความเบ้ ความโด่ง เพื่อแสดงลักษณะของข้อมูล นำเสนอค่าความเชื่อมั่น และดัชนีสอดคล้องระหว่างผู้ตรวจให้คะแนน เพื่อแสดงหลักฐานการมีคุณภาพของวิธีการตรวจให้คะแนน แบ่งตามชนิดเครื่องมือที่ใช้ในการวิจัย ดังตาราง 15 – 16

ตาราง 15 ค่าสถิติพื้นฐานของคะแนนการเขียนเรียงความ จำแนกตามวิธีการตรวจให้คะแนน

วิธีการตรวจ	สถิติ คะแนนเต็ม	\bar{X}	SD	S_k	K_u	α	RAI
Ana	5	2.342	0.926	0.479	-0.073	0.980	0.811
Hol	5	2.526	0.924	0.315	-0.237	0.988	0.823
Ann	5	2.426	0.950	0.368	-0.276	0.986	0.818

ผลการวิเคราะห์จากตาราง 15 พบว่า

ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน เมื่อตรวจโดยใช้วิธีการตรวจให้คะแนนแบบรวมองค์ประกอบ มีค่าเฉลี่ยสูงสุด คือ 2.526 คะแนน รองลงมาคือ แบบผสมผสาน และแบบแยกองค์ประกอบ มีค่า 2.426 และ 2.342 คะแนน ตามลำดับ การตรวจโดยใช้วิธีการตรวจให้คะแนนแบบผสมผสานมีส่วนเบี่ยงเบนมาตรฐานสูงสุด คือ 0.950 รองลงมา คือ แบบแยกองค์ประกอบ และแบบรวมองค์ประกอบ มีค่า 0.926 และ 0.924 ตามลำดับ แสดงให้เห็นว่า ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานมีค่าใกล้เคียงกัน

ความเบ้และความโด่ง เมื่อตรวจโดยใช้วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ มีลักษณะการแจกแจงแบบเบ้ขวาสูงสุดมีค่า 0.429 รองลงมา คือ แบบผสมผสาน และแบบรวมองค์ประกอบ มีค่า 0.368 และ 0.315 ตามลำดับ การตรวจโดยใช้วิธีการตรวจให้คะแนนแบบผสมผสานมีลักษณะการแจกแจงที่แบนมากที่สุดมีค่า -0.276 รองลงมา คือ แบบรวมองค์ประกอบ และแบบแยกองค์ประกอบ มีค่า -0.237 และ -0.073 ตามลำดับ แสดงให้เห็นว่า มีความเบ้ขวาใกล้เคียงกันและมีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ

ความเชื่อมั่นและดัชนีสอดคล้องของผู้ตรวจให้คะแนน เมื่อตรวจโดยใช้วิธีการตรวจให้คะแนนแบบรวมองค์ประกอบ มีค่าความเชื่อมั่นสูงสุด คือ 0.988 รองลงมาคือ แบบผสมผสาน และแบบแยกองค์ประกอบ มีค่า 0.986 และ 0.980 ตามลำดับ การตรวจโดยใช้วิธีการตรวจให้คะแนนแบบรวมองค์ประกอบ มีค่าดัชนีสอดคล้องของผู้ตรวจให้คะแนนสูงสุด คือ 0.823 รองลงมา คือ แบบผสมผสาน และแบบแยกองค์ประกอบ มีค่า 0.818 และ 0.811 ตามลำดับ แสดงให้เห็นว่า ค่าความเชื่อมั่นและค่าดัชนีสอดคล้องของผู้ตรวจให้คะแนนมีค่าสูงสอดคล้องกัน นั่นคือ วิธีการตรวจให้คะแนนการเขียนเรียงความมีความเชื่อมั่นตามทฤษฎีมาตรฐานเดิม

ตาราง 16 ค่าสถิติพื้นฐานของคะแนนการเขียนจดหมาย จำแนกตามวิธีการตรวจให้คะแนน

วิธีการตรวจ	สถิติ คะแนนเต็ม	\bar{X}	SD	S_k	K_u	α	RAI
Ana	5	2.608	1.155	0.387	-0.760	0.827	0.809
Hol	5	2.887	1.114	0.168	-0.796	0.918	0.810
Ann	5	2.533	0.999	0.554	-0.099	0.924	0.813

ผลการวิเคราะห์จากตาราง 16 พบว่า

ค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน เมื่อตรวจโดยใช้วิธีการตรวจให้คะแนนแบบรวมองค์ประกอบ มีค่าเฉลี่ยสูงสุดคือ 2.887 คะแนน รองลงมาคือ แบบแยกองค์ประกอบ และแบบผสมผสาน มีค่า 2.608 และ 2.533 คะแนน ตามลำดับ การตรวจโดยใช้วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบมีส่วนเบี่ยงเบนมาตรฐานสูงสุด คือ 1.155 รองลงมาคือ แบบรวมองค์ประกอบ และแบบผสมผสาน มีค่า 1.114 และ 0.999 ตามลำดับ แสดงให้เห็นว่าค่าเฉลี่ยและส่วนเบี่ยงเบนมาตรฐานมีค่าใกล้เคียงกัน

ความเบ้และความโด่ง เมื่อตรวจโดยใช้วิธีการตรวจให้คะแนนแบบผสมผสาน มีลักษณะการแจกแจงแบบเบ้ขวาสูงสุดมีค่า 0.554 รองลงมา คือ แบบแยกองค์ประกอบ และแบบรวมองค์ประกอบ มีค่า 0.387 และ 0.168 ตามลำดับ การตรวจโดยใช้วิธีการตรวจให้คะแนนแบบรวมองค์ประกอบมีลักษณะการแจกแจงที่แบนมากที่สุด มีค่า -0.796 รองลงมาคือ แบบแยกองค์ประกอบ และแบบผสมผสาน มีค่า -0.760 และ -0.099 ตามลำดับ แสดงให้เห็นว่า มีความเบ้ขวาแตกต่างกันเล็กน้อยและมีลักษณะการแจกแจงที่แบนกว่าโค้งปกติ

ความเชื่อมั่นและดัชนีสอดคล้องของผู้ตรวจให้คะแนน เมื่อตรวจโดยใช้วิธีการตรวจให้คะแนนแบบผสมผสาน มีค่าความเชื่อมั่นสูงที่สุด คือ 0.924 รองลงมาคือ แบบรวมองค์ประกอบ และแบบแยกองค์ประกอบ มีค่า 0.918 และ 0.827 ตามลำดับ การตรวจโดยใช้วิธีการตรวจให้คะแนนแบบผสมผสานมีค่าดัชนีสอดคล้องของผู้ตรวจ ให้คะแนนสูงสุด คือ 0.813 รองลงมา คือ แบบรวมองค์ประกอบ และแบบแยกองค์ประกอบ มีค่า 0.810 และ 0.809 ตามลำดับ แสดงให้เห็นว่า ค่าความเชื่อมั่นและค่าดัชนีสอดคล้องของผู้ตรวจให้คะแนนมีค่าสูงสอดคล้องกัน นั่นคือวิธีการตรวจการเขียนจดหมายมีความเชื่อมั่นตามทฤษฎีมาตรฐานเดิม

1.2 การวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการสรุปอ้างอิง ผู้วิจัยนำเสนอผลการวิเคราะห์ความแปรปรวน และผลจากปฏิสัมพันธ์จากแหล่งต่างๆ เพื่อแสดงหลักฐานการมีคุณภาพของวิธีการตรวจให้คะแนนตามทฤษฎีการสรุปอ้างอิง ผู้วิจัยใช้การออกแบบเอกภพสำหรับสององค์ประกอบ แบบ Two-Facet Crossed Design วิเคราะห์จากโปรแกรม EDUG 6.1-e โดยกำหนดให้ P คือ นักเรียน I คือ วิธีการตรวจให้คะแนน และ R คือ ผู้ตรวจให้คะแนน โดยที่ Design ที่ใช้สำหรับงานวิจัยนี้ คือ $p \times i \times r$ design

ประกอบไปด้วย นักเรียน 50 คน วิธีการตรวจให้คะแนน 3 รูปแบบ และผู้ตรวจให้คะแนน 3 คน แบ่งตามชนิดเครื่องมือที่ใช้ในการวิจัย ดังตาราง 17 – 18

ตาราง 17 ค่าความแปรปรวนจากแหล่งต่างๆ ของการเขียนเรียงความ

แหล่ง ความแปรปรวน	DF	Sum of Square	Mean Square	Estimated Variance Component	% of Total Variance
P	49	106.36331	2.17068	0.18296	29.30
I	2	4.45818	2.22909	0.00825	0.00
R	2	27.72664	13.86332	0.06749	10.80
PI	98	22.06162	0.22512	0.00827	1.30
PR	98	48.92650	0.49925	0.09965	16.00
IR	4	13.76464	3.44116	0.06482	10.40
PIR	196	39.25830	0.20030	0.20030	32.10
Total	449	262.55918			100%
Coefficients of Generalizability_Ralative				0.76	
Coefficients of Generalizability_Absolute				0.68	

จากตาราง 17 พบว่า ค่าปฏิสัมพันธ์ระหว่างนักเรียน วิธีการตรวจให้คะแนน และผู้ตรวจให้คะแนน (PIR) มีค่ามากที่สุด 0.20030 คิดเป็นร้อยละ 32.10 รองลงมาคือ ความแปรปรวนของนักเรียน (P) มีค่า 0.18296 คิดเป็นร้อยละ 29.30 แสดงว่า มีความแตกต่างของคะแนนที่มาจากความสามารถที่แตกต่างกันของนักเรียน วิธีการตรวจให้คะแนน และผู้ตรวจให้คะแนนสูงที่สุด และความแปรปรวนของวิธีการตรวจให้คะแนนมีค่าน้อยที่สุด 0.00825 คิดเป็นร้อยละ 0.00 แสดงว่า วิธีการตรวจให้คะแนนไม่ใช่แหล่งที่ทำให้คะแนนแตกต่างกัน ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์มีค่า 0.76 และค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์มีค่า 0.68 แสดงว่า วิธีการตรวจให้คะแนนการเขียนเรียงความมีความเชื่อมั่นตามทฤษฎีการสรุปอ้างอิง

ตาราง 18 ค่าความแปรปรวนจากแหล่งต่างๆ ของการเขียนจดหมาย

แหล่ง ความแปรปรวน	DF	Sum of Square	Mean Square	Estimated Variance Component	% of Total Variance
P	49	139.12998	2.83939	0.24416	21.20
I	2	10.39694	5.19847	0.01434	1.20
R	2	77.46039	38.73019	0.23487	20.40
PI	98	34.10163	0.34798	0.05251	0.00
PR	98	78.35119	0.79950	0.09800	8.50
IR	4	12.82184	3.20546	0.05400	4.70
PIR	196	99.07632	0.50549	0.50549	43.90
Total	449	451.33829			100%
Coefficients of Generalizability_Ralative				0.73	
Coefficients of Generalizability_Absolute				0.58	

จากตาราง 18 พบว่า ค่าปฏิสัมพันธ์ระหว่างนักเรียน วิธีการตรวจให้คะแนน และผู้ตรวจให้คะแนน (PIR) มีค่ามากที่สุด 0.50549 คิดเป็นร้อยละ 43.90 รองลงมาคือ ความแปรปรวนของนักเรียน (P) มีค่า 0.24416 คิดเป็นร้อยละ 21.20 แสดงว่า มีความแตกต่างของคะแนนที่มาจากความสามารถที่แตกต่างกันของนักเรียน วิธีการตรวจให้คะแนน และผู้ตรวจให้คะแนนสูงที่สุด และค่าปฏิสัมพันธ์ระหว่างนักเรียนกับวิธีการตรวจให้คะแนน (PI) มีค่าน้อยที่สุด 0.05251 คิดเป็นร้อยละ 0.00 แสดงว่านักเรียนกับวิธีการตรวจให้คะแนน ไม่ใช่แหล่งที่ทำให้คะแนนแตกต่างกัน ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ มีค่า 0.73 และค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ มีค่า 0.58 แสดงว่าวิธีการตรวจให้คะแนนการเขียนจดหมาย มีความเชื่อมั่นตามทฤษฎีการสรุปอ้างอิง

1.3 การวิเคราะห์คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบ เนื่องจากลักษณะของข้อมูลการให้คะแนนจดหมายไม่มีความเหมาะสมที่จะใช้วิเคราะห์ตามทฤษฎีการตอบสนองข้อสอบ ผู้วิจัยจึงวิเคราะห์โดยใช้โมเดลของราล์ชแบบหลายองค์ประกอบกับการแสดงหลักฐานการมีคุณภาพของวิธีการตรวจให้คะแนนเรียงความเพียงอย่างเดียวเท่านั้น การวิเคราะห์ในส่วนนี้ได้แบ่งการนำเสนอออกเป็น 3 ส่วนย่อย จำแนกตามพารามิเตอร์ คือ 1) พารามิเตอร์ของนักเรียนเพื่อแสดงให้เห็นค่าเฉลี่ยของความสามารถของนักเรียน จำนวน 50 คน 2) พารามิเตอร์ของข้อสอบเรียงความ เพื่อแสดงให้เห็นค่าความยากในการเปลี่ยนขั้นของการใช้วิธีการตรวจให้คะแนนทั้ง 3 รูปแบบ ว่าเป็นไปตามลำดับและมีความเที่ยงตรงในการเปลี่ยนขั้นหรือไม่ และ 3) พารามิเตอร์ผู้ตรวจให้คะแนน เพื่อแสดงให้เห็นค่าเฉลี่ยของความเข้มงวดโดยรวมของผู้ตรวจให้คะแนนทั้ง 130 คน ดังตาราง 19 – 21

ตาราง 19 แสดงค่าพารามิเตอร์ของนักเรียน จำแนกตามคุณลักษณะผู้ตรวจให้คะแนน

คุณลักษณะ		Logit	SE	Max.	Min.	Infit	Outfit	Sep.R	
S/L	Ana	-1.09	0.06	-4.02	0.52	1.00	1.00	.99	
	Hol	-0.95	0.14	-4.69	1.47	0.99	1.01	.99	
	Ann	-1.06	0.07	-4.50	1.00	1.00	1.01	.99	
	ชาย	-1.55	0.07	-4.13	-0.09	1.01	1.00	.99	
	หญิง	-0.65	0.06	-1.91	0.51	1.00	0.99	.99	
	รวม	-1.09	0.06	-4.02	0.52	1.00	1.00	.99	
DRF	Hol	ชาย	-1.71	0.15	-5.04	0.75	1.00	1.02	.99
		หญิง	-0.27	0.13	-2.32	1.48	0.99	0.99	.98
		รวม	-0.95	0.14	-4.69	1.47	0.99	1.01	.99
	Ann	ชาย	-1.70	0.08	-4.71	0.29	1.01	1.01	.99
		หญิง	-0.49	0.07	-1.85	1.00	1.00	1.00	.99
		รวม	-1.06	0.07	-4.50	1.00	1.00	1.01	.99
Ana	วันที่ 1	-1.09	0.06	-4.02	0.52	1.00	1.00	.99	
	วันที่ 2	-1.02	0.06	-4.02	0.30	1.00	1.00	.99	
	รวม	-0.97	0.04	-3.68	0.37	1.00	1.00	1.00	
DRIFT	Hol	วันที่ 1	-0.95	0.14	-4.69	1.47	0.99	1.01	.99
		วันที่ 2	-0.84	0.14	-4.92	1.41	1.00	1.00	.99
		รวม	-0.80	0.09	-4.34	1.23	1.00	1.01	.99
Ann	วันที่ 1	-1.06	0.07	-4.50	1.00	1.00	1.01	.99	
	วันที่ 2	-0.92	0.07	-4.20	0.77	1.00	1.00	.99	
	รวม	-0.92	0.05	-4.06	0.81	1.00	1.01	1.00	

จากตาราง 19 พบว่า ค่าพารามิเตอร์ความสามารถของนักเรียน (Student Ability : θ) เมื่อจำแนกตามคุณลักษณะผู้ตรวจให้คะแนน ผลปรากฏ ดังนี้

19.1 ความเข้มงวด/ใจดีของผู้ตรวจให้คะแนน แบบรวมองค์ประกอบนักเรียนมีค่าเฉลี่ยความสามารถสูงสุด คือ -0.95 รองลงมาคือ แบบผสมผสานและแบบรวมองค์ประกอบ มีค่า -1.06 และ -1.09 ตามลำดับ แบบรวมองค์ประกอบมีความแปรปรวนจากการวัดสูงสุด คือ 0.14 รองลงมาคือ แบบผสมผสานและแบบแยกองค์ประกอบ มีค่า 0.17 และ 0.16 ตามลำดับ และเมื่อพิจารณาเป็นรายบุคคลพบว่ามีนักเรียนที่มีค่าความสามารถไม่อยู่ในช่วงที่ยอมรับได้ จำนวน 2 คน และการเปรียบเทียบความเหมาะสมกลมกลืนของข้อมูลเชิงประจักษ์กับโมเดล Partial Credit Model ปรากฏว่า แบบแยกองค์ประกอบ แบบรวมองค์ประกอบและแบบผสมผสานมีค่าเฉลี่ย Infit และ Outfit อยู่ในช่วงที่ยอมรับได้ ความเชื่อมั่นแยกส่วนของนักเรียนเมื่อใช้วิธีการตรวจที่แตกต่างกัน 3 รูปแบบ มีค่า .99

19.2 ความลำเอียงด้านเพศ แบบรวมองค์ประกอบนักเรียนมีค่าเฉลี่ยความสามารถสูงสุด คือ -0.95 รองลงมาคือ แบบผสมผสานและแบบรวมองค์ประกอบ มีค่า -1.06 และ -1.09 ตามลำดับ แบบรวมองค์ประกอบมีความแปรปรวนจากการวัดสูงสุด คือ 0.14 รองลงมาคือ แบบผสมผสานและแบบแยกองค์ประกอบ มีค่า 0.17 และ 0.16 ตามลำดับ ระดับความสามารถเฉลี่ยของนักเรียนอยู่ในช่วงที่ยอมรับได้ และเมื่อพิจารณาเป็นรายบุคคลพบว่ามีนักเรียนที่มีค่าความสามารถไม่อยู่ในวิธีที่ยอมรับได้ จำนวน 1 คน และการเปรียบเทียบความเหมาะสมกลมกลืนของข้อมูลเชิงประจักษ์กับโมเดล Partial Credit Mode ปรากฏว่า แบบแยกองค์ประกอบ แบบรวมองค์ประกอบและแบบผสมผสานมีค่าเฉลี่ย Infit และ Outfit อยู่ในช่วงที่ยอมรับได้ ความเชื่อมั่นแยกส่วนของนักเรียน มีค่าอยู่ระหว่าง .98 – .99

19.3 การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป แบบรวมองค์ประกอบนักเรียนมีค่าเฉลี่ยความสามารถสูงสุด คือ -0.80 รองลงมาคือ แบบผสมผสานและแบบรวมองค์ประกอบ มีค่า -0.92 และ -0.97 ตามลำดับ แบบรวมองค์ประกอบมีความแปรปรวนจากการวัดสูงสุด คือ 0.09 รองลงมาคือ แบบผสมผสานและแบบแยกองค์ประกอบ มีค่า 0.05 และ 0.04 ตามลำดับ ระดับความสามารถเฉลี่ยของนักเรียนอยู่ในช่วงที่ยอมรับได้ และเมื่อพิจารณาเป็นรายบุคคลพบว่ามีนักเรียนที่มีค่าความสามารถไม่อยู่ในวิธีที่ยอมรับได้ จำนวน 2 คน และการเปรียบเทียบความเหมาะสมกลมกลืนของข้อมูลเชิงประจักษ์กับโมเดล Partial Credit Mode ปรากฏว่า แบบแยกองค์ประกอบ แบบรวมองค์ประกอบและแบบผสมผสานมีค่าเฉลี่ย Infit และ Outfit อยู่ในช่วงที่ยอมรับได้ ความเชื่อมั่นแยกส่วนของนักเรียน มีค่าอยู่ระหว่าง .98 – 1.00

ตาราง 20 แสดงค่าพารามิเตอร์ของข้อสอบการเขียนเรียงความ จำแนกตามคุณลักษณะผู้ตรวจให้คะแนน

คุณลักษณะ		ระดับการให้คะแนน	1	2	3	4	5	
S/L	Ana	ความยากในการเปลี่ยนขั้น		-2.71	-0.45	0.92	2.24	
		การควบคุมคุณภาพ	-2.29	-1.34	-0.48	0.21	0.60	
	Hol	ความยากในการเปลี่ยนขั้น		-3.40	-0.74	1.18	2.96	
		การควบคุมคุณภาพ	-3.04	-1.50	-0.29	0.67	1.22	
	Ann	ความยากในการเปลี่ยนขั้น		-2.90	-0.65	0.98	2.56	
		การควบคุมคุณภาพ	-2.64	-1.41	-0.44	0.42	0.95	
	Ana	เพศชาย	ความยากในการเปลี่ยนขั้น		-2.80	-0.46	1.00	2.27
			การควบคุมคุณภาพ	-2.68	-1.61	-0.72	-0.06	0.26
		เพศหญิง	ความยากในการเปลี่ยนขั้น		-2.71	-0.45	0.90	2.25
			การควบคุมคุณภาพ	-1.69	-1.06	-0.30	0.36	0.81
		รวม	ความยากในการเปลี่ยนขั้น		-2.71	-0.45	0.92	2.24
			การควบคุมคุณภาพ	-2.29	-1.34	-0.48	0.21	0.60
DRF	Hol	เพศชาย	ความยากในการเปลี่ยนขั้น		-3.68	-0.82	1.30	3.20
		การควบคุมคุณภาพ	-3.78	-1.88	-0.64	0.20	0.69	
	เพศหญิง	ความยากในการเปลี่ยนขั้น		-3.41	-0.73	1.16	2.98	
		การควบคุมคุณภาพ	-2.01	-1.12	-0.04	0.92	1.50	
	รวม	ความยากในการเปลี่ยนขั้น		-3.40	-0.74	1.18	2.96	
		การควบคุมคุณภาพ	-3.04	-1.50	-0.29	0.67	1.22	
	Ann	เพศชาย	ความยากในการเปลี่ยนขั้น		-3.08	-0.74	0.97	2.85
			การควบคุมคุณภาพ	-3.15	-1.78	-0.79	0.05	0.19
		เพศหญิง	ความยากในการเปลี่ยนขั้น		-2.88	-0.64	0.99	2.53
			การควบคุมคุณภาพ	-1.86	-1.07	-0.22	0.60	1.22
		รวม	ความยากในการเปลี่ยนขั้น		-2.90	-0.65	0.98	2.56
			การควบคุมคุณภาพ	-2.64	-1.41	-0.44	0.42	0.95

ตาราง 20 (ต่อ)

คุณลักษณะ		ระดับการให้คะแนน	1	2	3	4	5
Ana	วันที่1	ความยากในการเปลี่ยนชั้น		-2.71	-0.45	0.92	2.24
		การควบคุมคุณภาพ	-2.29	-1.34	-0.48	0.21	0.60
	วันที่2	ความยากในการเปลี่ยนชั้น		-2.95	-0.44	1.03	2.36
		การควบคุมคุณภาพ	-2.34	-1.37	-0.41	0.37	0.94
	รวม	ความยากในการเปลี่ยนชั้น		-2.60	-0.38	0.90	2.09
		การควบคุมคุณภาพ	-2.00	-1.21	-0.46	0.13	0.54
DRIFT	วันที่ 1	ความยากในการเปลี่ยนชั้น		-3.40	-0.74	1.18	2.96
		การควบคุมคุณภาพ	-3.04	-1.50	-0.29	0.67	1.22
	วันที่ 2	ความยากในการเปลี่ยนชั้น		-3.81	-1.67	1.34	3.54
		การควบคุมคุณภาพ	-3.38	-1.71	-0.23	0.97	1.59
	รวม	ความยากในการเปลี่ยนชั้น		-3.20	-0.80	1.12	2.88
		การควบคุมคุณภาพ	-2.62	-1.36	-0.30	0.53	1.12
Ann	วันที่ 1	ความยากในการเปลี่ยนชั้น		-2.90	-0.65	0.98	2.56
		การควบคุมคุณภาพ	-2.64	-1.41	-0.44	0.42	0.95
	วันที่ 2	ความยากในการเปลี่ยนชั้น		-2.94	-0.68	1.08	2.54
		การควบคุมคุณภาพ	-2.55	-1.37	-0.33	0.47	0.96
	รวม	ความยากในการเปลี่ยนชั้น		-2.69	-0.61	0.95	2.35
		การควบคุมคุณภาพ	-2.27	-1.25	-0.41	0.29	0.78

จากตาราง 20 พบว่า ค่าพารามิเตอร์ของเรียงความ (Step Difficulty : δ) จำแนกตามคุณลักษณะ ผู้ตรวจให้คะแนนผลปรากฏ ดังนี้

20.1 ความเข้มงวด/ใจดี แบบแยกองค์ประกอบ มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ คือ -2.71, -0.45, 0.92 และ 2.24 ตามลำดับ แบบรวมองค์ประกอบ มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ คือ -3.40, -0.74, 1.18 และ 2.96 ตามลำดับ และแบบผสมผสาน มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ คือ -2.90, -0.65, 0.98 และ 2.56 ตามลำดับ แสดงว่า แสดงว่า มีความเที่ยงตรงในการแบ่งชั้น และข้อมูลเชิงประจักษ์สอดคล้องกับโมเดล Partial Credit Model

20.2 ความลำเอียงด้านเพศ เมื่อพิจารณาเฉพาะนักเรียนที่เป็นเพศชาย แบบแยกองค์ประกอบ มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ คือ -2.80, -0.46, 1.00 และ 2.27 ตามลำดับ แบบรวม

องค์ประกอบ มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ คือ -3.68, -0.82, 1.30 และ 3.20 ตามลำดับ และแบบผสมผสาน มีความยากในชั้นที่ 1 – 4 เรียงลำดับกันดังนี้ คือ -3.08, -0.74, 0.97 และ 2.85 ตามลำดับ เมื่อพิจารณาเฉพาะนักเรียนที่เป็นเพศหญิง แบบแยกองค์ประกอบ มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ -2.71, -0.45, 0.90 และ 2.25 ตามลำดับ แบบรวมองค์ประกอบ มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ -3.41, -0.73, 1.16 และ 2.98 ตามลำดับ และแบบผสมผสาน มีความยากในชั้นที่ 1 – 4 เรียงลำดับกันดังนี้ -2.88, -0.64, 0.99 และ 2.53 ตามลำดับ เมื่อพิจารณาโดยรวม ทั้งเพศชายและเพศหญิง แบบแยกองค์ประกอบ มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ คือ -2.71, -0.45, 0.92 และ 2.24 ตามลำดับ แบบรวมองค์ประกอบ มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ คือ -3.40, -0.74, 1.18 และ 2.96 ตามลำดับ และแบบผสมผสาน มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ คือ -2.90, -0.65, 0.98 และ 2.56 ตามลำดับ แสดงว่า แสดงว่ามีความเที่ยงตรงในการแบ่งชั้น และข้อมูลเชิงประจักษ์สอดคล้องกับโมเดล Partial Credit Model

20.3 การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป เมื่อพิจารณาเฉพาะคะแนนสอบวันที่ 1 แบบแยกองค์ประกอบ มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ คือ -2.71, -0.45, 0.92 และ 2.24 ตามลำดับ แบบรวมองค์ประกอบ มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ คือ -3.40, -0.74, 1.18 และ 2.96 ตามลำดับ และแบบผสมผสาน มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ คือ -2.90, -0.65, 0.98 และ 2.56 ตามลำดับ เมื่อพิจารณาเฉพาะคะแนนสอบวันที่ 2 แบบแยกองค์ประกอบ มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ คือ -2.95, -0.44, 1.03 และ 2.63 ตามลำดับ แบบรวมองค์ประกอบ มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ คือ -3.81, -1.67, 1.34 และ 3.54 ตามลำดับ และแบบผสมผสาน มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ คือ -2.94, -0.68, 1.08 และ 2.54 ตามลำดับ เมื่อพิจารณารวมทั้งวันที่ 1 และวันที่ 2 แบบแยกองค์ประกอบ มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ คือ -2.71, -0.45, 0.92 และ 2.24 ตามลำดับ แบบรวมองค์ประกอบ มีความยากในชั้นที่ 1 – 4 เรียงลำดับกัน ดังนี้ คือ -3.40, -0.74, 1.18 และ 2.96 ตามลำดับ และแบบผสมผสาน มีความยากในชั้นที่ 1 – 4 เรียงลำดับกันดังนี้ คือ -2.90, -0.65, 0.98 และ 2.56 ตามลำดับ แสดงว่า มีความเที่ยงตรงในการแบ่งชั้น และข้อมูลเชิงประจักษ์สอดคล้องกับโมเดล Partial Credit Model

ตาราง 21 แสดงค่าพารามิเตอร์ของผู้ตรวจให้คะแนน จำแนกตามคุณลักษณะผู้ตรวจให้คะแนน

	คุณลักษณะ	Logit	SE	Max.	Min.	Infit	Outfit	Sep.R	
S/L	Ana	0.00	0.10	-2.29	2.21	1.00	1.00	.99	
	Hol	0.00	0.22	-2.31	2.54	1.00	1.01	.94	
	Ann	0.00	0.12	-2.08	2.29	1.00	1.01	.98	
	Ana ชาย	0.00	0.15	-2.35	2.43	1.00	1.00	.97	
	หญิง	0.00	0.16	-2.32	2.08	1.00	1.00	.97	
	รวม	0.00	0.10	-2.29	2.21	1.00	1.00	.99	
DRF	Hol ชาย	0.00	0.33	-2.79	3.86	1.01	1.02	.91	
	หญิง	0.00	0.30	-2.11	2.47	0.99	0.99	.89	
		รวม	0.00	0.22	-2.31	2.54	1.00	1.01	.94
	Ann ชาย	0.00	0.18	-2.33	2.60	1.00	1.01	.97	
	หญิง	0.00	0.16	-2.10	2.12	1.01	1.00	.97	
		รวม	0.00	0.12	-2.08	2.29	1.00	1.01	.98
Ana	วันที่ 1	0.00	0.10	-2.29	2.21	1.00	1.00	.99	
	วันที่ 2	0.00	0.10	-2.44	2.55	0.99	1.00	.99	
	รวม	0.00	0.07	-1.86	2.09	0.99	1.00	.99	
DRIFT	Hol วันที่ 1	0.00	0.22	-2.31	2.54	1.00	1.01	.94	
	วันที่ 2	0.00	0.23	-2.67	2.92	1.00	1.00	.96	
	รวม	0.00	0.15	-1.46	1.87	1.01	1.01	.96	
Ann	วันที่ 1	0.00	0.12	-2.08	2.29	1.00	1.01	.98	
	วันที่ 2	0.00	0.12	-2.15	2.60	1.00	1.01	.98	
	รวม	0.00	0.08	-1.93	2.30	1.00	1.01	.99	

จากตาราง 21 พบว่า พารามิเตอร์ของผู้ตรวจให้คะแนน (Rater Severity) จำแนกตามคุณลักษณะผู้ตรวจให้คะแนน ผลปรากฏ ดังนี้

21.1 ความเข้มงวด/ใจดี รูปแบบการตรวจให้คะแนนทั้ง 3 รูปแบบ มีค่าเฉลี่ยความเข้มงวดของผู้ตรวจให้คะแนนเท่ากันคือ 0.00 แบบรวมองค์ประกอบมีค่าเฉลี่ยความแปรปรวนสูงสุดคือ 0.22 รองลงมาคือ แบบผสมผสานและแบบแยกองค์ประกอบ มีค่า 0.12 และ 0.10 ตามลำดับ แบบรวมองค์ประกอบมีผู้ตรวจให้คะแนนใจดีที่สุด มีค่า -2.31 รองลงมาคือ แบบแยกองค์ประกอบและแบบผสมผสาน มีค่า -2.29 และ -2.08 ตามลำดับ แบบรวมองค์ประกอบมีผู้ตรวจให้คะแนนเข้มงวดที่สุด มีค่า 2.54 รองลงมาคือ แบบผสมผสานและแบบแยกองค์ประกอบ มีค่า 2.29 และ 2.21 ตามลำดับ แสดงว่า ความเข้มงวดของผู้ตรวจมีการกระจายแตกต่างกัน และการเปรียบเทียบความเหมาะสมกลมกลืนของข้อมูลเชิงประจักษ์กับโมเดล Partial Credit Mode ปรากฏว่า มีค่าเฉลี่ย Infit และ Outfit อยู่ในช่วงที่ยอมรับได้ และค่าความเชื่อมั่นแยกส่วนของผู้ตรวจให้คะแนน มีค่าอยู่ระหว่าง .94 – .99 แสดงว่า ผู้ตรวจให้คะแนน มีระดับความเข้มงวดที่แตกต่างกัน

21.2 ความลำเอียงด้านเพศ ทั้ง 3 รูปแบบ มีค่าเฉลี่ยความเข้มงวดของผู้ตรวจให้คะแนนเท่ากันคือ 0.00 แบบรวมองค์ประกอบมีค่าเฉลี่ยความแปรปรวนสูงสุดคือ 0.22 รองลงมาคือ แบบผสมผสานและแบบแยกองค์ประกอบ มีค่า 0.12 และ 0.10 ตามลำดับ แบบรวมองค์ประกอบมีผู้ตรวจให้คะแนนใจดีที่สุด มีค่า -2.31 รองลงมาคือ แบบแยกองค์ประกอบและแบบผสมผสาน มีค่า -2.29 และ -2.08 ตามลำดับ แบบรวมองค์ประกอบมีผู้ตรวจให้คะแนนเข้มงวดที่สุด มีค่า 2.54 รองลงมาคือ แบบผสมผสานและแบบแยกองค์ประกอบ มีค่า 2.29 และ 2.21 ตามลำดับ แสดงว่า ความเข้มงวดของผู้ตรวจมีการกระจายแตกต่างกัน และการเปรียบเทียบความเหมาะสมกลมกลืนของข้อมูลเชิงประจักษ์กับโมเดล Partial Credit Mode ปรากฏว่า มีค่าเฉลี่ย Infit และ Outfit อยู่ในช่วงที่ยอมรับได้ และค่าความเชื่อมั่นแยกส่วนของผู้ตรวจให้คะแนน มีค่าอยู่ระหว่าง .89 – .99 แสดงว่า ผู้ตรวจให้คะแนนมีระดับความเข้มงวดที่แตกต่างกัน

21.3 การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป ทั้ง 3 รูปแบบ มีค่าเฉลี่ยความเข้มงวดของผู้ตรวจให้คะแนนเท่ากัน คือ 0.00 แบบรวมองค์ประกอบมีค่าเฉลี่ยความแปรปรวนสูงสุดคือ 0.15 รองลงมาคือ แบบผสมผสานและแบบแยกองค์ประกอบ มีค่า 0.08 และ 0.07 ตามลำดับ แบบผสมผสานมีผู้ตรวจให้คะแนนใจดีที่สุด มีค่า -1.93 รองลงมาคือแบบแยกองค์ประกอบและแบบรวมองค์ประกอบ มีค่า -1.86 และ -1.46 ตามลำดับ แบบผสมผสานมีผู้ตรวจให้คะแนนเข้มงวดที่สุด มีค่า 2.30 รองลงมาคือ แบบแยกองค์ประกอบ และแบบรวมองค์ประกอบมีค่า 2.09 และ 1.87 ตามลำดับ แสดงว่า ความเข้มงวดของผู้ตรวจมีการกระจายแตกต่างกัน และการเปรียบเทียบความเหมาะสมกลมกลืนของข้อมูลเชิงประจักษ์กับโมเดล Partial Credit Mode ปรากฏว่า มีค่าเฉลี่ย Infit และ Outfit อยู่ในช่วงที่ยอมรับได้ และค่าความเชื่อมั่นแยกส่วนของผู้ตรวจให้คะแนน มีค่าอยู่ระหว่าง .94 – .99 แสดงว่า ผู้ตรวจให้คะแนนมีระดับความเข้มงวดที่แตกต่างกัน

ตอนที่ 2 การศึกษาคุณลักษณะของผู้ตรวจให้คะแนน

ในส่วนนี้เป็นการวิเคราะห์เพื่อตอบคำถามการวิจัยข้อที่ 2 ที่ว่า มีผู้ตรวจให้คะแนนที่มีคุณลักษณะเข้มงวด และใจดี หรือไม่ มีผู้ตรวจให้คะแนนที่มีคุณลักษณะลำเอียงเข้าข้างเพศชาย และลำเอียง เข้าข้างเพศหญิงหรือไม่ และมีผู้ตรวจให้คะแนนที่มีคุณลักษณะการตรวจให้คะแนนไม่คงที่ ข้ามเวลาหรือไม่ และเมื่อผู้ตรวจให้คะแนนใช้วิธีการตรวจให้คะแนนที่แตกต่างกัน 3 รูปแบบแล้วคุณลักษณะของผู้ตรวจจะแตกต่างกันหรือไม่ ผู้วิจัยแบ่งนำเสนอออกเป็น 4 ส่วนย่อย คือ 2.1 การระบุความเข้มงวด/ใจดี 2.2 การระบุความลำเอียงด้านเพศ 2.3 การระบุการทำหน้าที่เบี่ยงเบนของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป และ 2.4 หาค่าความสอดคล้องกันในการระบุคุณลักษณะของผู้ตรวจให้คะแนน ดังปรากฏในตาราง 22 – 26

2.1 การระบุความเข้มงวด/ใจดี ในขั้นแรกนั้น ต้องพิจารณาว่าผู้ตรวจให้คะแนนแต่ละคนให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล Partial Credit Model หรือไม่ โดยพิจารณา ค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนที่มีการถ่วงน้ำหนัก Weighted mean square residual: INFIT และค่าเฉลี่ยกำลังสองของความคลาดเคลื่อนที่ไม่มีการถ่วงน้ำหนัก Unweighted mean square residual : OUTFIT ที่กำหนดขอบเขตของการยอมรับได้ของโมเดลที่เหมาะสมกับข้อมูลอยู่ระหว่าง 0.5 ถึง 1.5 จากนั้นจึงระบุความเข้มงวด/ใจดี จากค่าศูนย์ (0) เครื่องหมายบวก (+) ลบ (-) ของค่า Logit ผู้ตรวจให้คะแนนที่เข้มงวดจะมีค่า Logit มากกว่า +1.00 ให้คะแนนเป็นกลาง $-1 < \text{Logit} < +1$ และให้คะแนนใจดีจะมีค่า Logit น้อยกว่า -1.00 รายละเอียดดูที่ภาคผนวก ข

ตาราง 22 แสดงเส้นภาพแสดงความเข้มงวด/ใจดีของผู้ตรวจให้คะแนน จากโปรแกรม FACETS

วิธีการตรวจ	1) แบบแยกองค์ประกอบ	2) แบบรวมองค์ประกอบ	3) แบบผสมผสาน
เข้มงวด	Measr)-Teacher	ESSAY	Measr)-Teacher
	ESSAY	ESSAY	ESSAY
	3+	+(5)	3+
	---	---	---
	3	3 92	72 83
	2+83	2+25	2+14
	40 72	46	3
	14	11 99	40
	42	101 72	115
	47	107 40	11 42
เป็นกลาง	1 48 62	17	15 17 66
	11 70	---	25 46 74 76 77
	1+15 25 39 52 76	1+14 21 34 42 70 76 83 85 90	1+101 39 5
	---	15 65	---
	111 115 46 77	115 126 23 33	108 126 16 47 48 79
	112 16 31 74 79	117 12 128 29 31 43 53 66 81 95	107 111 34 70 81
	109 117 126 43 60 80 93	48 5 86	38 86
	22 34 38 53 55 63 81	118 125 22 32 57 74 77 89 99	128 22 43 52 80
	107 51 68 75 95	124 38 44 50 80 9	105 12 26 31 51 89
	108 57 65 66 85 98	123 26 63 68	18 29 33 55 60 85 93 98
104 118 120 86 89 99	---	103 112 117 23 30 49 57 62 65 97 3	
ใจดี	* 0 * 102 105 12 128 33 37 97	* 0 * 1 103 104 105 52 64 75 97	* 0 * 119 124 129 32 6 63 90
	119 18 21 23 26 29 30 32 59 8 84 87	109 111 127 16 18 30 60 79 8 84 87 98	130 35 64 68
	101 103 5 6 64 67 71	10 120 54 55 6	118 123 36 50 53 67 78 8 87
	122 124 125 127 49 73	108 110 112 129 24 51 96	100 109 114 21 44 92
	129 130 17 41 50 92 99	116 19 35 37 59 62 91	125 37 59 75 96
	110 13 35 36 44 45 54 56	2 20 78	104 110 122 24 91
	100 121 19 20 82 96	122 13 73 82	120 54 71 73 95 99
	106 114 123 4 61 78 91	100 113 119 171 41 58 62	28 45 7 82 84
	---	---	---
	-1+7	-1+28 39	-1+10 58 9
28 88	114 130 27 36 49 56	102 121 13 4 56	
10 24	71	106 127 19 2 20 27 41 61 88	
2 27 69 9	45 7 88	69 94	
58	102 69	---	
113	47 61	---	
116	4	---	
---	94	---	
-2+	-2+	-2+113	
94	106	116	
---	---	---	
-3+	-3+	-3+	
Measr)-Teacher	ESSAY	Measr)-Teacher	
ESSAY	ESSAY	ESSAY	
+(1)	+(1)	+(1)	

หมายเหตุ เส้นประ--- แสดงเส้นแบ่งการระบุความเข้มงวด,ความเป็นกลาง และความใจดี

ตาราง 23 แสดงการระบุความเข้มงวด/ใจดีของผู้ตรวจให้คะแนน จำแนกตามวิธีการตรวจให้คะแนน

วิธีการ ตรวจ	ผู้ตรวจให้คะแนน			
	ไม่สอดคล้อง กับโมเดล	เข้มงวด	เป็นกลาง	ใจดี
Ana	10, 12, 36, 24, 27, 46, 55, 56, 85, 94, 100, 117, 125, 127, 129	1, 3, 11, 14, 15, 25, 39, 40, 42, 47, 48, 52, 62, 70, 72, 76, 83	4, 5, 6, 8, 13, 16, 17, 18, 19, 20, 21, 22, 23, 26, 29, 30, 31, 32, 33, 34, 35, 37, 38, 41, 43, 44, 45, 49, 50, 51, 53, 54, 57, 59, 60, 61, 63, 64, 65, 66, 67, 68, 71, 73, 74, 75, 77, 78, 79, 80, 81, 82, 84, 86, 87, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, 99, 101, 102, 103, 105, 106, 107, 108, 109, 110, 112, 114, 115, 118, 119, 120, 121, 122, 123, 124, 126, 128, 130	2, 7, 9, 28, 58, 69, 88, 113, 116
Hol	2, 3, 10, 11, 12, 29, 30, 40, 44, 60, 61, 70, 92, 117, 125, 127, 128, 130	14, 17, 25, 42, 46, 72, 83, 93, 101, 107	1, 5, 6, 8, 9, 13, 15, 16, 18, 19, 20, 21, 22, 23, 24, 26, 27, 31, 32, 33, 34, 35, 37, 38, 39, 41, 43, 48, 50, 51, 52, 53, 54, 55, 57, 58, 59, 62, 63, 64, 65, 66, 67, 68, 73, 74, 75, 76, 77, 78, 79, 80, 81, 82, 84, 85, 86, 87, 89, 90, 91, 95, 96, 97, 98, 99, 100, 103, 104, 105, 108, 109, 110, 111, 112, 113, 115, 116, 118, 119, 120, 121, 122, 123, 124, 126, 129	4, 7, 28, 36, 45, 47, 49, 56, 69, 71, 88, 94, 102, 106, 114
Ann	10, 14, 27, 30, 36, 43, 44, 46, 47, 48, 55, 56, 60, 91, 94, 119, 125, 129	1, 3, 11, 15, 17, 25, 39, 40, 42, 66, 72, 74, 76, 77, 83, 101, 115	5, 6, 7, 8, 9, 12, 16, 18, 21, 22, 23, 24, 26, 28, 29, 31, 32, 33, 34, 35, 37, 38, 45, 49, 50, 51, 52, 53, 54, 57, 59, 62, 63, 64, 65, 67, 68, 70, 71, 73, 75, 78, 79, 80, 81, 82, 84, 85, 86, 87, 89, 90, 92, 93, 95, 96, 97, 98, 99, 100, 103, 104, 105, 107, 108, 109, 110, 111, 112, 114, 117, 118, 120, 122, 123, 124, 126, 128, 130	2, 4, 13, 19, 20, 41, 58, 61, 69, 88, 102, 106, 113, 116, 121, 127

จากตาราง 22 – 23 พบว่า การระบุความเข้มงวด/ใจดีจำแนกตามวิธีการตรวจให้คะแนนได้ผล ดังนี้

1. แบบแยกองค์ประกอบ จากผู้ตรวจ 130 คน มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล จำนวน 115 คน คิดเป็นร้อยละ 88.46 ผู้ตรวจให้คะแนนมีความเข้มงวด 17 คน คิดเป็นร้อยละ 13.08 เป็นกลาง 89 คน คิดเป็นร้อยละ 68.46 และมีความใจดี 9 คน คิดเป็นร้อยละ 6.92

2. แบบรวมองค์ประกอบ จากผู้ตรวจ 130 คน มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล จำนวน 112 คน คิดเป็นร้อยละ 86.15 ผู้ตรวจให้คะแนนมีความเข้มงวด 10 คน คิดเป็นร้อยละ 7.69 เป็นกลาง 83 คน คิดเป็นร้อยละ 63.85 และมีความใจดี 15 คน คิดเป็นร้อยละ 11.54

3. แบบผสมผสาน จากผู้ตรวจ 130 คน มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล จำนวน 112 คน คิดเป็นร้อยละ 86.46 ผู้ตรวจให้คะแนนมีความเข้มงวด 17 คน คิดเป็นร้อยละ 13.08 เป็นกลาง 79 คน คิดเป็นร้อยละ 60.77 มีความใจดี 16 คน คิดเป็นร้อยละ 12.31

เมื่อพิจารณาโดยรวม วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ และแบบผสมผสาน ทำให้ผู้ตรวจให้คะแนนแสดงความเข้มงวดมากที่สุด จำนวน 17 คน แบบผสมผสาน ทำให้ผู้ตรวจให้คะแนนแสดงความใจดีมากที่สุด จำนวน 16 คน

2.2 การระบุความลำเอียงด้านเพศของผู้ตรวจให้คะแนน ในขั้นแรกนั้นต้องพิจารณาว่า ผู้ตรวจให้คะแนนแต่ละคนให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล Partial Credit Model หรือไม่โดยพิจารณาค่า INFIT และ OUTFIT ต่อมาจึงระบุความลำเอียงด้านเพศของผู้ตรวจให้คะแนน จากค่า Z-Score ที่มีนัยสำคัญทางสถิติ สุดท้ายจึงพิจารณาว่าผู้ตรวจให้คะแนนคนนั้นๆ ลำเอียงเข้าข้างเพศชาย หรือเพศหญิง คะแนนที่มีแนวโน้มใจดี (Logit ต่ำกว่า) กับเพศใดแสดงว่าลำเอียงเข้าข้างเพศนั้น รายละเอียดดูที่ภาคผนวก ข

ตาราง 24 แสดงการระบุความล่าเอียงด้านเพศ จำแนกตามวิธีการตรวจให้คะแนน

วิธีการ ตรวจ	ผู้ตรวจให้คะแนน			
	ไม่สอดคล้อง กับโมเดล	เข้าข้าง เพศชาย	เป็นกลาง	เข้าข้างเพศ หญิง
Ana	10, 12, 24,	2, 9, 26	1, 3, 4, 5, 6, 7, 8, 9, 11, 13, 15, 16, 17,	14, 20, 25,
	27, 36, 46,	,28, 42,	18, 21, 22, 23, 30, 31, 32, 33, 34, 35, 37,	29, 44, 45,
	55, 56, 85,	65, 120,	38, 39, 40, 41, 43, 47, 48, 49, 50, 51, 53,	52, 54, 66,
	94, 100,	122	57, 58, 59, 60, 61, 62, 63, 64, 67, 68, 69,	76, 101,
Hol	117, 125,		70, 71, 72, 73, 74, 75, 77, 78, 79, 80, 81,	104, 115,
	127, 129		82, 83, 84, 86, 87, 88, 89, 90, 91, 92, 93,	119, 123
			95, 96, 97, 98, 99, 102, 103, 105, 106,	
			107, 108, 109, 110, 111, 112, 113, 114,	
Ann			116, 118, 121, 124, 126, 128, 130	
	2, 3, 10, 11,	5, 7, 89,	1, 4, 6, 8, 9, 13, 14, 15, 16, 17, 18, 19,	25, 50, 53,
	12, 29, 30,	99, 120	20, 21, 22, 23, 24, 26, 27, 28, 31, 32, 33,	54, 59, 66,
	40, 44, 60,		34, 35, 36, 37, 38, 39, 41, 42, 43, 45, 46,	69, 115
Ann	61, 67, 70,		47, 48, 49, 51, 52, 55, 56, 57, 58, 62, 63,	
	92, 117,		64, 65, 68, 71, 72, 73, 74, 75, 76, 77, 78,	
	125, 126,		79, 80, 81, 82, 83, 84, 85, 86, 87, 88, 90,	
	127, 128,		91, 93, 94, 95, 96, 97, 98, 100, 101, 102,	
Ann	130		103, 104, 105, 106, 107, 108, 109, 110,	
			111, 112, 113, 114, 116, 118, 119, 121,	
			122, 123, 124, 129	
Ann	10, 14, 27,	4, 7, 20,	1, 3, 5, 6, 9, 11, 12, 13, 15, 16, 17, 19,	2,8, 18, 22,
	30, 36, 43,	33, 39, 40,	21, 23, 24, 25, 26, 28, 29, 31, 32, 34, 35,	54, 66, 69,
	44, 46, 47,	58, 82, 92,	37, 38, 41, 42, 45, 49, 50, 52, 53, 57, 59,	76, 81, 103,
	48, 51, 55,	120	61, 62, 63, 64, 65, 67, 68, 70, 71, 72, 73,	115
Ann	56, 60, 91,		74, 75, 77, 78, 79, 80, 83, 84, 85, 86, 87,	
	119, 125,		88, 89, 90, 93, 94, 95, 96, 97, 98, 99, 100,	
	129		101, 102, 104, 105, 106, 107, 108, 109,	
			110, 111, 112, 113, 114, 116, 117, 118,	
		121, 122, 123, 124, 126, 127, 128, 130		

จากตาราง 24 พบว่า การระบุความลำเอียงด้านเพศ จำแนกตามวิธีการตรวจให้คะแนนได้ผล ดังนี้

1. แบบแยกองค์ประกอบ จากผู้ตรวจ 130 คน มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล 115 คน คิดเป็นร้อยละ 88.46 มีผู้ตรวจให้คะแนนลำเอียงเข้าข้างเพศชาย 8 คน คิดเป็นร้อยละ 6.15 ไม่ลำเอียงด้านเพศ 92 คน คิดเป็นร้อยละ 70.77 และเข้าข้างเพศหญิง 15 คน คิดเป็นร้อยละ 11.54

2. แบบรวมองค์ประกอบ จากผู้ตรวจ 130 คน มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล 110 คน คิดเป็นร้อยละ 84.62 มีผู้ตรวจให้คะแนนลำเอียงเข้าข้างเพศชาย 5 คน คิดเป็นร้อยละ 3.85 ไม่ลำเอียงด้านเพศ 97 คน คิดเป็นร้อยละ 74.62 และเข้าข้างเพศหญิง 8 คน คิดเป็นร้อยละ 6.15

3. แบบผสมผสาน จากผู้ตรวจ 130 คน มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล 112 คน คิดเป็นร้อยละ 86.15 มีผู้ตรวจให้คะแนนลำเอียงเข้าข้างเพศชาย 10 คน คิดเป็นร้อยละ 7.69 ไม่ลำเอียงด้านเพศ 91 คน คิดเป็นร้อยละ 70.00 และเข้าข้างเพศหญิง 11 คน คิดเป็นร้อยละ 8.46

เมื่อพิจารณาโดยรวมพบว่า วิธีการตรวจให้คะแนนแบบผสมผสานทำให้ผู้ตรวจให้คะแนนแสดงความลำเอียงเข้าข้างเพศชายมากที่สุด จำนวน 10 คน แบบแยกองค์ประกอบทำให้ผู้ตรวจให้คะแนนแสดงความลำเอียงเข้าข้างเพศหญิงมากที่สุด จำนวน 15 คน

2.3 การระบุการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป ในขั้นแรกนั้น ต้องพิจารณาว่าผู้ตรวจให้คะแนนแต่ละคนให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล Partial Credit Model หรือไม่โดยพิจารณาค่า INFIT และ OUTFIT ต่อมาจึงระบุการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป จากค่า Z-Score โดยที่ค่า Z-Score ที่มีนัยสำคัญทางสถิติแสดงว่าผู้ตรวจให้คะแนนให้คะแนนไม่คงที่ข้ามเวลา ถ้า Z-Score ไม่มีนัยสำคัญทางสถิติแสดงว่า ผู้ตรวจให้คะแนนคงที่ข้ามเวลา รายละเอียดดูที่ภาคผนวก ข

ตาราง 25 แสดงการระบุการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป จำแนกตามวิธีการตรวจให้คะแนน

วิธีการตรวจ	ไม่สอดคล้องกับโมเดล	ผู้ตรวจที่ให้คะแนนไม่คงที่ข้ามเวลา	ผู้ตรวจที่ให้คะแนนคงที่ข้ามเวลา
Ana	3, 4, 5, 10, 11, 12, 27, 30, 46, 94, 99, 100, 115, 127	1, 2, 6, 7, 8, 9, 13, 14, 15, 16, 17, 19, 21, 22, 23, 25, 26, 32, 33, 34, 36, 37, 39, 40, 41, 43, 44, 45, 49, 50, 51, 52, 53, 55, 56, 57, 58, 59, 60, 61, 62, 63, 64, 65, 66, 67, 68, 70, 71, 72, 73, 74, 76, 77, 79, 80, 81, 83, 84, 85, 86, 87, 89, 90, 93, 96, 101, 103, 104, 105, 106, 108, 112, 113, 114, 116, 119, 120, 122, 123, 124, 129, 130	18, 20, 24, 28, 29, 31, 35, 38, 42, 47, 48, 54, 69, 75, 78, 82, 88, 91, 92, 95, 97, 98, 102, 107, 109, 110, 111, 117, 118, 121, 125, 126, 128
Hol	2, 3, 10, 12, 15, 27, 30, 43, 44, 64, 67, 69, 70, 73, 91, 92, 94, 115, 117, 127, 128	4, 6, 7, 11, 14, 16, 17, 20, 23, 26, 29, 33, 34, 36, 37, 40, 46, 53, 56, 57, 58, 60, 63, 66, 68, 72, 76, 87, 90, 98, 101, 102, 103, 105, 106, 107, 108, 113, 114, 122, 123	1, 5, 8, 9, 13, 18, 19, 21, 22, 24, 25, 28, 31, 32, 35, 38, 39, 41, 42, 45, 47, 48, 49, 50, 51, 52, 54, 55, 59, 61, 62, 65, 71, 74, 75, 77, 78, 79, 80, 81, 82, 83, 84, 85, 86, 88, 89, 93, 95, 96, 97, 99, 100, 104, 109, 110, 111, 112, 116, 118, 119, 120, 121, 124, 125, 126, 129, 130
Ann	3,10,14,22, 27,30,43, 44,46,48, 51,55,73, 91,94,97, 100,127	1, 2, 4, 5, 6, 7, 8, 9, 11, 12, 13, 15, 17, 18, 19, 21, 23, 24, 26, 31, 33, 34, 35, 37, 38, 39, 40, 47, 49, 50, 53, 57, 58, 59, 60, 63, 65, 67, 68, 70, 75, 76, 77, 78, 79, 80, 81, 83, 84, 85, 86, 87, 88, 90, 92, 93, 95, 99, 101, 102, 103, 104, 105, 106, 109, 112, 116, 118, 120, 122, 128	16, 20, 25, 28, 29, 32, 36, 41, 42, 45, 52, 54, 56, 61, 62, 64, 66, 69, 71, 72, 74, 82, 89, 96, 98, 107, 108, 110, 111, 113, 114, 115, 117, 119, 121, 123, 124, 125, 126, 129, 130

จากตาราง 25 พบว่า การระบุการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป จำแนกตามวิธีการตรวจให้คะแนนได้ผล ดังนี้

1. แบบแยกองค์ประกอบ จากผู้ตรวจ 130 คน มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล จำนวน 116 คน คิดเป็นร้อยละ 89.23 มีผู้ตรวจให้คะแนนไม่คงที่ข้ามเวลา 83 คน คิดเป็นร้อยละ 63.85 และผู้ตรวจให้คะแนนคงที่ข้ามเวลา 33 คน คิดเป็นร้อยละ 25.38

2. แบบรวมองค์ประกอบ จากผู้ตรวจ 130 คน มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล จำนวน 109 คน คิดเป็นร้อยละ 83.85 มีผู้ตรวจให้คะแนนไม่คงที่ข้ามเวลา 41 คน คิดเป็นร้อยละ 31.54 และผู้ตรวจให้คะแนนคงที่ข้ามเวลา 68 คน คิดเป็นร้อยละ 52.31

3. แบบผสมผสาน จากผู้ตรวจ 130 คน มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล จำนวน 112 คน คิดเป็นร้อยละ 86.15 มีผู้ตรวจให้คะแนนไม่คงที่ข้ามเวลา 71 คน คิดเป็นร้อยละ 54.62 และผู้ตรวจให้คะแนนคงที่ข้ามเวลา 41 คน คิดเป็นร้อยละ 31.54

เมื่อพิจารณาโดยรวม พบว่า วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ ทำให้ผู้ตรวจให้คะแนนแสดงการตรวจให้คะแนนไม่คงที่ข้ามเวลามากที่สุด จำนวน 85 คน

2.4 หาค่าความสอดคล้องกันในการระบุคุณลักษณะของผู้ตรวจให้คะแนน เพื่อศึกษาว่าเมื่อผู้ตรวจให้คะแนนใช้วิธีการตรวจให้คะแนนแตกต่างกัน 3 รูปแบบ แล้วจะมีคุณลักษณะในแต่ละแบบตรงกันหรือไม่ นั่นคือ คุณลักษณะของผู้ตรวจให้คะแนนคงที่ ข้ามวิธีการตรวจให้คะแนนหรือไม่ โดยทดสอบความแตกต่างจากการใช้ χ^2 จาก Friedman Test

ตาราง 26 ค่าความสอดคล้องกันในการระบุคุณลักษณะของผู้ตรวจให้คะแนน เมื่อใช้วิธีการตรวจให้คะแนนเรียงความแตกต่างกัน

วิธีการตรวจ	จำนวนผู้ตรวจที่ระบุตรงกัน	χ^2	p
S/L	61	4.000	.135
DRF	61	1.032	.597
DRIFT	32	27.875	.000

จากตาราง 26 พบว่า ค่าความสอดคล้องกันในการระบุคุณลักษณะของผู้ตรวจให้คะแนน เมื่อใช้วิธีการตรวจให้คะแนนเรียงความแตกต่างกัน 3 รูปแบบ ผลปรากฏ ดังนี้

26.1 การระบุความเข้มงวด/ใจดี มีผู้แสดงความเข้มงวด ทั้ง 3 รูปแบบ จำนวน 5 คน คิดเป็นร้อยละ 3.85 แสดงความใจดี ทั้ง 3 รูปแบบ จำนวน 2 คน คิดเป็นร้อยละ 1.54 แสดงความเป็นกลาง ทั้ง 3 รูปแบบ จำนวน 54 คน คิดเป็นร้อยละ 41.54 รวมแสดงความเข้มงวด/ใจดีตรงกันทั้งหมด จำนวน 61 คน คิดเป็นร้อยละ 46.92 เมื่อทดสอบความแตกต่าง พบว่า เมื่อใช้วิธีการตรวจให้คะแนนแตกต่างกัน

การระบุคุณลักษณะระบุความเข้มงวด/ใจดีไม่แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ($\chi^2_{0.1,1} = 4.000$) แสดงว่า การระบุคุณลักษณะระบุความเข้มงวด/ใจดีมีความสอดคล้องกัน

26.2 ความลำเอียงด้านเพศ มีผู้แสดงความลำเอียงเข้าข้างเพศชายตรงกัน 1 คน คิดเป็นร้อยละ 0.77 ไม่ลำเอียงด้านเพศตรงกัน 57 คน คิดเป็นร้อยละ 43.85 และลำเอียงเข้าข้างเพศหญิงตรงกัน 3 คน คิดเป็นร้อยละ 2.31 รวมแสดงความลำเอียงด้านเพศตรงกันทั้งหมด 61 คน คิดเป็นร้อยละ 46.92 เมื่อทดสอบความแตกต่าง พบว่า เมื่อใช้วิธีการตรวจให้คะแนนแตกต่างกัน การระบุคุณลักษณะความลำเอียงด้านเพศไม่แตกต่างกันอย่างไม่มีนัยสำคัญทางสถิติ ($\chi^2_{0.1,1} = 1.032$) แสดงว่า การระบุคุณลักษณะความลำเอียงด้านเพศมีความสอดคล้องกัน

26.3 การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป มีผู้ให้คะแนนไม่คงที่ทั้ง 3 รูปแบบ จำนวน 24 คน คิดเป็นร้อยละ 18.46 และคงที่ทั้ง 3 รูปแบบ จำนวน 8 คน คิดเป็นร้อยละ 6.15 รวมแสดงการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไปตรงกันทั้งหมด จำนวน 32 คน คิดเป็นร้อยละ 24.62 เมื่อทดสอบความแตกต่าง พบว่า เมื่อใช้วิธีการตรวจให้คะแนนแตกต่างกัน การระบุคุณลักษณะการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนน เมื่อเวลาผ่านไปแตกต่างกัน อย่างมีนัยสำคัญทางสถิติที่ระดับ .01 ($\chi^2_{0.1,1} = 27.875$) แสดงว่า การระบุคุณลักษณะการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไปไม่สอดคล้องกัน

ตอนที่ 3 การตรวจสอบความเที่ยงตรงของความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3

ในส่วนนี้ผู้วิจัยวิเคราะห์ข้อมูลเพื่อตอบคำถามการวิจัยข้อที่ 3 ที่ว่า วิธีการตรวจให้คะแนนที่ผู้วิจัยสร้างขึ้น เพื่อวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 มีความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย หรือไม่ ดังตาราง 27 และมีความเที่ยงตรงตามสภาพ ภายใต้การใช้วิธีการตรวจให้คะแนนแตกต่างกัน ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนนที่มีคุณลักษณะแตกต่างกัน หรือไม่ ดังตาราง 28 – 29

3.1 การวิเคราะห์หาความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย โดยผู้วิจัยนำคะแนนที่ได้จากการตรวจให้คะแนนโดยใช้วิธีการตรวจให้คะแนนเรียงความ 3 รูปแบบ (Methods) คือ แบบแยกองค์ประกอบ แบบรวมองค์ประกอบ และแบบผสมผสาน และคุณลักษณะผู้ตรวจให้คะแนน 3 คุณลักษณะ (Traits) คือ ความเข้มงวด/ใจดี ความลำเอียงด้านเพศ และการทำหน้าที่ต่างกัน เมื่อเวลาผ่านไป มาหาค่าความเที่ยงตรงเชิงเหมือนและค่าความเที่ยงตรงเชิงจำแนก

ตาราง 27 ความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย

วิธีการตรวจ	S/L			DRF			DRIFT			
	Ana	Hol	Ann	Ana	Hol	Ann	Ana	Hol	Ann	
S/L	Ana	1	.597**	.846**	.170	.101	.033	.366**	.200*	.279**
	Hol		1	.665**	.214*	.142	.140	.199*	.481**	.252**
	Ann			1	.225**	.072	.111	.233**	.219*	.433**
DRF	Ana				1	.525**	.603**	.065	.004	.059
	Hol					1	.590**	.157	.039	.086
	Ann						1	.165	.001	.144
DRIFT	Ana							1	.604**	.714**
	Hol								1	.646**
	Ann									1

จากตาราง 27 พบว่า ความเที่ยงตรงเชิงเหมือน ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการวัดคุณลักษณะเดียวกันด้วยวิธีการตรวจให้คะแนนต่างกัน ความเข้มงวด/ใจดี ค่าสัมประสิทธิ์สหสัมพันธ์มีค่าอยู่ระหว่าง .597** - .846** ความลำเอียงด้านเพศ ค่าสัมประสิทธิ์สหสัมพันธ์ มีค่าอยู่ระหว่าง .525** - .603** และการทำหน้าที่ต่างกันเมื่อเวลาผ่านไป ค่าสัมประสิทธิ์สหสัมพันธ์ มีค่าอยู่ระหว่าง .604** - .714** จะเห็นได้ว่าค่าสหสัมพันธ์มีค่าอยู่ระหว่าง 0.525** ถึง 0.846** แสดงว่า มีความเที่ยงตรงเชิงเหมือนสูง

ความเที่ยงตรงเชิงจำแนก ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการวัดคุณลักษณะต่างกันด้วยวิธีการตรวจให้คะแนนต่างกัน ค่าสหสัมพันธ์มีค่า .001 - .279** หรือค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการวัดคุณลักษณะต่างกันด้วยวิธีการตรวจให้คะแนนเดียวกัน ค่าสหสัมพันธ์มีค่า .0039 - .481** จะเห็นได้ว่า ค่าสหสัมพันธ์มีค่าอยู่ระหว่าง 0.481** ถึง 0.001 จะเห็นได้ว่า มีความเที่ยงตรงเชิงจำแนกต่ำกว่าความเที่ยงตรงเชิงเหมือน แสดงว่า มีความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย

3.2 การวิเคราะห์หาความเที่ยงตรงตามสภาพ ผู้วิจัยสนใจว่าเงื่อนไขใดที่มีความเที่ยงตรงตามสภาพสูงสุด ภายใต้การใช้วิธีการตรวจให้คะแนนแตกต่างกัน ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนนที่มีคุณลักษณะแตกต่างกัน โดยใช้สูตรสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน หาค่าสหสัมพันธ์ r จากนั้นแปลงสหสัมพันธ์ r แบบ Fisher' Z เป็น z' แล้วจึงเปรียบเทียบว่าเงื่อนไขใดมีความเที่ยงตรงสูงสุด โดยแบ่งการวิเคราะห์ออกเป็น 2 ส่วนย่อย คือ เมื่อผู้ตรวจมีคุณลักษณะลำเอียง และเมื่อผู้ตรวจมีคุณลักษณะเป็นกลาง

ตาราง 28 แสดงค่าสหสัมพันธ์และค่าสหสัมพันธ์แปลงของความเที่ยงตรงตามสภาพ เมื่อผู้ตรวจมี
คุณลักษณะลำเอียง

วิธีการตรวจ	คุณลักษณะ	สัดส่วนจำนวนผู้ตรวจให้คะแนน	r	z'	
Ana	S/L	เข้มงวด 0 คน : ใจดี 2 คน	0.744**	0.962	
		เข้มงวด 1 คน : ใจดี 1 คน	0.781**	1.045	
		เข้มงวด 2 คน : ใจดี 0 คน	0.789**	1.071	
		ค่าเฉลี่ย S/L 2 คน	0.770**	1.026	
		เข้มงวด 0 คน : ใจดี 3 คน	0.795**	1.085	
		เข้มงวด 1 คน : ใจดี 2 คน	0.745**	0.962	
		เข้มงวด 2 คน : ใจดี 1 คน	0.798**	1.099	
		เข้มงวด 3 คน : ใจดี 0 คน	0.775**	0.984	
		ค่าเฉลี่ย S/L 3 คน	0.775**	1.033	
		เพศชาย 0 คน : เพศหญิง 2 คน	0.761**	0.996	
	DRF	เพศชาย 1 คน : เพศหญิง 1 คน	0.747**	0.962	
		เพศชาย 2 คน : เพศหญิง 0 คน	0.533**	0.597	
		ค่าเฉลี่ย DRF 2 คน	0.695**	0.852	
		เพศชาย 0 คน : เพศหญิง 3 คน	0.821**	1.157	
		เพศชาย 1 คน : เพศหญิง 2 คน	0.824**	1.172	
		เพศชาย 2 คน : เพศหญิง 1 คน	0.814**	1.142	
		เพศชาย 3 คน : เพศหญิง 0 คน	0.717**	0.897	
		ค่าเฉลี่ย DRF 3 คน	0.800**	1.092	
		DRIFT	คงตัว 0 คน : ไม่คงตัว 2 คน	0.752**	0.973
			คงตัว 0 คน : ไม่คงตัว 3 คน	0.774**	1.033
Hol	S/L	เข้มงวด 0 คน : ใจดี 2 คน	0.574**	0.655	
		เข้มงวด 1 คน : ใจดี 1 คน	0.524**	0.583	
		เข้มงวด 2 คน : ใจดี 0 คน	0.564**	0.640	
		ค่าเฉลี่ย S/L 2 คน	0.555**	0.626	
		เข้มงวด 0 คน : ใจดี 3 คน	0.722**	0.908	
		เข้มงวด 1 คน : ใจดี 2 คน	0.690**	0.848	

ตาราง 28 (ต่อ)

วิธีการตรวจ	คุณลักษณะ	สัดส่วนจำนวนผู้ตรวจให้คะแนน	r	z'
Hol	S/L	เข้ม้งวด 2 คน : ใจดี 1 คน	0.567**	0.640
		เข้ม้งวด 3 คน : ใจดี 0 คน	0.569**	0.648
		ค่าเฉลี่ย S/L 3 คน	0.640**	0.761
	DRF	เพศชาย 0 คน : เพศหญิง 2 คน	0.688**	0.848
		เพศชาย 1 คน : เพศหญิง 1 คน	0.752**	0.973
		เพศชาย 2 คน : เพศหญิง 0 คน	0.563**	0.540
		ค่าเฉลี่ย DRF 2 คน	0.655**	0.787
		เพศชาย 0 คน : เพศหญิง 3 คน	0.807**	1.113
		เพศชาย 1 คน : เพศหญิง 2 คน	0.681**	0.829
		เพศชาย 2 คน : เพศหญิง 1 คน	0.619**	0.725
		เพศชาย 3 คน : เพศหญิง 0 คน	0.591**	0.678
		ค่าเฉลี่ย DRF 3 คน	0.685**	0.836
		DRIFT	คงตัว 0 คน : ไม่คงตัว 2 คน	0.568**
	คงตัว 0 คน : ไม่คงตัว 3 คน		0.721**	0.908
	เข้ม้งวด 0 คน : ใจดี 2 คน		0.660**	0.793
	เข้ม้งวด 1 คน : ใจดี 1 คน		0.761**	0.996
	เข้ม้งวด 2 คน : ใจดี 0 คน		0.528**	0.590
	ค่าเฉลี่ย S/L 2 คน		0.660**	0.793
	S/L		เข้ม้งวด 0 คน : ใจดี 3 คน	0.636**
		เข้ม้งวด 1 คน : ใจดี 2 คน	0.657**	0.784
เข้ม้งวด 2 คน : ใจดี 1 คน		0.649**	0.775	
เข้ม้งวด 3 คน : ใจดี 0 คน		0.651**	0.775	
ค่าเฉลี่ย S/L 3 คน		0.650**	0.771	
DRF		เพศชาย 0 คน : เพศหญิง 2 คน	0.528**	0.590
		เพศชาย 1 คน : เพศหญิง 1 คน	0.639**	0.758
		เพศชาย 2 คน : เพศหญิง 0 คน	0.577**	0.655
		ค่าเฉลี่ย DRF 2 คน	0.575**	0.668
Ann				

ตาราง 28 (ต่อ)

วิธีการตรวจ	คุณลักษณะ	สัดส่วนจำนวนผู้ตรวจให้คะแนน	r	z'
	DRF	เพศชาย 0 คน : เพศหญิง 3 คน	0.623**	0.725
		เพศชาย 1 คน : เพศหญิง 2 คน	0.595**	0.685
		เพศชาย 2 คน : เพศหญิง 1 คน	0.682**	0.829
		เพศชาย 3 คน : เพศหญิง 0 คน	0.463**	0.504
		ค่าเฉลี่ย DRF 3 คน	0.595	0.686
Ann	DRIFT	คงตัว 0 คน : ไม่คงตัว 2 คน	0.714**	0.897
		คงตัว 0 คน : ไม่คงตัว 3 คน	0.711**	0.887

จากตาราง 28 พบว่า การใช้วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนจำนวนผู้ตรวจให้คะแนน 3 คน โดยผู้ตรวจให้คะแนนมีความลำเอียงด้านเพศ เข้าข้างเพศชาย 1 คน และลำเอียงเข้าข้างเพศหญิง 2 คน มีความความเที่ยงตรงตามสภาพสูงที่สุด ($r = 0.824$, $z' = 1.172$) รองลงมาคือ การใช้วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนจำนวนผู้ตรวจให้คะแนน 3 คน โดยผู้ตรวจให้คะแนนมีความลำเอียงด้านเพศ เข้าข้างเพศชาย 0 คน และลำเอียงเข้าข้างเพศหญิง 3 คน ($r = 0.821$, $z' = 1.157$) และการใช้วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนจำนวนผู้ตรวจให้คะแนน 3 คน โดยผู้ตรวจให้คะแนนมีความลำเอียงด้านเพศ เข้าข้างเพศชาย 2 คน และลำเอียงเข้าข้างเพศหญิง 1 คน ($r = 0.814$, $z' = 1.142$) ตามลำดับ และการใช้วิธีการตรวจให้คะแนนแบบรวมองค์ประกอบ ในสัดส่วนจำนวนผู้ตรวจให้คะแนน 2 คน โดยผู้ตรวจให้คะแนนมีความเข้มงวด 1 คน และใจดี 1 คน มีความความเที่ยงตรงตามสภาพต่ำที่สุด ($r = 0.524$, $z' = 0.583$)

ตาราง 29 แสดงค่าสหสัมพันธ์และค่าสหสัมพันธ์แปลงของความเที่ยงตรงตามสภาพ เมื่อผู้ตรวจมี
คุณลักษณะเป็นกลาง

วิธีการตรวจ ให้คะแนน	คุณลักษณะ ผู้ตรวจให้คะแนน	สัดส่วนจำนวนผู้ตรวจให้คะแนน	r	z'
Ana	S/L	กรณีผู้ตรวจให้คะแนน 2 คน	0.752**	0.973
		กรณีผู้ตรวจให้คะแนน 3 คน	0.861**	1.293
	DRF	กรณีผู้ตรวจให้คะแนน 2 คน	0.678**	0.829
		กรณีผู้ตรวจให้คะแนน 3 คน	0.759**	0.996
	DRIFT	กรณีผู้ตรวจให้คะแนน 2 คน	0.718**	0.908
		กรณีผู้ตรวจให้คะแนน 3 คน	0.795**	1.085
Hol	S/L	กรณีผู้ตรวจให้คะแนน 2 คน	0.692**	0.848
		กรณีผู้ตรวจให้คะแนน 3 คน	0.734**	0.940
	DRF	กรณีผู้ตรวจให้คะแนน 2 คน	0.620**	0.725
		กรณีผู้ตรวจให้คะแนน 3 คน	0.657**	0.784
	DRIFT	กรณีผู้ตรวจให้คะแนน 2 คน	0.473**	0.517
		กรณีผู้ตรวจให้คะแนน 3 คน	0.675**	0.820
Ann	S/L	กรณีผู้ตรวจให้คะแนน 2 คน	0.522**	0.576
		กรณีผู้ตรวจให้คะแนน 3 คน	0.624**	0.733
	DRF	กรณีผู้ตรวจให้คะแนน 2 คน	0.703**	0.877
		กรณีผู้ตรวจให้คะแนน 3 คน	0.624**	0.733
	DRIFT	กรณีผู้ตรวจให้คะแนน 2 คน	0.690**	0.848
		กรณีผู้ตรวจให้คะแนน 3 คน	0.657**	0.784

จากตาราง 29 พบว่า การใช้วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนจำนวนผู้ตรวจให้คะแนน 3 คน โดยผู้ตรวจให้คะแนนเป็นกลาง ไม่เข้มงวด/ใจดี มีความความเที่ยงตรงตามสภาพสูงที่สุด ($r = 0.861$, $z' = 1.293$) รองลงมาคือ การใช้วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนจำนวนผู้ตรวจให้คะแนน 3 คน โดยผู้ตรวจให้คะแนนเป็นกลาง ไม่มีความไม่คงที่ข้ามเวลา ($r = 0.795$, $z' = 1.085$) และการใช้วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนจำนวนผู้ตรวจให้คะแนน 2 คน โดยผู้ตรวจให้คะแนนเป็นกลาง ไม่เข้มงวด/ใจดี ($r = 0.752$, $z' = 0.973$) ตามลำดับ และการใช้วิธีการตรวจให้คะแนนแบบรวมองค์ประกอบ ในสัดส่วนจำนวนผู้ตรวจให้คะแนน 2 คน โดยผู้ตรวจให้คะแนนโดยผู้ตรวจให้คะแนนเป็นกลาง ไม่มีความไม่คงที่ข้ามเวลา มีความความเที่ยงตรงตามสภาพต่ำที่สุด ($r = 0.473$, $z' = 0.517$)

บทที่ 5

สรุป อภิปรายผล และข้อเสนอแนะ

การศึกษาวิจัยในครั้งนี้เป็นการศึกษาคุณภาพของเครื่องมือวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 เพื่อนำไปศึกษาคุณลักษณะของผู้ตรวจให้คะแนน คือ ความเข้มงวด/ใจดี ความลำเอียงด้านเพศ และการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป และจากนั้นตรวจสอบความเที่ยงตรงของการวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 โดยมีความมุ่งหมายในการวิจัย สมมติฐานและวิธีดำเนินการวิจัยดังนี้ คือ

ความมุ่งหมายของการวิจัย

การวิจัยครั้งนี้มีความมุ่งหมายสำคัญเพื่อ

1. สร้างและพัฒนาคุณภาพของเครื่องมือวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3
2. เพื่อศึกษาคุณลักษณะของผู้ตรวจให้คะแนน คือ ความเข้มงวด/ใจดี ความลำเอียงด้านเพศ และการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป
3. เพื่อตรวจสอบความเที่ยงตรงของการวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3

คำถามของการวิจัย

เพื่อให้การดำเนินงานศึกษาวิจัยครั้งนี้บรรลุตามความมุ่งหมายข้างต้น ผู้วิจัยจึงได้ทำการรวบรวมและวิเคราะห์ข้อมูลเพื่อตอบคำถามการวิจัยต่อไปนี้

1. เครื่องมือวัดความสามารถในการเขียนของนักเรียน ชั้นประถมศึกษาปีที่ 3 คือ วิธีการให้คะแนนการเขียนเรียงความ และข้อสอบการเขียนจดหมายและวิธีการตรวจให้คะแนนการเขียนจดหมายที่ผู้วิจัยสร้างขึ้นมีคุณภาพหรือไม่ อย่างไร
2. การระบุคุณลักษณะของผู้ตรวจให้คะแนน จากการวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 ในครั้งนี้ จะพบผู้ตรวจให้คะแนนที่มีคุณลักษณะต่อไปนี้หรือไม่
 - 2.1 มีผู้ตรวจให้คะแนนที่มีคุณลักษณะเข้มงวด และใจดี หรือไม่ อย่างไร
 - 2.2 มีผู้ตรวจให้คะแนนที่มีคุณลักษณะลำเอียงเข้าข้างเพศชาย และลำเอียงเข้าข้างเพศหญิงหรือไม่ อย่างไร
 - 2.3 มีผู้ตรวจให้คะแนนที่มีคุณลักษณะการให้คะแนนไม่คงที่เมื่อเวลาผ่านไปหรือไม่ อย่างไร
 - 2.4 ผู้ตรวจคนเดียวกันจะมีคุณลักษณะแตกต่างกันหรือไม่เมื่อใช้วิธีการตรวจแตกต่างกัน

3. การวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 จากการวิจัยครั้งนี้ มีหลักฐานแสดงความเที่ยงตรงในด้านต่าง ๆ ต่อไปนี้หรือไม่

3.1 มีความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย คือ ความเที่ยงตรง เชิงเหมือนและมีความเที่ยงตรงเชิงจำแนก หรือไม่ อย่างไร

3.2 มีความเที่ยงตรงตามสภาพภายใต้การใช้วิธีการตรวจให้คะแนนแตกต่างกัน ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนนที่มีคุณลักษณะแตกต่างกัน หรือไม่ (เงื่อนไขใดที่มีความเที่ยงตรงตามสภาพสูงสุด)

สมมติฐานของการวิจัย

เพื่อให้การดำเนินงานศึกษาวิจัยครั้งนี้บรรลุตามความมุ่งหมายข้างต้น ผู้วิจัยจึงได้ทำการรวบรวมและวิเคราะห์ข้อมูลเพื่อหาคำตอบในสมมติฐาน ดังต่อไปนี้

1. การระบุคุณลักษณะของผู้ตรวจให้คะแนน

1.1 มีผู้ตรวจให้คะแนนที่มีคุณลักษณะเข้มงวด และใจดี โดยมีค่า Logit สูงกว่าเกณฑ์ที่ตั้งไว้

1.2 มีผู้ตรวจให้คะแนนที่มีคุณลักษณะลำเอียงเข้าข้างเพศชาย และลำเอียงเข้าข้างเพศหญิง อย่างมีนัยสำคัญทางสถิติ

1.3 มีผู้ตรวจให้คะแนนที่มีคุณลักษณะการให้คะแนนไม่คงที่ ข้ามเวลา อย่างมีนัยสำคัญทางสถิติ

1.4 ผู้ตรวจคนเดียวกันมีคุณลักษณะที่ไม่แตกต่างกันเมื่อใช้วิธีการตรวจแตกต่างกัน อย่างมีนัยสำคัญทางสถิติ

2. ความเที่ยงตรงของการวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3

2.1 มีความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย โดยที่ความเที่ยงตรงเชิงเหมือน คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการวัดคุณลักษณะเดียวกันด้วยวิธีการตรวจให้คะแนนต่างกัน มีค่าสูง อย่างมีนัยสำคัญทางสถิติ และความเที่ยงตรงเชิงจำแนก คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการวัดคุณลักษณะต่างกันด้วยวิธีการตรวจให้คะแนนต่างกัน หรือระหว่างการวัดคุณลักษณะต่างกันด้วยวิธีการตรวจให้คะแนนเดียวกัน มีค่าต่ำ

2.2 มีความเที่ยงตรงตามสภาพภายใต้การใช้วิธีการตรวจให้คะแนนแตกต่างกัน ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนนที่มีคุณลักษณะแตกต่างกัน โดยมีเงื่อนไขคือ 1) เมื่อผู้ตรวจมีคุณลักษณะลำเอียง การให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนน 2 คน ที่มีคุณลักษณะเข้มงวด 1 คน และใจดี 1 คน มีความเที่ยงตรงตามสภาพสูงสุด และ 2) เมื่อผู้ตรวจมีคุณลักษณะเป็นกลาง การให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนน 2 คน ที่มีคุณลักษณะเป็นกลางไม่เข้มงวด/ใจดี มีความเที่ยงตรงตามสภาพสูงสุด

การดำเนินการวิจัย

กลุ่มตัวอย่างที่ใช้ในการวิจัยมี 2 กลุ่ม คือ 1) นักเรียนชั้นประถมศึกษาปีที่ 3 ภาคการเรียนที่ 2 ปีการศึกษา 2553 ของโรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน สำนักงานเขตพื้นที่การศึกษาประถมศึกษาสุรินทร์ เขต 1 ที่ได้มาจากการสุ่มแบบหลายขั้นตอนจากนักเรียนที่เข้าสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐานเพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 ในวันที่ 23 กุมภาพันธ์ 2554 จำนวน 50 คน และ 2) นักศึกษาคณะศึกษาศาสตร์ ที่เรียนวิชาเอกภาษาไทย ชั้นปีที่ 4 ภาคการเรียนที่ 1 ปีการศึกษา 2555 ใน มหาวิทยาลัยราชภัฏสุรินทร์ บุรีรัมย์ และร้อยเอ็ด ที่ได้มาจากการสุ่มอย่างง่าย จำนวน 130 คน

เครื่องมือที่ใช้ในการวิจัยมีอยู่ 2 ส่วน 1) วิธีการตรวจให้คะแนนการเขียนเรียงความในส่วนของ การเขียน จากข้อสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 และ 2) ข้อสอบการเขียนจดหมายที่ผู้วิจัยสร้างขึ้นเอง โดยเกณฑ์การตรวจให้คะแนน 3 รูปแบบ ที่ผู้วิจัยสร้างขึ้น มีความเชื่อมั่นอยู่ระหว่าง .813 – .921 จากนั้นนำมาปรับปรุงแก้ไข เพื่อนำไปทดลองกับกลุ่มตัวอย่างตามลำดับขั้น ดังนี้

- นำข้อสอบการเขียนจดหมายที่ผู้วิจัยสร้างขึ้นไปทดสอบกับนักเรียนชั้นประถมศึกษาปีที่ 4 ปีการศึกษา 2554 ซึ่งเป็นนักเรียนที่เลื่อนชั้นจากนักเรียนชั้นประถมศึกษาปีที่ 3 กลุ่มเดียวกับกับกลุ่มตัวอย่างที่ใช้ในการวิจัยที่เข้าสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 แล้วตรวจให้คะแนนโดยครูประจำชั้น จากนั้นนำคะแนนที่ได้มาใช้เป็นคะแนนวิธีที่ได้จากการสอบการเขียนของนักเรียน เพื่อใช้เป็นส่วนหนึ่งในการหาค่าความเที่ยงตรงตามสภาพต่อไป

- นำกระดาษคำตอบในส่วนของ การเขียนที่ได้คัดสำเนาจากข้อสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 และคู่มือการตรวจให้คะแนนที่ผ่านการหาคุณภาพแล้ว มาคัดสำเนา จัดชุดตามจำนวนนักศึกษามหาวิทยาลัยราชภัฏ วิชาเอกภาษาไทย ชั้นปีที่ 4 ที่เป็นกลุ่มตัวอย่าง จำนวน 185 ชุด เพื่อนำไปทดลองครั้งที่ 1 ระหว่างวันที่ 8 – 15 มิถุนายน 2555 ในครั้งนี้ มีนักศึกษาเข้าร่วมการวิจัย จำนวน 140 คน ไม่มาเข้าร่วมการวิจัย จำนวน 45 คน

- จัดอบรมการตรวจให้คะแนนแก่นักศึกษาที่เป็นกลุ่มตัวอย่าง ตามวันและเวลาที่ได้นัดหมาย เริ่มจากชี้แจงให้นักศึกษาได้ทราบถึงความมุ่งหมายและความสำคัญของการวิจัย จากนั้นฝึกสอนนักศึกษาให้มีความรู้ความเข้าใจในการใช้วิธีการตรวจให้คะแนนเรียงความที่ถูกต้องทั้ง 3 รูปแบบ

- ทำการทดลองครั้งที่ 1 โดยแบ่งนักศึกษาที่เป็นกลุ่มทดลองออกเป็น 3 กลุ่ม ด้วยวิธีการสุ่มอย่างง่าย แล้วดำเนินการตามรูปแบบการทดลองแบบหมุนเวียนสมดุล จากนั้นนัดหมายวันและเวลาเพื่อทำการทดลองครั้งต่อไปในอีก 5 สัปดาห์

5. นำข้อมูลที่ได้จากการวิจัยครั้งที่ 1 มาหาค่าสถิติพื้นฐาน คือ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ความเบ้ ความโด่ง สัมประสิทธิ์แอลฟาของครอนบาค และดัชนีความสอดคล้องของผู้ตรวจให้คะแนน ตามทฤษฎีมาตรฐานเดิม หาค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ และค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ เพื่อแสดงคุณภาพของเครื่องมือตามทฤษฎีการสรุปอ้างอิง จากโปรแกรม EDUG 6.1-e ผู้วิจัยใช้การออกแบบเอกภพสำหรับสององค์ประกอบ แบบ Two-Facet Crossed Design โดยกำหนดให้ P คือ นักเรียน I คือ วิธีการตรวจให้คะแนน และ R คือ ผู้ตรวจให้คะแนน Design ที่ใช้สำหรับงานวิจัยนี้ คือ $p \times i \times r$ design ประกอบไปด้วย นักเรียน 50 คน วิธีการตรวจให้คะแนน 3 รูปแบบ และผู้ตรวจให้คะแนน 3 คน (ข้อสอบเรียงความใช้วิธีสุ่มอย่างง่ายในการคัดเลือกผู้ตรวจให้คะแนน ข้อสอบการเขียนจดหมาย ใช้ครู 3 คน ที่เป็นครูประจำชั้น) และหาค่าพารามิเตอร์ต่างๆ ตามทฤษฎีการตอบสนองข้อสอบ โดยในที่นี้จะวิเคราะห์โดยใช้ Partial Credit Model ซึ่งจะได้พารามิเตอร์ของนักเรียน พารามิเตอร์ของเรียงความ และพารามิเตอร์ของผู้ตรวจให้คะแนน จากโปรแกรม FACETS เพื่อใช้เป็นหลักฐานในการแสดงคุณภาพของเครื่องมือสำหรับการวิจัยตามความมุ่งหมายของการวิจัยข้อที่ 1

6. วิเคราะห์คุณลักษณะผู้ตรวจให้คะแนนแต่ละคนโดยใช้โปรแกรม FACETS ซึ่งจะได้คุณลักษณะของผู้ตรวจให้คะแนนในครั้งนี้จำนวน 2 คุณลักษณะ คือ ความเข้มงวด/ใจดีของผู้ตรวจให้คะแนน และความลำเอียงด้านเพศ ตามความมุ่งหมายของการวิจัยข้อที่ 2

7. นำกระดาษคำตอบและวิธีการตรวจให้คะแนนมาจัดชุดตามจำนวนของนักศึกษา วิชาเอกภาษาไทย ชั้นปีที่ 4 ที่เป็นกลุ่มตัวอย่าง จำนวน 140 ชุด เพื่อนำไปทดลองครั้งที่ 2 ระหว่างวันที่ 6 – 17 สิงหาคม 2555 ผู้วิจัยต้องการให้นักศึกษาลืมการให้คะแนนครั้งที่ 1 จึงทิ้งระยะในการเก็บข้อมูลครั้งที่ 2 ไว้ประมาณ 5 สัปดาห์ เนื่องจากการศึกษาของนักจิตวิทยาเกี่ยวกับการจำและการลืมของมนุษย์ พบว่าเราจะสามารถจำเรื่องราวที่ได้อ่านไปประมาณครึ่งหนึ่ง และจะลดลงไปอีกครึ่งหนึ่งของที่เหลือในทุกๆ 7 วัน จนในที่สุดจะนึกไม่ออกเลย (จตุรงค์ ลังกาพินธุ์. ออนไลน์: 1 – 2) ในครั้งนี้มีนักศึกษามาเข้าร่วมการวิจัย จำนวน 130 คน ไม่มาเข้าร่วมการวิจัย จำนวน 10 คน โดยกระทำซ้ำขั้นตอนดังข้อที่ 3

8. นำข้อมูลที่ได้จากการทดลองครั้งที่ 2 ที่มีลักษณะผู้ตรวจให้คะแนนตรวจ ให้คะแนนสมบูรณ์ ทั้ง 3 รูปแบบ จำนวน 130 คน มาวิเคราะห์คุณลักษณะผู้ตรวจให้คะแนนแต่ละคน ซึ่งจะได้คุณลักษณะการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไปของผู้ตรวจให้คะแนน

9. นำการระบุคุณลักษณะมาศึกษาความสอดคล้องของการระบุความเข้มงวด/ใจดี ความลำเอียงด้านเพศ และการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป เมื่อใช้วิธีการตรวจให้คะแนนเรียงความแตกต่างกัน 3 รูปแบบ โดยใช้ค่า χ^2 จาก Friedman Test

10. หาคความเที่ยงตรงของความสามารถในการวัดการเขียนเรียงความของนักเรียน ตามความมุ่งหมายของการวิจัยข้อที่ 3 โดยมีขั้นตอน ดังนี้

10.1 หาคความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย โดยผู้วิจัยนำค่า Logit และ Z-Score ที่ได้จากการตรวจให้คะแนนโดยใช้วิธีการตรวจให้คะแนนเรียงความ 3 รูปแบบ (Methods) คือ แบบแยกองค์ประกอบ แบบรวมองค์ประกอบ และแบบผสมผสาน และคุณลักษณะผู้ตรวจให้คะแนน

3 ชนิด (Traits) คือ ความขี้มวด/ใจดี ความลำเอียงด้านเพศ และการทำหน้าที่ที่ต่างกันเมื่อเวลาผ่านไป มาหาความสัมพันธ์โดยใช้สูตรสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน

10.2 หาความเที่ยงตรงตามสภาพ โดยผู้วิจัยนำคุณลักษณะของผู้ตรวจให้คะแนนที่ได้ จากข้อ 6 และข้อ 9 มาจัดชุดตามสัดส่วนโดยใช้วิธีการสุ่มอย่างง่าย แยกเป็น กรณีเมื่อผู้ตรวจมีคุณลักษณะ ลำเอียง และเมื่อผู้ตรวจมีคุณลักษณะเป็นกลาง นำค่าที่ได้มาเปรียบเทียบความเที่ยงตรงตามสภาพ โดยการหาความสัมพันธ์โดยใช้สูตรสัมประสิทธิ์สหสัมพันธ์แบบเพียร์สันจากนั้นแปลงค่า r ที่ได้เป็นพิชเชอร์ z'

สรุปผลการวิจัย

การศึกษาวิจัยครั้งนี้ สรุปผลการวิจัยได้ ดังนี้

1. เครื่องมือวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 ที่ผู้วิจัยสร้างขึ้น มีคุณภาพตามทฤษฎีมาตรฐานเดิม ทฤษฎีการสรุปอ้างอิง และทฤษฎีการตอบสนองข้อสอบ

1.1 คุณภาพข้อสอบตามทฤษฎีมาตรฐานเดิม

1.1.1 การเขียนเรียงความ ค่าความเชื่อมั่น มีค่าระหว่าง 0.980 ถึง 0.988 ค่าดัชนี สอดคล้องของผู้ตรวจให้คะแนน มีค่าระหว่าง 0.811 ถึง 0.823 แสดงให้เห็นว่าค่าความเชื่อมั่นและค่าดัชนี สอดคล้องของผู้ตรวจให้คะแนนมีค่าสูง

1.1.2 การเขียนจดหมาย ค่าความเชื่อมั่น มีค่าระหว่าง 0.827 ถึง 0.924 ตามลำดับ ส่วนค่าดัชนีสอดคล้องของผู้ตรวจให้คะแนน มีค่าระหว่าง 0.809 ถึง 0.813 แสดงให้เห็นว่า ค่าความเชื่อมั่น และค่าดัชนีสอดคล้องของผู้ตรวจให้คะแนนมีค่าสูง

1.2 คุณภาพข้อสอบตามทฤษฎีการสรุปอ้างอิง

1.2.1 การเขียนเรียงความ ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์มีค่า 0.76 และ ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์มีค่า 0.68 แสดงว่า การตรวจให้คะแนนการเขียนเรียงความ มีความเชื่อมั่นตามทฤษฎีการสรุปอ้างอิง

1.2.2 การเขียนจดหมาย ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมพัทธ์ มีค่า 0.73 และ ค่าสัมประสิทธิ์การสรุปอ้างอิงเชิงสัมบูรณ์ มีค่า 0.58 แสดงว่าวิธีการตรวจให้คะแนนการเขียนจดหมาย มีความเชื่อมั่นตามทฤษฎีการสรุปอ้างอิง

1.3 คุณภาพข้อสอบตามทฤษฎีการตอบสนองข้อสอบ

1.3.1 ค่าพารามิเตอร์ความสามารถของนักเรียน (Student Ability : θ) พบว่า ความขี้มวด/ใจดี ความลำเอียงด้านเพศ และการทำหน้าที่ที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป มีความเชื่อมั่นแยกส่วนของนักเรียนในทุกๆ วิธีการตรวจให้คะแนน ซึ่งมีค่าอยู่ระหว่าง .98 ถึง 1.00 แสดงว่า มีความแตกต่างระหว่างความสามารถของผู้สอบแสดงว่า มีความสามารถในการจำแนกความสามารถ ของผู้สอบแต่ละคนออกจากกัน

1.3.2 ค่าพารามิเตอร์ของเรียงความ (Step Difficulty : δ) ความขี้มวด/ใจดี ความลำเอียง ด้านเพศ และการทำหน้าที่ที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป ในทุกๆ วิธีการตรวจให้คะแนน

มีค่าความยากในการเปลี่ยนขั้นเรียงลำดับจากน้อยไปมาก และการทดสอบความสอดคล้องของโมเดลของการตรวจให้คะแนนเรียงความ พบว่า ค่าความยากในการเปลี่ยนขั้นของวิธีการตรวจให้คะแนนเรียงความ มีความเที่ยงตรงในการแบ่งขั้น แสดงว่า ข้อมูลเชิงประจักษ์สอดคล้องกับโมเดล Partial Credit Model

1.3.3 พารามิเตอร์ของผู้ตรวจให้คะแนน (Rater Severity) ความเข้มงวด/ใจดี ความลำเอียง ด้านเพศ การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป ในทุกๆ วิธีการตรวจให้คะแนน มีค่าเฉลี่ยความเข้มงวดของผู้ตรวจให้คะแนนเท่ากันคือ 0.00 ความเข้มงวดของผู้ตรวจ มีการกระจายแตกต่างกัน และค่าความเชื่อมั่นแยกส่วนของผู้ตรวจให้คะแนน มีค่าอยู่ระหว่าง .89 ถึง .99 แสดงว่า ผู้ตรวจมีความเข้มงวดในการตรวจให้คะแนนแตกต่างกัน

2. ผลการระบุคุณลักษณะของผู้ตรวจให้คะแนน จากการวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3

2.1 ความเข้มงวด/ใจดี 1) แบบแยกองค์ประกอบ มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล จำนวน 115 คน ผู้ตรวจให้คะแนนมีความเข้มงวด 17 คน เป็นกลาง 89 คน และมีความใจดี 9 คน 2) แบบรวมองค์ประกอบ มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล จำนวน 112 คน ผู้ตรวจให้คะแนนมีความเข้มงวด 10 คน เป็นกลาง 83 คน และมีความใจดี 15 คน 3) แบบผสมผสาน มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล จำนวน 112 คน ผู้ตรวจให้คะแนนมีความเข้มงวด 17 คน เป็นกลาง 79 คน มีความใจดี 16 คน

2.2 ความลำเอียงด้านเพศ 1) แบบแยกองค์ประกอบ มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล 115 คน มีผู้ตรวจให้คะแนนลำเอียงเข้าข้างเพศชาย 8 คน ไม่ลำเอียงด้านเพศ 92 คน และเข้าข้างเพศหญิง 15 คน 2) แบบรวมองค์ประกอบ มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล 110 คน มีผู้ตรวจให้คะแนนลำเอียงเข้าข้างเพศชาย 5 คน ไม่ลำเอียงด้านเพศ 97 คน และเข้าข้างเพศหญิง 8 คน 3) แบบผสมผสาน มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล 112 คน มีผู้ตรวจให้คะแนนลำเอียงเข้าข้างเพศชาย 10 คน ไม่ลำเอียงด้านเพศ 91 คน และเข้าข้างเพศหญิง 11 คน

2.3 การทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป 1) แบบแยกองค์ประกอบ มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล จำนวน 116 คน มีผู้ตรวจให้คะแนนไม่คงที่ข้ามเวลา 83 คน และผู้ตรวจให้คะแนนคงที่ข้ามเวลา 33 คน 2) แบบรวมองค์ประกอบ มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล จำนวน 109 คน มีผู้ตรวจให้คะแนนไม่คงที่ข้ามเวลา 41 คน และผู้ตรวจให้คะแนนคงที่ข้ามเวลา 68 คน 3) แบบผสมผสาน มีผู้ตรวจให้คะแนนเชิงประจักษ์เหมาะสมกลมกลืนกับโมเดล จำนวน 112 คน มีผู้ตรวจให้คะแนนไม่คงที่ข้ามเวลา 71 คน และผู้ตรวจให้คะแนนคงที่ข้ามเวลา 41 คน

2.4 ค่าความสอดคล้องกันในการระบุคุณลักษณะของผู้ตรวจให้คะแนน 1) การระบุคุณลักษณะระบุความเข้มงวด/ใจดีมีความสอดคล้องกัน ($\chi^2_{0.1,1} = 4.000, p > .01$) 2) การระบุคุณลักษณะความลำเอียงด้านเพศมีความสอดคล้องกัน ($\chi^2_{0.1,1} = 1.032, p > .01$) 3) การระบุคุณลักษณะการทำหน้าที่ต่างกัน

ของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไปไม่สอดคล้องกัน ($\chi^2_{0.1,1} = 27.875, p < .01$)

3. ผลการหาความเที่ยงตรงของการวัดความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3

3.1 มีความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย โดยมีความเที่ยงตรงเชิงเหมือนมีค่าสหสัมพันธ์กันสูง มีค่าอยู่ระหว่าง 0.525** ถึง 0.846** และค่าสหสัมพันธ์ของความเที่ยงตรงเชิงจำแนก มีค่าอยู่ระหว่าง 0.481** ถึง 0.001 ค่าความเที่ยงตรงเชิงเหมือนมีค่าสูงกว่าความเที่ยงตรงเชิงจำแนก

3.2 มีความเที่ยงตรงตามสภาพภายใต้การใช้วิธีการตรวจให้คะแนนแตกต่างกัน ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนนที่มีคุณลักษณะแตกต่างกัน 1) เมื่อผู้ตรวจมีคุณลักษณะลำเอียง การใช้วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนจำนวนผู้ตรวจให้คะแนน 3 คน ผู้ตรวจให้คะแนนมีความลำเอียงด้านเพศ เข้าข้างเพศชาย 1 คน และลำเอียงเข้าข้างเพศหญิง 2 คน มีความความเที่ยงตรงตามสภาพสูงที่สุด ($r = 0.824, z' = 1.172$) และ 2) เมื่อผู้ตรวจมีคุณลักษณะเป็นกลาง การใช้วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนจำนวนผู้ตรวจให้คะแนน 3 คน โดยผู้ตรวจให้คะแนนเป็นกลางไม่เข้มงวด/ใจดี มีความความเที่ยงตรงตามสภาพสูงที่สุด ($r = 0.861, z' = 1.293$)

อภิปรายผล

1. เกณฑ์การตรวจให้คะแนนการเขียนทั้ง 3 รูปแบบที่ผู้วิจัยสร้างและพัฒนาขึ้นคือ แบบแยกองค์ประกอบ แบบรวมองค์ประกอบ และแบบผสมผสาน มีคุณภาพที่เหมาะสมสำหรับการนำไปใช้ในการวัดความสามารถในการเขียนเรียงความ และความสามารถในการเขียนจดหมายกับนักเรียน ชั้นประถมศึกษาปีที่ 3 ของโรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน สำนักงานเขตพื้นที่การศึกษาประถมศึกษาสุรินทร์ ทั้งในทฤษฎีมาตรฐานเดิม ทฤษฎีการสรุปอ้างอิง และทฤษฎีการตอบสนองข้อสอบ ทั้งนี้เนื่องจากการกำหนดวิธีการตรวจให้คะแนนที่ผู้วิจัยสร้างขึ้น มีจำนวนหัวข้อในการตรวจให้คะแนนน้อย แต่ครอบคลุมความสามารถในการวัดความสามารถในการเขียน และสอดคล้องกับวุฒิภาวะของเด็กชั้นประถมศึกษาปีที่ 3 นอกจากนี้การตรวจให้คะแนนเรียงความนักเรียนแต่ละชั้นมีความยาวนานพอ ประมาณ 10 บรรทัด/นักเรียน 1 คน ดังนั้นผู้ตรวจให้คะแนนจึงสามารถตรวจให้คะแนนได้อย่างละเอียดโดยไม่ล้าจนเกินไป จึงทำให้คะแนนที่ได้มีความเชื่อมั่นสูงใกล้เคียงกัน

2. การศึกษาคุณลักษณะของผู้ตรวจให้คะแนน คือ ความเข้มงวด/ใจดี ความลำเอียงด้านเพศ และการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป

2.1 ความเข้มงวด/ใจดี ผู้ตรวจให้คะแนนที่มีความเข้มงวด มีจำนวนมากกว่าใจดี และค่อนข้างไม่เปลี่ยนแปลง เมื่อใช้วิธีการตรวจให้คะแนนแตกต่างกัน สอดคล้องกับสมมติฐานการวิจัยข้อที่ 1.1 ที่ว่า มีผู้ตรวจให้คะแนนที่มีคุณลักษณะเข้มงวด และใจดี โดยมีค่า Logit สูงกว่าเกณฑ์ที่ตั้งไว้ แต่ไม่สอดคล้องกับงานวิจัยของ วูฟ (Wolfe, 2004) ที่ได้ศึกษาการระบุผลกระทบจากผู้ตรวจให้คะแนนโดยใช้โมเดลตัวแปรแฝง ซึ่งพบว่า ผู้ตรวจให้คะแนนที่มีความใจดีมีจำนวนมากกว่าผู้ตรวจที่มีความเข้มงวด ทั้งนี้

อาจเนื่องจากนักศึกษา วิชาเอกภาษาไทย ชั้นปีที่ 4 ที่เป็นกลุ่มตัวอย่าง ขาดประสบการณ์ในการสอนในสภาพจริง และไม่เคยตรวจให้คะแนนเรียงความมาก่อน จึงมีความคาดหวังผลว่านักเรียนชั้นประถมศึกษาปีที่ 3 จะเขียนเรียงความได้ดีกว่า เป็นผลทำให้เกิดการให้คะแนนที่มีค่าเฉลี่ยต่ำกว่าที่ควร และมีผู้ตรวจให้คะแนนที่มีคุณลักษณะเข้มงวดเป็นจำนวนมาก สอดคล้องกับงานวิจัยของ อุษณีย์ บัวศิริพันธ์ (2543) ที่ได้ศึกษาเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจจำนวนผู้ตรวจ และประสบการณ์การตรวจแตกต่างกัน โดยกลุ่มแรกมีประสบการณ์การตรวจให้คะแนนต่ำกว่า 5 ปี เปรียบเทียบกับกลุ่มที่ 2 ที่มีประสบการณ์การตรวจสูงกว่า 5 ปี ซึ่งผลการวิจัยพบว่า ผู้ตรวจให้คะแนนที่มีประสบการณ์ต่างกันส่งผลให้คะแนนมีค่าแตกต่างกัน

2.2 ความลำเอียงด้านเพศ ผู้ตรวจให้คะแนนลำเอียงเข้าข้างเพศหญิงมากกว่าลำเอียงเข้าข้างเพศชาย และค่อนข้างไม่เปลี่ยนแปลงเมื่อใช้วิธีการตรวจให้คะแนนแตกต่างกัน และเมื่อพิจารณาเพศของผู้ตรวจที่ถูกระบุว่า มีความลำเอียงด้านเพศ พบว่า มีทั้งเพศชายและเพศหญิง ทั้งนี้อาจเนื่องจากภาพที่เป็นโจทย์ ซึ่งเป็นการเดินทางตลาดซึ่งอาจเป็นมิติรอง (Secondary dimension) ในการวัด (Shealy; & Stout. 1993: 163) นักเรียนหญิงอาจจะถนัดในการเลือกใช้คำ เนื่องจากการคุ้นเคยกับการเดินตลาดมากกว่านักเรียนชาย และนอกจากนี้การใช้ภาษาเพื่อการสื่อสารนักเรียนหญิงมักจะทำคะแนนได้ดีกว่านักเรียนชาย สอดคล้องกับงานวิจัยของ เอกเคส (Eckes. 2005) ที่ได้ศึกษาการวิเคราะห์รหัสแบบหลายองค์ประกอบ ในการตรวจสอบผลกระทบบจากผู้ตรวจให้คะแนนในการสอบภาษาเยอรมันในฐานะภาษาต่างประเทศ ซึ่งพบว่า ผู้ตรวจให้คะแนนส่วนใหญ่ให้คะแนนในส่วนของการพูดและการเขียนลำเอียงเข้าข้างเพศหญิง

2.3 การระบุการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป ผู้ตรวจให้คะแนนให้คะแนนไม่คงที่ข้ามเวลาเป็นจำนวนมาก และค่อนข้างเปลี่ยนแปลงเมื่อใช้วิธีการตรวจให้คะแนนแตกต่างกัน สอดคล้องกับสมมติฐานการวิจัยข้อที่ 1.3 ที่ว่า มีผู้ตรวจให้คะแนนที่มีคุณลักษณะการตรวจให้คะแนนไม่คงที่ข้ามเวลา อย่างมีนัยสำคัญทางสถิติ แต่ไม่สอดคล้องกับงานวิจัยของ เมฟอร์ด และ วูฟ (Myford; & Wolfe. 2009) ที่ได้ศึกษาการเฝ้าตรวจสอบการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนนเมื่อเวลาผ่านไป ที่พบว่า ผู้ตรวจให้คะแนนให้คะแนนไม่คงที่มีจำนวนน้อยมาก สาเหตุอาจมาจากกลุ่มตัวอย่างที่ใช้ในงานวิจัยของเมฟอร์ดและวูฟมีประสบการณ์การตรวจข้อสอบอย่างต่ำ 4 ปี จึงมีความคุ้นเคยกับวิธีการตรวจให้คะแนนมากกว่า ดังนั้นเมื่อต้องตรวจให้คะแนนเรียงความจำนวนมาก จึงมีความแม่นยำและคงเส้นคงวา ในการตรวจให้คะแนนมากกว่านักศึกษาปี 4 มหาวิทยาลัยราชภัฏที่มีประสบการณ์การตรวจ ให้คะแนนเรียงความเพียง 2 ครั้ง อันเป็นผลมาจากข้อจำกัดของการวิจัยที่ต้องการใช้กลุ่มตัวอย่างที่เป็นผู้ตรวจให้คะแนนจำนวนมาก ประกอบกับการทิ้งระยะการเก็บข้อมูลเป็นเวลานานกว่างานวิจัยของเมฟอร์ด และ วูฟ ซึ่งเป็นการเก็บข้อมูลโดยใช้องค์ประกอบของเวลา 8 ช่วง ภายใน 4 วัน ดังนั้นโอกาสที่ผู้ตรวจให้คะแนนที่เป็นนักศึกษาราชภัฏที่เป็นกลุ่มตัวอย่างในการวิจัยครั้งนี้ จะให้คะแนนไม่คงที่ข้ามเวลา จึงมีสูงกว่า

3. การตรวจสอบความเที่ยงตรงของความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3

3.1 การตรวจสอบความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย มีค่าความเที่ยงตรงเชิงเหมือนมีค่าสูงกว่าความเที่ยงตรงเชิงจำแนก ดังนั้นวิธีการตรวจให้คะแนนการเขียนเรียงความมีความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย สอดคล้องกับสมมติฐานการวิจัยข้อที่ 2.1 ที่ว่ามีความเที่ยงตรงเชิงโครงสร้างแบบลักษณะหลากหลาย-วิธีหลาย โดยที่ความเที่ยงตรงเชิงเหมือน คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการวัดคุณลักษณะเดียวกันด้วยวิธีการตรวจให้คะแนนต่างกัน มีค่าสูงอย่างมีนัยสำคัญทางสถิติ และความเที่ยงตรงเชิงจำแนก คือ ค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างการวัดคุณลักษณะต่างกันด้วยวิธีการตรวจให้คะแนนต่างกัน หรือระหว่างการวัดคุณลักษณะต่างกันด้วยวิธีการตรวจให้คะแนนเดียวกัน มีค่าต่ำ และเป็นไปตามเกณฑ์ของ แคมป์เบลล์ และ ฟิสต์ (ศิริชัย กาญจนวาสี. 2544: 97; อ้างอิงจาก Campbell; & Fiske. 1959) ที่กล่าวไว้ว่า คะแนนจากการวัดลักษณะเดียวกัน ถึงแม้จะใช้การวัดต่างกันย่อมมีความสัมพันธ์กันสูง และสำหรับการวัดลักษณะต่างกัน ซึ่งเรียกว่า ความเที่ยงตรงเชิงจำแนกย่อมมีความสัมพันธ์กันต่ำ โดยมีค่าต่ำกว่า ค่าสัมประสิทธิ์สหสัมพันธ์จากการวัดลักษณะเดียวกัน แสดงให้เห็นว่าวิธีการตรวจให้คะแนนเรียงความที่ผู้วิจัยสร้างขึ้นมีค่าความเที่ยงตรงเชิงโครงสร้าง

แต่อย่างไรก็ตาม ในการศึกษาครั้งนี้ ค่าความเที่ยงตรงเชิงจำแนกบางค่ามีความสัมพันธ์กันอยู่ค่อนข้างมากในบางคุณลักษณะ ซึ่งอาจเป็นไปได้เนื่องจากโครงสร้างของสิ่งที่กำลังวัดทั้งสองนั้นมิได้แยกกันอย่างชัดเจน (สำเร็จ บุญเรืองรัตน์. 2547: 120) สาเหตุมาจากในการคำนวณคุณลักษณะการทำหน้าที่ต่างกันของผู้ให้คะแนนเมื่อเวลาผ่านไปใช้คะแนนที่มาจากผู้ตรวจและนักเรียนคนเดียวกันครั้งที่ 1 มาเปรียบเทียบกับคะแนนที่ได้มาใหม่ครั้งที่ 2 ซึ่งคะแนนครั้งที่ 1 ได้ใช้ในการหาคุณลักษณะความเข้มงวด/ใจดี และความลำเอียงด้านเพศ ดังนั้นข้อมูลที่ได้ จึงมีการซ้อนทับกันบางอย่างจุด ความเที่ยงตรงเชิงจำแนกจึงอาจสูงไปแต่ยังถือว่าน้อยกว่าความเที่ยงตรงเชิงเหมือน

3.2 การตรวจสอบความเที่ยงตรงตามสภาพของความสามารถในการเขียนของนักเรียนชั้นประถมศึกษาปีที่ 3 พบว่า เมื่อผู้ตรวจมีคุณลักษณะลำเอียง การใช้วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนจำนวนผู้ตรวจให้คะแนน 3 คน ผู้ตรวจให้คะแนนมีความลำเอียงด้านเพศ เข้าข้างเพศชาย 1 คน และลำเอียงเข้าข้างเพศหญิง 2 คน มีความความเที่ยงตรงตามสภาพสูงที่สุด และเมื่อผู้ตรวจมีคุณลักษณะเป็นกลาง การใช้วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนจำนวนผู้ตรวจให้คะแนน 3 คน โดยผู้ตรวจให้คะแนนเป็นกลาง ไม่เข้มงวด/ใจดี มีความความเที่ยงตรงตามสภาพสูงที่สุด ซึ่งไม่สอดคล้องกับสมมติฐานการวิจัยข้อที่ 2.2 ที่ว่า มีความเที่ยงตรงตามสภาพภายใต้การใช้วิธีการตรวจให้คะแนนแตกต่างกัน ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนนที่มีคุณลักษณะแตกต่างกัน โดยมีเงื่อนไข เมื่อผู้ตรวจมีคุณลักษณะลำเอียง การให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนน 2 คน ที่มีคุณลักษณะเข้มงวด 1 คน และใจดี 1 คน มีความเที่ยงตรงตามสภาพสูงที่สุด และเมื่อผู้ตรวจมีคุณลักษณะเป็นกลาง การให้คะแนนแบบแยกองค์ประกอบ ในสัดส่วนของจำนวนของผู้ตรวจให้คะแนน 2 คน ที่มีคุณลักษณะไม่เข้มงวด/ใจดี มีความเที่ยงตรงตามสภาพสูงที่สุด จะเห็นได้ว่า เงื่อนไขที่ทำให้ความเที่ยงตรงตามสภาพสูงที่สุดทั้งในเงื่อนไขที่ผู้ตรวจมีบุคลิกภาพลำเอียงและผู้ตรวจมีคุณลักษณะเป็นกลางต่างก็มีจำนวนผู้ตรวจให้คะแนน 3 คน ทั้งนี้อาจเกิดจากจำนวนของ

ผู้ตรวจให้คะแนนเท่ากับจำนวนของครู จึงทำให้ความแปรปรวนที่ได้มีค่าใกล้เคียงกันเมื่อนำมาหาความสัมพัทธ์ จึงทำให้ค่าสหสัมพันธ์สูงสุด ซึ่งอาจเป็นสาเหตุสำคัญอีกประการหนึ่งที่ทำให้เงื่อนไขดังกล่าวมีค่าสหสัมพันธ์สูงสุด ผลการวิจัยนี้สอดคล้องกับงานวิจัยของ อุษณีย์ บัวศิริพันธ์ (2543: 53 – 54) ที่ได้ศึกษาเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจจำนวนผู้ตรวจ และประสบการณ์การตรวจแตกต่างกัน ซึ่งพบว่า เมื่อมีจำนวนผู้ตรวจแตกต่างกัน คะแนนที่ได้จะมีต่างกัน ยังมีจำนวนผู้ตรวจให้คะแนนจำนวนมากความสัมพัทธ์จะยิ่งสูง

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 คู่มือการตรวจให้คะแนนเรียงความ และข้อสอบการเขียนจดหมายที่ผู้วิจัยสร้างขึ้น มีค่าความเชื่อมั่น และความเที่ยงตรงสูงสามารถนำไปประยุกต์ใช้กับการสอบการเขียนเรียงความ ชั้นประถมศึกษาปีที่ 3 โดยเฉพาะโรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน สำนักงานเขตพื้นที่การศึกษาประถมศึกษาสุรินทร์ เขต 1 ถ้านำไปใช้ในบริบทอื่น หรือกลุ่มตัวอย่างอื่น อาจต้องนำไปทดลองใช้ใหม่

1.2 ควรอ่านและทำความเข้าใจคู่มือการตรวจให้คะแนนอย่างละเอียด โดยเฉพาะขั้นตอนในการให้คะแนนในแต่ละวิธีการตรวจให้คะแนน เนื่องจากผู้ตรวจให้คะแนนต้องทำตามลำดับขั้นตอน และใช้วิจารณญาณในการพิจารณาคะแนนในแต่ละประเด็นอย่างเคร่งครัด คู่มือฉบับนี้จึงจะมีประโยชน์สูงสุด

1.3 การตรวจให้คะแนนเรียงความถึงแม้ว่าจะมีผู้ให้คะแนนมีบุคลิกภาพลำเอียง แต่เมื่อจำนวนผู้ตรวจมากกว่าความเที่ยงตรงจะสูงขึ้น แต่ถ้าในกรณีที่ผู้ตรวจให้คะแนนเพียง 2 คน ถ้าต้องการให้คะแนนมีความเที่ยงตรงควรเลือกผู้ตรวจให้คะแนนที่มีคุณลักษณะเข้มงวด 2 คนในการให้คะแนน

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

2.1 ควรมีการทำวิจัยเชิงคุณภาพถึงภูมิหลังที่ทำให้ผู้ตรวจให้คะแนนมีคุณลักษณะที่เข้มงวดหรือใจดี เช่น ผลสัมฤทธิ์ทางการเรียนของผู้ตรวจให้คะแนน ประสบการณ์ในการสอน การสังกัดในโรงเรียนที่มีขนาดต่างกัน

2.2 ควรมีการศึกษาว่า สัดส่วนจำนวนผู้ตรวจให้คะแนนที่มีเพศสภาพต่างกัน จะส่งผลอย่างไรต่อการให้คะแนนความสามารถในการเขียนเรียงความ โดยถ้าใช้ผู้ตรวจให้คะแนนที่เป็นเพศชายจำนวนเท่ากับผู้ตรวจให้คะแนนที่เป็นเพศหญิง จะมีผู้แสดงคุณลักษณะลำเอียงเข้าข้างเพศใดเพศหนึ่งหรือไม่ ถ้าใช้ผู้ตรวจให้คะแนนที่เป็นเพศชายจำนวนมากกว่าผู้ตรวจให้คะแนนที่เป็นเพศหญิง จะมีผู้แสดงคุณลักษณะลำเอียงเข้าข้างเพศใดเพศหนึ่งหรือไม่ อย่างไร

2.3 ควรมีการใช้เกณฑ์การตรวจให้คะแนนที่มีการให้น้ำหนักไม่เท่ากัน เพื่อศึกษาว่า ลักษณะการให้คะแนนของผู้ตรวจให้คะแนนจะแตกต่างกันไปจากงานวิจัยนี้หรือไม่ อย่างไร

2.4 ควรมีการศึกษาการทำหน้าที่ต่างกันของผู้ตรวจให้คะแนน เมื่อเวลาผ่านไป โดยใช้ องค์ประกอบของเวลา (Time Facet) ในรูปแบบอื่นๆ เช่น แบ่งออกเป็น ช่วงเช้า-ช่วงบ่าย เพื่อศึกษาว่า เมื่อองค์ประกอบของเวลาเปลี่ยน ผลการระบุคุณลักษณะการให้คะแนนไม่คงที่ข้ามเวลาจะเปลี่ยนแปลงหรือไม่

2.5 ควรมีการศึกษาเพื่อดูพัฒนาการ (Longitudinal Study) ของการตรวจให้คะแนนของผู้ตรวจให้คะแนนแต่ละคนว่าคะแนนที่ได้นั้น จะทำให้คุณลักษณะการตรวจให้คะแนนผันแปรไปหรือไม่อย่างไร





ปรรณานุกรม

บรรณานุกรม

- กนกวรรณ รัตนชน. (2544). การเปรียบเทียบความตรงตามสภาพและความคงที่ในการประมาณค่าความสามารถของผู้สอบในการทดสอบแบบปรับเหมาะกับความสามารถของผู้สอบด้วยคอมพิวเตอร์เมื่อระดับความสามารถของผู้สอบและอัตราการใช้ข้อสอบซ้ำแตกต่างกัน. วิทยานิพนธ์ ค.ม. (การศึกษาและการสอน). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- กรมวิชาการ. (2546). การจัดสาระการเรียนรู้กลุ่มสาระการเรียนรู้ภาษาไทย ตามหลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2551. กรุงเทพฯ: โรงพิมพ์คุรุสภาลาดพร้าว.
- กระทรวงศึกษาธิการ. (2552ก). คู่มือการเรียนการสอนภาษาไทย คิดและเขียนอย่างสร้างสรรค์: เรียงความ ย่อความ การสรุปความ ช่วงชั้นที่ 2-4. กรุงเทพฯ: โรงพิมพ์คุรุสภาลาดพร้าว.
- . (2552ข). ตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้ภาษาไทยตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551. กรุงเทพฯ: โรงพิมพ์คุรุสภาลาดพร้าว.
- จตุรงค์ ลังกาพินธุ์. (ม.ม.ป.: ออนไลน์). การเพิ่มศักยภาพการจำ. สืบค้นเมื่อ 4 กรกฎาคม 2556, จาก http://www.203.158.253.84/weblibrary/DOCUMENT/book_01/09.pdf
- จินตนา ธนวิบูลชัย. (2535). ความเที่ยงตรง : ปัญหาและแนวคิดใหม่. วารสารวัดผลการศึกษา. 13(39): 63 – 90.
- ฉัตรศิริ ปิยะพิมลสิทธิ์. (ม.ม.ป.). การแปลงฟิซเซอร์ 'z' และการเปรียบเทียบความแตกต่างระหว่างค่าสหสัมพันธ์. สืบค้นเมื่อ 4 กุมภาพันธ์ 2554, จาก <http://www.watpon.com>.
- ชัยนันท์ นันทพันธ์. (2526). ศิลปะการเขียนเรียงความและรวมเรียงความรางวัลพระราชทาน. กรุงเทพฯ: ประกายพริก.
- ชัยฤทธิ์ ศิลาเดช. (2544). คู่มือการเขียนแผนการสอนที่เน้นผู้เรียนเป็นสำคัญระดับประถมศึกษา. กรุงเทพฯ: แม็ค.
- ชาญวิทย์ จรัสสุทธิอิศร. (2550). การแสดงหลักฐานความเที่ยงตรงตามโครงสร้างของคะแนนการประเมินทักษะกระบวนการทางวิทยาศาสตร์โดยใช้ทฤษฎีมาตรฐานเดิมและทฤษฎีการตอบข้อคำถามตามโมเดลราล์ชแบบหลายองค์ประกอบ. ปรินญาณีพนธ์ กศ.ด. (การทดสอบและการวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- ชำนาญ สุภานิตย์; และ สุทิน พูลสวัสดิ์. (2547). เตรียมสอบการเขียน Essay TOFEL GMAT GRE ฉบับสมบูรณ์. กรุงเทพฯ: ที.จี.อาร์.อี.
- ชวลีพร ดัดงาม. (2552). การพัฒนาเกณฑ์การตรวจให้คะแนนการเขียนเรียงความของนักเรียนช่วงชั้นที่ 2. ปรินญาณีพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.

- เดือนใจ เกตุษา. (2549). การสร้างแบบทดสอบ1 แบบทดสอบวัดผลสัมฤทธิ์ TEST CONSTRUCTION I. พิมพ์ครั้งที่ 8. กรุงเทพฯ: สำนักพิมพ์มหาวิทยาลัยรามคำแหง.
- ถวัลย์ มาตจรัส. (2545). การเขียนเชิงสร้างสรรค์เพื่อการศึกษาและอาชีพ. กรุงเทพฯ: ชารอักษร.
- ธิดารัตน์ ธนะคำดี. (2552). การพัฒนาแบบวัดความมีวินัยในตนเองสำหรับนักเรียนช่วงชั้นที่ 2 โรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษา กรุงเทพมหานคร เขต 1. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- นงลักษณ์ วิรัชชัย. (2538). วิธีวิทยาการขั้นสูงด้านการวิจัย สถิติ การวัดผลและประเมินผลการศึกษา. สืบค้นเมื่อ 17 กันยายน 2553, จาก http://onknow.blogspot.com/2006/03/blog-post_6728.html.
- นเรศ สุรสิทธิ์. (2550). การเขียนเรียงความ (Essay Writing). กรุงเทพฯ: พี.เอส.เพรส.
- น้ำผึ้ง อินทะเนตร. (2554). การศึกษาคูณลักษณะของคะแนนแบบทดสอบปลายเปิดวิชาคณิตศาสตร์ เมื่อจำนวนผู้ตรวจและรูปแบบการตรวจให้คะแนนต่างกัน โดยใช้โมเดลการสรุปอ้างอิงและโมเดลหลายองค์ประกอบของราล์ช. วิทยานิพนธ์ กศ.ด. (การทดสอบและวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- นิตารัตน์ คงสวัสดิ์. (2544). ผลการวิเคราะห์ค่าความเชื่อมั่นของผู้ตรวจแบบทดสอบความเรียงที่มีจำนวนผู้ตรวจ และวิธีการตรวจต่างกัน. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- บุญเชิด ภิญโญนนตพงษ์. (2521). การวัดและประเมินผลการการศึกษา: ทฤษฎีประยุกต์. กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ.
- . (2544). รายงานการวิจัยการประเมินการเรียนรู้ที่เน้นผู้เรียนเป็นสำคัญ: แนวคิดและวิธีการ. กรุงเทพฯ: อมรินทร์พรินติ้งแอนด์พับลิชชิ่ง.
- . (2547). การวัดประเมินการเรียนรู้ (การวัดประเมินแนวใหม่). กรุงเทพฯ: มหาวิทยาลัยศรีนครินทรวิโรฒ.
- บุญลักษณ์ เอี่ยมสำอางค์; เกื้อกมล พฤษประมุข; และ โสภิต พิทักษ์. (2551). ภาษาไทย หลักการใช้ภาษาและการใช้ภาษา ม.4. กรุงเทพฯ: อักษรเจริญทัศน์.
- ประพนธ์ เรืองณรงค์. (2545). กลุ่มสาระการเรียนรู้ ภาษาไทย ช่วงชั้นที่ 2 (ป.4 – 6) สาระที่ 1 – 5. กรุงเทพฯ: สำนักพิมพ์ประสานมิตร.
- ปวีณา ปีอาทิตย์. (2545). การศึกษาจำนวนผู้ประเมินและจำนวนงานเขียนที่เหมาะสมเมื่อใช้เกณฑ์การให้คะแนนที่ต่างกัน. วิทยานิพนธ์ ค.ม. (การศึกษาและการสอน). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- ปิยะรัตน์ ประมวลทรัพย์. (2546). การพัฒนาแบบสอบอัตนัยประยุกต์วัดสมรรถนะทางวิชาชีพพยาบาล. วิทยานิพนธ์ ค.ม. (การศึกษาและการสอน). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

- เปลื้อง ณ นคร. (2512). *ตำราเรียงความชั้นสูง*. กรุงเทพฯ: ไทยวัฒนาพานิช.
- ผจงวาด พูลแก้ว; พัทธณี ปี่เพระ; และ สุวรรณ โสประดิษฐ์. (2551). *การเขียนเรียงความ จดหมาย*. กรุงเทพฯ: สุวีริยาสาส์น.
- พรณี เจียมสมบูรณ์. (2543). *การเปรียบเทียบความเชื่อมั่นของแบบทดสอบวัดความสามารถในการแก้โจทย์ปัญหาทางคณิตศาสตร์ที่มีจำนวนผู้ตรวจและวิธีการตรวจต่างกัน*. ปรินญาณิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- พวงรัตน์ ทวีรัตน์. (2530). *การสร้างและพัฒนาแบบทดสอบวัดผลสัมฤทธิ์*. กรุงเทพฯ: สำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ.
- ไพศาล หวังพานิช. (2545). *การวัดและประเมินผลการเรียน*. กรุงเทพฯ: สำนักทดสอบทางการศึกษาและจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ.
- มหาวิทยาลัยสุโขทัยธรรมมาธิราช. (2549). *เอกสารการสอนชุดวิชาการพัฒนาแบบทดสอบวัดผลสัมฤทธิ์ทางการเรียน Development of Achievement Test /สาขาวิชาศึกษาศาสตร์*. พิมพ์ครั้งที่ 7. กรุงเทพฯ: สำนักพิมพ์สุโขทัยธรรมมาธิราช.
- เยาวดี วิบูลย์ศรี. (2552). *การวัดผลและการสร้างแบบสอบผลสัมฤทธิ์ = Measurement and Achievement Test Construction*. กรุงเทพฯ: สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- รองรัตน์ อิศรภักดี; และ เทือก กุสุมา ณ อยู่ชยา. (2513). *ภาษาไทยตอน 1 วิธีการสอนภาษาไทย ชั้นมัธยม*. กรุงเทพฯ: โรงพิมพ์คุรุสภาลาดพร้าว.
- ล้วน สายยศ; และ อังคณา สายยศ. (2543). *เทคนิคการวัดผลการเรียนรู้*. พิมพ์ครั้งที่ 2. กรุงเทพฯ: สุวีริยาสาส์น.
- ศิริชัย กาญจนวาสี. (2550). *ทฤษฎีการทดสอบแนวใหม่ (MODERN TEST THEORIES)*. กรุงเทพฯ: โรงพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.
- ส. วาสนา ประवालพฤกษ์. (2533). *การพิจารณาความเชื่อมั่นของแบบทดสอบการปฏิบัติ*. *การวัดผลการศึกษา*. 12(35).
- สำเร็จ บุญเรืองรัตน์. (2547). *การวิเคราะห์ลักษณะหลากหลาย-วิธีหลาย*. *การวัดและประเมินผลการศึกษา: โครงการสารานุกรมศึกษาศาสตร์ คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ ฉบับรวมเล่มเฉพาะเรื่อง อันดับที่ 3*. กรุงเทพฯ: พัฒนาศึกษา.
- เสาวนีย์ เกรียร์. (2540). *คู่มือการอบรมการใช้แฟ้มสะสมผลงานนักเรียน*. กรุงเทพฯ: เนติกุลการพิมพ์.
- องอาจ นัยพัฒน์. (2548). *วิธีวิทยาการวิจัยเชิงปริมาณและเชิงคุณภาพทางพฤติกรรมศาสตร์และสังคมศาสตร์*. กรุงเทพฯ: สามลดา.
- อัญญารัตน์ เจริญพุดผิง. (2546). *การพัฒนาแบบประเมินทักษะการอ่าน คิดวิเคราะห์ เขียน ของนักเรียนชั้นประถมศึกษา*. วิทยานิพนธ์ ค.ม. (การศึกษาและการสอน). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.

- อุษณีย์ บัวศิริพันธ์. (2543). *การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่มีวิธีการตรวจจำนวนผู้ตรวจและประสบการณ์ของผู้ตรวจแตกต่างกัน*. วิทยานิพนธ์ กศ.ม. (การวัดผลการศึกษา). กรุงเทพฯ: บัณฑิตวิทยาลัย มหาวิทยาลัยศรีนครินทรวิโรฒ. ถ่ายเอกสาร.
- เอมอร จังศิริพรพรรณ. (2545). *การเปรียบเทียบคุณภาพของแบบสอบเลือกตอบ เมื่อตรวจด้วยวิธีการให้คะแนนความรู้บางส่วนกับวิธีประเพณีนิยม*. วิทยานิพนธ์ ค.ม. (การศึกษาและการสอน). กรุงเทพฯ: บัณฑิตวิทยาลัย จุฬาลงกรณ์มหาวิทยาลัย. ถ่ายเอกสาร.
- Carmines, Edward G.; & Zeller, Richard A. (1987). *Reliability and Validity Assessment*. 9th ed. Beverly Hills: SAGE Publications.
- Clauser, R. E.; & Mazor, K. M. (1998). Using Statistical Procedures to Identify Differentially Functioning Test Items. *Educational Measurement : Issues and Practice*. 17(1): 31 – 44.
- Eckes, Thomas. (2005). Examining Rater Effects in TestDaF Writing and Speaking Performance Assessments : A Many-Facet Rasch Analysis. *Language Assessment Quarterly*. 2(3): 197 – 221.
- Engelhard, George Jr.; & Myford, Carrol M. (2003). *Monitoring Faculty Consultant Performance in the Advance Placement English literature and Composition Program with Many-Faceted Rasch Model*. USA.: College Entrance Examination Board.
- Gruijter, Dato N.M. de; & Leo, J. Th. Van Der Kamp. (2008). *Statistical Test Theory for the Behavioral Sciences*. Boca Raton: Chapman & Hall.
- Iramaneerat, Cherdasak; & Yudkowsky Rachel. (2006, February). How Good are Our Raters? Rater Errors in Clinical Skills Assessment. *Paper Presented at the 2006 Graduate Educational Conference : Education and the Public Good : Interdisciplinary Trends in Graduate Scholarship*. Chicago, Illinois USA.
- IRD – NEUCHATEL – SWITZERLAND. (2010). *EDUGUSER GUIDE*. Switzerland: Swiss Society for Research in Education Working Group.
- Lawson, Douglas M.; & Brailovsky Carlos. (2006, October). The Presence and Impact of Local Item Dependence on Objective Structured Clinical Examinations Scores and The Potential Use of the Polytomous, Many-Facet Rasch Model. National University of Health Sciences. *Journal of Manipulative Physiol Ther*. 29(8): 651 – 7.
- Linacre, John M. (1994). *MANY-TACET RASCH MEASUREMENT*. Chicago: MESA Press.
- Linacre, John M. (2012). *A User's Guide to FACETS Rasch-Model Computer Programs*. USA.:
- Linderman, Richard H.; & Merenda, Peter F. (1979). *Education Measurement*. Illinois: Scott, Foresman Winsteps.

- Lyman, Howard B. (1991). *TEST SCORES & WHAT THEY MEAN*. 5th ed. New Jersey: Prentice-Hall.
- McMillan, James H. (2004). *CLASSROOM ASSESSMENT Principles and Practice for Effective Instruction*. USA.: Peason Education.
- Miller, M. David; Linn, Robert L.; & Gronlund, Norman E. (2009). *Measurement and Assessment in Teaching*. USA.: Peason Education.
- Mulqueen, Casey; Baker, David P.; & Dismukes, R. Key. (2000). Running Head : Multifacet Rasch Analysis of Rater Training Program. *Poster Paper Presented at the 15th Annual Meeting of the Society for Industrial /Organizational Psychology (SIOP)*. LA.: New Orleans.
- Myford, Carol M.; & Wolfe, Edward W. (2009, Winter). Monitoring Rater Performance Over Time : A Framework for Detecting Differential Accuracy and Differential Scale Category Use. *Journal of Educational Measurement*. 46(4): 371 – 389.
- Nakamura, Yuji. (2002, May). An Application of a Many-faceted Rasch Model of Writing Test Analysis. *Curriculum Innovation, Testing and Evaluation : Proceedings of the 1st Annual JALT Pan-SIG Conference*. pp. 11 – 12. Japan: Kyoto Institute of Technology.
- Nitko, Anthony J. (1996). *Educational Assessment of Students*. New Jersey: Merrill.
- Park, Taejoon. (2003). An Investigation of an ESL Placement Test of Writing Using Manyfacet Rasch Measurement. Teachers College, Columbia University Working Papers. *TESOL & Applied Linguistics*. 4(1).
- Payne, David A. (1997). *Applied Educational Assessment*. Belmont, California: Wadsworth.
- Saal; Downey; & Lahey. (1980). Rating the Ratings : Assessing the Psychometric Quality of Rating Data. *Psychological Bulletin*. 88(2): 413 – 428.
- Shealy, R.; & Stout, W. (1993, June). A Model-based Standardization Approach Thatseparates True Bias/DIF from Group Ability Differences and Detects Test Bias/DTF as Well as Item Bias/DIF. *Psychometrika*. 58(2): 159 – 194.
- Tragarn, Kalchayanant. (2552). *Writing Essays*. Bangkok: Thummasart University.
- Wolfe, Edward W. (2004). Identifying Rater Effects Using Latent Trait Models. *Psychology Science*. 46(1): 35 – 51.
- Wright, Benjamin D.; & Master, Geoffrey N. (1982). *Rating Scale Analysis*. Chicago: MESA Press.
- Yen, M.; & Fitzpatrick. (2006). *Educational Measurement*. by Robert, L Brennan; National Council on Measurement in Education (Washington (D.C.), Estados Unidos); American Council on Education (Washington (D.C.): Estados Unidos.



ภาคผนวก



ภาคผนวก ก
คู่มือการใช้เกณฑ์การตรวจให้คะแนน

แผนการฝึกอบรมการตรวจให้คะแนนการเขียนเรียงความ

เรื่อง การตรวจให้คะแนนการเขียนเรียงความ
 กลุ่มผู้เรียน นักศึกษาคณะครุศาสตร์ สาขาวิชาภาษาไทย ชั้นปีที่ 4
 เวลาที่ใช้สอน จำนวน 6 ชั่วโมง

จุดมุ่งหมาย

เพื่อให้ผู้เรียน

1. มีความรู้ความเข้าใจในการใช้เกณฑ์การตรวจให้คะแนนการเขียนเรียงความใน 3 รูปแบบ

คือ

- 1.1 การตรวจให้คะแนนแบบแยกองค์ประกอบ (Analytic Scoring Rubrics)
- 1.2 การตรวจให้คะแนนแบบรวมองค์ประกอบ (Holistic Scoring Rubrics)
- 1.3 การตรวจให้คะแนนแบบผสมผสาน (Annotated Scoring Rubrics)
2. สามารถใช้เกณฑ์การตรวจให้คะแนนการเขียนเรียงความใน 3 รูปแบบ

สาระการเรียนรู้

1. การเขียนเรียงความ หมายถึง การเขียนร้อยแก้วที่เรียบเรียงข้อความเป็นเรื่องราวที่เขียนถึงบุคคล สถานที่ เรื่อง หรือสิ่งใดสิ่งหนึ่ง โดยอาศัยความรู้ ความเข้าใจ ความคิดเห็น หรือประสบการณ์ของผู้เขียนที่เกี่ยวข้องกับเรื่องนั้นๆ ถ่ายทอดออกมา เพื่อแสดงความสามารถของผู้เขียน วัดโดยใช้แบบทดสอบการเขียนเรียงความ

2. วิธีการตรวจให้คะแนน หมายถึง ชุดของแนวทางในการให้คะแนนความสามารถในการเขียนเรียงความของนักเรียน แบ่งออกเป็น 3 แบบ คือ

- 2.1 การตรวจให้คะแนนแบบแยกองค์ประกอบ (Analytic Scoring Rubrics) คือ การให้คะแนนงานชิ้นใดชิ้นหนึ่ง โดยมีการแยกองค์ประกอบของการให้คะแนน และอธิบายคุณภาพของงานในแต่ละองค์ประกอบเป็นระดับ

- 2.2 การตรวจให้คะแนนแบบรวมองค์ประกอบ (Holistic Scoring Rubrics) คือ การให้คะแนนงานชิ้นใดชิ้นหนึ่ง โดยดูภาพรวมของชิ้นงานว่ามีความเข้าใจในความคิดรวบยอด การสื่อความหมาย กระบวนการที่ใช้ และผลงาน แล้วเขียนอธิบายคุณภาพของงาน หรือความสำเร็จของงานเป็นชิ้นๆ

- 2.3 การตรวจให้คะแนนแบบผสมผสาน (Annotated Scoring Rubrics) เป็นการตรวจให้คะแนนที่เพิ่มขึ้นมาใหม่ จะรวม Holistic และ Analytic ไว้ด้วยกัน เริ่มด้วยการให้คะแนนในภาพรวมของการปฏิบัติงานด้วย Holistic แล้วเลือกให้คะแนนอีกเพียงบางคุณลักษณะของงานแบบ Analytic

กระบวนการจัดการเรียนรู้

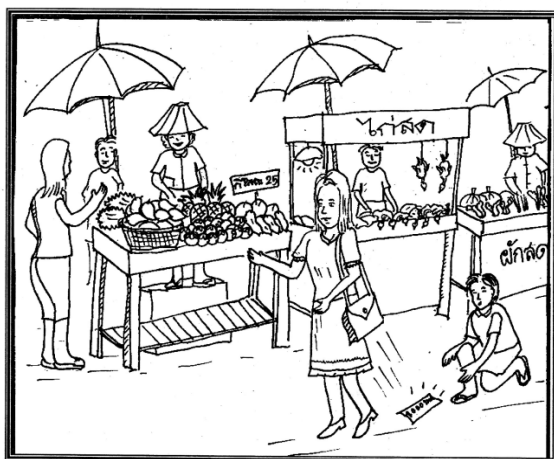
ขั้นนำเข้าสู่บทเรียน

นักศึกษาร่วมกันอภิปรายถึงการเขียนเรียงความตามความรู้เดิม เพื่อให้ทราบถึงพื้นฐานความรู้โดยผู้วิจัยเป็นผู้ตั้งหัวข้ออภิปราย เช่น

1. นักศึกษาเคยเขียนเรียงความหรือไม่ อย่างไร (นักศึกษาตอบตามประสบการณ์ของตนเอง)
2. นักศึกษารู้จักแบบแผนตามลักษณะการเขียนเรียงความหรือไม่ (นักศึกษาตอบตามประสบการณ์ของตนเอง)

ขั้นสอน

1. ผู้วิจัยร่วมสนทนากับนักศึกษา โดยใช้คำถามต่อไปนี้
 - 1.1 นักศึกษาเคยตรวจให้คะแนนการเขียนเรียงความหรือไม่ อย่างไร (นักศึกษาตอบตามประสบการณ์ของตนเอง)
 - 1.2 นักศึกษาคิดว่าการตรวจให้คะแนนเรียงความ ควรมีเกณฑ์ในการให้คะแนนอะไรบ้าง อย่างไร (นักศึกษาตอบตามประสบการณ์ของตนเอง)
2. นักศึกษาฟังการบรรยายประกอบ Power Point ชุดการฝึกการตรวจให้คะแนนการเขียนเรียงความ
 - 2.1 ผู้วิจัยอธิบายถึงที่มาของเรียงความที่นักศึกษาจะใช้ตรวจให้คะแนน ว่ามาจากกระดาษคำตอบในส่วนของ การเขียน ที่ได้คัดสำเนาจากข้อสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 โรงเรียนเมืองสุรินทร์ จังหวัดสุรินทร์
 - 2.2 ผู้วิจัยอธิบายถึงที่มาของเกณฑ์การให้คะแนนว่ามีขั้นตอนการสร้างอย่างไร และชี้แจงให้นักศึกษาทราบว่าในการทดลองในครั้งนี้อยู่ในช่วงของการทดลองการใช้เครื่องมือ (Try Out)
 - 2.3 ผู้วิจัยเสนอรูปภาพที่เป็นโจทย์จากข้อสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553 ดังนี้



ขั้นฝึกฝนและนำไปใช้

1. ผู้วิจัยถามนักศึกษาว่าประเด็นสำคัญของรูปภาพนี้คืออะไร (ผู้หญิงทำเงินตก เด็กผู้ชาย เก็บได้)

2. ผู้วิจัยถามนักศึกษาว่าจากรูปภาพอะไรบ้างที่ควรอยู่ในเนื้อเรื่องเรียงความ (ผู้หญิง เงิน เด็กผู้ชาย ตลาดสด ไม้ขีด ผลไม้ ผักสด มะม่วง ทูเรียน มังคุด สับปะรด มะละกอ แม่ค้า พ่อค้า และคนซื้อ)

3. ผู้วิจัยนำเสนอนิยามศัพท์เฉพาะของเกณฑ์การตรวจให้คะแนนเรียงความ
การตรวจให้คะแนนเรียงความแบบแยกองค์ประกอบ

1. ผู้วิจัยชี้แจงขั้นตอนการตรวจให้คะแนนแบบแยกองค์ประกอบ 1) อ่านเรียงความ 2) ให้คะแนนนักเรียนทีละประเด็นจนครบ 4 ประเด็น)

2. ผู้วิจัยอธิบายพร้อมยกตัวอย่างการตรวจให้คะแนนเรียงความแบบแยกองค์ประกอบ ทีละประเด็นโดยชี้ให้นักศึกษาเห็นว่า งานเขียนเรียงความควรจะได้คะแนนเท่าไร จำนวน 2 ตัวอย่าง (ยกตัวอย่างงานเขียนเรียงความของครองขวัญ และภาวิณี)

3. ผู้วิจัยร่วมกับนักศึกษอภิปรายถึงคะแนนของงานเรียงความว่า ควรจะได้คะแนนเท่าไร จำนวน 3 ตัวอย่าง (ยกตัวอย่างงานเขียนเรียงความของกมลชนก ปฏิภาณ และจิตกร)

4. ผู้วิจัยให้นักศึกษาฝึกให้คะแนนงานเขียนเรียงความของ กาญจนา ขวัญใจ นพพร พรเทพ และ วิจิตร จากนั้นนำคะแนนมาอภิปรายถึงความเหมาะสมตามความคิดเห็นของนักศึกษา

การตรวจให้คะแนนเรียงความแบบรวมองค์ประกอบ

1. ผู้วิจัยชี้แจงขั้นตอนการตรวจให้คะแนนแบบรวมองค์ประกอบ 1) อ่านเรียงความ 2) วางเรียงความของนักเรียนลงในกองคะแนน 5 กอง แยกตามระดับคุณภาพของเรียงความ)

2. ผู้วิจัยยกตัวอย่างการให้คะแนนแบบแยกองค์ประกอบโดยใช้เรียงความของ กมลชนก ครองขวัญ จิตกร ปฏิภาณและภาวิณี

3. ผู้วิจัยร่วมกับนักศึกษอภิปรายถึงคะแนนของเรียงความว่าควรจะได้คะแนนเท่าไร โดยใช้เรียงความของกาญจนา ขวัญใจ นพพร พรเทพ และ วิจิตร

การตรวจให้คะแนนเรียงความแบบผสมผสาน

1. ผู้วิจัยถามนักศึกษาว่าประเด็นใดที่น่าจะนำมาเป็นคุณลักษณะที่สะท้อนกลับเพื่อนำไปพัฒนาการเขียนในอนาคตสำหรับนักเรียนผู้เขียนเรียงความ
2. ผู้วิจัยชี้แจงขั้นตอนการตรวจให้คะแนนแบบผสมผสาน 1) อ่านเรียงความแล้ววางเรียงความของนักเรียนลงในกองคะแนน 5 กอง แยกตามระดับคุณภาพของเรียงความ 2) เลือกประเด็นนำมาเป็นคุณลักษณะที่สะท้อนกลับเพื่อนำไปพัฒนาการเขียนในอนาคตสำหรับนักเรียน ผู้เขียนเรียงความ 3) ให้คะแนนนักเรียนที่ละประเด็นที่ได้เลือกไว้จนครบทุกประเด็น)
3. ผู้วิจัยยกตัวอย่างการให้คะแนนแบบผสมผสานโดยใช้เรียงความของกมลชนก ครอบขวัญ จิตกร ปฏิภาณ และ ภาวินี
4. ผู้วิจัยร่วมกับนักศึกษอภิปรายถึงคะแนนของเรียงความว่าควรจะได้คะแนนเท่าไร โดยใช้เรียงความของกาญจนา ขวัญใจ นพพร พรเทพ และ วิจิตร

ขั้นสรุป

ผู้วิจัยและนักศึกษาร่วมกันสรุปบทเรียน โดยผู้วิจัยใช้คำถามต่อไปนี้

1. การตรวจให้คะแนนเรียงความในที่มีเกณฑ์การให้คะแนนที่ประเด็น อะไรบ้าง (การตรวจให้คะแนนเรียงความมี 4 ประเด็นหลัก คือ ประเด็นสำคัญของเรียงความ รูปแบบการเขียนเรียงความ การนำเสนอ และการใช้ภาษา)
 - 1.1 ในประเด็นที่ 1 ประเด็นสำคัญของเรียงความ ระดับคุณภาพของคะแนนจะลดหลั่นกันไป เนื่องจากประเด็นใด (ความน่าสนใจของชื่อเรื่องและความสอดคล้องของเนื้อเรื่องกับชื่อเรื่อง)
 - 1.2 ในประเด็นที่ 2 รูปแบบการเขียนเรียงความ ระดับคุณภาพของคะแนนจะลดหลั่นกันไปเนื่องจากประเด็นใด 1) ความครบถ้วนของคำนำ เนื้อเรื่อง สรุป 2) ความเหมาะสมของความยาวในแต่ละส่วน และ 3) จำนวนบรรทัดที่นักเรียนเขียน)
 - 1.3 ในประเด็นที่ 3 การนำเสนอ ระดับคุณภาพของคะแนนจะลดหลั่นกันไปเนื่องจากประเด็นใด 1) การเรียบเรียงเนื้อเรื่องเป็นลำดับต่อเนื่องสัมพันธ์กัน 2) ความเชื่อมโยงในแต่ละแนวคิด 3) การสอดแทรกแนวคิดที่เป็นประโยชน์ต่อสังคม และ 4) การแสดงความคิดเห็นสนับสนุนความคิดเห็น)
 - 1.4 ในประเด็นที่ 4 การใช้ภาษา ระดับคุณภาพของคะแนนจะลดหลั่นกันไปเนื่องจากประเด็นใด 1) การสะกดคำ 2) การเลือกใช้ถ้อยคำ 3) การฉีกคำ 4) ลายมือ และ 5) รอยลบขีดฆ่า)
2. ขั้นตอนการใช้เกณฑ์การตรวจให้คะแนนเรียงความแบบแยกองค์ประกอบเป็นอย่างไร 1) อ่านเรียงความ 2) ให้คะแนนนักเรียนที่ละประเด็นจนครบ 4 ประเด็น)
3. ขั้นตอนการใช้เกณฑ์การตรวจให้คะแนนเรียงความแบบรวมองค์ประกอบเป็นอย่างไร 1) อ่านเรียงความ 2) วางเรียงความของนักเรียนลงในกองคะแนน 5 กอง แยกตามระดับคุณภาพของเรียงความ)
4. ขั้นตอนการใช้เกณฑ์การตรวจให้คะแนนเรียงความแบบผสมผสานเป็นอย่างไร 1) อ่านเรียงความแล้ววางเรียงความของนักเรียนลงในกองคะแนน 5 กอง แยกตามระดับคุณภาพของเรียงความ

2) เลือกประเด็นนำมาเป็นคุณลักษณะที่สะท้อนกลับเพื่อนำไปพัฒนาการเขียนในอนาคตสำหรับนักเรียนผู้เขียนเรียงความ 3) ให้คะแนนนักเรียนที่ละประเด็นที่ได้เลือกไว้จนครบทุกประเด็น)

ขั้นประเมินผล

นักศึกษาดูงานให้คะแนนการเขียนเรียงความ จำนวน 50 ชิ้น โดยแบ่งนักศึกษาออกเป็น 3 กลุ่ม ด้วยวิธีการสุ่มอย่างง่าย โดยในแต่ละกลุ่มใช้รูปแบบการตรวจ 3 แบบที่แตกต่างกัน (x_1 , x_2 , และ x_3) แล้วดำเนินการตามรูปแบบการทดลองแบบ Counterbalanced Designs โดยมีรายละเอียด ดังนี้

ตารางแบบแผนการทดลองแบบ Counterbalanced Designs

กลุ่มที่	ระยะเวลาการให้สิ่งทดลอง (X) และการวัดค่าตัวแปรตาม (O)		
	ครั้งที่ 1	ครั้งที่ 2	ครั้งที่ 3
1	X_1O	X_2O	X_3O
2	X_3O	X_1O	X_2O
3	X_2O	X_3O	X_1O

สื่อ/แหล่งการเรียนรู้

1. ชุดฝึกการตรวจให้คะแนนการเขียนเรียงความ
2. Power Point ชุดการฝึกการตรวจให้คะแนนการเขียนเรียงความ

การวัดและประเมินผลการเรียนรู้

วิธีการ	เครื่องมือ	เกณฑ์การประเมิน
1. สังเกตพฤติกรรมการเรียน	การสังเกต	ความสนใจในการเรียน
2. การตรวจผลงาน	แบบฝึกในชุดการตรวจให้คะแนนเรียงความ	ให้คะแนนตรงกันร้อยละ 80
	การตรวจให้คะแนนเรียงความ	ค่า Inter-Rater มีค่าสูง

บันทึกผลการเรียนรู้

กระบวนการเรียนรู้

.....

.....

.....

.....

ปัญหา/อุปสรรค

.....

.....

.....

.....

สิ่งที่ผู้สอนต้องปรับปรุงพัฒนา

.....

.....

.....

.....



คู่มือ

การตรวจให้คะแนนเรียงความ

① การตรวจให้คะแนนแบบแยกองค์ประกอบ
(Analytic Scoring Rubrics)

② การตรวจให้คะแนนแบบรวมองค์ประกอบ
(Holistic Scoring Rubrics)

③ การตรวจให้คะแนนแบบผสมผสาน
(Annotated Scoring Rubrics)

โดย

บุษวรรษ์ แสนปลื้ม

การทดสอบและวัดผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ

สารบัญ

รายการ

หน้า

ภูมิหลัง	
ความมุ่งหมาย	
เรียงความ คืออะไร?	
เกณฑ์การให้คะแนน (Rubrics) คืออะไร?	
รูปภาพที่ใช้เป็นโจทย์ในการเขียนเรียงความ	
นิยามศัพท์เฉพาะ	
ชุดการฝึกใช้เกณฑ์การให้คะแนนเรียงความแบบแยกองค์ประกอบ.....	
ตัวอย่างการตรวจให้คะแนนเรียงความแบบแยกองค์ประกอบ.....	
ชุดการฝึกใช้เกณฑ์การให้คะแนนเรียงความแบบแยกองค์ประกอบ	
ตัวอย่างการตรวจให้คะแนนเรียงความแบบรวมองค์ประกอบ	
ชุดการฝึกใช้เกณฑ์การให้คะแนนเรียงความแบบผสมผสาน	
ตัวอย่างการตรวจให้คะแนนเรียงความแบบผสมผสาน	
แบบฝึกหัด.....	
บรรณานุกรม.....	

เกณฑ์การตรวจให้คะแนนเรียงความ

ภาษาไทยเป็นเอกลักษณ์ประจำชาติไทย เป็นวัฒนธรรมของการใช้ถ้อยคำเพื่อสื่อสารให้เกิดความเข้าใจที่ตรงกันและตรงตามจุดมุ่งหมายของการสื่อสาร ไม่ว่าจะเป็นการแสดงความคิด ความต้องการ ความรู้สึก ดังนั้นผู้ที่เป็นคนไทยจึงมีความจำเป็นต้องอ่านและเขียนภาษาไทยได้จึงจะอยู่ในสังคมอย่างปกติสุข ในปัจจุบันทุกคนที่เกิดเป็นคนไทย จะมีสิทธิขั้นพื้นฐานในการรับการศึกษาในระบบโรงเรียน โดยมีจุดประสงค์หลักเพื่อให้ทุกคนอ่านออกเขียนได้และสามารถนำไปต่อยอดเพื่อการเรียนต่อหรือประกอบอาชีพในอนาคต (ชานาญ สุภนิตย์; และ สุทิน พูลสวัสดิ์. 2547: 257) การศึกษาในระดับประถมศึกษาชั้นนั้น ผู้เรียนต้องมีผลการประเมินการอ่าน คิดวิเคราะห์ และเขียนในระดับผ่านเกณฑ์การประเมินตามที่สถานศึกษากำหนด โดยที่หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ได้กำหนดมาตรฐานการเรียนรู้ไว้ในกลุ่มสาระการเรียนรู้ภาษาไทย สาระที่ 2 การเขียน มาตรฐาน ท 2.1 ใช้กระบวนการเขียน เขียนสื่อสาร เขียนเรียงความ ย่อความ และเขียนเรื่องราวในรูปแบบต่างๆ เขียนรายงานข้อมูลสารสนเทศ และรายงานการศึกษาค้นคว้าอย่างมีประสิทธิภาพ

ในการทดสอบทุกรูปแบบมีหน้าที่เหมือนกันคือมุ่งวัดผลสัมฤทธิ์ในการเรียนของนักเรียน แต่ในกรณีที่ใช้ข้อสอบเรียงความนั้นมีส่วนที่ดีเพิ่มจากแบบอื่นๆ หลายประการ แต่ด้วยความจริงที่ว่าไม่มีข้อสอบชนิดใดดีพร้อมทุกอย่าง ดังนั้นเป็นปัญหาของครูในฐานะที่เป็นที่ผู้ให้คะแนนความสามารถ ในการเขียนเรียงความของนักเรียน จึงต้องคำนึงถึงและแก้ไขข้อบกพร่องนั้นให้มันน้อยที่สุด ข้อจำกัดที่สำคัญในการใช้ข้อสอบเรียงความ คือ ตรวจยาก การตรวจให้คะแนนถือว่ามีความยุ่งยากกว่าแบบทดสอบอื่นๆ สาเหตุสำคัญของปัญหานี้มีหลายกรณี เช่น คำตอบของผู้สอบแตกต่างกันไปตามความคิดของแต่ละคน ซึ่งผู้ออกข้อสอบอาจไม่ได้กำหนดขอบเขตของการตอบไว้อย่างชัดเจนเพียงพอ ผู้ตรวจให้คะแนนไม่มีเกณฑ์การให้คะแนนที่แน่นอน หรือผู้ตรวจให้คะแนนมีภาวะทางอารมณ์ที่ส่งผลกระทบต่อผลการตรวจให้คะแนนได้ง่าย ถ้ามได้จำกัด เนื่องจากข้อสอบแบบนี้ผู้สอบต้องใช้เวลาในการคิดและเขียนเป็นเหตุให้ไม่สามารถใช้คำถามได้มากข้อ ต้องจำกัดการถามเฉพาะในบางเนื้อหาเท่านั้น ลักษณะเช่นนี้ อาจทำให้การวัดไม่ครอบคลุมครบถ้วน ต้องใช้เวลาในการตอบและการตรวจให้คะแนน ผู้สอบต้องใช้เวลาสำหรับคิดและเขียน และผลจากการตอบเหล่านี้ ครูผู้ตรวจให้คะแนนต้องใช้เวลาในการตรวจมากพอสมควร และผลที่ได้มักไม่มีความเป็นปรนัย ก่อให้เกิดการเขียนที่วกวน แม้ว่าจะขจัดปัญหาในการเดาได้แต่ไม่สามารถป้องกันผู้สอบบางคนที่นิยมการตอบวกวนไม่ตรงจุดหรือการตอบแบบครอบจักรวาล ซึ่งเกิดจากที่ผู้สอบคิดว่า ขอให้เขียนอะไรก็ได้บางอย่างลงไปดีกว่าไม่เขียนอะไรเลย และคะแนนมักจะขึ้นอยู่กับความสามารถในการเขียน ซึ่งจะพบว่าความสามารถในการเขียนของผู้สอบมักมีความสัมพันธ์กันสูงกับคุณภาพสาระที่เขียน ดังนั้นคะแนนส่วนหนึ่งมักจะได้จากความสามารถทางภาษาของผู้สอบ (ไพศาล หวังพานิช. 2545)

จากปัญหาดังที่กล่าวมาข้างต้นผู้วิจัย จึงสนใจที่จะใช้การเขียนเรียงความ เป็นเครื่องมือที่ใช้ในการวิจัย โดยใช้กลุ่มตัวอย่างเป็นนักเรียนชั้นประถมศึกษาปีที่ 3 เพราะนอกจากจะเป็นการส่งเสริมความสามารถในการเขียนเรียงความของนักเรียนที่เรียนในระดับพื้นฐาน โดยการพัฒนาวิธีการวัดเพื่อ

พัฒนาผู้เรียนแล้ว ยังเป็นการพัฒนาวิธีการตรวจให้คะแนนให้กับครูผู้ตรวจให้คะแนนเรียงความในอนาคตอีกด้วย

ความมุ่งหมาย

เพื่อให้ผู้เรียน

1. มีความรู้ความเข้าใจในการใช้เกณฑ์การตรวจให้คะแนนการเขียนเรียงความใน 3 รูปแบบ คือ
 - 1.1 การตรวจให้คะแนนแบบแยกองค์ประกอบ (Analytic Scoring Rubrics)
 - 1.2 การตรวจให้คะแนนแบบรวมองค์ประกอบ (Holistic Scoring Rubrics)
 - 1.3 การตรวจให้คะแนนแบบผสมผสาน (Annotated Scoring Rubrics)
2. สามารถใช้เกณฑ์การตรวจให้คะแนนการเขียนเรียงความใน 3 รูปแบบ
3. มีเจตคติที่ดีต่อการตรวจให้คะแนนการเขียนเรียงความ

เรียงความ คืออะไร?

การเขียนเรียงความ คือ การเขียนร้อยแก้วที่เรียบเรียงข้อความเป็นเรื่องราว ที่เขียนถึงบุคคล สถานที่ เรื่อง หรือสิ่งใดสิ่งหนึ่ง โดยอาศัยความรู้ ความเข้าใจ ความคิดเห็น หรือประสบการณ์ของผู้เขียนที่เกี่ยวข้องกับเรื่องนั้นๆ ถ่ายทอดออกมาเพื่อแสดงความสามารถของผู้เขียน การเขียนเรียงความ มีความสำคัญอย่างมาก นอกจากเป็นเครื่องวัดความรู้ความสามารถของนักเรียน (เปลื้อง ณ นคร. 2512) แล้ว การเขียนเรียงความยังเป็น วิธีการขั้นพื้นฐานที่สำคัญยิ่งของการใช้ภาษาไทย ซึ่งสัญลักษณ์การเขียนของคนไทยก็คืออักษรไทย ดังนั้นการเขียนเรียงความจึงเป็นส่วนหนึ่งที่ใช้ในการเรียนใน

โรงเรียนทุกระดับ (ชัยนันท์ นันทพันธ์. 2526: 1; ชำนาญ สุภินทรีย์; และ สุทิน พูลสวัสดิ์. 2547; และ กรมวิชาการ. 2546: 21 – 22) โดยที่ข้อสอบเรียงความนั้นมีส่วนที่ดีเพิ่มจากแบบอื่นๆ หลายประการ เช่น 1) ให้อิสระในการตอบ เปิดโอกาสให้ผู้สอบได้เสนอความคิด และรายละเอียดต่างๆ ได้อย่างเสรี ภายใต้ขอบเขตที่กำหนดไว้ โดยไม่จำกัดทิศทางที่ผู้สอบกำหนดให้เท่านั้น 2) แก้ไขปัญหาการเดา ซึ่งในกรณีที่ผู้ตอบไม่ทราบเรื่องนั้นๆ ย่อมไม่มีโอกาสในการทำข้อสอบถูก 3) สะดวกที่จะใช้กับผู้สอบจำนวนน้อยๆ 4) ประหยัดเวลาในการทำข้อสอบ 5) ส่งเสริมความคิดนอกเนกนัย (Divergent Thinking) โดยผู้ตอบสามารถเสนอแนวคิดแปลกใหม่และริเริ่มในแนวทางใหม่ๆ (ไพศาล หวังพานิช. 2545)

เกณฑ์การให้คะแนน (Rubrics) คืออะไร?

เกณฑ์การให้คะแนน (Rubrics) คือ ชุดของแนวทางในการให้คะแนนผลการปฏิบัติเรื่องใดเรื่องหนึ่ง สำหรับใช้ประเมินคุณภาพการปฏิบัติงานของผู้เรียน ประกอบด้วยประเด็นต่างๆ (Criteria) ที่จะใช้พิจารณางานหนึ่งๆ และคำอธิบายระดับคุณภาพของแต่ละประเด็นประเมิน ซึ่งเกณฑ์การให้คะแนนจะเน้นการแสดงความสามารถของนักเรียนในด้านความเข้าใจในความคิดรวบยอด ความรู้เกี่ยวกับกระบวนการ

หรือขั้นตอน ทักษะในการแก้ปัญหา การใช้เทคโนโลยีใหม่ๆ ในการประมวลผลข้อมูล และการสื่อความหมายกับผู้อื่น

ประเภทของวิธีการให้คะแนน

สำหรับงานวิจัยชิ้นนี้ ผู้วิจัยแบ่งประเภทของวิธีการให้คะแนนไว้ 3 แบบ แต่ผู้เชี่ยวชาญส่วนใหญ่ได้แบ่งวิธีการให้คะแนนเป็น 2 แบบ (บุญเชิด ภิญญอนันตพงษ์. 2544: 90; สำนักทดสอบทางการศึกษา กรมวิชาการ. 2544: 35 – 38; และ ส. วาสนา ประवालพฤกษ์. 2533: 39 – 42) คือ

1. การตรวจให้คะแนนแบบแยกองค์ประกอบ (Analytic Scoring Rubrics) เพื่อให้การมองคุณภาพของงานหรือความสามารถของนักเรียนได้อย่างชัดเจน จึงได้มีการแยกองค์ประกอบของการให้คะแนน และอธิบายคุณภาพของงานในแต่ละองค์ประกอบเป็นระดับ

2. การตรวจให้คะแนนแบบรวมองค์ประกอบ (Holistic Scoring Rubrics) คือ การให้คะแนนงานชิ้นใดชิ้นหนึ่ง โดยดูภาพรวมของชิ้นงานว่ามีความเข้าใจในความคิดรวบยอด การสื่อความหมาย กระบวนการที่ใช้ และผลงาน แล้วเขียนอธิบายคุณภาพของงาน หรือความสำเร็จของงานเป็นชิ้นๆ

ส่วนแบบที่ 3 การตรวจให้คะแนนแบบผสมผสาน (Annotated Scoring Rubrics) เป็นการตรวจให้คะแนนที่เพิ่มขึ้นมาใหม่ (ฉัตรศิริ ปิยะพิมลสิทธิ์. 2544: ออนไลน์; อ้างอิงจาก Nitko. 1996) จะรวม Holistic และ Analytic ไว้ด้วยกัน เริ่มด้วยการให้คะแนนในภาพรวมของการปฏิบัติงานด้วย Holistic แล้วเลือกให้คะแนนอีกเพียงบางคุณลักษณะของงานแบบ Analytic ซึ่งการให้คะแนนเพียงบางคุณลักษณะนี้ จะไม่มีผลต่อการเปลี่ยนแปลงคะแนนที่ประเมินแบบ Holistic ประโยชน์ก็คือ จะมีความรวดเร็วในการให้คะแนนและเป็นการให้ผู้ตรวจให้คะแนนได้เลือกเฉพาะบางคุณลักษณะที่โดดเด่นเพียงไม่กี่องค์ประกอบเพื่อเป็นผลสะท้อน (Feedback) ให้แก่ผู้เรียน แต่ไม่มีประโยชน์ในการวินิจฉัยผู้เรียนว่าบกพร่องในคุณลักษณะใด เพราะหลายๆ คุณลักษณะไม่ได้ถูกประเมิน

รูปภาพที่ใช้เป็นโจทย์ในการเขียนเรียงความ

จากข้อสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3

คำชี้แจง: ให้นักเรียนเขียนเรื่องจากภาพ พร้อมตั้งชื่อเรื่อง เขียนความยาวไม่ต่ำกว่า 10 บรรทัด ลงกระดาษที่แจกให้ภายในเวลา 40 นาที



นิยามศัพท์เฉพาะ

1. ประเด็นสำคัญของเรียงความ หมายถึง การนำแนวคิดของรูปภาพมาตั้งชื่อเรื่อง ให้น่าสนใจ มีความคิดแปลกใหม่ สื่อความหมายได้ถูกต้องชัดเจนตามประเด็นสำคัญของภาพและใช้ถ้อยคำกะทัดรัด สอดคล้องกับประเด็นสำคัญของรูปภาพ และนำแนวคิดจากภาพและชื่อเรื่องที่ตั้งไว้มาเขียนเรื่องที่มีขอบข่ายครอบคลุมและตรงกับจุดมุ่งหมายที่วิเคราะห์ได้ ซึ่งประเด็นสำคัญของภาพคือ ผู้หญิงเงินตก 1,000 บาท เด็กผู้ชายเก็บเงินได้ และตลาดสด

2. รูปแบบการเขียนเรียงความ หมายถึง งานเขียนมีจำนวนบรรทัดตามกำหนด และมีแบบแผนตามลักษณะการเขียนเรียงความ ประกอบด้วย คำนำ เนื้อหา สรุป ซึ่งมีรายละเอียด ดังนี้

2.1 คำนำ หมายถึง บทเริ่มต้นของการเขียน เพื่อนำให้ผู้อ่านสนใจติดตามเรื่องไปโดยตลอด ด้วยวิธีการต่างๆ เช่น การยกคำพังเพย สุภาษิตที่เกี่ยวข้องการขึ้นต้นด้วยคำถามหรือปัญหาที่น่าสนใจ การบอกจุดประสงค์หรือความสำคัญที่เขียน

2.2 เนื้อเรื่อง หมายถึง การเขียนเนื้อเรื่องได้สอดคล้องชื่อเรื่อง และครอบคลุมกับภาพอย่างชัดเจน ทุกตอนของเรื่องสัมพันธ์กันโดยตลอด มีแนวคิดอย่างเดีวตลอดเรื่อง เนื้อเรื่องน่าสนใจชวนติดตาม มีการเน้นข้อคิดที่สำคัญให้เห็นอย่างเด่นชัด สอดแทรกแนวคิดที่เป็นประโยชน์ในเชิงสร้างสรรค์ แสดงเหตุผลสนับสนุนความคิดเห็นอย่างถูกต้อง

2.3 สรุป หมายถึง การกลั่นกรอง การรวบรวมใจความสำคัญ หรือการฝากข้อคิดเห็นแก่ผู้อ่านเป็นการปิดเรื่อง อาจทำได้โดยการยกสุภาษิต คำคม คำกล่าว หรือบทกวี และเป็นย่อหน้าสุดท้ายของเรื่อง

3. การนำเสนอ หมายถึง เรียบเรียงเนื้อเรื่องเป็นลำดับต่อเนื่องสัมพันธ์กันชัดเจน ตลอดตั้งแต่การเปิดเรื่อง การดำเนินเรื่อง และการปิดเรื่อง เสนอความคิดในแง่มุมต่างๆ ทุกตอนของเรื่องสัมพันธ์กันโดยตลอด มีเนื้อเรื่องน่าสนใจชวนติดตาม มีการเน้นข้อคิดที่สำคัญให้เห็นอย่างเด่นชัด สอดแทรกแนวคิดที่เป็นประโยชน์ต่อสังคม แสดงเหตุผลสนับสนุนความคิดเห็นอย่างถูกต้อง

4. การใช้ภาษา หมายถึง การเขียนสะกดคำถูกต้องทุกคำ เลือกใช้ถ้อยคำได้ตรงกับความหมาย ใช้คำสุภาพ ไม่กำกวม มีความหมายชัดเจน เข้าใจง่าย ไม่ใช้ภาษาพูดในภาษาเขียน ไม่มีการเขียนฉีกคำหรือแยกคำ และลักษณะของชิ้นงานโดยภาพรวมมีความสะอาดเรียบร้อย ไม่มีรอยลบขีดฆ่า ลายมือเป็นระเบียบสวยงาม อ่านง่ายทำงานเสร็จทันเวลาที่กำหนด โดยมีจำนวนบรรทัดทั้งหมดของงานเขียนไม่ต่ำกว่าที่โจทย์กำหนดไว้ 10 บรรทัด

ชุดการฝึกใช้เกณฑ์การตรวจให้คะแนนเรียงความแบบแยกองค์ประกอบ (Analytic Scoring Rubrics)

เกณฑ์การให้คะแนนการเขียนเรียงความในส่วนของงานเขียน สร้างจากข้อสอบการอ่านออกเสียง การเขียน และการคิดคำนวณ ชั้นประถมศึกษาปีที่ 3 ในโครงการประเมินคุณภาพการศึกษาขั้นพื้นฐาน เพื่อประกันคุณภาพผู้เรียน ปีการศึกษา 2553

วิธีการใช้เกณฑ์การตรวจให้คะแนนเรียงความแบบแยกองค์ประกอบ

1. อ่านเรียงความให้จบ 1 รอบ
2. พิจารณาให้คะแนนในประเด็นที่ 1 – 4 ที่ละประเด็นจนครบ
3. ถ้าไม่แน่ใจในประเด็นใดให้วางพักแยกไว้ก่อนแล้วจึงนำมาพิจารณาอีกครั้งภายหลัง

เกณฑ์การตรวจให้คะแนนเรียงความแบบแยกองค์ประกอบ

1. ประเด็นสำคัญของเรียงความ หมายถึง การนำแนวคิดของภาพมาตั้งชื่อเรื่องให้น่าสนใจ มีความคิดแปลกใหม่ สื่อความหมายได้ถูกต้องชัดเจนตามประเด็นสำคัญของภาพ และใช้ถ้อยคำกะทัดรัด สอดคล้องกับประเด็นสำคัญของภาพ และนำแนวคิดจากภาพและชื่อเรื่องที่ตั้งไว้มาเขียนเรื่องที่มีขอบข่ายครอบคลุมและตรงกับจุดมุ่งหมายที่วิเคราะห์ได้

ประเด็น การตรวจ ระดับคุณภาพ	1. ประเด็นสำคัญของเรียงความ
5	1. ชื่อเรื่องแปลกใหม่ น่าสนใจ ใช้ถ้อยคำกะทัดรัด สอดคล้องกับประเด็นสำคัญของภาพ 2. เนื้อเรื่องเขียนได้สอดคล้องกับชื่อเรื่องและครอบคลุมประเด็นสำคัญของภาพอย่างชัดเจน
4	1. ชื่อเรื่องน่าสนใจ ใช้ถ้อยคำเข้าใจง่าย สอดคล้องกับประเด็นสำคัญของภาพ 2. เนื้อเรื่องเขียนได้สอดคล้องกับชื่อเรื่องและครอบคลุมประเด็นสำคัญของภาพเป็นส่วนใหญ่
3	1. ชื่อเรื่องใช้ถ้อยคำไม่กระชับรัดกุม แต่สอดคล้องกับประเด็นสำคัญของภาพ 2. เนื้อเรื่องเขียนได้สอดคล้องกับชื่อเรื่องและครอบคลุมประเด็นสำคัญของภาพบางส่วน
2	1. ชื่อเรื่องใช้ถ้อยคำไม่กระชับรัดกุม แต่ชื่อเรื่องบางส่วนสอดคล้องกับประเด็นสำคัญของภาพ 2. เนื้อเรื่องเขียนได้สอดคล้องกับชื่อเรื่องแต่ไม่ครอบคลุมประเด็นสำคัญของภาพ
1	1. ตั้งชื่อเรื่องไม่สอดคล้องกับประเด็นสำคัญของภาพ 2. เนื้อเรื่องเขียนไม่สอดคล้องกับชื่อเรื่องและไม่ครอบคลุมประเด็นสำคัญของภาพ

2. รูปแบบการเขียนเรียงความ หมายถึง งานเขียนมีจำนวนบรรทัดตามกำหนดและมีแบบแผนตามลักษณะการเขียนเรียงความ ประกอบด้วย คำนำ เนื้อหา สรุป ซึ่งมีรายละเอียด ดังนี้

2.1 คำนำ หมายถึง บทเริ่มต้นของการเขียน เพื่อนำให้ผู้อ่านสนใจติดตามเรื่องไปโดยตลอดด้วยวิธีการต่างๆ เช่น การยกคำพังเพย สุภาษิตที่เกี่ยวข้อง การขึ้นต้นด้วยคำถามหรือปัญหาที่น่าสนใจ การบอกจุดประสงค์หรือความสำคัญที่เขียน

2.2 เนื้อเรื่อง หมายถึง การเขียนเนื้อเรื่องได้สอดคล้องชื่อเรื่องและครอบคลุมกับภาพอย่างชัดเจน ทุกตอนของเรื่องสัมพันธ์กันโดยตลอด มีแนวคิดอย่างเดียวตลอดเรื่อง เนื้อเรื่องน่าสนใจชวนติดตาม มีการเน้นข้อคิดที่สำคัญให้เห็นอย่างเด่นชัด สอดแทรกแนวคิดที่เป็นประโยชน์ในเชิงสร้างสรรค์ แสดงเหตุผลสนับสนุนความคิดเห็นอย่างถูกต้อง

2.3 สรุป หมายถึง การกลั่นกรอง การรวบรวมใจความสำคัญ หรือการฝากข้อคิดเห็นแก่ผู้อ่านเป็นการปิดเรื่อง อาจทำได้โดยการยกสุภาษิต คำคม คำกล่าว หรือบทกวี และเป็นย่อหน้าสุดท้ายของเรื่อง

ประเด็น การตรวจ ระดับคุณภาพ	2. รูปแบบการเขียนเรียงความ
5	1. เขียนได้ครบถ้วนทั้ง 3 ส่วน ได้แก่ คำนำ เนื้อเรื่อง สรุป 2. ความยาวของแต่ละส่วนเป็นสัดส่วนอย่างเหมาะสม 3. จำนวนบรรทัดทั้งหมดมีไม่ต่ำกว่าที่กำหนด
4	1. เขียนได้ครบถ้วนทั้ง 3 ส่วน ได้แก่ คำนำ เนื้อเรื่อง สรุป 2. ความยาวของแต่ละส่วนไม่เป็นสัดส่วนเหมาะสม 3. จำนวนบรรทัดทั้งหมดมีไม่ต่ำกว่าที่กำหนด
3	1. เขียนมีเพียง 2 ส่วน เช่น มีการเขียนเนื้อเรื่องและสรุปแต่ไม่มีการเขียนคำนำ หรือ เขียนคำนำและเนื้อเรื่องแต่ไม่มีการเขียนสรุป 2. ความยาวของแต่ละส่วนไม่เป็นสัดส่วนเหมาะสม 3. จำนวนบรรทัดทั้งหมดมีไม่ต่ำกว่าที่กำหนด
2	1. เขียนได้เพียงส่วนเดียว คือ คำนำ เนื้อเรื่อง หรือสรุป 2. ความยาวของแต่ละส่วนไม่เป็นสัดส่วนเหมาะสม 3. มีจำนวนบรรทัดทั้งหมดต่ำกว่าที่กำหนด
1	1. เขียนแล้วแยกส่วนไม่ได้ 2. ความยาวของแต่ละส่วนไม่เป็นสัดส่วนเหมาะสม 3. มีจำนวนบรรทัดทั้งหมดต่ำกว่าที่กำหนด

3. การนำเสนอ หมายถึง เรียบเรียงเนื้อเรื่องเป็นลำดับต่อเนื่องสัมพันธ์กันชัดเจนตลอดตั้งแต่การเปิดเรื่อง การดำเนินเรื่อง และการปิดเรื่อง เสนอความคิดเห็นในแง่มุมต่างๆ ทุกตอนของเรื่องสัมพันธ์กันโดยตลอด มีเนื้อเรื่องน่าสนใจชวนติดตาม มีการเน้นข้อคิดที่สำคัญให้เห็นอย่างเด่นชัด สอดแทรกแนวคิดที่เป็นประโยชน์ต่อสังคม แสดงเหตุผลสนับสนุนความคิดเห็นอย่างถูกต้อง

ประเด็น การตรวจ ระดับคุณภาพ	3. การนำเสนอ
5	1. เรียบเรียงเนื้อเรื่องเป็นลำดับต่อเนื่องสัมพันธ์กันชัดเจน 2. เสนอความคิดในแง่มุมต่างๆ เชื่อมโยงทุกตอนของเรื่องให้สัมพันธ์กันโดยตลอด 3. สอดแทรกแนวคิดที่เป็นประโยชน์ต่อสังคม 4. แสดงเหตุผลสนับสนุนความคิดเห็นอย่างถูกต้อง
4	1. เรียบเรียงเนื้อเรื่องเป็นลำดับต่อเนื่องสัมพันธ์กันค่อนข้างชัดเจน 2. เสนอความคิดในแง่มุมต่างๆ เชื่อมโยงทุกตอนของเรื่องให้สัมพันธ์กันโดยตลอด 3. สอดแทรกแนวคิดที่เป็นประโยชน์ต่อสังคม 4. แสดงเหตุผลสนับสนุนความคิดเห็นชัดเจน
3	1. เรียบเรียงเนื้อเรื่องสับสน วากวน แต่ยังอยู่ในกรอบ 2. เสนอความคิดในบางตอนไม่เชื่อมโยงสัมพันธ์กัน 3. สอดแทรกแนวคิดที่เป็นประโยชน์เล็กน้อย 4. แสดงเหตุผลสนับสนุนความคิดเห็นบางส่วน
2	1. เรียบเรียงเนื้อเรื่องสับสน วากวน 2. เสนอความคิดไม่เชื่อมโยงสัมพันธ์กัน 3. สอดแทรกแนวคิดที่เป็นประโยชน์น้อย 4. ขาดการแสดงเหตุผลสนับสนุนความคิดเห็น
1	1. เรียบเรียงเนื้อเรื่องสับสน ทำให้ผู้อ่านเข้าใจผิด 2. เสนอความคิดไม่เชื่อมโยงสัมพันธ์กัน 3. ไม่ได้สอดแทรกแนวคิดที่เป็นประโยชน์ 4. ขาดการแสดงเหตุผลสนับสนุนความคิดเห็น

4. การใช้ภาษา หมายถึง การเขียนสะกดคำถูกต้องทุกคำ เลือกใช้ถ้อยคำได้ตรงกับความหมาย ใช้คำสุภาพ ไม่กำกวม มีความหมายชัดเจน เข้าใจง่าย ไม่ใช้ภาษาพูดในภาษาเขียน ไม่มีการเขียนฉีกคำหรือแยกคำ และลักษณะของชิ้นงานโดยภาพรวมมีความสะอาดเรียบร้อย ไม่มีรอยลบขีดฆ่า ลายมือเป็นระเบียบสวยงาม อ่านง่ายทำงานเสร็จทันเวลาที่กำหนด โดยมีจำนวนบรรทัดทั้งหมดของงานเขียนไม่ต่ำกว่าที่กำหนด

ประเด็น การตรวจ ระดับคุณภาพ	4. การใช้ภาษา
5	<ol style="list-style-type: none"> 1. สะกดคำถูกต้องทุกคำ 2. เลือกใช้ถ้อยคำได้ตรงกับความหมายใช้คำสุภาพไม่ใช้ภาษาพูดในภาษาเขียน 3. ไม่มีการเขียนฉีกคำหรือแยกคำเลยตลอดทั้งเรื่อง 4. ลายมือเป็นระเบียบสวยงาม อ่านง่าย สะอาด 5. ไม่มีรอยลบขีดฆ่า
4	<ol style="list-style-type: none"> 1. สะกดคำผิด 1 – 3 คำ 2. เลือกใช้ถ้อยคำได้ตรงความหมาย ใช้คำสุภาพ 3. ฉีกคำหรือแยกคำ 1 – 3 แห่ง 4. ลายมืออ่านง่าย 5. มีรอยลบขีดฆ่า 1 – 3 แห่ง
3	<ol style="list-style-type: none"> 1. สะกดคำผิด 4 – 5 คำ 2. ใช้ถ้อยคำไม่เหมาะสมบางแห่ง 3. ฉีกคำหรือแยกคำ 4 – 5 แห่ง 4. ลายมืออ่านง่าย 5. มีรอยลบขีดฆ่า 4 – 5 แห่ง
2	<ol style="list-style-type: none"> 1. เขียนสะกดคำผิดตั้งแต่ 6 – 9 คำ 2. ใช้ถ้อยคำไม่เหมาะสมบางแห่ง 3. ฉีกคำหรือแยกคำ 6 – 9 แห่ง 4. ลายมือพออ่านออก 5. มีรอยลบขีดฆ่า 6 – 9 แห่ง
1	<ol style="list-style-type: none"> 1. เขียนสะกดคำผิดตั้งแต่ 10 คำขึ้นไป 2. ใช้ถ้อยคำไม่เหมาะสมเป็นส่วนใหญ่ 3. ฉีกคำหรือแยกคำตั้งแต่ 10 แห่งขึ้นไป 4. ลายมืออ่านไม่ออก 5. มีรอยลบขีดฆ่าตั้งแต่ 10 แห่ง ขึ้นไป

ตัวอย่าง

การตรวจให้คะแนนเรียงความแบบแยกองค์ประกอบ (Analytic Scoring Rubrics)

ผู้ตรวจให้คะแนนคือ บุชวรินทร์ แสนปลื้ม

ชื่อ - นามสกุล	คะแนน					หมายเหตุ
	1. ประเด็นสำคัญของเรียงความ	2. รูปแบบการเขียนเรียงความ	3. การนำเสนอ	4. การใช้ภาษา	รวม	
1. ด.ญ.กมลชนก	2	1	2	4	10	
2. ด.ญ.ครองขวัญ	5	5	5	5	20	
3. ด.ช.ปฎิภาณ	2	3	2	2	9	
4. ด.ญ.ภาวิณี	1	1	2	4	8	
5. ด.ช.ฤทธิ์ไกร	3	3	4	3	13	
6. ด.ญ.จิตาภา						
7. ด.ญ.แตงอ่อน						
8. ด.ญ.ภูษณิษา						
9. ด.ญ.วนัทปรียา						
10. ด.ช.จิรจรัส						

ความสามารถในการเขียนเรื่องจากภาพ ชั้นประถมศึกษาปีที่ ๓

ชื่อ.ณ.ญ. กมลชนก นามสกุล.....

เลขที่.๒๑ เลขประจำตัวประชาชน

โรงเรียน.....อำเภอ.เสด็จ จังหวัด.สุรินทร์

ชื่อเรื่อง

คนที่มีหัวใจ

ชื่อเรื่อง

มีอยู่วันหนึ่งมีคนคนหนึ่งมาเห็นเงินค่า
ขนมหนึ่งพันบาทอยู่ในตะกร้าของเขา
เขาก็มีผู้หนึ่งมาบอกให้เขารวยมาก
และเป็นคนรวยในตำบลเล็ก ๆ และ
ยังมีหัวใจดีใจอีกต่างหากและมีผู้
คนหนึ่งมาบอกให้เขาเลิกกินเนื้อ
หมูและเลิกกินอาหารที่มีไขมันสูง
คนคนหนึ่งหัวใจดีมาก ๆ และยังมีหัวใจ
ดีอีก

ตัวอย่างที่ 1

- 1 ประเด็นสำคัญของเรียงความ 2 คะแนน**
- 1) ชื่อเรื่องใช้ถ้อยคำไม่กระชับรัดกุม แต่ชื่อเรื่องบางส่วนสอดคล้องกับประเด็นสำคัญของภาพ 2) เนื้อเรื่องเขียนได้สอดคล้องกับชื่อเรื่องแต่ไม่ครอบคลุมประเด็นสำคัญของภาพ

- 2 รูปแบบการเขียนเรียงความ 1 คะแนน**
- 1) เขียนแล้วแยกส่วนไม่ได้ 2) ความยาวของแต่ละส่วนไม่เป็นสัดส่วนเหมาะสม 3) มีจำนวนบรรทัดทั้งหมดต่ำกว่าที่กำหนด

- 3 การนำเสนอ 2 คะแนน**
- 1) เรียบเรียงเนื้อเรื่องสับสน วกวน 2) เสนอความคิดไม่เชื่อมโยงสัมพันธ์กัน 3) สอดแทรกแนวคิดที่เป็นประโยชน์น้อย 4) ขาดการแสดงเหตุผลสนับสนุนความคิดเห็น

- 4 การใช้ภาษา 4 คะแนน**
- 1) สะกดคำผิด 1 – 3 คำ 2) เลือกใช้ถ้อยคำได้ตรงความหมายใช้คำสุภาพ 3) ฉีกคำหรือแยกคำ 1 – 3 แห่ง 4) ลายมืออ่านง่าย 5) มีรอยลบขีดฆ่า 1 – 3 แห่ง

ความสามารถในการเขียนเรื่องจากภาพ ชั้นประถมศึกษาปีที่ ๓

ชื่อสกุล.....นามสกุล.....

เลขที่.....เลขประจำตัวประชาชน.....

โรงเรียน.....อำเภอ.....จังหวัด.....

ชื่อเรื่อง.....
คนอีขของเรา

ตัวอย่างที่ 2

ที่ตลาดมีร้านค้ามากมายมีผลไม้
หน้ากินผักสดสะอาดมากมีไก่สดขาย
มีคนซื้อผลไม้ไปหุงต้มมีน้ำใ้ใจมาก
แม่ฉันไปตลาดจะไปซื้อผลไม้แม่
ฉันกำลังเดินไปซื้อผลไม้เห็นแม่ฉัน
ตกมีคนเก็บคนที่เก็บแม่คือคนที่
แม่ฉันชอบคนเฒ่าฉันก็ซื้อผลไม้
ให้แม่ซื้อผลไม้ให้แม่ซื้อไก่สดให้กิน
แม่ฉันก็ซื้อผลไม้มากมาขายมาก
บ้านและฉันก็น้องก็กินกันอร่อย
มาก ๆ และครอบครัวฉันก็มีความสุข

๑ ประเด็นสำคัญของเรียงความ 5 คะแนน

1) ชื่อเรื่องแปลกใหม่ น่าสนใจ ใช้ถ้อยคำกะทัดรัด สอดคล้องกับประเด็นสำคัญของภาพ 2) เนื้อเรื่องเขียนได้สอดคล้องกับชื่อเรื่องและครอบคลุมประเด็นสำคัญของภาพอย่างชัดเจน

๒ รูปแบบการเขียนเรียงความ 5 คะแนน

1) เขียนได้ครบถ้วนทั้ง 3 ส่วน ได้แก่ คำนำ เนื้อเรื่อง สรุป 2) ความยาวของแต่ละส่วนเป็นสัดส่วนอย่างเหมาะสม 3) จำนวนบรรทัดทั้งหมดมีไม่ต่ำกว่าที่กำหนด

๓ การนำเสนอ 5 คะแนน

1) เรียบเรียงเนื้อเรื่องเป็นลำดับต่อเนื่องสัมพันธ์กันชัดเจน 2) เสนอความคิดในแง่มุมต่างๆ เชื่อมโยงทุกตอนของเรื่องให้สัมพันธ์กันโดยตลอด 3) สอดแทรกแนวคิดที่เป็นประโยชน์ต่อสังคม 4) แสดงเหตุผลสนับสนุนความคิดเห็นอย่างถูกต้อง

๔ การใช้ภาษา 5 คะแนน

1. สะกดคำถูกต้องทุกคำ 2. เลือกใช้ถ้อยคำได้ตรงกับความหมายใช้คำสุภาพ ไม่ใช้ภาษาพูดในภาษาเขียน 3. ไม่มีการเขียนฉีกคำหรือแยกคำเลยตลอดทั้งเรื่อง 4. ลายมือเป็นระเบียบสวยงาม อ่านง่าย สะอาด 5. ไม่มีรอยลบขีดฆ่า

ความสามารถในการเขียนเรื่องจากภาพ ชั้นประถมศึกษาปีที่ ๓

ชื่อ...ศ.ศ.ปฏิภาณ.....นามสกุล.....

เลขที่...๖.๖. เลขประจำตัวประชาชน

โรงเรียน..... ีเภอ...เมือง.....จังหวัด...สุรินทร์.....

ชื่อเรื่อง...ศตวรรษที่ ๖๖..... ชื่อเรื่อง

วันหนึ่งผมไปศตวรรษที่ ๖๖
 ซึ่งกำลังมีชื่อพักพิงที่ ผมมาจาก
 ศตวรรษที่ ๖๖ ซึ่งผมพักพิงที่มีผมจะ
 สักผม ๖๖ ๖๖ ๖๖ สักผม
 ผมระก้อ (๖๖) พักพัก
 ๖๖ ๖๖ ๖๖ ๖๖ ๖๖ ๖๖
 ๖๖ ๖๖ ๖๖ ๖๖ ๖๖ ๖๖
 ๖๖ ๖๖ ๖๖ ๖๖ ๖๖ ๖๖
 ๖๖ ๖๖ ๖๖ ๖๖ ๖๖ ๖๖
 ๖๖ ๖๖ ๖๖ ๖๖ ๖๖ ๖๖
 ๖๖ ๖๖ ๖๖ ๖๖ ๖๖ ๖๖

ตัวอย่างที่ 3

1 ประเด็นสำคัญของเรียงความ 2 คะแนน
 1) ชื่อเรื่องใช้ถ้อยคำไม่กระชับรัดกุม แต่ชื่อเรื่องบางส่วนสอดคล้องกับประเด็นสำคัญของภาพ 2) เนื้อเรื่องเขียนได้สอดคล้องกับชื่อเรื่องแต่ไม่ครอบคลุมประเด็นสำคัญของภาพ***

2 รูปแบบการเขียนเรียงความ 3 คะแนน
 1) เขียนมีเพียง 2 ส่วน เขียนคำนำและเนื้อเรื่องแต่ไม่มีการเขียนสรุป 2) ความยาวของแต่ละส่วนไม่เป็นสัดส่วนเหมาะสม 3) จำนวนบรรทัดทั้งหมดมีไม่ต่ำกว่าที่กำหนด

3 การนำเสนอ 2 คะแนน
 1) เรียบเรียงเนื้อเรื่องสับสน วกวน 2) เสนอความคิดไม่เชื่อมโยงสัมพันธ์กัน 3) สอดแทรกแนวคิดที่เป็นประโยชน์น้อย 4) ขาดการแสดงเหตุผลสนับสนุนความคิดเห็น

4 การใช้ภาษา 2 คะแนน
 1) เขียนสะกดคำผิด 6 - 9 คำ 2) ใช้ถ้อยคำไม่เหมาะสมบางแห่ง 3) ฉีกคำหรือแยกคำ 6-9 แห่ง 4) ลายมือพ้ออ่านออก 5) มีรอยลบขีดฆ่า 6 - 9*** แห่ง

ความสามารถในการเขียนเรื่องจากภาพ ชั้นประถมศึกษาปีที่ ๓

ชื่อ.....นามสกุล.....

เลขที่.....เลขประจำตัวประชาชน.....

โรงเรียน.....อำเภอ.....จังหวัด.....

ชื่อเรื่อง.....

เช้าวันหนึ่งอากาศสดใสมีนกหลายตัวบินเล่นในสวน
 เห็นปลาหลายตัวว่ายเล่นในบ่อน้ำในสวน
 คนที่สวนเห็นนกก็เอาน้ำไปรดต้นไม้ในสวน
 นกก็บินมาดื่มน้ำที่รดต้นไม้
 ปลาเห็นคนก็ว่ายมาใกล้คน
 คนก็เอาน้ำไปรดต้นไม้
 นกก็บินมาดื่มน้ำที่รดต้นไม้
 ปลาเห็นคนก็ว่ายมาใกล้คน

.....

ตัวอย่างที่ 4

1 ประเด็นสำคัญของเรียงความ 1 คะแนน

1) ตั้งชื่อเรื่องไม่สอดคล้องกับประเด็นสำคัญของภาพ 2) เนื้อเรื่องเขียนไม่สอดคล้องกับชื่อเรื่องและไม่ครอบคลุมประเด็นสำคัญของภาพ

2 รูปแบบการเขียนเรียงความ 1 คะแนน

1) เขียนแล้วแยกส่วนไม่ได้ 2) ความยาวของแต่ละส่วนไม่เป็นสัดส่วนเหมาะสม

3 การนำเสนอ 2 คะแนน

1) เรียบเรียงเนื้อเรื่องสับสน ทำให้ผู้อ่านเข้าใจผิด 2) เสนอความคิดไม่เชื่อมโยงสัมพันธ์กัน 3) ไม่ได้สอดแทรกแนวคิดที่เป็นประโยชน์ 4) ขาดการแสดงเหตุผลสนับสนุนความคิดเห็น

4 การใช้ภาษา 4 คะแนน

1) สะกดคำผิด 1 – 3 คำ 2) เลือกใช้ถ้อยคำได้ตรงความหมายใช้คำสุภาพ 3) ฉีกคำหรือแยกคำ 1 – 3 แห่ง 4) ลายมืออ่านง่าย 5) มีรอยลบขีดฆ่า 1 – 3 แห่ง

ความสามารถในการเขียนเรื่องจากภาพ ชั้นประถมศึกษาปีที่ ๓

ชื่อ... นามสกุล...

เลขที่... เลขประจำตัวประชาชน

โรงเรียน... อำเภอ... จังหวัด...

ชื่อเรื่อง...

มีผู้หญิงคนหนึ่งเดินลุยฝนในตลาดแล้วแบกทำเงิน
ตก 1000 บาท พี่ชายที่เดินตามหลังเขาก็เห็นพอดี
เขาก็เลยจะเก็บให้ผู้หญิงคนนั้นและเขาก็เดิน
มอริไม่ลงมาลงเห็นแม่ค้าก็พูดในว่าเอาพี่กส่าไม่
คะเต้ไก่สอไม่ครั้นเตาผลสไมไมคะผู้หญิงคนนั้น
เห็นมะละกอกก็เลยผูกมากินก็เลยหนีไปกระเฝ้าเงิน
มาแล้วผู้หญิงคนนั้นก็บอกกับแม่ค้าว่าฉันทำเงิน
ตก 1000 บาทแล้วพี่ชายคนนั้นก็เห็นผู้หญิงคนนั้นก็
ทำเงินตกพี่ชายคนนั้นก็เลยวิ่งไปหาและบอก
นี่สิเงินของคุณผู้หญิงคนนั้นก็เลยมีเงินที่
จะซื้อมะละกอกกินอย่างอร่อยมาก ๆ เลย

ตัวอย่างที่ 5

1 ประเด็นสำคัญของเรียงความ 3 คะแนน

- 1) ชื่อเรื่องใช้ถ้อยคำไม่กระชับรัดกุม แต่สอดคล้องกับประเด็นสำคัญของภาพ 2) เนื้อเรื่องเขียนได้สอดคล้องกับชื่อเรื่องและครอบคลุมประเด็นสำคัญของภาพบางส่วน

2 รูปแบบการเขียนเรียงความ 3 คะแนน

- 1) เขียนมีเพียง 2 ส่วน เขียนคำนำและเนื้อเรื่องแต่ไม่มีการเขียนสรุป 2) ความยาวของแต่ละส่วนไม่เป็นสัดส่วนเหมาะสม 3) จำนวนบรรทัดทั้งหมดมีไม่ต่ำกว่าที่กำหนด

3 การนำเสนอ 4 คะแนน

- 1) เรียบเรียงเนื้อเรื่องเป็นลำดับต่อเนื่องสัมพันธ์กันค่อนข้างชัดเจน 2) เสนอความคิดในแง่มุมต่างๆ เชื่อมโยงทุกตอนของเรื่องให้สัมพันธ์กันโดยตลอด 3) สอดแทรกแนวคิดที่เป็นประโยชน์ต่อสังคม 4) แสดงเหตุผลสนับสนุนความคิดเห็นชัดเจน

4 การใช้ภาษา 3 คะแนน

- 1) สะกดคำผิด 4 - 5 คำ 2) ใช้ถ้อยคำไม่เหมาะสมบางแห่ง 3) ฉีกคำหรือแยกคำ 4 - 5 แห่ง 4) ลายมืออ่านง่าย 5) มีรอยลบขีดฆ่า 4 - 5 แห่ง

ชุดการฝึกใช้เกณฑ์การตรวจให้คะแนนเรียงความแบบรวมองค์ประกอบ (Holistic Scoring Rubrics)

วิธีการใช้เกณฑ์การตรวจให้คะแนนเรียงความแบบรวมองค์ประกอบ

1. พิจารณาภาพรวมและแบ่งกระดาษออกเป็น 5 กอง คือ กอง 5 คะแนน กอง 4 คะแนน กอง 3 คะแนน กอง 2 คะแนน กอง 1 คะแนน
2. เรียงความชิ้นใดถ้าผู้ตรวจไม่แน่ใจว่าจะอยู่กลุ่มใดให้ผู้ตรวจทำเครื่องหมายไว้และวางไว้กองที่น่าจะอยู่กับกองนั้นๆ หลังจากนั้นจึงนำมาพิจารณาอีกครั้ง

คำอธิบาย
<p>ระดับคุณภาพ 5 คะแนน</p> <ol style="list-style-type: none"> 1. ชื่อเรื่องแปลกใหม่ น่าสนใจ ใช้ถ้อยคำกะทัดรัด สอดคล้องกับประเด็นสำคัญของภาพ และเนื้อเรื่องเขียนได้สอดคล้องกับชื่อเรื่องและครอบคลุมประเด็นสำคัญอย่างชัดเจน 2. เขียนได้ครบถ้วนทั้ง 3 ส่วน ได้แก่ คำนำ เนื้อเรื่อง สรุป โดยที่ความยาวของแต่ละส่วนเป็นสัดส่วนอย่างเหมาะสม และมีจำนวนบรรทัดทั้งหมดไม่ต่ำกว่าที่กำหนด 3. เรียบเรียงเนื้อเรื่องเป็นลำดับต่อเนื่องสัมพันธ์กันชัดเจน เชื่อมโยงทุกตอนของเรื่องให้สัมพันธ์กันโดยตลอด สอดแทรกแนวคิดแสดงเหตุผลที่เป็นประโยชน์ต่อสังคม 4. สะกดคำถูกต้องทุกคำ เลือกใช้ถ้อยคำได้ตรงกับความหมาย ใช้คำสุภาพ ไม่ใช้ภาษาพูดในภาษาเขียน ไม่มีการเขียนฉีกคำหรือแยกคำเลยตลอดทั้งเรื่อง ลายมือเป็นระเบียบสวยงาม อ่านง่าย สะอาด ไม่มีรอยลบขีดฆ่า
<p>ระดับคุณภาพ 4 คะแนน</p> <ol style="list-style-type: none"> 1. ชื่อเรื่องน่าสนใจ ใช้ถ้อยคำเข้าใจง่าย สอดคล้องกับประเด็นสำคัญของภาพ และเนื้อเรื่อง เขียนได้สอดคล้องกับชื่อเรื่องและครอบคลุมประเด็นสำคัญเป็นส่วนมาก 2. เขียนได้ครบถ้วนทั้ง 3 ส่วน ได้แก่ คำนำ เนื้อเรื่อง สรุป โดยที่ความยาวของแต่ละส่วนไม่เป็นสัดส่วนเหมาะสม และมีจำนวนบรรทัดทั้งหมดไม่ต่ำกว่าที่กำหนด 3. เรียบเรียงเนื้อเรื่องเป็นลำดับต่อเนื่องสัมพันธ์กัน ค่อนข้างชัดเจน เชื่อมโยงทุกตอนของเรื่อง ให้สัมพันธ์กันโดยตลอด สอดแทรกแนวคิดและแสดงเหตุผลที่เป็นประโยชน์ต่อสังคม 4. สะกดคำผิด 1 – 3 คำ เลือกใช้ถ้อยคำได้ตรงความหมาย ใช้คำสุภาพ เขียนฉีกคำหรือแยกคำ 1 – 3 แห่ง ลายมืออ่านง่าย มีรอยลบขีดฆ่า 1 – 3 แห่ง

คำอธิบาย

ระดับคุณภาพ 3 คะแนน

1. ชื่อเรื่องใช้ถ้อยคำไม่กระชับรัดกุม แต่สอดคล้องกับประเด็นสำคัญของภาพ และเนื้อเรื่องเขียนได้สอดคล้องกับชื่อเรื่องและครอบคลุมประเด็นสำคัญบางส่วน
2. เขียนมีเพียง 2 ส่วน เช่น มีการเขียนเนื้อเรื่องและสรุป แต่ไม่มีการเขียนคำนำ หรือเขียนคำนำ และเนื้อเรื่องแต่ไม่มีการเขียนสรุป โดยที่ความยาวของแต่ละส่วนไม่เป็นสัดส่วนเหมาะสม มีจำนวนบรรทัดทั้งหมดไม่ต่ำกว่าที่กำหนด
3. เรียบเรียงเนื้อเรื่องสับสน วกวน แต่ยังอยู่ในกรอบ เสนอความคิดในบางตอนไม่เชื่อมโยงสัมพันธ์กัน สอดแทรกแนวคิดและแสดงเหตุผลที่เป็นประโยชน์เล็กน้อย
4. สะกดคำผิด 4 – 5 คำ ใช้ถ้อยคำไม่เหมาะสมบางแห่ง เขียนฉีกคำหรือแยกคำ 4 – 5 แห่ง ลายมืออ่านง่าย มีรอยลบขีดฆ่า 4 – 5 แห่ง

ระดับคุณภาพ 2 คะแนน

1. ชื่อเรื่องใช้ถ้อยคำไม่กระชับรัดกุม แต่ชื่อเรื่องบางส่วนสอดคล้องกับประเด็นสำคัญของภาพ และเนื้อเรื่องเขียนได้สอดคล้องกับชื่อเรื่องแต่ไม่ครอบคลุมประเด็นสำคัญ
2. เขียนได้เพียงส่วนเดียว คือ คำนำ เนื้อเรื่อง หรือสรุป และมีจำนวนบรรทัดทั้งหมดต่ำกว่าที่กำหนด
3. เรียบเรียงเนื้อเรื่องสับสน วกวน เสนอความคิดไม่เชื่อมโยงสัมพันธ์กัน สอดแทรกแนวคิดที่เป็นประโยชน์น้อย และขาดการแสดงเหตุผลสนับสนุนความคิดเห็น
4. เขียนสะกดคำผิด 6 – 9 คำ ใช้ถ้อยคำไม่เหมาะสมบางแห่ง เขียนฉีกคำหรือแยกคำ 6 – 9 แห่ง ลายมือพออ่านออก มีรอยลบขีดฆ่า 6 – 9 แห่ง

ระดับคุณภาพ 1 คะแนน

1. ตั้งชื่อเรื่องไม่สอดคล้องกับประเด็นสำคัญของภาพ และเนื้อเรื่องเขียนไม่สอดคล้องกับชื่อเรื่อง และไม่ครอบคลุมประเด็นสำคัญ
2. เขียนแล้วแยกส่วนไม่ได้ และมีจำนวนบรรทัดทั้งหมดต่ำกว่าที่กำหนด
3. เรียบเรียงเนื้อเรื่องสับสน ทำให้ผู้อ่านเข้าใจผิด เสนอความคิดไม่เชื่อมโยงสัมพันธ์กัน ไม่ได้สอดแทรกแนวคิดที่เป็นประโยชน์ และขาดการแสดงเหตุผลสนับสนุนความคิดเห็น
4. เขียนสะกดคำผิดตั้งแต่ 10 คำ ขึ้นไป ใช้ถ้อยคำไม่เหมาะสมเป็นส่วนใหญ่ เขียนฉีกคำหรือแยกคำตั้งแต่ 10 แห่งขึ้นไป ลายมืออ่านไม่ออก มีรอยลบขีดฆ่าตั้งแต่ 10 แห่ง ขึ้นไป

ชุดการฝึกใช้เกณฑ์การตรวจให้คะแนนเรียงความแบบผสมผสาน (Annotated Scoring Rubrics)

วิธีการใช้เกณฑ์การตรวจให้คะแนนเรียงความแบบผสมผสาน

ขั้นตอนที่ 1 ตรวจให้คะแนนโดยใช้วิธีการให้คะแนนแบบรวมองค์ประกอบ

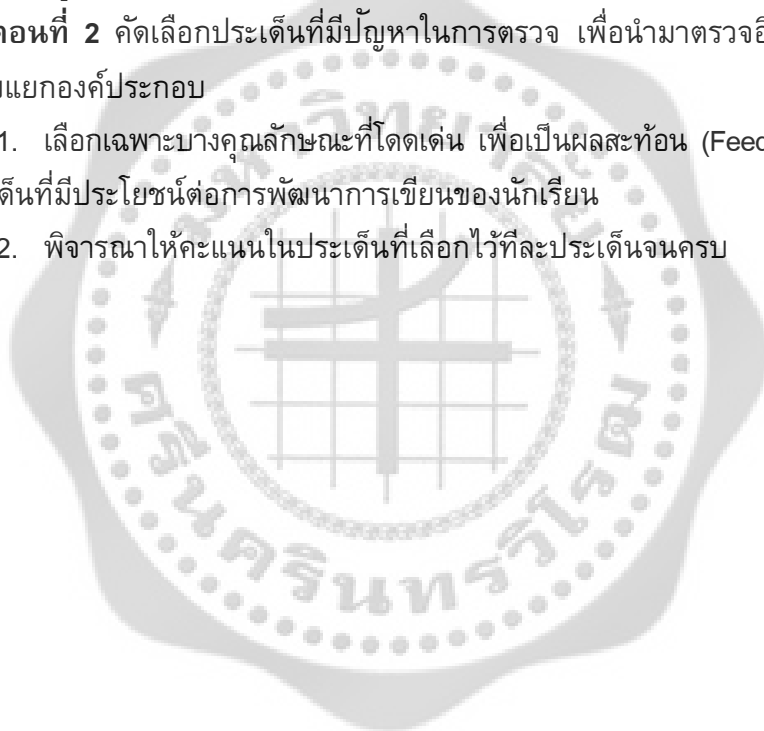
1. พิจารณาภาพรวมและแบ่งกระดาษออกเป็น 5 กอง คือ กอง 5 คะแนน กอง 4 คะแนน กอง 3 คะแนน กอง 2 คะแนน กอง 1 คะแนน

2. เรียงความขึ้นไต่ถ้าผู้ตรวจไม่แน่ใจว่าจะอยู่กลุ่มใดให้ผู้ตรวจทำเครื่องหมายไว้ และวางไว้กองที่น่าจะอยู่กับกองนั้นๆ หลังจากนั้นจึงนำมาพิจารณาอีกครั้ง

ขั้นตอนที่ 2 คัดเลือกประเด็นที่มีปัญหาในการตรวจ เพื่อนำมาตรวจอีกครั้งโดยใช้วิธีการให้คะแนนแบบแยกองค์ประกอบ

1. เลือกเฉพาะบางคุณลักษณะที่โดดเด่น เพื่อเป็นผลสะท้อน (Feedback) ให้แก่ผู้เรียน โดยเลือกประเด็นที่มีประโยชน์ต่อการพัฒนาการเขียนของนักเรียน

2. พิจารณาให้คะแนนในประเด็นที่เลือกไว้ทีละประเด็นจนครบ



**ชุดการฝึกใช้เกณฑ์การตรวจให้คะแนนเรียงความแบบผสมผสาน
(Annotated Scoring Rubrics)**

ผู้ตรวจให้คะแนนคือ บุชวรินทร์ แส่นปลื้ม

ชื่อ-นามสกุล	คะแนนที่ได้จาก Holistic	คะแนน				หมายเหตุ
		1. ประเด็นสำคัญของเรียงความ	2. รูปแบบการเขียนเรียงความ	3. การนำเสนอ	4. การใช้ภาษา	
1. ด.ญ.กมลชนก	3	2		2		
2. ด.ญ.ครองขวัญ	5	5		5		
3. ด.ช.ปฎิภาณ	2	2		2		
4. ด.ญ.ภาวิณี	1	1		2		
5. ด.ช.ฤทธิ์ไกร	4	3		4		
6. ด.ญ.จิตาภา						
7. ด.ญ.แตงอ่อน						
8. ด.ญ.ภูษนิษา						
9. ด.ญ.วนัทปรียา						
10. ด.ช.จิรจรัส						

แบบฝึกหัด

คำสั่ง: จงตรวจให้คะแนนการเขียนเรียงความต่อไปนี้ โดยบันทึกคะแนนลงในคู่มือที่ปรากฏในบทเรียน

ความสามารถในการเขียนเร็ว ม.ประถมศึกษาปีที่ ๓

ชื่อ. น. น. กมลชนก นามสกุล.....

เลขที่. ๒๑ เลขประจำตัวประชาชน

โรงเรียน อำเภอ. เมือง จังหวัด. สุรินทร์

ชื่อเรื่อง คนที่มีหัวใจ ชื่อเรื่อง

มีอยู่วันหนึ่งมีคนคนหนึ่งมาเห็นหมาเห็นคน
 คนหนึ่งตั้งหน้าตาอยู่ในศาลาตาตาแห่ง
 ใด ๆ มีผู้หญิงคนหนึ่งช่วยมาก
 และเป็นคนช่วยในต่าง ๆ ได้ ๆ และ
 ยังมีหัวใจดีใจเสียอีกต่างหากและมีผู้
 คนคนหนึ่งมาอีกทั้งหาว่ามีคนรักแล้ว
 คนหนึ่งคน ลูกคนตาตาคนมีตาคน
 คนคนหนึ่งหัวใจมาก ๆ และยังมีใจ
 ดีอีก

ความสามารถในการเขียนเรื่องจากภาพ ชั้นประถมศึกษาปีที่ ๓

ชื่อ.....นามสกุล.....

เลขที่..... เลขประจำตัวประชาชน.....

โรงเรียน.....อำเภอ.....จังหวัด.....

ชื่อเรื่อง..... คนอีขงเรา..... ชื่อเรื่อง.....

ที่ตลาดมีร้านค้ามากมายมีผลไม้
หน้าก็หนักก็สดสะอาดมากมีไก่สดขาย
มีคนซื้อก็ไปหมดมีคนหันหน้าใจมาก
แม่ฉันไปตลาดจะไปซื้อผลไม้แม่
ฉันกำลังเดินไปซื้อผลไม้เห็นแม่ฉัน
ตกมีคนเก็บคนที่เก็บแม่คือคนอีขง
แม่ฉันชอบคนอีขงแม่ฉันก็ซื้อผลไม้
ให้และซื้อของให้และซื้อไก่สดให้
แม่ฉันก็ซื้อผลไม้มากมายมาก
บ้านและฉันก็นั่งกินกันอย่าง
มากและครอบครัวฉันก็มีความสุข

ความสามารถในการเขียนเรื่องจากภาพ ชั้นประถมศึกษาปีที่ ๓

ชื่อ... ๓๒ ปรากฏาณ.....นามสกุล.....

เลขที่... ๒๗... เลขประจำตัวประชาชน

โรงเรียน..... อำเภอ... ๖๕๑๑ จังหวัด... สุรินทร์.....

ชื่อเรื่อง

..... ตลาด๖๖สี่นสี่น

ชื่อเรื่อง

วันหนึ่งฉันเดินไปตลาด๖๖สี่น
สี่นก็เข้าไปซื้อผักผลไม้ มาจาก
ตลาด๖๖สี่นสี่นผลไม้มีมะม่วง
ส้ม กล้วย สับปะรด
มะระกาว กล้วย ฟักทอง
แตงหอม ผักชี กระแตสี่นสี่น
๖๖สี่นสี่นสี่นสี่น ๖๖ ๖๖สี่นสี่น
๖๖ สี่นสี่นสี่นสี่นสี่นสี่นสี่น
สี่นสี่นสี่นสี่นสี่นสี่นสี่นสี่น
สี่นสี่นสี่นสี่นสี่นสี่นสี่นสี่น
สี่นสี่นสี่นสี่นสี่นสี่นสี่นสี่น
สี่นสี่นสี่นสี่นสี่นสี่นสี่นสี่น

ความสามารถในการเขียนเรื่องจากภาพ ชั้นประถมศึกษาปีที่ ๓

ชื่อ.ญาติไกร.....นามสกุล.....

เลขที่.....๙.....เลขประจำตัวประชาชน

โรงเรียน.....อำเภอเมืองสุรินทร์.....จังหวัดสุรินทร์

ชื่อเรื่อง ตลาดสดของเมืองไทยอยู่สูงสงโงาชนชาติอื่น ชื่อเรื่อง

มีผู้หญิงคนหนึ่งเดินดูสิ่งในตลาดแล้วเขาทำเงิน
 ตก 1000 บาทผู้ชายที่เดินตามหลังเขาก็เห็นพอ
 ดิเขาก็เลขจะเก็บได้ผู้หญิงคนนั้นและเขาก็เดิน
 มุ่งไปมองมาคนนั้นแม่ค้าก็พูดไปว่าเขาพักสตาไม
 ค่ะเขาไปสตาไมครั้นเขาผลไม้มันคะผู้หญิงคนนั้น
 เห็นมะละกอกก็เลขออกมาก็นักเลขหิมกระเข้าเห็น
 มาแล้วผู้หญิงคนนั้นก็ขสกันแม่ค้าว่าฉันทำเงิน
 ตก 1000 บาทแล้วผู้ชายคนนั้นก็เห็นผู้หญิงคนที่
 ทำเงินตกผู้ชายคนนั้นก็เลขวิ่งไปหาและบอกว
 ้นครั้นเงินของคุณผู้หญิงคนนั้นก็เลขมีเงินที่
 จะซื้อมะละกอกก็เลขมากรรณมาก ๆ เลข

ความสามารถในการเขียนเรื่องจากภาพ ชั้นประถมศึกษาปีที่ ๓

ชื่อ.....นามสกุล.....

เลขที่.....เลขประจำตัวประชาชน.....

โรงเรียน.....อำเภอ.....จังหวัด.....

ชื่อเรื่อง ความกล้าหาญ ชื่อเรื่อง

วันหนึ่งแม่ไปตลาดจะซื้อผัก
ผักผลไม้ และ ก็นั่งข้างข้าง
ข้างนั้น ผัก

แล้วแม่ก็ทำคนง้อสิบกว่า
บาท มีผู้ทำคนง้อหนึ่งคน
สิบกว่าบาทจึงถือมาแม่
ขอบคุณแม่แม่บอกว่าเป็น
คนดีพวกเธอควรทำแบบอย่าง
ความดีของเธอ

ความกล้าหาญอยู่ที่ไหนความ
ดีอยู่ที่ไหน

ความสามารถในการเขียนเรื่องจากภาพ ชั้นประถมศึกษาปีที่ ๓

ชื่อ.....นามสกุล.....

เลขที่..... เลขประจำตัวประชาชน

โรงเรียน.....อำเภอ.....จังหวัด.....

ชื่อเรื่อง..... คนตีมิวห่วย..... ชื่อเรื่อง

มีผู้ชื้อสินค้าคนหนึ่งชื่อน้ำวันชื่อน้ำต๋องไปตลาด
น้ำวันชื้อข้าวน้ำต๋องชื้อผลไม้ไปชื้ออาหารเพื่อ
ทำอาหารเย็นตอนนั้นน้ำวันชื้อมาก น้ำต๋องไป
ชื้อขนมกระเป๋าคันโตกระเป๋ามันขาดแล้วน้ำต๋องก็
ออกไปชื้อน้ำจืดไปซื้อตลาดในตลาดมีทั้งผัก
สด กล้วย ผลไม้ต่างๆ แล้วน้ำวันคิดว่า
กระเป๋าคันน้ำวันนั้นได้กระเป๋ามันหนักแล้วเดิน
ที่น้ำวันใส่ไว้มันหลุดออกไปแล้วมีชายคนหนึ่งเดิน
เข้ามาแล้วชายคนนั้นเห็นเงินจำนวน 1000 บาท
เขาก็รีบเก็บแล้วเขารู้ว่าต้องเป็นของผู้อื่น
นั้นเขาเลยเอาไปคืนน้ำวันมาองแล้วขอบคุณผู้ชื้อคนนั้น

ความสามารถในการเขียนเรื่องจากภาพ ชั้นประถมศึกษาปีที่ ๓

ชื่อ...ด.ช.จิรจรัส.....นามสกุล.....

เลขที่...1... เลขประจำตัวประชาชน

โรงเรียน..... อำเภอ...บึงสามพันจังหวัด...สิงห์บุรี.....

ชื่อเรื่อง แม่สาวคนทำเงินหลวม ชื่อเรื่อง

มีอยู่มาวันหนึ่งมีแม่สาวคนหนึ่งชื่อนาง
หงษ์เธอมาตลาดเพื่อมาซื้อของ มีของ
เยอะแยะมากมายเช่นทุเรียน มะม่วง
มังคุด ลำปราง มะละกอ มะพร้าว
กล้วย ฟักทอง ตะขอส
เป็ดต้ม แม่สาวจึงคิดว่าซื้อ
ผลไม้ กล้วย และฟักทอง แม่สาว
จึงห่อเอาเงินมาหนึ่งพัน บาท
ที่อยู่เหนกระเป๋าก็แต่มีผล
แต่แม่สาวก็ห่ออย่างอื่นลงเลย
ขโมยไป

บรรณานุกรม

- กรมวิชาการ. (2546). การจัดการเรียนรู้อุ้กลุ่มสาระการเรียนรู้ภาษาไทย ตามหลักสูตรการศึกษา
ขั้นพื้นฐาน พุทธศักราช 2551. กรุงเทพฯ: โรงพิมพ์คุรุสภาลาดพร้าว.
- ฉัตรศิริ ปิยะพิมลสิทธิ์. (2544). เกณฑ์การให้คะแนน (Scoring Rubrics). สืบค้นเมื่อ 4 กุมภาพันธ์
2554. จาก <http://www.watpon.com/Elearning/mea5.htm>.
- ชัยนันท์ นันทพันธ์. (2526). ศิลปะการเขียนเรียงความและรวมเรียงความรางวัลพระราชทาน. กรุงเทพฯ:
ประกายพริก.
- ชำนาญ สุภนิตย์; และ สุทิน พูลสวัสดิ์. (2547). เตรียมสอบการเขียน Essay TOFEL GMAT GRE
ฉบับสมบูรณ์. กรุงเทพฯ: ที.จี.อาร์.อี.
- บุญเชิด ภิญญอนันตพงษ์. (2544). รายงานการวิจัยการประเมินการเรียนรู้ที่เน้นผู้เรียนเป็นสำคัญ:
แนวคิดและวิธีการ. กรุงเทพฯ: อมรินทร์พริ้นติ้งแอนด์พับลิชชิ่ง.
- เปลื้อง ณ นคร. (2512). ตำราเรียงความชั้นสูง. กรุงเทพฯ: ไทยวัฒนาพานิช.
- ไพศาล หวังพานิช. (2545). การวัดและประเมินผลการเรียน. กรุงเทพฯ: สำนักทดสอบทางการศึกษา
และจิตวิทยา มหาวิทยาลัยศรีนครินทรวิโรฒ.
- ส. วาสนา ประवालพฤษ์. (2533). การพิจารณาความเชื่อมั่นของแบบทดสอบการปฏิบัติ. การวัดผล
การศึกษา. 12(35).

คู่มือ การตรวจให้คะแนนจดหมาย

① การตรวจให้คะแนนแบบแยกองค์ประกอบ
(Analytic Scoring Rubrics)

② การตรวจให้คะแนนแบบรวมองค์ประกอบ
(Holistic Scoring Rubrics)

③ การตรวจให้คะแนนแบบผสมผสาน
(Annotated Scoring Rubrics)

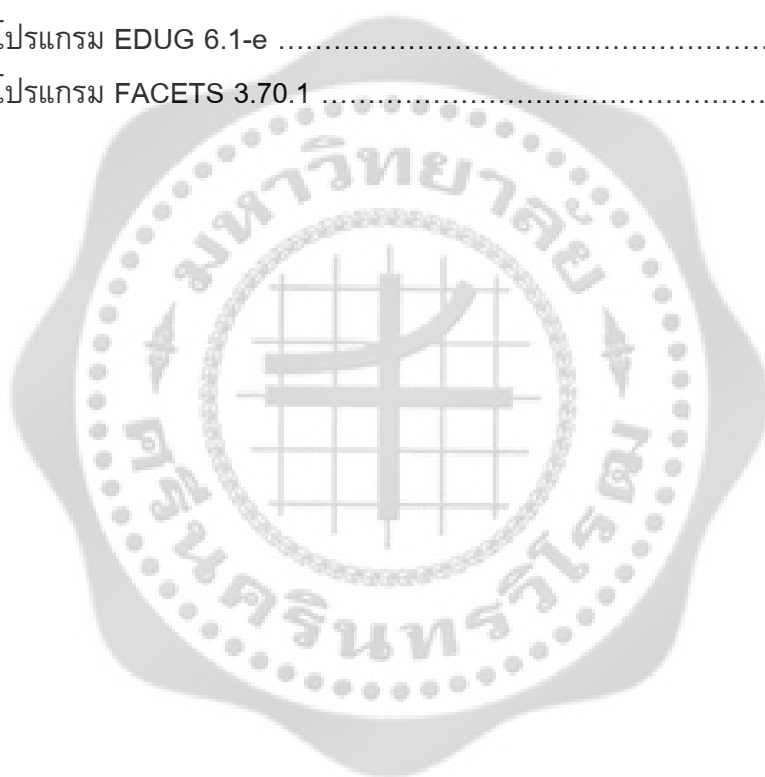
โดย

บุษวรรษ์ แสนปลื้ม

การทดสอบและวัดผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยศรีนครินทรวิโรฒ

สารบัญ

รายการ	หน้า
แบบทดสอบวัดความสามารถในการเขียนจดหมาย	184
ตัวอย่างข้อสอบที่นักเรียนจะได้รับ	185
นิยามศัพท์เฉพาะ	186
การตรวจให้คะแนนการเขียนจดหมายแบบแยกองค์ประกอบ	187
การตรวจให้คะแนนการเขียนจดหมายแบบรวมองค์ประกอบ	191
การตรวจให้คะแนนการเขียนจดหมายแบบผสมผสาน	191
คู่มือการใช้โปรแกรม EDUG 6.1-e	210
คู่มือการใช้โปรแกรม FACETS 3.70.1	216



เรื่อง การเขียนจดหมายบรรยายการออกกำลังกาย

คำชี้แจง: สมมติให้นักเรียนมีเพื่อนเก่าคนหนึ่งชื่อว่า กัมปนาท ซึ่งตอนนี้กำลังเรียนที่โรงเรียนอื่น และต้องการเขียนจดหมายติดต่อกับนักเรียน ให้นักเรียนเขียนตอบจดหมายของกัมปนาท โดยเขียนจดหมายตอบตามคำถามที่กัมปนาทถาม โดยมีความยาวไม่ต่ำกว่า 20 บรรทัด ลงกระดาษที่แจกภายในเวลา 60 นาที

31 หมู่ 1 หมู่บ้านสลักไต่
ตำบลสลักไต่ อำเภอเมือง
จังหวัดสุรินทร์ 32000
25 มิถุนายน 2554

ถึงA.....เพื่อนรัก

นานแล้วนะที่เราไม่ได้เจอกัน สุขภาพของ...A...เป็นอย่างไรบ้าง ส่วนเราสบายดี เรียนที่โรงเรียนเมืองสนุกใหม่ ผลการเรียนของเธอดีหรือเปล่า ช่วงนี้อากาศเปลี่ยนแปลงบ่อย อย่าลืมรักษาสุขภาพด้วยการรับประทานอาหารที่ดีและมีประโยชน์และที่สำคัญอย่าลืมออกกำลังกายด้วยนะ ...A... ชอบออกกำลังกายหรือเปล่า เธอออกกำลังกายบ้างไหมทั้งที่บ้านและที่โรงเรียน และชอบเล่นกีฬาอะไรเป็นพิเศษช่วยเขียนเล่าให้เราฟังหน่อยนะ

รีบเขียนตอบจดหมายมานะ เราจะตั้งใจรออ่านจดหมายตอบของ...A....

รักและคิดถึงเสมอ
กัมปนาท

ตัวอย่างข้อสอบที่นักเรียนจะได้รับ

<p>ถึง จารุวรรณเพื่อนรัก</p> <p>นานแล้วนะที่เราไม่ได้เจอกัน สุขภาพของจารุวรรณเป็นอย่างไรบ้าง ส่วนเราสบายดี เรียนที่โรงเรียนเมืองสนุกใหม่ ผลการเรียนของเธอดีหรือเปล่า ช่วงนี้อากาศเปลี่ยนแปลงบ่อย อย่าลืมนักศึกษาสุขภาพด้วยการรับประทานอาหารที่ดีและมีประโยชน์ และที่สำคัญอย่าลืมออกกำลังกายด้วยนะ จารุวรรณชอบออกกำลังกายหรือเปล่า เธอออกกำลังกายบ้างไหมทั้งที่บ้านและที่โรงเรียน และชอบเล่นกีฬาอะไรเป็นพิเศษช่วยเขียนเล่าให้เราฟังหน่อยนะ</p> <p>รีบเขียนตอบจดหมายมานะ เราจะตั้งใจอ่านจดหมายตอบของจารุวรรณ</p>	<p>31 หมู่ 1 หมู่บ้านสลักได ตำบลสลักได อำเภอเมือง จังหวัดสุรินทร์ 32000 25 มิถุนายน 2554</p> <p>รักและคิดถึงเสมอ กัมปนาท</p>
<p>ถึง ทรงสมรเพื่อนรัก</p> <p>นานแล้วนะที่เราไม่ได้เจอกัน สุขภาพของทรงสมรเป็นอย่างไรบ้าง ส่วนเราสบายดี เรียนที่โรงเรียนเมืองสนุกใหม่ ผลการเรียนของเธอดีหรือเปล่า ช่วงนี้อากาศเปลี่ยนแปลงบ่อย อย่าลืมนักศึกษาสุขภาพด้วยการรับประทานอาหารที่ดีและมีประโยชน์ และที่สำคัญอย่าลืมออกกำลังกายด้วยนะ ทรงสมรชอบออกกำลังกายหรือเปล่า เธอออกกำลังกายบ้างไหมทั้งที่บ้านและที่โรงเรียน และชอบเล่นกีฬาอะไรเป็นพิเศษช่วยเขียนเล่าให้เราฟังหน่อยนะ</p> <p>รีบเขียนตอบจดหมายมานะ เราจะตั้งใจอ่านจดหมายตอบของทรงสมร</p>	<p>31 หมู่ 1 หมู่บ้านสลักได ตำบลสลักได อำเภอเมือง จังหวัดสุรินทร์ 32000 25 มิถุนายน 2554</p> <p>รักและคิดถึงเสมอ กัมปนาท</p>

นियามศัพท์เฉพาะ

1. **ประเด็นสำคัญของจดหมาย** หมายถึง การนำแนวคิดของจดหมายมาตอบอย่างตรงประเด็น ในที่นี้หมายถึงการออกกำลังกาย

2. **รูปแบบการเขียนจดหมาย** หมายถึง งานเขียนมีจำนวนบรรทัดตามกำหนดและมีแบบแผน ตามลักษณะการเขียนจดหมาย คือ ที่อยู่ วันที่ คำขึ้นต้น และคำลงท้าย และเนื้อเรื่องประกอบด้วย คำนำ เนื้อหา สรุป ซึ่งมีรายละเอียด ดังนี้

2.1 คำนำ หมายถึง บทเริ่มต้นของการเขียนเพื่อช่วยให้ผู้อ่านสนใจติดตามเรื่องไปโดยตลอด ด้วยวิธีการต่างๆ เช่น การขึ้นต้นด้วยคำถามหรือปัญหาที่น่าสนใจ การบอกจุดประสงค์หรือความสำคัญ ที่เขียน

2.2 เนื้อเรื่อง หมายถึง การเขียนเนื้อเรื่องได้สอดคล้องชื่อเรื่องและครอบคลุมกับภาพอย่างชัดเจน ทุกตอนของเรื่องสัมพันธ์กันโดยตลอด มีแนวคิดอย่างเดี่ยวตลอดเรื่อง เนื้อเรื่องน่าสนใจ ชวนติดตาม มีการเน้นข้อคิดที่สำคัญให้เห็นอย่างเด่นชัด สอดแทรกแนวคิดที่เป็นประโยชน์ในเชิงสร้างสรรค์ แสดง เหตุผลสนับสนุนความคิดเห็นอย่างถูกต้อง

2.3 สรุป หมายถึง การกลั่นกรอง การรวบรวมใจความสำคัญ หรือการฝากข้อคิดเห็นแก่ ผู้อ่านเป็นการปิดเรื่อง อาจทำได้โดยการยกสุภาษิต คำคม คำกล่าว หรือบทกวี และเป็นย่อหน้าสุดท้าย ของเรื่อง

3. **การนำเสนอ** หมายถึง เรียบเรียงเนื้อเรื่องเป็นลำดับต่อเนื่องสัมพันธ์กันชัดเจนตลอด ตั้งแต่ การเปิดเรื่อง การดำเนินเรื่อง และการปิดเรื่อง เสนอความคิดในแง่มุมต่างๆ ทุกตอนของเรื่องสัมพันธ์ กันโดยตลอด มีเนื้อเรื่องน่าสนใจชวนติดตาม มีการเน้นข้อคิดที่สำคัญให้เห็นอย่างเด่นชัด สอดแทรก แนวคิดที่เป็นประโยชน์ต่อสังคม แสดงเหตุผลสนับสนุนความคิดเห็นอย่างถูกต้อง

4. **การใช้ภาษา** หมายถึง การเขียนสะกดคำถูกต้องทุกคำ เลือกใช้ถ้อยคำได้ตรงกับความหมาย ใช้คำสุภาพ ไม่กำกวม มีความหมายชัดเจน เข้าใจง่าย ไม่ใช้ภาษาพูดในภาษาเขียน ไม่มีการเขียนฉีกคำ หรือแยกคำ และลักษณะของชิ้นงานโดยภาพรวม มีความสะอาดเรียบร้อย ไม่มีรอยลบขีดฆ่า ลายมือ เป็นระเบียบสวยงาม อ่านง่ายทำงานเสร็จทันเวลาที่กำหนด โดยมีจำนวนบรรทัดทั้งหมดของงานเขียน ไม่ต่ำกว่าที่โจทย์กำหนดไว้ 10 บรรทัด

การตรวจให้คะแนนการเขียนจดหมายแบบแยกองค์ประกอบ (Analytic Scoring Rubrics)

วิธีการใช้เกณฑ์การตรวจให้คะแนนการเขียนจดหมายแบบแยกองค์ประกอบ

1. อ่านเรียงความให้จบ 1 รอบ
2. พิจารณาให้คะแนนในประเด็นที่ 1 – 4 ทีละประเด็นจนครบ
3. ถ้าไม่แน่ใจในประเด็นใดให้วางพักแยกไว้ก่อนแล้วจึงนำมาพิจารณาอีกครั้งภายหลัง

เกณฑ์การตรวจให้คะแนนการเขียนจดหมายแบบแยกองค์ประกอบ

1. ประเด็นสำคัญของจดหมาย หมายถึง การนำแนวคิดของจดหมายมาตอบอย่างตรงประเด็น ในที่นี้หมายถึงการออกกำลังกาย

ประเด็น การตรวจ ระดับคุณภาพ	1. ประเด็นสำคัญของเรียงความ
5	เนื้อเรื่องเขียนได้สอดคล้องกับการออกกำลังกาย และครอบคลุมประเด็นสำคัญของจดหมายอย่างชัดเจน
4	เนื้อเรื่องเขียนได้สอดคล้องกับการออกกำลังกาย และครอบคลุมประเด็นสำคัญของจดหมายเป็นส่วนใหญ่
3	เนื้อเรื่องเขียนได้สอดคล้องกับการออกกำลังกาย และครอบคลุมประเด็นสำคัญของจดหมายบางส่วน
2	เนื้อเรื่องเขียนได้สอดคล้องกับการออกกำลังกาย แต่ไม่ครอบคลุมประเด็นสำคัญของจดหมาย
1	เนื้อเรื่องเขียนไม่สอดคล้องกับการออกกำลังกาย และไม่ครอบคลุมประเด็นสำคัญของจดหมาย

2. รูปแบบการเขียนจดหมาย หมายถึง งานเขียนมีจำนวนบรรทัดตามกำหนดและมีแบบแผนตามลักษณะการเขียนจดหมาย คือ ที่อยู่ วันที่ คำขึ้นต้น และคำลงท้าย และเนื้อเรื่องประกอบด้วย คำนำ เนื้อหา สรุป ซึ่งมีรายละเอียดดังนี้

2.1 คำนำ หมายถึง บทเริ่มต้นของการเขียนเพื่อนำให้ผู้อ่านสนใจติดตามเรื่องไปโดยตลอดด้วยวิธีการต่างๆ เช่น การขึ้นต้นด้วยคำถามหรือปัญหาที่น่าสนใจ การบอกจุดประสงค์หรือความสำคัญที่เขียน

2.2 เนื้อเรื่อง หมายถึง การเขียนเนื้อเรื่องได้สอดคล้องชื่อเรื่องและครอบคลุมกับภาพอย่างชัดเจน ทุกตอนของเรื่องสัมพันธ์กันโดยตลอด มีแนวคิดอย่างเดี่ยวตลอดเรื่อง เนื้อเรื่องน่าสนใจชวนติดตาม มีการเน้นข้อคิดที่สำคัญให้เห็นอย่างเด่นชัด สอดแทรกแนวคิดที่เป็นประโยชน์ในเชิงสร้างสรรค์ แสดง เหตุผลสนับสนุนความคิดเห็นอย่างถูกต้อง

2.3 สรุป หมายถึง การกลั่นกรอง การรวบรวมใจความสำคัญ หรือการฝากข้อคิดเห็นแก่ ผู้อ่านเป็นการปิดเรื่อง อาจทำได้โดยการยกสุภาษิต คำคม คำกล่าว หรือบทกวี และเป็นย่อหน้าสุดท้าย ของเรื่อง

ประเด็น การตรวจ ระดับคุณภาพ	2. รูปแบบการเขียนเรียงความ
5	1. มีแบบแผนตามลักษณะการเขียนจดหมาย 2. เนื้อเรื่องเขียนได้ครบถ้วนทั้ง 3 ส่วน ได้แก่ คำนำ เนื้อเรื่อง สรุป 3. ความยาวของแต่ละส่วนเป็นสัดส่วนอย่างเหมาะสม 4. จำนวนบรรทัดทั้งหมดมีไม่ต่ำกว่าที่กำหนด
4	1. มีแบบแผนตามลักษณะการเขียนจดหมาย 2. เนื้อเรื่องเขียนได้ครบถ้วนทั้ง 3 ส่วน ได้แก่ คำนำ เนื้อเรื่อง สรุป 3. ความยาวของแต่ละส่วนไม่เป็นสัดส่วนเหมาะสม 4. จำนวนบรรทัดทั้งหมดมีไม่ต่ำกว่าที่กำหนด
3	1. มีแบบแผนตามลักษณะการเขียนจดหมาย 2. เนื้อเรื่องเขียนมีเพียง 2 ส่วน เช่น มีการเขียนเนื้อเรื่องและสรุปแต่ไม่มีการเขียน คำนำ หรือ เขียนคำนำและเนื้อเรื่องแต่ไม่มีการเขียนสรุป 3. ความยาวของแต่ละส่วนไม่เป็นสัดส่วนเหมาะสม 4. จำนวนบรรทัดทั้งหมดมีไม่ต่ำกว่าที่กำหนด
2	1. มีแบบแผนตามลักษณะการเขียนจดหมายบ้าง 2. เนื้อเรื่องเขียนได้เพียงส่วนเดียว คือ คำนำ เนื้อเรื่อง หรือสรุป 3. ความยาวของแต่ละส่วนไม่เป็นสัดส่วนเหมาะสม 4. มีจำนวนบรรทัดทั้งหมดต่ำกว่าที่กำหนด
1	1. ไม่มีแบบแผนตามลักษณะการเขียนจดหมาย 2. เนื้อเรื่องเขียนแล้วแยกส่วนไม่ได้ 3. ความยาวของแต่ละส่วนไม่เป็นสัดส่วนเหมาะสม 4. มีจำนวนบรรทัดทั้งหมดต่ำกว่าที่กำหนด

3. การนำเสนอ หมายถึง เรียบเรียงเนื้อเรื่องเป็นลำดับต่อเนื่องสัมพันธ์กันชัดเจนตลอด ตั้งแต่การเปิดเรื่อง การดำเนินเรื่อง และการปิดเรื่อง เสนอความคิดในแง่มุมต่างๆ ทุกตอนของเรื่องสัมพันธ์กัน โดยตลอด มีเนื้อเรื่องน่าสนใจชวนติดตาม มีการเน้นข้อคิดที่สำคัญให้เห็นอย่างเด่นชัด สอดแทรกแนวคิด ที่เป็นประโยชน์ต่อสังคม แสดงเหตุผลสนับสนุนความคิดเห็นอย่างถูกต้อง

ประเด็น การตรวจ ระดับคุณภาพ	3. การนำเสนอ
5	1. เรียบเรียงเนื้อเรื่องเป็นลำดับต่อเนื่องสัมพันธ์กันชัดเจน 2. เสนอความคิดในแง่มุมต่างๆ เชื่อมโยงทุกตอนของเรื่องให้สัมพันธ์กันโดยตลอด 3. สอดแทรกแนวคิดที่เป็นประโยชน์ต่อสังคม 4. แสดงเหตุผลสนับสนุนความคิดเห็นอย่างถูกต้อง
4	1. เรียบเรียงเนื้อเรื่องเป็นลำดับต่อเนื่องสัมพันธ์กันค่อนข้างชัดเจน 2. เสนอความคิดในแง่มุมต่างๆ เชื่อมโยงทุกตอนของเรื่องให้สัมพันธ์กันโดยตลอด 3. สอดแทรกแนวคิดที่เป็นประโยชน์ต่อสังคม 4. แสดงเหตุผลสนับสนุนความคิดเห็นชัดเจน
3	1. เรียบเรียงเนื้อเรื่องสับสน วกวน แต่ยังอยู่ในกรอบ 2. เสนอความคิดในบางตอนไม่เชื่อมโยงสัมพันธ์กัน 3. สอดแทรกแนวคิดที่เป็นประโยชน์เล็กน้อย 4. แสดงเหตุผลสนับสนุนความคิดเห็นบางส่วน
2	1. เรียบเรียงเนื้อเรื่องสับสน วกวน 2. เสนอความคิดไม่เชื่อมโยงสัมพันธ์กัน 3. สอดแทรกแนวคิดที่เป็นประโยชน์น้อย 4. ขาดการแสดงเหตุผลสนับสนุนความคิดเห็น
1	1. เรียบเรียงเนื้อเรื่องสับสน ทำให้ผู้อ่านเข้าใจผิด 2. เสนอความคิดไม่เชื่อมโยงสัมพันธ์กัน 3. ไม่ได้สอดแทรกแนวคิดที่เป็นประโยชน์ 4. ขาดการแสดงเหตุผลสนับสนุนความคิดเห็น

4. การใช้ภาษา หมายถึง การเขียนสะกดคำถูกต้องทุกคำ เลือกใช้ถ้อยคำได้ตรงกับความหมาย ใช้คำสุภาพ ไม่กำกวม มีความหมายชัดเจน เข้าใจง่าย ไม่ใช้ภาษาพูดในภาษาเขียน ไม่มีการเขียนฉีกคำหรือแยกคำ และลักษณะของชิ้นงานโดยภาพรวม มีความสะอาดเรียบร้อย ไม่มีรอยลบขีดฆ่า ลายมือเป็นระเบียบสวยงาม อ่านง่ายทำงานเสร็จทันเวลาที่กำหนด โดยมีจำนวนบรรทัดทั้งหมดของงานเขียนไม่ต่ำกว่าที่กำหนด

ประเด็น การตรวจ ระดับคุณภาพ	4. การใช้ภาษา
5	<ol style="list-style-type: none"> 1. สะกดคำถูกต้องทุกคำ 2. เลือกใช้ถ้อยคำได้ตรงกับความหมาย ใช้คำสุภาพไม่ใช้ภาษาพูดในภาษาเขียน 3. ไม่มีการเขียนฉีกคำหรือแยกคำเลยตลอดทั้งเรื่อง 4. ลายมือเป็นระเบียบสวยงาม อ่านง่าย สะอาด 5. ไม่มีรอยลบขีดฆ่า
4	<ol style="list-style-type: none"> 1. สะกดคำผิด 1 – 3 คำ 2. เลือกใช้ถ้อยคำได้ตรงความหมาย ใช้คำสุภาพ 3. ฉีกคำหรือแยกคำ 1 – 3 แห่ง 4. ลายมืออ่านง่าย 5. มีรอยลบขีดฆ่า 1 – 3 แห่ง
3	<ol style="list-style-type: none"> 1. สะกดคำผิด 4 – 5 คำ 2. ใช้ถ้อยคำไม่เหมาะสมบางแห่ง 3. ฉีกคำหรือแยกคำ 4 – 5 แห่ง 4. ลายมืออ่านง่าย 5. มีรอยลบขีดฆ่า 4 – 5 แห่ง
2	<ol style="list-style-type: none"> 1. เขียนสะกดคำผิด 6 – 9 คำ 2. ใช้ถ้อยคำไม่เหมาะสมบางแห่ง 3. ฉีกคำหรือแยกคำ 6 – 9 แห่ง 4. ลายมือพออ่านออก 5. มีรอยลบขีดฆ่า 6 – 9 แห่ง
1	<ol style="list-style-type: none"> 1. เขียนสะกดคำผิดตั้งแต่ 10 คำ ขึ้นไป 2. ใช้ถ้อยคำไม่เหมาะสมเป็นส่วนใหญ่ 3. ฉีกคำหรือแยกคำตั้งแต่ 10 แห่ง ขึ้นไป 4. ลายมืออ่านไม่ออก 5. มีรอยลบขีดฆ่าตั้งแต่ 10 แห่ง ขึ้นไป

การตรวจให้คะแนนการเขียนจดหมายแบบรวมองค์ประกอบ (Holistic Scorings)

วิธีการใช้เกณฑ์การตรวจให้คะแนนการเขียนจดหมายแบบรวมองค์ประกอบ

1. พิจารณาภาพรวมและแบ่งกระดาษออกเป็น 5 กอง คือ กอง 5 คะแนน กอง 4 คะแนน กอง 3 คะแนน กอง 2 คะแนน กอง 1 คะแนน
2. จดหมายชิ้นใดถ้าผู้ตรวจไม่แน่ใจว่าจะอยู่กลุ่มใดให้ผู้ตรวจทำเครื่องหมายไว้และวางไว้กองที่น่าจะอยู่กับกองนั้นๆ หลังจากนั้นจึงนำมาพิจารณาอีกครั้ง

การตรวจให้คะแนนการเขียนจดหมายแบบผสมผสาน (Annotated Scoring Rubrics)

วิธีการใช้เกณฑ์การตรวจให้คะแนนการเขียนจดหมายแบบผสมผสาน

ขั้นตอนที่ 1 ตรวจให้คะแนนโดยใช้วิธีการให้คะแนนแบบรวมองค์ประกอบ

1. พิจารณาภาพรวมและแบ่งกระดาษออกเป็น 5 กอง คือ กอง 5 คะแนน กอง 4 คะแนน กอง 3 คะแนน กอง 2 คะแนน กอง 1 คะแนน
2. จดหมายชิ้นใดถ้าผู้ตรวจไม่แน่ใจว่าจะอยู่กลุ่มใดให้ผู้ตรวจทำเครื่องหมายไว้และวางไว้กองที่น่าจะอยู่กับกองนั้นๆ หลังจากนั้นจึงนำมาพิจารณาอีกครั้ง

ขั้นตอนที่ 2 คัดเลือกประเด็นที่มีปัญหาในการตรวจ เพื่อนำมาตรวจอีกครั้งโดยใช้วิธีการให้คะแนนแบบแยกองค์ประกอบ

1. เลือกเฉพาะบางคุณลักษณะที่โดดเด่น เพื่อเป็นผลสะท้อน (Feedback) ให้แก่ผู้เรียน โดยเลือกประเด็นที่มีประโยชน์ต่อการพัฒนาการเขียนของนักเรียน
2. พิจารณาให้คะแนนในประเด็นที่เลือกไว้ที่ละประเด็นจนครบ



ภาคผนวก ข
ตารางแสดงคุณลักษณะผู้ตรวจให้คะแนน

ตาราง 30 แสดงค่าความเข้มงวด/ใจดีของผู้ตรวจให้คะแนน จำแนกเป็นรายบุคคล

ID	แยกองค์ประกอบ				รวมองค์ประกอบ				ผสมผสาน			
	Logit	Infit	Outfit	ผล	Logit	Infit	Outfit	ผล	Logit	Infit	Outfit	ผล
1	1.28	1.01	1.07	เข้มงวด	0.06	0.62	0.63	กลาง	1.34	1.12	1.10	เข้มงวด
2	-1.38	1.18	1.21	ใจดี	-0.56	2.16	2.28	กลาง	-1.26	1.09	1.10	ใจดี
3	2.21	0.92	0.86	เข้มงวด	2.47	1.62	1.35	เข้มงวด	1.82	1.00	0.98	เข้มงวด
4	-0.85	1.05	1.02	กลาง	-1.73	0.52	0.52	ใจดี	-1.10	1.00	0.99	ใจดี
5	-0.21	1.49	1.49	กลาง	0.49	0.64	0.73	กลาง	0.99	1.15	1.24	กลาง
6	-0.26	0.48	0.49	กลาง	-0.30	1.29	1.32	กลาง	0.05	0.73	0.75	กลาง
7	-1.00	0.90	0.90	กลาง	-1.32	0.93	0.94	ใจดี	-0.87	1.19	1.40	กลาง
8	-0.11	0.71	0.71	กลาง	-0.08	0.73	0.75	กลาง	-0.31	1.11	1.25	กลาง
9	-1.39	0.70	0.70	ใจดี	0.20	0.73	0.70	กลาง	-0.95	0.87	0.87	กลาง
10	-1.28	1.98	1.96	ใจดี	-0.30	2.09	2.11	กลาง	-1.03	1.97	1.98	ใจดี
11	1.14	1.07	1.15	เข้มงวด	1.60	1.46	1.99	เข้มงวด	1.45	0.99	1.00	เข้มงวด
12	0.02	1.79	1.73	กลาง	0.64	2.54	2.37	กลาง	0.33	1.26	1.28	กลาง
13	-0.67	0.77	0.77	กลาง	-0.69	1.01	1.03	กลาง	-1.14	0.75	0.75	ใจดี
14	1.69	1.05	0.97	เข้มงวด	1.05	0.72	0.68	เข้มงวด	2.03	1.60	1.37	เข้มงวด
15	1.01	0.53	0.57	เข้มงวด	0.84	0.54	0.65	กลาง	1.22	0.81	0.80	เข้มงวด
16	0.69	0.83	0.85	กลาง	-0.12	0.98	0.97	กลาง	0.83	0.74	0.74	กลาง
17	-0.50	1.33	1.32	กลาง	1.10	0.71	0.70	เข้มงวด	1.20	0.77	0.76	เข้มงวด
18	-0.09	0.90	0.87	กลาง	-0.12	0.48	0.49	กลาง	0.26	0.67	0.67	กลาง
19	-0.69	1.38	1.37	กลาง	-0.48	1.38	1.38	กลาง	-1.22	1.17	1.17	ใจดี
20	-0.71	0.60	0.61	กลาง	-0.65	0.68	0.71	กลาง	-1.22	0.86	0.87	ใจดี
21	-0.14	0.62	0.58	กลาง	0.99	1.06	0.99	กลาง	-0.35	0.64	0.63	กลาง
22	0.51	1.28	1.27	กลาง	0.34	1.50	1.43	กลาง	0.48	1.39	1.36	กลาง
23	-0.09	0.68	0.67	กลาง	0.79	0.73	0.74	กลาง	0.10	0.60	0.59	กลาง
24	-1.20	0.43	0.44	ใจดี	-0.35	0.63	0.64	กลาง	-0.58	0.49	0.49	กลาง
25	1.00	1.25	1.13	เข้มงวด	1.95	1.09	0.97	เข้มงวด	1.07	0.90	0.88	เข้มงวด
26	-0.11	0.78	0.79	กลาง	0.11	0.57	0.54	กลาง	0.39	0.57	0.56	กลาง
27	-1.39	0.42	0.45	ใจดี	-1.16	0.48	0.49	ใจดี	-1.26	0.40	0.43	ใจดี
28	-1.09	0.86	0.86	ใจดี	-1.03	1.36	1.37	ใจดี	-0.84	1.26	1.22	กลาง
29	-0.14	0.71	0.70	กลาง	0.64	1.61	1.51	กลาง	0.23	0.82	0.82	กลาง
30	-0.10	1.46	1.54	กลาง	-0.08	2.24	2.17	กลาง	0.07	1.68	1.62	กลาง
31	0.69	0.89	0.89	กลาง	0.64	0.66	0.64	กลาง	0.44	0.48	0.49	กลาง
32	-0.10	1.04	1.03	กลาง	0.39	0.65	0.65	กลาง	0.03	0.86	0.84	กลาง
33	0.01	0.79	0.80	กลาง	0.74	1.13	1.04	กลาง	0.21	0.74	0.75	กลาง
34	0.48	0.52	0.58	กลาง	0.94	0.37	0.37	กลาง	0.77	0.72	0.71	กลาง
35	-0.56	0.80	0.80	กลาง	-0.48	0.62	0.64	กลาง	-0.18	0.55	0.56	กลาง

ตาราง 30 (ต่อ)

ID	แยกองค์ประกอบ				รวมองค์ประกอบ				ผสมผสาน			
	Logit	Infit	Outfit	ผล	Logit	Infit	Outfit	ผล	Logit	Infit	Outfit	ผล
36	-0.58	1.63	1.68	กลาง	-1.16	1.01	1.05	ใจดี	-0.31	1.64	1.84	กลาง
37	0.03	0.88	0.89	กลาง	-0.48	0.81	0.88	กลาง	-0.48	0.78	0.80	กลาง
38	0.54	0.49	0.51	กลาง	0.30	0.76	0.76	กลาง	0.63	0.68	0.69	กลาง
39	1.01	0.79	0.85	เข้มงวด	-0.99	1.11	1.12	กลาง	1.00	0.96	1.00	เข้มงวด
40	1.89	0.71	0.71	เข้มงวด	1.31	0.41	0.48	เข้มงวด	1.78	0.93	0.97	เข้มงวด
41	-0.52	0.97	0.97	กลาง	-0.82	0.53	0.55	กลาง	-1.25	0.92	0.92	ใจดี
42	1.46	1.00	1.04	เข้มงวด	1.05	0.87	0.87	เข้มงวด	1.45	0.92	1.08	เข้มงวด
43	0.66	1.32	1.30	กลาง	0.59	1.51	1.43	กลาง	0.48	2.57	2.44	กลาง
44	-0.67	1.27	1.22	กลาง	0.25	2.41	2.31	กลาง	-0.39	2.07	2.01	กลาง
45	-0.63	0.92	0.91	กลาง	-1.40	0.65	0.66	ใจดี	-0.92	0.73	0.76	กลาง
46	0.88	1.72	1.63	กลาง	1.89	1.38	1.37	เข้มงวด	1.10	1.83	1.87	เข้มงวด
47	1.42	1.03	1.02	เข้มงวด	-1.57	0.71	0.77	ใจดี	0.83	1.64	1.66	กลาง
48	1.26	1.40	1.41	เข้มงวด	0.44	1.35	1.32	กลาง	0.88	1.72	1.66	กลาง
49	-0.37	1.25	1.25	กลาง	-1.16	0.68	0.70	ใจดี	0.10	1.33	1.36	กลาง
50	-0.44	0.49	0.49	กลาง	0.20	1.20	1.14	กลาง	-0.31	0.73	0.71	กลาง
51	0.40	0.49	0.51	กลาง	-0.39	0.54	0.60	กลาง	0.42	0.38	0.37	กลาง
52	1.00	1.28	1.21	เข้มงวด	-0.03	0.96	1.00	กลาง	0.51	1.15	1.14	กลาง
53	0.52	0.85	0.84	กลาง	0.69	1.08	1.08	กลาง	-0.22	0.81	0.81	กลาง
54	-0.68	1.14	1.12	กลาง	-0.26	1.03	1.03	กลาง	-0.74	0.92	0.89	กลาง
55	0.56	1.85	1.79	กลาง	-0.21	1.38	1.36	กลาง	0.29	1.69	1.67	กลาง
56	-0.68	2.10	2.06	กลาง	-1.11	0.80	0.79	ใจดี	-1.06	1.61	1.62	ใจดี
57	0.25	0.62	0.61	กลาง	0.39	0.79	0.78	กลาง	0.18	0.70	0.69	กลาง
58	-1.44	0.79	0.81	ใจดี	-0.90	0.78	0.77	กลาง	-1.00	1.18	1.21	กลาง
59	-0.15	0.76	0.74	กลาง	-0.52	1.08	1.07	กลาง	-0.45	1.14	1.13	กลาง
60	0.61	1.42	1.40	กลาง	-0.08	1.92	1.83	กลาง	0.22	1.60	1.54	กลาง
61	-0.87	1.05	1.04	กลาง	-1.65	1.59	1.58	ใจดี	-1.26	1.40	1.39	ใจดี
62	1.23	0.74	0.86	เข้มงวด	-0.48	0.85	0.87	กลาง	0.12	1.04	1.04	กลาง
63	0.45	1.43	1.43	กลาง	0.15	0.96	0.98	กลาง	0.01	1.31	1.33	กลาง
64	-0.29	1.41	1.36	กลาง	0.02	1.48	1.44	กลาง	-0.18	0.83	0.83	กลาง
65	0.29	0.75	0.72	กลาง	0.84	0.75	0.73	กลาง	0.10	1.04	1.02	กลาง
66	0.27	1.47	1.40	กลาง	0.59	1.35	1.33	กลาง	1.31	1.21	1.11	เข้มงวด
67	-0.23	0.59	0.60	กลาง	-0.82	0.35	0.37	กลาง	-0.24	0.78	0.78	กลาง
68	0.33	0.69	0.70	กลาง	0.11	0.78	0.78	กลาง	-0.10	0.69	0.68	กลาง
69	-1.38	0.79	0.79	ใจดี	-1.53	1.16	1.18	ใจดี	-1.53	0.97	0.97	ใจดี
70	1.15	1.11	1.12	เข้มงวด	0.94	1.82	1.89	กลาง	0.69	1.25	1.25	กลาง
71	-0.26	1.02	1.08	กลาง	-1.24	0.60	0.60	ใจดี	-0.74	1.06	1.07	กลาง
72	1.91	0.68	0.71	เข้มงวด	1.54	0.65	0.61	เข้มงวด	2.29	0.62	0.61	เข้มงวด
73	-0.41	1.35	1.38	กลาง	-0.69	0.85	0.85	กลาง	-0.77	1.26	1.30	กลาง
74	0.72	0.92	0.94	กลาง	0.39	0.67	0.68	กลาง	1.15	1.04	0.99	เข้มงวด

ตาราง 30 (ต่อ)

ID	แยกองค์ประกอบ				รวมองค์ประกอบ				ผสมผสาน			
	Logit	Infit	Outfit	ผล	Logit	Infit	Outfit	ผล	Logit	Infit	Outfit	ผล
75	0.35	0.94	0.94	กลาง	0.02	0.74	0.72	กลาง	-0.48	1.08	1.09	กลาง
76	1.03	0.92	0.90	เข้มงวด	0.94	1.23	1.20	กลาง	1.07	1.09	1.08	เข้มงวด
77	0.83	0.93	0.92	กลาง	0.39	0.67	0.67	กลาง	1.15	0.98	0.98	เข้มงวด
78	-0.90	0.78	0.77	กลาง	-0.65	0.70	0.70	กลาง	-0.28	0.71	0.74	กลาง
79	0.73	0.57	0.57	กลาง	-0.08	0.59	0.57	กลาง	0.81	0.87	0.83	กลาง
80	0.64	0.54	0.55	กลาง	0.25	0.76	0.76	กลาง	0.51	0.85	0.82	กลาง
81	0.56	0.98	0.98	กลาง	0.59	0.82	0.76	กลาง	0.77	0.82	0.77	กลาง
82	-0.71	1.02	1.02	กลาง	-0.74	1.18	1.18	กลาง	-0.93	0.78	0.79	กลาง
83	2.06	0.93	0.94	เข้มงวด	1.05	0.94	0.93	เข้มงวด	2.20	1.17	1.12	เข้มงวด
84	-0.14	1.07	1.06	กลาง	-0.08	0.66	0.65	กลาง	-0.82	0.93	0.92	กลาง
85	0.24	0.43	0.42	กลาง	0.94	0.56	0.52	กลาง	0.22	0.69	0.67	กลาง
86	0.10	0.57	0.60	กลาง	0.54	1.09	1.07	กลาง	0.57	0.73	0.73	กลาง
87	-0.08	0.55	0.55	กลาง	-0.12	0.66	0.68	กลาง	-0.31	0.67	0.69	กลาง
88	-1.11	0.45	0.49	ใจดี	-1.32	0.46	0.49	ใจดี	-1.20	0.56	0.57	ใจดี
89	0.10	1.27	1.28	กลาง	0.34	0.63	0.69	กลาง	0.42	0.89	0.95	กลาง
90	0.16	0.73	0.75	กลาง	0.94	1.20	1.15	กลาง	-0.04	1.16	1.16	กลาง
91	-0.81	0.54	0.54	กลาง	-0.48	0.49	0.50	กลาง	-0.62	0.44	0.46	กลาง
92	-0.55	0.74	0.76	กลาง	2.54	3.23	4.44	เข้มงวด	-0.43	0.99	1.12	กลาง
93	0.65	0.88	0.86	กลาง	1.65	0.60	0.60	เข้มงวด	0.26	0.85	0.90	กลาง
94	-2.29	1.58	1.52	ใจดี	-1.94	1.11	1.08	ใจดี	-1.46	1.78	1.79	ใจดี
95	0.32	1.16	1.15	กลาง	0.64	0.68	0.64	กลาง	-0.77	0.54	0.55	กลาง
96	-0.80	0.52	0.52	กลาง	-0.39	0.85	0.84	กลาง	-0.44	0.49	0.54	กลาง
97	0.01	0.48	0.48	กลาง	0.06	0.58	0.58	กลาง	0.11	0.51	0.60	กลาง
98	0.20	0.81	0.81	กลาง	-0.08	0.90	0.86	กลาง	0.31	0.60	0.58	กลาง
99	-0.49	1.53	1.51	กลาง	0.39	0.58	0.58	กลาง	-0.81	1.15	1.17	กลาง
100	-0.74	1.84	1.81	กลาง	-0.90	0.97	0.98	กลาง	-0.40	1.11	1.09	กลาง
101	-0.30	1.25	1.24	กลาง	1.54	1.23	1.12	เข้มงวด	1.04	1.08	1.08	เข้มงวด
102	0.00	0.94	0.93	กลาง	-1.49	0.67	0.66	ใจดี	-1.19	0.97	0.98	ใจดี
103	-0.21	0.62	0.62	กลาง	-0.03	0.58	0.59	กลาง	0.11	1.38	1.39	กลาง
104	0.12	0.85	0.83	กลาง	-0.03	1.32	1.32	กลาง	-0.64	0.69	0.70	กลาง
105	0.01	0.46	0.49	กลาง	-0.03	0.77	0.83	กลาง	0.41	0.68	0.72	กลาง
106	-0.91	0.78	0.77	กลาง	-2.31	0.46	0.48	ใจดี	-1.21	0.71	0.71	ใจดี
107	0.33	0.76	0.83	กลาง	1.31	0.95	0.93	เข้มงวด	0.77	0.71	0.73	กลาง
108	0.20	1.09	1.10	กลาง	-0.43	0.85	0.82	กลาง	0.91	1.03	1.07	กลาง
109	0.62	0.73	0.75	กลาง	-0.12	0.45	0.47	กลาง	-0.35	0.60	0.61	กลาง
110	-0.68	0.92	0.90	กลาง	-0.43	0.78	0.77	กลาง	-0.60	0.77	0.80	กลาง
111	0.90	1.42	1.52	กลาง	-0.12	0.75	0.79	กลาง	0.74	1.23	1.33	กลาง
112	0.72	1.09	1.06	กลาง	-0.39	1.03	0.97	กลาง	0.16	0.89	0.92	กลาง
113	-1.67	0.89	0.91	ใจดี	-0.86	1.10	1.08	กลาง	-2.01	0.78	0.77	ใจดี

ตาราง 30 (ต่อ)

D	แยกองค์ประกอบ				รวมองค์ประกอบ				ผสมผสาน			
	Logit	Infit	Outfit	ผล	Logit	Infit	Outfit	ผล	Logit	Infit	Outfit	ผล
114	-0.85	0.71	0.73	กลาง	-1.07	0.92	0.89	ใจดี	-0.32	1.09	1.11	กลาง
115	0.83	1.37	1.28	กลาง	0.79	1.43	1.35	กลาง	1.57	1.32	1.15	เข้มงวด
116	-1.73	1.05	1.09	ใจดี	-0.52	1.24	1.23	กลาง	-2.08	0.84	0.86	ใจดี
117	0.60	1.80	1.73	กลาง	0.69	2.03	2.06	กลาง	0.10	1.15	1.16	กลาง
118	0.06	0.81	0.82	กลาง	0.39	0.84	0.83	กลาง	-0.28	0.70	0.70	กลาง
119	-0.10	1.32	1.32	กลาง	-0.82	1.05	1.02	กลาง	0.03	1.59	1.55	กลาง
120	0.13	0.84	0.89	กลาง	-0.30	0.89	0.89	กลาง	-0.76	0.50	0.53	กลาง
121	-0.73	0.52	0.52	กลาง	-0.86	0.57	0.57	กลาง	-1.14	0.52	0.54	ใจดี
122	-0.34	0.72	0.73	กลาง	-0.69	1.04	1.07	กลาง	-0.62	0.92	0.91	กลาง
123	-0.91	1.33	1.34	กลาง	0.11	1.13	1.10	กลาง	-0.31	1.23	1.22	กลาง
124	-0.33	0.84	0.84	กลาง	0.20	0.93	0.97	กลาง	-0.03	0.64	0.63	กลาง
125	-0.37	1.92	1.90	กลาง	0.39	1.58	1.55	กลาง	-0.45	1.55	1.55	กลาง
126	0.64	1.01	0.99	กลาง	0.74	0.44	0.42	กลาง	0.91	0.62	0.61	กลาง
127	-0.42	2.39	2.35	กลาง	-0.08	1.82	1.78	กลาง	-1.20	1.40	1.42	ใจดี
128	0.03	1.17	1.15	กลาง	0.69	1.82	1.75	กลาง	0.54	1.32	1.27	กลาง
129	-0.49	1.79	1.82	กลาง	-0.43	1.03	1.04	กลาง	-0.01	1.58	1.57	กลาง
130	-0.45	1.41	1.41	กลาง	-1.07	1.88	1.99	ใจดี	-0.11	1.26	1.27	กลาง

ตาราง 31 แสดงค่าความลำเอียงด้านเพศของผู้ตรวจให้คะแนน จำแนกเป็นรายบุคคล

ID	แยกองค์ประกอบ								รวมองค์ประกอบ								ผสมผสาน							
	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความต่าง	Z	ผล	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความต่าง	Z	ผล	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความต่าง	Z	ผล
1	1.28	1.01	1.07	1.13	1.42	-0.29	-1.27	กลาง	0.06	0.62	0.63	-0.17	0.29	-0.45	-1.06	กลาง	1.34	1.12	1.10	1.28	1.39	-0.10	-0.40	กลาง
2	-1.38	1.18	1.21	-1.65	-1.10	-0.55	-3.07**	ชาย	-0.56	2.16	2.28	-1.05	-0.07	-0.98	-2.37*	ชาย	-1.26	1.09	1.10	-1.03	-1.48	0.45	2.04*	หญิง
3	2.21	0.92	0.86	2.39	2.09	0.30	1.08	กลาง	2.47	1.62	1.35	3.68	1.78	1.90	3.08**	หญิง	1.82	1.00	0.98	2.10	1.62	0.48	1.65	กลาง
4	-0.85	1.05	1.02	-0.89	-0.81	-0.08	-0.43	กลาง	-1.73	0.52	0.52	-1.47	-2.02	0.56	1.36	กลาง	-1.10	1.00	0.99	-1.35	-0.85	-0.50	-2.26*	ชาย
5	-0.21	1.49	1.49	-0.21	-0.21	0.00	0.01	กลาง	0.49	0.64	0.73	0.02	0.95	-0.93	-2.10*	ชาย	0.99	1.15	1.24	0.99	0.98	0.01	0.04	กลาง
6	-0.26	0.48	0.49	-0.34	-0.19	-0.15	-0.80	กลาง	-0.30	1.29	1.32	0.02	-0.59	0.61	1.44	กลาง	0.05	0.73	0.75	0.11	0.01	0.10	0.43	กลาง
7	-1.00	0.90	0.90	-1.07	-0.92	-0.15	-0.83	กลาง	-1.32	0.93	0.94	-1.79	-0.84	-0.94	-2.34*	ชาย	-0.87	1.19	1.40	-1.25	-0.48	-0.77	-3.48**	ชาย
8	-0.11	0.71	0.71	-0.03	-0.19	0.16	0.84	กลาง	-0.08	0.73	0.75	0.02	-0.16	0.18	0.42	กลาง	-0.31	1.11	1.25	0.14	-0.70	0.84	3.62**	หญิง
9	-1.39	0.70	0.70	-1.59	-1.18	-0.40	-2.26*	ชาย	0.20	0.73	0.70	-0.07	0.47	-0.54	-1.25	กลาง	-0.95	0.87	0.87	-0.98	-0.92	-0.06	-0.27	กลาง
10	-1.28	1.98	1.96	-1.32	-1.23	-0.09	-0.48	กลาง	-0.30	2.09	2.11	-0.53	-0.07	-0.46	-1.10	กลาง	-1.03	1.97	1.98	-1.15	-0.90	-0.25	-1.15	กลาง
11	1.14	1.07	1.15	1.13	1.15	-0.02	-0.10	กลาง	1.60	1.46	1.99	1.29	1.89	-0.60	-1.25	กลาง	1.45	0.99	1.00	1.40	1.49	-0.09	-0.33	กลาง
12	0.02	1.79	1.73	0.23	-0.15	0.38	1.92	กลาง	0.64	2.54	2.37	0.73	0.56	0.17	0.37	กลาง	0.33	1.26	1.28	0.28	0.38	-0.10	-0.40	กลาง
13	-0.67	0.77	0.77	-0.62	-0.71	0.09	0.46	กลาง	-0.69	1.01	1.03	-0.71	-0.67	-0.04	-0.09	กลาง	-1.14	0.75	0.75	-1.03	-1.24	0.21	0.94	กลาง
14	1.69	1.05	0.97	1.98	1.47	0.50	2.01*	หญิง	1.05	0.72	0.68	1.06	1.05	0.01	0.03	กลาง	2.03	1.60	1.37	2.42	1.76	0.66	2.17*	หญิง
15	1.01	0.53	0.57	0.94	1.08	-0.14	-0.62	กลาง	0.84	0.54	0.65	0.52	1.15	-0.63	-1.38	กลาง	1.22	0.81	0.80	0.96	1.45	-0.49	-1.90	กลาง
16	0.69	0.83	0.85	0.69	0.70	0.00	-0.02	กลาง	-0.12	0.98	0.97	0.02	-0.25	0.27	0.63	กลาง	0.83	0.74	0.74	0.69	0.95	-0.26	-1.05	กลาง
17	-0.50	1.33	1.32	-0.42	-0.57	0.15	0.82	กลาง	1.10	0.71	0.70	0.95	1.25	-0.30	-0.65	กลาง	1.20	0.77	0.76	1.21	1.20	0.01	0.04	กลาง
18	-0.09	0.90	0.87	0.08	-0.24	0.32	1.64	กลาง	-0.12	0.48	0.49	0.02	-0.25	0.27	0.63	กลาง	0.26	0.67	0.67	0.59	-0.02	0.61	-1.85*	หญิง
19	-0.69	1.38	1.37	-0.73	-0.66	-0.07	-0.39	กลาง	-0.48	1.38	1.38	-0.80	-0.16	-0.64	-1.53	กลาง	-1.22	1.17	1.17	-1.23	-1.22	-0.01	-0.05	กลาง
20	-0.71	0.60	0.61	-0.91	-0.52	-0.38	-2.06*	ชาย	-0.65	0.68	0.71	-0.88	-0.42	-0.46	-1.12	กลาง	-1.22	0.86	0.87	-1.51	-0.92	-0.59	-2.68*	ชาย
21	-0.14	0.62	0.58	-0.15	-0.14	-0.01	-0.06	กลาง	0.99	1.06	0.99	1.29	0.75	0.54	1.16	กลาง	-0.35	0.64	0.63	-0.23	-0.46	0.22	0.98	กลาง
22	0.51	1.28	1.27	0.59	0.44	0.15	0.72	กลาง	0.34	1.50	1.43	0.73	0.02	0.72	1.62	กลาง	0.48	1.39	1.36	0.89	0.14	0.75	3.06**	หญิง
23	-0.09	0.68	0.67	-0.17	-0.02	-0.14	-0.72	กลาง	0.79	0.73	0.74	0.95	0.66	0.29	0.64	กลาง	0.10	0.60	0.59	0.02	0.16	-0.14	-0.61	กลาง
24	-1.20	0.43	0.44	-1.29	-1.12	-0.17	-0.93	กลาง	-0.35	0.63	0.64	-0.26	-0.42	0.16	0.37*	กลาง	-0.58	0.49	0.49	-0.58	-0.58	0.00	0.01	กลาง
25	1.00	1.25	1.13	1.60	0.55	1.05	4.58**	หญิง	1.95	1.09	0.97	2.82	1.35	1.47	2.77*	หญิง	1.07	0.90	0.88	1.32	0.86	0.46	1.76	กลาง

ตาราง 31 (ต่อ)

ID	แยกองค์ประกอบ									รวมองค์ประกอบ						ผสมผสาน								
	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความต่าง	Z	ผล	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความต่าง	Z	ผล	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความต่าง	Z	ผล
26	-0.11	0.78	0.79	-0.32	0.09	-0.41	-2.11*	ชาย	0.11	0.57	0.54	0.22	0.02	0.20	0.46	กลาง	0.39	0.57	0.56	0.41	0.38	0.03	0.11	กลาง
27	-1.39	0.42	0.45	-1.63	-1.14	-0.50	-2.80*	ชาย	-1.16	0.48	0.49	-1.22	-1.09	-0.13	-0.31	กลาง	-1.26	0.40	0.43	-1.49	-1.02	-0.47	-2.13*	ชาย
28	-1.09	0.86	0.86	-1.29	-0.89	-0.40	-2.19*	ชาย	-1.03	1.36	1.37	-1.30	-0.76	-0.54	-1.33	กลาง	-0.84	1.26	1.22	-0.98	-0.70	-0.28	-1.26	กลาง
29	-0.14	0.71	0.70	0.10	-0.35	0.45	2.29*	หญิง	0.64	1.61	1.51	1.65	-0.16	1.81	3.93*	หญิง	0.23	0.82	0.82	0.34	0.14	0.21	0.87	กลาง
30	-0.10	1.46	1.54	-0.27	0.05	-0.32	-1.61	กลาง	-0.08	2.24	2.17	0.02	-0.16	0.18	0.42	กลาง	0.07	1.68	1.62	0.34	-0.18	0.52	2.20*	หญิง
31	0.69	0.89	0.89	0.52	0.85	-0.33	-1.52	กลาง	0.64	0.66	0.64	0.52	0.75	-0.23	-0.52	กลาง	0.44	0.48	0.49	0.41	0.46	-0.06	-0.24	กลาง
32	-0.10	1.04	1.03	0.06	-0.24	0.30	1.53	กลาง	0.39	0.65	0.65	0.42	0.38	0.04	0.09	กลาง	0.03	0.86	0.84	0.11	-0.05	0.15	0.66	กลาง
33	0.01	0.79	0.80	0.04	-0.01	0.04	0.22	กลาง	0.74	1.13	1.04	0.52	0.95	-0.43	-0.95	กลาง	0.21	0.74	0.75	-0.04	0.44	-0.47	-1.99*	ชาย
34	0.48	0.52	0.58	0.45	0.50	-0.05	-0.24	กลาง	0.94	0.37	0.37	0.84	1.05	-0.21	-0.45	กลาง	0.77	0.72	0.71	0.86	0.69	0.17	0.67	กลาง
35	-0.56	0.80	0.80	-0.68	-0.45	-0.22	-1.19	กลาง	-0.48	0.62	0.64	-0.71	-0.25	-0.46	-1.11	กลาง	-0.18	0.55	0.56	-0.34	-0.02	-0.32	-1.39	กลาง
36	-0.58	1.63	1.68	-0.82	-0.35	-0.47	-2.51*	ชาย	-1.16	1.01	1.05	-1.14	-1.18	0.04	0.10	กลาง	-0.31	1.64	1.84	-0.60	-0.02	-0.58	-2.56*	ชาย
37	0.03	0.88	0.89	-0.05	0.11	-0.15	-0.77	กลาง	-0.48	0.81	0.88	-0.71	-0.25	-0.46	-1.11	กลาง	-0.48	0.78	0.80	-0.63	-0.33	-0.30	-1.33	กลาง
38	0.54	0.49	0.51	0.34	0.72	-0.38	-1.79	กลาง	0.30	0.76	0.76	-0.07	0.66	-0.73	-1.68	กลาง	0.63	0.68	0.69	0.44	0.81	-0.37	-1.50	กลาง
39	1.01	0.79	0.85	0.81	1.20	-0.38	-1.73	กลาง	-0.99	1.11	1.12	-1.38	-0.59	-0.80	-1.95	กลาง	1.00	0.96	1.00	0.72	1.26	-0.54	-2.12*	ชาย
40	1.89	0.71	0.71	1.87	1.91	-0.04	-0.16	กลาง	1.31	0.41	0.48	0.73	1.89	-1.16	-2.48*	ชาย	1.78	0.93	0.97	1.44	2.08	-0.65	-2.32*	ชาย
41	-0.52	0.97	0.97	-0.61	-0.44	-0.17	-0.89	กลาง	-0.82	0.53	0.55	-0.53	-1.09	0.56	1.35	กลาง	-1.25	0.92	0.92	-1.15	-1.34	0.18	0.83	กลาง
42	1.46	1.00	1.04	1.10	1.80	-0.69	-2.95*	ชาย	1.05	0.87	0.87	1.06	1.05	0.01	0.03	กลาง	1.45	0.92	1.08	1.17	1.69	-0.52	-1.94	กลาง
43	0.66	1.32	1.30	0.57	0.74	-0.17	-0.79	กลาง	0.59	1.51	1.43	0.95	0.29	0.66	1.48	กลาง	0.48	2.57	2.44	0.47	0.49	-0.02	-0.10	กลาง
44	-0.67	1.27	1.22	-0.29	-1.01	0.72	3.81**	หญิง	0.25	2.41	2.31	1.29	-0.59	1.88	4.18	กลาง	-0.39	2.07	2.01	0.44	-1.09	1.53	6.51**	หญิง
45	-0.63	0.92	0.91	-0.38	-0.86	0.48	2.53*	หญิง	-1.40	0.65	0.66	-1.14	-1.68	0.54	1.32	กลาง	-0.92	0.73	0.76	-1.11	-0.73	-0.38	-1.71	กลาง
46	0.88	1.72	1.63	1.05	0.74	0.31	1.40	กลาง	1.89	1.38	1.37	2.05	1.78	0.27	0.54	กลาง	1.10	1.83	1.87	1.32	0.92	0.40	1.53	กลาง
47	1.42	1.03	1.02	1.51	1.35	0.16	0.68	กลาง	-1.57	0.71	0.77	-1.71	-1.43	-0.28	-0.69	กลาง	0.83	1.64	1.66	0.66	0.98	-0.33	-1.30	กลาง
48	1.26	1.40	1.41	1.63	0.96	0.67	2.85	กลาง	0.44	1.35	1.32	0.52	0.38	0.14	0.32	กลาง	0.88	1.72	1.66	0.56	1.17	-0.60	-2.41*	ชาย
49	-0.37	1.25	1.25	-0.51	-0.24	-0.27	-1.41	กลาง	-1.16	0.68	0.70	-1.38	-0.93	-0.46	-1.13	กลาง	0.10	1.33	1.36	-0.09	0.27	-0.36	-1.55	กลาง

ตาราง 31 (ต่อ)

ID	แยกองค์ประกอบ								รวมองค์ประกอบ								ผสมผสาน							
	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความต่าง	Z	ผล	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความต่าง	Z	ผล	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความต่าง	Z	ผล
50	-0.44	0.49	0.49	-0.57	-0.31	-0.25	-1.34	กลาง	0.20	1.20	1.14	0.73	-0.25	0.98	2.22*	หญิง	-0.31	0.73	0.71	-0.26	-0.35	0.10	0.42	กลาง
51	0.40	0.49	0.51	0.41	0.40	0.01	0.03	กลาง	-0.39	0.54	0.60	-0.53	-0.25	-0.29	-0.68	กลาง	0.42	0.38	0.37	0.34	0.49	-0.15	-0.61	กลาง
52	1.00	1.28	1.21	1.38	0.70	0.69	3.04**	หญิง	-0.03	0.96	1.00	0.12	-0.16	0.28	0.65	กลาง	0.51	1.15	1.14	0.69	0.35	0.34	1.38	กลาง
53	0.52	0.85	0.84	0.48	0.57	-0.09	-0.43	กลาง	0.69	1.08	1.08	1.29	0.20	1.10	2.40*	หญิง	-0.22	0.81	0.81	-0.31	-0.12	-0.19	-0.82	กลาง
54	-0.68	1.14	1.12	-0.44	-0.89	0.45	2.41*	หญิง	-0.26	1.03	1.03	0.22	-0.67	0.89	2.08*	หญิง	-0.74	0.92	0.89	-0.45	-1.02	0.57	2.55*	หญิง
55	0.56	1.85	1.79	0.43	0.67	-0.24	-1.15	กลาง	-0.21	1.38	1.36	-0.26	-0.16	-0.10	-0.24	กลาง	0.29	1.69	1.67	0.08	0.49	-0.41	-1.73	กลาง
56	-0.68	2.10	2.06	-0.71	-0.64	-0.07	-0.38	กลาง	-1.11	0.80	0.79	-1.05	-1.18	0.12	0.30	กลาง	-1.06	1.61	1.62	-0.96	-1.17	0.21	0.94	กลาง
57	0.25	0.62	0.61	0.10	0.38	-0.28	-1.38	กลาง	0.39	0.79	0.78	0.02	0.75	-0.73	-1.67	กลาง	0.18	0.70	0.69	0.14	0.22	-0.08	-0.34	กลาง
58	-1.44	0.79	0.81	-1.57	-1.31	-0.26	-1.45	กลาง	-0.90	0.78	0.77	-1.14	-0.67	-0.46	-1.13	กลาง	-1.00	1.18	1.21	-1.30	-0.70	-0.60	-2.70**	ชาย
59	-0.15	0.76	0.74	-0.09	-0.21	0.12	0.62	กลาง	-0.52	1.08	1.07	0.02	-1.01	1.03	2.43*	หญิง	-0.45	1.14	1.13	-0.29	-0.60	0.32	1.40	กลาง
60	0.61	1.42	1.40	0.48	0.74	-0.26	-1.24	กลาง	-0.08	1.92	1.83	0.02	-0.16	0.18	0.42	กลาง	0.22	1.60	1.54	0.02	0.41	-0.39	-1.63	กลาง
61	-0.87	1.05	1.04	-0.80	-0.94	0.14	0.76	กลาง	-1.65	1.59	1.58	-1.63	-1.80	0.05	0.13	กลาง	-1.26	1.40	1.39	-1.18	-1.34	0.16	0.72	กลาง
62	1.23	0.74	0.86	1.07	1.37	-0.30	-1.30	กลาง	-0.48	0.85	0.87	-0.35	-0.59	0.24	0.56	กลาง	0.12	1.04	1.04	0.08	0.16	-0.09	-0.36	กลาง
63	0.45	1.43	1.43	0.34	0.55	-0.21	-0.99	กลาง	0.15	0.96	0.98	0.02	0.29	-0.26	-0.61	กลาง	0.01	1.31	1.33	-0.04	0.06	-0.09	-0.40	กลาง
64	-0.29	1.41	1.36	-0.15	-0.42	0.27	1.41	กลาง	0.02	1.48	1.44	-0.17	0.20	-0.36	-0.85	กลาง	-0.18	0.83	0.83	-0.23	-0.12	-0.11	-0.46	กลาง
65	0.29	0.75	0.72	0.08	0.48	-0.40	-1.98*	ชาย	0.84	0.75	0.73	0.52	1.15	-0.63	-1.38	กลาง	0.10	1.04	1.02	-0.01	0.19	-0.20	-0.84	กลาง
66	0.27	1.47	1.40	0.87	-0.21	1.07	5.15**	หญิง	0.59	1.35	1.33	1.41	-0.07	1.48	3.25*	หญิง	1.31	1.21	1.11	1.95	0.83	1.12	4.04**	หญิง
67	-0.23	0.59	0.60	-0.29	-0.17	-0.11	-0.59	กลาง	-0.82	0.35	0.37	-1.05	-0.59	-0.46	-1.13	กลาง	-0.24	0.78	0.78	-0.23	-0.25	0.02	0.09	กลาง
68	0.33	0.69	0.70	0.25	0.40	-0.15	-0.73	กลาง	0.11	0.78	0.78	0.32	-0.07	0.39	0.90	กลาง	-0.10	0.69	0.68	0.08	-0.25	0.33	1.42	กลาง
69	-1.38	0.79	0.79	-1.21	-1.55	0.35	1.93	กลาง	-1.53	1.16	1.18	-0.97	-2.11	1.14	2.74*	หญิง	-1.53	0.97	0.97	-1.13	-1.96	0.83	3.71**	หญิง
70	1.15	1.11	1.12	0.99	1.30	-0.30	-1.34	กลาง	0.94	1.82	1.89	0.84	1.05	-0.21	-0.45	กลาง	0.69	1.25	1.25	0.56	0.81	-0.24	-0.98	กลาง
71	-0.26	1.02	1.08	-0.42	-0.12	-0.30	-1.57	กลาง	-1.24	0.60	0.60	-1.05	-1.43	0.37	0.91	กลาง	-0.74	1.06	1.07	-0.76	-0.73	-0.03	-0.14	กลาง
72	1.91	0.68	0.71	2.17	1.71	0.46	1.76	กลาง	1.54	0.65	0.61	1.53	1.56	-0.03	-0.07	กลาง	2.29	0.62	0.61	2.54	2.12	0.42	1.32	กลาง
73	-0.41	1.35	1.38	-0.38	-0.44	0.06	0.29	กลาง	-0.69	0.85	0.85	-0.80	-0.59	-0.21	-0.50	กลาง	-0.77	1.26	1.30	-0.93	-0.60	-0.33	-1.48	กลาง
74	0.72	0.92	0.94	0.81	0.63	0.18	0.85	กลาง	0.39	0.67	0.68	0.63	0.20	0.43	0.97	กลาง	1.15	1.04	0.99	1.44	0.92	0.51	1.96	กลาง

ตาราง 31 (ต่อ)

ID	แยกองค์ประกอบ									รวมองค์ประกอบ						ผสมผสาน								
	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความต่าง	Z	ผล	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความต่าง	Z	ผล	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความต่าง	Z	ผล
75	0.35	0.94	0.94	0.45	0.26	0.19	0.93	กลาง	0.02	0.74	0.72	0.02	0.02	0.01	0.01	กลาง	-0.48	1.08	1.09	-0.34	-0.60	0.26	1.16	กลาง
76	1.03	0.92	0.90	1.30	0.81	0.49	2.18*	หญิง	0.94	1.23	1.20	1.18	0.75	0.42	0.92	กลาง	1.07	1.09	1.08	1.56	0.69	0.87	3.29**	หญิง
77	0.83	0.93	0.92	0.94	0.74	0.20	0.93	กลาง	0.39	0.67	0.67	0.02	0.75	-0.73	-1.67	กลาง	1.15	0.98	0.98	0.92	1.36	-0.43	-1.67	กลาง
78	-0.90	0.78	0.77	-0.71	-1.07	0.36	1.94	กลาง	-0.65	0.70	0.70	-0.80	-0.50	-0.29	-0.71	กลาง	-0.28	0.71	0.74	-0.20	-0.35	0.15	0.66	กลาง
79	0.73	0.57	0.57	0.74	0.72	0.02	0.10	กลาง	-0.08	0.59	0.57	0.22	-0.33	0.55	1.28	กลาง	0.81	0.87	0.83	0.89	0.75	0.14	0.57	กลาง
80	0.64	0.54	0.55	0.45	0.81	-0.35	-1.65	กลาง	0.25	0.76	0.76	-0.07	0.56	-0.64	-1.47	กลาง	0.51	0.85	0.82	0.38	0.63	-0.26	-1.06	กลาง
81	0.56	0.98	0.98	0.74	0.40	0.34	1.59	กลาง	0.59	0.82	0.76	0.95	0.29	0.66	1.48	กลาง	0.77	0.82	0.77	1.17	0.44	0.74	2.91**	หญิง
82	-0.71	1.02	1.02	-0.82	-0.61	-0.21	-1.13	กลาง	-0.74	1.18	1.18	-0.97	-0.50	-0.46	-1.13	กลาง	-0.93	0.78	0.79	-1.15	-0.70	-0.45	-2.04*	ชาย
83	2.06	0.93	0.94	2.02	2.09	-0.07	-0.28	กลาง	1.05	0.94	0.93	1.06	1.05	0.01	0.03	กลาง	2.20	1.17	1.12	2.42	2.05	0.37	1.21	กลาง
84	-0.14	1.07	1.06	-0.11	-0.17	0.07	0.33	กลาง	-0.08	0.66	0.65	0.12	-0.25	0.37	0.85	กลาง	-0.82	0.93	0.92	-0.66	-0.97	0.32	1.41	กลาง
85	0.24	0.43	0.42	0.06	0.40	-0.34	-1.69	กลาง	0.94	0.56	0.52	0.63	1.25	-0.62	-1.37	กลาง	0.22	0.69	0.67	0.38	0.08	0.29	1.22	กลาง
86	0.10	0.57	0.60	0.21	0.01	0.20	0.97	กลาง	0.54	1.09	1.07	0.42	0.66	-0.24	-0.54	กลาง	0.57	0.73	0.73	0.66	0.49	0.17	0.68	กลาง
87	-0.08	0.55	0.55	-0.13	-0.04	-0.08	-0.42	กลาง	-0.12	0.66	0.68	-0.35	0.11	-0.46	-1.08	กลาง	-0.31	0.67	0.69	-0.42	-0.20	-0.22	-0.96	กลาง
88	-1.11	0.45	0.49	-1.41	-0.81	-0.61	-3.36	กลาง	-1.32	0.46	0.49	-1.38	-1.26	-0.12	-0.31	กลาง	-1.20	0.56	0.57	-1.27	-1.12	-0.16	-0.71	กลาง
89	0.10	1.27	1.28	0.08	0.13	-0.05	-0.23	กลาง	0.34	0.63	0.69	-0.17	0.85	-1.02	-2.33*	ชาย	0.42	0.89	0.95	0.25	0.58	-0.32	-1.33	กลาง
90	0.16	0.73	0.75	0.21	0.13	0.08	0.40	กลาง	0.94	1.20	1.15	1.18	0.75	0.42	0.92	กลาง	-0.04	1.16	1.16	0.08	-0.15	0.23	0.98	กลาง
91	-0.81	0.54	0.54	-0.92	-0.71	-0.21	-1.16	กลาง	-0.48	0.49	0.50	-0.53	-0.42	-0.11	-0.27	กลาง	-0.62	0.44	0.46	-0.86	-0.38	-0.48	-2.14*	ชาย
92	-0.55	0.74	0.76	-0.68	-0.44	-0.24	-1.28	กลาง	2.54	3.23	4.44	2.65	2.48	0.17	0.32	กลาง	-0.43	0.99	1.12	-0.76	-0.10	-0.66	-2.91**	ชาย
93	0.65	0.88	0.86	0.69	0.61	0.08	0.38	กลาง	1.65	0.60	0.60	1.41	1.89	-0.48	-1.00	กลาง	0.26	0.85	0.90	0.28	0.24	0.04	0.17	กลาง
94	-2.29	1.58	1.52	-2.27	-2.31	0.04	0.23	กลาง	-1.94	1.11	1.08	-2.02	-1.85	-0.17	-0.43	กลาง	-1.46	1.78	1.79	-1.13	-1.81	0.68	3.05	กลาง
95	0.32	1.16	1.15	0.50	0.16	0.34	1.63	กลาง	0.64	0.68	0.64	0.42	0.85	-0.43	-0.97	กลาง	-0.77	0.54	0.55	-0.78	-0.75	-0.03	-0.14	กลาง
96	-0.80	0.52	0.52	-0.91	-0.69	-0.21	-1.15	กลาง	-0.39	0.85	0.84	-0.71	-0.07	-0.64	-1.53	กลาง	-0.44	0.49	0.54	-0.60	-0.28	-0.33	-1.44	กลาง
97	0.01	0.48	0.48	0.10	-0.06	0.16	0.82	กลาง	0.06	0.58	0.58	0.32	-0.16	0.48	1.10	กลาง	0.11	0.51	0.60	0.17	0.06	0.11	0.46	กลาง
98	0.20	0.81	0.81	0.32	0.09	0.23	1.13	กลาง	-0.08	0.90	0.86	-0.17	0.02	-0.18	-0.43	กลาง	0.31	0.60	0.58	0.25	0.35	-0.10	-0.41	กลาง
99	-0.49	1.53	1.51	-0.57	-0.42	-0.15	-0.79	กลาง	0.39	0.58	0.58	-0.07	0.85	-0.92	-2.11*	ชาย	-0.81	1.15	1.17	-0.96	-0.65	-0.30	-1.37	กลาง

ตาราง 31 (ต่อ)

ID	แยกองค์ประกอบ						รวมองค์ประกอบ						ผสมผสาน											
	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความ ต่าง	Z	ผล	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความ ต่าง	Z	ผล	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความ ต่าง	Z	ผล
100	-0.74	1.84	1.81	-0.64	-0.83	0.18	0.99	กลาง	-0.90	0.97	0.98	-1.05	-0.76	-0.29	-0.72	กลาง	-0.40	1.11	1.09	-0.20	-0.58	0.38	1.64	กลาง
101	-0.30	1.25	1.24	0.04	-0.59	0.63	3.23**	หญิง	1.54	1.23	1.12	1.91	1.25	0.66	1.37	กลาง	1.04	1.08	1.08	1.10	0.98	0.12	0.46	กลาง
102	0.00	0.94	0.93	0.19	-0.15	0.34	1.70	กลาง	-1.49	0.67	0.66	-1.22	-1.76	0.54	1.33	กลาง	-1.19	0.97	0.98	-1.13	-1.24	0.11	0.50	กลาง
103	-0.21	0.62	0.62	-0.17	-0.24	0.08	0.39	กลาง	-0.03	0.58	0.59	-0.07	0.02	-0.09	-0.21	กลาง	0.11	1.38	1.39	0.38	-0.12	0.50	2.11*	หญิง
104	0.12	0.85	0.83	0.39	-0.10	0.48	2.39*	หญิง	-0.03	1.32	1.32	0.12	-0.16	0.28	0.65	กลาง	-0.64	0.69	0.70	-0.63	-0.65	0.02	0.11	กลาง
105	0.01	0.46	0.49	-0.03	0.05	-0.08	-0.38	กลาง	-0.03	0.77	0.83	-0.26	0.20	-0.46	-1.07	กลาง	0.41	0.68	0.72	0.38	0.44	-0.06	-0.25	กลาง
106	-0.91	0.78	0.77	-0.84	-0.99	0.15	0.84	กลาง	-2.31	0.46	0.48	-2.57	-2.02	-0.55	-1.35	กลาง	-1.21	0.71	0.71	-1.23	-1.19	-0.03	-0.16	กลาง
107	0.33	0.76	0.83	0.25	0.40	-0.15	-0.73	กลาง	1.31	0.95	0.93	1.41	1.25	0.16	0.34	กลาง	0.77	0.71	0.73	0.63	0.89	-0.27	-1.08	กลาง
108	0.20	1.09	1.10	0.04	0.34	-0.30	-1.50	กลาง	-0.43	0.85	0.82	-0.44	-0.42	-0.02	-0.06	กลาง	0.91	1.03	1.07	0.89	0.92	-0.03	-0.13	กลาง
109	0.62	0.73	0.75	0.48	0.76	-0.28	-1.34	กลาง	-0.12	0.45	0.47	-0.17	-0.07	-0.10	-0.23	กลาง	-0.35	0.60	0.61	-0.50	-0.20	-0.30	-1.31	กลาง
110	-0.68	0.92	0.90	-0.59	-0.76	0.17	0.92	กลาง	-0.43	0.78	0.77	-0.26	-0.59	0.33	0.78	กลาง	-0.60	0.77	0.80	-0.50	-0.70	0.20	0.90	กลาง
111	0.90	1.42	1.52	0.79	1.01	-0.22	-1.00	กลาง	-0.12	0.75	0.79	-0.26	0.02	-0.28	-0.65	กลาง	0.74	1.23	1.33	0.56	0.89	-0.33	-1.34	กลาง
112	0.72	1.09	1.06	0.67	0.76	-0.09	-0.44	กลาง	-0.39	1.03	0.97	-0.35	-0.42	0.07	0.16	กลาง	0.16	0.89	0.92	0.20	0.14	0.06	0.24	กลาง
113	-1.67	0.89	0.91	-1.65	-1.70	0.05	0.29	กลาง	-0.86	1.10	1.08	-0.88	-0.84	-0.04	-0.10	กลาง	-2.01	0.78	0.77	-1.93	-2.11	0.19	0.84	กลาง
114	-0.85	0.71	0.73	-1.01	-0.69	-0.31	-1.72	กลาง	-1.07	0.92	0.89	-1.22	-0.93	-0.29	-0.72	กลาง	-0.32	1.09	1.11	-0.47	-0.18	-0.30	-1.30	กลาง
115	0.83	1.37	1.28	1.27	0.48	0.78	3.54**	หญิง	0.79	1.43	1.35	1.65	0.11	1.55	3.34**	หญิง	1.57	1.32	1.15	2.31	1.07	1.23	4.22**	หญิง
116	-1.73	1.05	1.09	-1.86	-1.59	-0.28	-1.56	กลาง	-0.52	1.24	1.23	-0.80	-0.25	-0.55	-1.32	กลาง	-2.08	0.84	0.86	-2.20	-1.93	-0.27	-1.22	กลาง
117	0.60	1.80	1.73	0.72	0.50	0.21	0.99	กลาง	0.69	2.03	2.06	1.41	0.11	1.30	2.85**	หญิง	0.10	1.15	1.16	0.20	0.01	0.19	0.80	กลาง
118	0.06	0.81	0.82	0.12	0.01	0.11	0.55	กลาง	0.39	0.84	0.83	0.73	0.11	0.63	1.41	กลาง	-0.28	0.70	0.70	-0.15	-0.40	0.26	1.11	กลาง
119	-0.10	1.32	1.32	0.39	-0.51	0.89	4.47**	หญิง	-0.82	1.05	1.02	-0.53	-1.09	0.56	1.35	กลาง	0.03	1.59	1.55	0.34	-0.25	0.60	2.53*	หญิง
120	0.13	0.84	0.89	-0.19	0.44	-0.63	-3.14**	ชาย	-0.30	0.89	0.89	-0.80	0.20	-0.99	-2.36*	ชาย	-0.76	0.50	0.53	-1.06	-0.46	-0.60	-2.71*	ชาย
121	-0.73	0.52	0.52	-0.78	-0.68	-0.11	-0.58	กลาง	-0.86	0.57	0.57	-0.88	-0.84	-0.04	-0.10	กลาง	-1.14	0.52	0.54	-1.18	-1.09	-0.08	-0.38	กลาง
122	-0.34	0.72	0.73	-0.66	-0.02	-0.63	-3.32**	ชาย	-0.69	1.04	1.07	-0.71	-0.67	-0.04	-0.09	กลาง	-0.62	0.92	0.91	-1.06	-0.18	-0.88	-3.93**	กลาง
123	-0.91	1.33	1.34	-0.71	-1.09	0.37	2.03*	หญิง	0.11	1.13	1.10	0.12	0.11	0.01	0.03	กลาง	-0.31	1.23	1.22	-0.09	-0.51	0.41	1.79	กลาง
124	-0.33	0.84	0.84	-0.29	-0.37	0.08	0.43	กลาง	0.20	0.93	0.97	0.02	0.38	-0.36	-0.82	กลาง	-0.03	0.64	0.63	-0.01	-0.05	0.04	0.17	กลาง

ตาราง 31 (ต่อ)

ID	แยกองค์ประกอบ							รวมองค์ประกอบ							ผสมผสาน									
	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความต่าง	Z	ผล	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความต่าง	Z	ผล	Logit	Infit	Outfit	Logit ชาย	Logit หญิง	ความต่าง	Z	ผล
125	-0.37	1.92	1.90	-0.51	-0.23	-0.29	-1.51	กลาง	0.39	1.58	1.55	0.22	0.56	-0.35	-0.79	กลาง	-0.45	1.55	1.55	-0.71	-0.20	-0.51	-2.23*	ชาย
126	0.64	1.01	0.99	0.84	0.46	0.38	1.76	กลาง	0.74	0.44	0.42	0.42	1.05	-0.63	-1.40	กลาง	0.91	0.62	0.61	0.72	1.07	-0.35	-1.39	กลาง
127	-0.42	2.39	2.35	-0.05	-0.74	0.70	3.62**	หญิง	-0.08	1.82	1.78	0.22	-0.33	0.55	1.28	กลาง	-1.20	1.40	1.42	-1.01	-1.39	0.38	1.71	กลาง
128	0.03	1.17	1.15	-0.07	0.13	-0.19	-0.97	กลาง	0.69	1.82	1.75	0.63	0.75	-0.13	-0.29	กลาง	0.54	1.32	1.27	0.47	0.60	-0.14	-0.56	กลาง
129	-0.49	1.79	1.82	-0.44	-0.54	0.10	0.54	กลาง	-0.43	1.03	1.04	-0.62	-0.25	-0.38	-0.90	กลาง	-0.01	1.58	1.57	-0.09	0.06	-0.15	-0.64	กลาง
130	-0.45	1.41	1.41	-0.55	-0.35	-0.20	-1.06	กลาง	-1.07	1.88	1.99	-0.80	-1.34	0.55	1.32	กลาง	-0.11	1.26	1.27	-0.09	-0.12	0.03	0.14	กลาง

ตาราง 32 แสดงการทำหน้าที่ต่างกันของผู้ให้คะแนนเมื่อเวลาผ่านไป จำแนกเป็นรายบุคคล

ID	แยกองค์ประกอบ								รวมองค์ประกอบ								ผสมผสาน							
	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล
			t	วัน 1	วัน 2	ต่าง						วัน 1	วัน 2	ต่าง						วัน 1	วัน 2	ต่าง		
1	0.85	0.66	0.68	1.25	0.50	0.75	4.91**	ไม่คงที่	0.20	0.51	0.51	0.07	0.33	-0.26	-0.86	คงที่	0.76	0.78	0.76	1.27	0.31	0.97	5.61*	ไม่คงที่
2	-1.45	0.97	0.98	-1.31	-1.58	0.27	2.21**	ไม่คงที่	-1.20	1.72	1.72	-0.53	-1.85	1.32	4.61*	ไม่คงที่	-1.39	0.96	0.98	-1.17	-1.61	0.43	2.89*	ไม่คงที่
3	1.52	1.57	1.86	2.16	1.01	1.15	6.71**	ไม่คงที่	1.36	2.80	3.14	2.35	0.55	1.79	5.40**	ไม่คงที่	1.36	1.85	2.19	1.73	1.04	0.69	3.76*	ไม่คงที่
4	-0.37	1.46	1.55	-0.81	0.09	-0.90	-6.80**	ไม่คงที่	-1.20	0.67	0.68	-1.66	-0.73	-0.94	-3.31*	ไม่คงที่	-0.61	0.92	0.92	-1.02	-0.19	-0.83	-5.42*	ไม่คงที่
5	0.50	2.03	2.35	-0.20	1.31	-1.51	-10.21*	ไม่คงที่	0.63	1.08	1.23	0.48	0.78	-0.30	-0.99	คงที่	0.67	1.14	1.25	0.94	0.43	0.51	3.04*	ไม่คงที่
6	-0.72	0.55	0.56	-0.25	-1.15	0.90	6.97**	ไม่คงที่	-0.73	0.98	1.00	-0.27	-1.17	0.89	3.11*	ไม่คงที่	-0.59	0.80	0.82	0.06	-1.19	1.25	8.04*	ไม่คงที่
7	-0.16	1.23	1.22	-0.95	0.78	-1.73	-12.53*	ไม่คงที่	-0.29	1.12	1.12	-1.26	0.74	-2.00	-6.80*	ไม่คงที่	0.03	1.28	1.50	-0.80	0.98	-1.78	-10.93*	ไม่คงที่
8	-0.73	1.15	1.15	-0.10	-1.29	1.18	9.16**	ไม่คงที่	-0.25	0.77	0.77	-0.06	-0.44	0.38	1.30	คงที่	-0.70	1.14	1.17	-0.28	-1.10	0.82	5.34*	ไม่คงที่
9	-0.82	0.75	0.75	-1.32	-0.30	-1.02	-8.03*	ไม่คงที่	0.18	0.53	0.53	0.21	0.15	0.05	0.18	คงที่	-0.52	0.78	0.79	-0.89	-0.15	-0.73	-4.75*	ไม่คงที่
10	-1.04	1.63	1.62	-1.21	-0.86	-0.35	-2.82*	ไม่คงที่	-0.10	1.65	1.64	-0.27	0.07	-0.34	-1.17	คงที่	-0.76	1.58	1.58	-0.95	-0.56	-0.39	-2.59*	ไม่คงที่
11	0.46	1.40	1.56	1.11	-0.09	1.20	8.19*	ไม่คงที่	0.61	1.30	1.34	1.52	-0.19	1.71	5.52*	ไม่คงที่	0.81	1.08	1.09	1.37	0.32	1.05	6.07*	ไม่คงที่
12	0.13	1.56	1.51	0.03	0.23	-0.20	-1.44	คงที่	0.59	2.00	1.92	0.62	0.55	0.07	0.22	คงที่	0.15	1.28	1.28	0.32	-0.02	0.34	2.11*	ไม่คงที่
13	-1.42	0.99	0.99	-0.64	-2.19	1.55	12.23*	ไม่คงที่	-0.55	1.41	1.40	-0.65	-0.44	-0.21	-0.74	คงที่	-1.42	1.11	1.12	-1.06	-1.78	0.72	4.76*	ไม่คงที่
14	1.19	0.92	0.91	1.65	0.79	0.86	5.35**	ไม่คงที่	0.47	1.14	1.10	1.01	-0.02	1.03	3.38*	ไม่คงที่	0.96	1.78	1.66	1.92	0.19	1.73	9.56*	ไม่คงที่
15	0.67	0.78	0.81	0.99	0.38	0.60	4.04**	ไม่คงที่	-0.77	1.55	1.56	0.81	-2.22	3.03	10.11*	ไม่คงที่	0.69	1.23	1.21	1.16	0.27	0.89	5.23*	ไม่คงที่
16	0.86	0.77	0.79	0.67	1.05	-0.38	-2.50*	ไม่คงที่	0.29	0.81	0.78	-0.10	0.69	-0.79	-2.65*	ไม่คงที่	0.85	0.69	0.69	0.79	0.90	-0.11	-0.66	คงที่
17	-1.05	1.19	1.21	-0.48	-1.59	1.11	8.80*	ไม่คงที่	0.18	1.37	1.36	1.06	-0.60	1.66	5.51*	ไม่คงที่	0.44	1.07	1.05	1.14	-0.18	1.32	7.85*	ไม่คงที่
18	0.00	0.90	0.88	-0.09	0.08	-0.17	-1.23	คงที่	0.00	0.65	0.67	-0.10	0.11	-0.21	-0.72	คงที่	-0.05	0.88	0.89	0.26	-0.35	0.60	3.79*	ไม่คงที่
19	-0.30	1.22	1.22	-0.66	0.07	-0.73	-5.53*	ไม่คงที่	-0.17	1.40	1.41	-0.44	0.11	-0.55	-1.90	คงที่	-0.61	1.23	1.25	-1.14	-0.07	-1.07	-6.94*	ไม่คงที่
20	-0.74	0.72	0.75	-0.68	-0.81	0.13	1.02	คงที่	-1.01	0.82	0.85	-0.61	-1.41	0.80	2.79*	ไม่คงที่	-1.26	0.81	0.85	-1.14	-1.38	0.24	1.62	คงที่
21	0.18	0.63	0.60	-0.13	0.50	-0.63	-4.50*	ไม่คงที่	0.80	0.83	0.80	0.96	0.64	0.31	1.02	คงที่	-0.02	0.71	0.70	-0.32	0.28	-0.60	-3.74*	ไม่คงที่
22	0.35	1.50	1.45	0.50	0.22	0.28	1.98*	ไม่คงที่	0.61	1.35	1.29	0.34	0.88	-0.53	-1.75	คงที่	0.56	1.71	1.68	0.46	0.66	-0.20	-1.18	คงที่
23	-0.38	0.77	0.75	-0.09	-0.65	0.56	4.26*	ไม่คงที่	0.27	0.78	0.77	0.76	-0.19	0.95	3.17*	ไม่คงที่	-0.14	0.66	0.65	0.10	-0.36	0.46	2.90*	ไม่คงที่
24	-1.20	0.36	0.38	-1.14	-1.26	0.12	0.95	คงที่	-0.48	0.54	0.54	-0.32	-0.64	0.33	1.13	คงที่	-1.01	0.72	0.75	-0.53	-1.47	0.94	6.15*	ไม่คงที่
25	0.81	0.95	0.91	0.97	0.66	0.32	2.10*	ไม่คงที่	1.71	1.01	0.98	1.86	1.57	0.29	0.88	คงที่	0.85	0.85	0.84	1.02	0.69	0.33	1.92	คงที่

ตาราง 32 (ต่อ)

ID	แยกองค์ประกอบ								รวมองค์ประกอบ								ผสมผสาน							
	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล
			t	วัน 1	วัน 2	ต่าง						วัน 1	วัน 2	ต่าง						วัน 1	วัน 2	ต่าง		
26	0.40	0.85	0.84	-0.10	0.96	-1.06	-7.36*	ไม่คงที่	0.84	0.95	0.89	0.12	1.62	-1.50	-4.83*	ไม่คงที่	0.82	0.76	0.74	0.38	1.29	-0.91	-5.27*	ไม่คงที่
27	-1.23	0.43	0.44	-1.32	-1.15	-0.17	-1.39	คงที่	-0.99	0.41	0.41	-1.10	-0.89	-0.22	-0.76	คงที่	-1.13	0.39	0.40	-1.17	-1.10	-0.07	-0.47	คงที่
28	-0.92	0.73	0.73	-1.03	-0.81	-0.22	-1.76	คงที่	-1.06	0.91	0.92	-0.98	-1.13	0.15	0.52	คงที่	-0.65	0.90	0.88	-0.78	-0.52	-0.26	-1.68	คงที่
29	-0.20	0.94	0.95	-0.13	-0.27	0.14	1.03	คงที่	-0.04	1.26	1.27	0.62	-0.64	1.27	4.27*	ไม่คงที่	0.27	0.91	0.92	0.23	0.31	-0.07	-0.46	คงที่
30	-0.39	1.61	1.67	-0.09	-0.65	0.56	4.25*	ไม่คงที่	0.03	1.89	1.83	-0.06	0.11	-0.17	-0.57	คงที่	0.04	1.85	1.82	0.07	0.01	0.06	0.40	คงที่
31	0.64	0.76	0.77	0.67	0.60	0.07	0.48	คงที่	0.84	0.59	0.58	0.62	1.07	-0.45	-1.45	คงที่	0.61	0.45	0.46	0.42	0.80	-0.38	-2.27*	ไม่คงที่
32	-0.28	0.97	0.95	-0.09	-0.45	0.35	2.66*	ไม่คงที่	0.14	0.71	0.70	0.39	-0.10	0.49	1.66	คงที่	-0.05	0.87	0.84	0.03	-0.13	0.16	1.02	คงที่
33	0.28	0.79	0.81	0.02	0.55	-0.53	-3.77*	ไม่คงที่	1.04	1.15	1.09	0.72	1.36	-0.65	-2.07*	ไม่คงที่	0.72	0.96	0.96	0.20	1.27	-1.07	-6.24*	ไม่คงที่
34	0.12	0.98	0.96	0.47	-0.20	0.67	4.79*	ไม่คงที่	0.43	0.70	0.68	0.91	-0.02	0.93	3.06*	ไม่คงที่	0.54	0.99	0.98	0.73	0.36	0.37	2.23*	ไม่คงที่
35	-0.65	0.63	0.64	-0.54	-0.77	0.23	1.81	คงที่	-0.57	0.65	0.66	-0.44	-0.68	0.24	0.84	คงที่	-0.46	0.60	0.60	-0.16	-0.76	0.60	3.88*	ไม่คงที่
36	-0.13	1.15	1.20	-0.55	0.32	-0.88	-6.47*	ไม่คงที่	-0.32	1.11	1.11	-1.10	0.51	-1.61	-5.51*	ไม่คงที่	-0.20	0.99	1.10	-0.28	-0.13	-0.15	-0.97	คงที่
37	-0.35	0.85	0.86	0.04	-0.69	0.73	5.52*	ไม่คงที่	-1.04	0.74	0.75	-0.44	-1.61	1.16	4.06*	ไม่คงที่	-0.84	0.72	0.74	-0.44	-1.23	0.80	5.22*	ไม่คงที่
38	0.52	0.63	0.64	0.52	0.53	-0.01	-0.06	คงที่	0.22	0.88	0.87	0.30	0.15	0.14	0.48	คงที่	0.35	0.70	0.70	0.60	0.11	0.49	2.99*	ไม่คงที่
39	0.43	0.88	0.93	0.99	-0.06	1.04	7.16*	ไม่คงที่	-1.14	0.77	0.77	-0.94	-1.33	0.39	1.37	คงที่	0.40	1.07	1.07	0.96	-0.09	1.05	6.28*	ไม่คงที่
40	1.03	1.10	1.05	1.84	0.37	1.47	9.15*	ไม่คงที่	0.73	0.54	0.55	1.26	0.24	1.02	3.30*	ไม่คงที่	0.83	1.20	1.16	1.69	0.12	1.57	8.85*	ไม่คงที่
41	-0.67	0.88	0.88	-0.49	-0.84	0.34	2.70*	ไม่คงที่	-0.65	0.66	0.67	-0.78	-0.52	-0.26	-0.89	คงที่	-1.12	0.81	0.81	-1.16	-1.08	-0.08	-0.55	คงที่
42	1.35	0.81	0.91	1.42	1.29	0.13	0.82	คงที่	1.09	0.71	0.70	1.01	1.16	-0.16	-0.50	คงที่	1.21	0.73	0.87	1.37	1.05	0.32	1.79	คงที่
43	0.35	1.47	1.46	0.64	0.09	0.55	3.83*	ไม่คงที่	0.63	1.61	1.55	0.57	0.69	-0.12	-0.38	คงที่	0.44	2.23	2.14	0.46	0.43	0.04	0.22	คงที่
44	-0.87	1.37	1.35	-0.64	-1.09	0.46	3.63*	ไม่คงที่	0.49	1.57	1.52	0.25	0.74	-0.48	-1.60	คงที่	-0.49	1.62	1.60	-0.35	-0.63	0.28	1.79	คงที่
45	-0.32	0.79	0.79	-0.60	-0.03	-0.57	-4.34*	ไม่คงที่	-1.46	0.66	0.67	-1.34	-1.57	0.22	0.79	คงที่	-0.72	0.59	0.63	-0.85	-0.58	-0.27	-1.75	คงที่
46	0.03	1.80	1.80	0.86	-0.65	1.51	10.75*	ไม่คงที่	0.75	1.22	1.24	1.80	-0.15	1.95	6.18*	ไม่คงที่	0.39	1.56	1.60	1.05	-0.19	1.24	7.41*	ไม่คงที่
47	1.40	1.04	1.02	1.38	1.42	-0.04	-0.23	คงที่	-1.24	0.57	0.58	-1.50	-0.97	-0.54	-1.89	คงที่	0.62	1.38	1.46	0.79	0.45	0.34	2.01*	ไม่คงที่
48	1.09	1.41	1.39	1.22	0.96	0.26	1.68	คงที่	0.38	1.07	1.05	0.43	0.33	0.11	0.35	คงที่	0.89	1.57	1.54	0.83	0.95	-0.11	-0.65	คงที่
49	-0.86	1.26	1.29	-0.36	-1.33	0.97	7.63*	ไม่คงที่	-0.99	0.79	0.80	-1.10	-0.89	-0.22	-0.76	คงที่	-0.56	1.28	1.33	0.10	-1.18	1.28	8.20*	ไม่คงที่
50	-0.18	0.50	0.49	-0.42	0.07	-0.49	-3.65*	ไม่คงที่	0.25	0.89	0.87	0.21	0.29	-0.08	-0.26	คงที่	-0.07	0.66	0.64	-0.28	0.14	-0.42	-2.63*	ไม่คงที่

ตาราง 32 (ต่อ)

ID	แยกองค์ประกอบ								รวมองค์ประกอบ								ผสมผสาน							
	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล
			t	วัน 1	วัน 2	ต่าง						วัน 1	วัน 2	ต่าง						วัน 1	วัน 2	ต่าง		
51	0.12	0.45	0.47	0.39	-0.14	0.53	3.82*	ไม่คงที่	-0.10	0.49	0.54	-0.36	0.15	-0.51	-1.75	คงที่	0.22	0.37	0.39	0.41	0.03	0.37	2.29*	ไม่คงที่
52	0.63	1.07	1.04	0.97	0.31	0.66	4.45*	ไม่คงที่	0.25	0.98	0.98	-0.01	0.51	-0.52	-1.75	คงที่	0.46	0.98	0.97	0.49	0.44	0.05	0.30	คงที่
53	0.30	0.86	0.85	0.51	0.11	0.40	2.81*	ไม่คงที่	-0.17	1.31	1.32	0.67	-0.93	-1.60	5.39*	ไม่คงที่	-0.66	1.10	1.09	-0.19	-1.10	0.91	5.88*	ไม่คงที่
54	-0.71	0.91	0.91	-0.64	-0.77	0.13	1.03	คงที่	-0.34	0.77	0.77	-0.23	-0.44	0.21	0.71	คงที่	-0.64	0.78	0.77	-0.69	-0.58	-0.10	-0.68	คงที่
55	0.05	1.23	1.24	0.54	-0.39	0.93	6.72*	ไม่คงที่	-0.36	1.18	1.18	-0.19	-0.52	0.33	1.15	คงที่	0.69	1.62	1.59	0.28	1.13	-0.84	-4.97*	ไม่คงที่
56	-0.84	1.52	1.49	-0.64	-1.02	0.38	3.01*	ไม่คงที่	-0.42	1.12	-1.09	-1.06	0.24	-1.30	-4.49*	ไม่คงที่	-1.06	1.43	1.42	-0.99	-1.12	0.13	0.89	คงที่
57	0.68	0.65	0.66	0.24	1.16	-0.92	-6.17*	ไม่คงที่	0.77	0.65	0.66	0.39	1.16	-0.78	-2.52*	ไม่คงที่	0.58	0.67	0.68	0.18	1.01	-0.83	-4.93*	ไม่คงที่
58	-0.45	1.40	1.42	-1.37	0.63	-2.00	-14.84*	ไม่คงที่	-0.12	0.92	0.93	-0.86	0.64	-1.50	-5.10*	ไม่คงที่	-0.28	1.14	1.17	-0.93	0.41	-1.34	-8.51*	ไม่คงที่
59	0.13	0.88	0.84	-0.14	0.41	-0.55	-3.93*	ไม่คงที่	-0.29	1.04	1.02	-0.49	-0.10	-0.38	-1.32	คงที่	-0.22	0.95	0.93	-0.41	-0.03	-0.39	-2.46*	ไม่คงที่
60	0.15	1.15	1.15	0.60	-0.25	0.85	6.05*	ไม่คงที่	-0.71	1.49	1.47	-0.06	-1.33	1.27	4.39*	ไม่คงที่	-0.30	1.27	1.25	0.22	-0.77	0.99	6.27*	ไม่คงที่
61	-1.28	1.23	1.22	-0.83	-1.71	0.88	7.06*	ไม่คงที่	-1.44	1.36	1.36	-1.58	-1.29	-0.30	-1.05	คงที่	-1.24	1.44	1.43	-1.17	-1.30	0.13	0.88	คงที่
62	0.88	0.64	0.68	1.20	0.60	0.59	3.88*	ไม่คงที่	-0.40	0.69	0.69	-0.44	-0.36	-0.09	-0.31	คงที่	0.18	0.82	0.81	0.13	0.23	-0.10	-0.63	คงที่
63	0.16	1.31	1.34	0.44	-0.09	0.53	3.78*	ไม่คงที่	-0.57	1.07	1.06	0.16	-1.25	1.41	4.85*	ไม่คงที่	-0.33	1.19	1.20	0.02	-0.65	0.68	4.32*	ไม่คงที่
64	-0.09	1.14	1.09	-0.28	0.10	-0.38	-2.79*	ไม่คงที่	0.14	1.76	1.91	0.03	0.24	-0.21	-0.71	คงที่	-0.18	0.81	0.81	-0.16	-0.21	0.06	0.37	คงที่
65	0.51	0.69	0.67	0.28	0.75	-0.47	-3.25*	ไม่คงที่	0.68	0.71	0.68	0.81	0.55	0.26	0.85	คงที่	0.49	0.91	0.89	0.10	0.90	-0.80	-4.82*	ไม่คงที่
66	0.78	1.33	1.26	0.26	1.36	-1.10	-7.27*	ไม่คงที่	1.42	1.37	1.31	0.57	2.35	-1.78	-5.42*	ไม่คงที่	1.36	1.10	1.03	1.24	1.48	-0.24	-1.34	คงที่
67	-0.65	0.63	0.64	-0.21	-1.05	0.83	6.46*	ไม่คงที่	-1.10	0.40	0.41	-0.78	-1.41	0.63	2.22*	ไม่คงที่	-0.62	0.67	0.68	-0.22	-1.01	0.79	5.13*	ไม่คงที่
68	0.08	0.66	0.70	0.32	-0.15	0.47	3.39*	ไม่คงที่	-0.61	0.87	0.89	0.12	-1.29	1.41	4.84*	ไม่คงที่	-0.29	0.72	0.72	-0.08	-0.49	0.41	2.62*	ไม่คงที่
69	-1.23	1.44	1.45	-1.31	-1.15	-0.16	-1.33	คงที่	-1.22	2.39	2.38	-1.46	-0.97	-0.50	-1.75	คงที่	-1.34	1.43	1.43	-1.43	-1.25	-0.18	-1.23	คงที่
70	1.47	0.95	0.98	1.12	1.84	-0.72	-4.32*	ไม่คงที่	1.55	1.53	1.60	0.91	2.23	-1.32	-4.02*	ไม่คงที่	1.26	1.17	1.21	0.66	1.93	-1.27	-6.95*	ไม่คงที่
71	-0.51	1.19	1.22	-0.25	-0.75	0.50	3.86*	ไม่คงที่	-1.42	0.55	0.57	-1.18	-1.65	0.46	1.64	คงที่	-0.63	1.30	1.32	-0.69	-0.57	-0.12	-0.76	คงที่
72	2.09	0.69	0.68	1.86	2.33	-0.47	-2.57*	ไม่คงที่	1.87	0.81	0.77	1.47	2.29	-0.82	-2.44*	ไม่คงที่	2.30	0.59	0.57	2.17	2.44	-0.26	-1.24	คงที่
73	0.20	1.45	1.44	-0.39	0.88	-1.27	-8.95*	ไม่คงที่	0.80	1.96	1.94	-0.65	2.54	-3.19	-9.84*	ไม่คงที่	0.14	1.63	1.63	-0.71	1.10	-1.81	-11.00*	ไม่คงที่
74	0.87	0.87	0.90	0.70	1.05	-0.36	-2.35*	ไม่คงที่	0.49	1.03	1.03	0.39	0.60	-0.21	-0.69	คงที่	1.10	0.97	0.94	1.09	1.10	0.00	-0.02	คงที่
75	0.35	0.78	0.78	0.34	0.35	-0.01	-0.08	คงที่	0.09	0.79	0.78	0.03	0.15	-0.12	-0.42	คงที่	0.02	0.89	0.89	-0.44	0.51	-0.94	-5.89*	ไม่คงที่

ตาราง 32 (ต่อ)

ID	แยกองค์ประกอบ								รวมองค์ประกอบ								ผสมผสาน							
	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล
			t	วัน 1	วัน 2	ต่าง						วัน 1	วัน 2	ต่าง						วัน 1	วัน 2	ต่าง		
76	0.46	0.90	0.89	1.00	-0.01	1.01	6.91*	ไม่คงที่	0.47	1.09	1.07	0.91	0.07	0.84	2.77*	ไม่คงที่	0.40	1.04	1.02	1.02	-0.14	1.16	6.93*	ไม่คงที่
77	1.06	0.72	0.72	0.81	1.31	-0.50	-3.23*	ไม่คงที่	0.56	0.63	0.63	0.39	0.74	-0.35	-1.15	คงที่	1.31	0.81	0.82	1.09	1.53	-0.44	-2.42*	ไม่คงที่
78	-0.89	0.70	0.70	-0.85	-0.92	0.07	0.55	คงที่	-0.87	0.64	0.65	-0.61	-1.13	0.52	1.81	คงที่	-0.52	0.64	0.67	-0.26	-0.78	0.53	3.40*	ไม่คงที่
79	0.48	0.55	0.54	0.71	0.27	0.44	3.05*	ไม่คงที่	-0.21	0.55	0.54	-0.06	-0.36	0.30	1.02	คงที่	0.46	0.77	0.75	0.78	0.16	0.61	3.69*	ไม่คงที่
80	0.98	0.56	0.61	0.62	1.36	-0.75	-4.82*	ไม่คงที่	0.43	0.64	0.67	0.25	0.60	-0.35	-1.15	คงที่	1.07	0.91	0.91	0.49	1.73	-1.24	-6.93*	ไม่คงที่
81	0.21	0.97	0.98	0.54	-0.08	0.63	4.44*	ไม่คงที่	0.70	0.75	0.72	0.57	0.83	-0.26	-0.84	คงที่	0.40	0.90	0.90	0.73	0.09	0.65	3.91*	ไม่คงที่
82	-0.78	0.80	0.81	-0.68	-0.88	0.21	1.65	คงที่	-0.69	0.88	0.88	-0.69	-0.68	-0.01	-0.03	คงที่	-0.76	0.64	0.65	-0.86	-0.67	-0.20	-1.30	คงที่
83	1.76	0.89	0.96	2.25	1.36	0.88	4.98*	ไม่คงที่	0.70	0.67	0.68	1.01	0.42	0.59	1.92	คงที่	1.65	1.10	1.09	2.09	1.27	0.82	4.24*	ไม่คงที่
84	-0.32	0.97	0.97	-0.13	-0.50	0.37	2.78*	ไม่คงที่	-0.10	0.70	0.69	-0.06	-0.15	0.09	0.30	คงที่	-0.48	0.88	0.87	-0.76	-0.19	-0.57	-3.68*	ไม่คงที่
85	0.76	0.68	0.69	0.23	1.35	-1.12	-7.40*	ไม่คงที่	1.21	0.59	0.57	0.91	1.52	-0.61	-1.92	คงที่	0.72	0.77	0.75	0.22	1.25	-1.04	-6.07*	ไม่คงที่
86	-0.08	0.47	0.49	0.10	-0.24	0.35	2.56*	ไม่คงที่	0.34	0.67	0.69	0.53	0.15	0.37	1.24	คงที่	0.01	0.69	0.72	0.54	-0.48	1.02	6.35*	ไม่คงที่
87	0.53	0.87	0.89	-0.08	1.22	-1.30	-8.82*	ไม่คงที่	0.63	0.81	0.87	-0.10	1.41	-1.52	-4.94*	ไม่คงที่	0.57	1.04	1.09	-0.28	1.57	-1.85	-10.75*	ไม่คงที่
88	-0.94	0.47	0.49	-1.06	-0.81	-0.24	-1.95	คงที่	-1.10	0.46	0.47	-1.26	-0.93	-0.34	-1.19	คงที่	-0.94	0.55	0.56	-1.11	-0.77	-0.34	-2.28*	ไม่คงที่
89	0.26	0.96	0.99	0.10	0.42	-0.31	-2.21*	ไม่คงที่	0.22	0.67	0.69	0.34	0.11	0.23	0.78	คงที่	0.40	0.78	0.81	0.41	0.40	0.01	0.05	คงที่
90	-0.17	0.83	0.82	0.16	-0.47	0.63	4.65*	ไม่คงที่	0.20	1.32	1.32	0.91	-0.44	1.35	4.49*	ไม่คงที่	-0.36	1.08	1.07	-0.03	-0.68	0.65	4.15*	ไม่คงที่
91	-0.70	0.50	0.50	-0.77	-0.62	-0.15	-1.19	คงที่	-0.63	0.40	0.41	-0.44	-0.81	0.36	1.26	คงที่	-0.69	0.43	0.44	-0.57	-0.82	0.25	1.61	คงที่
92	-0.47	0.70	0.73	-0.53	-0.42	-0.11	-0.86	คงที่	0.63	3.06	3.30	2.41	-0.77	3.18	9.73*	ไม่คงที่	-0.64	0.86	0.93	-0.39	-0.88	0.49	3.22*	ไม่คงที่
93	1.02	0.95	0.93	0.63	1.44	-0.81	-5.22*	ไม่คงที่	1.26	0.58	0.58	1.58	0.97	0.61	1.90	คงที่	0.60	0.81	0.82	0.26	0.95	-0.69	-4.11*	ไม่คงที่
94	-1.63	1.57	1.56	-2.17	-1.10	-1.07	-8.62*	ไม่คงที่	-1.14	1.69	1.66	-1.86	-0.40	-1.47	-5.14*	ไม่คงที่	-1.16	1.66	1.67	-1.36	-0.95	-0.41	-2.73*	ไม่คงที่
95	0.31	0.76	0.76	0.31	0.30	0.01	0.05	คงที่	0.87	0.68	0.68	0.62	1.12	-0.50	-1.60	คงที่	-0.26	0.57	0.57	-0.71	0.21	-0.93	-5.89*	ไม่คงที่
96	-0.42	0.56	0.55	-0.76	-0.06	-0.70	-5.35*	ไม่คงที่	-0.34	0.62	0.62	-0.36	-0.31	-0.05	-0.16	คงที่	-0.30	0.49	0.50	-0.40	-0.20	-0.20	-1.29	คงที่
97	0.12	0.53	0.53	0.02	0.23	-0.21	-1.51	คงที่	-0.12	0.55	0.56	0.07	-0.31	0.39	1.32	คงที่	0.18	0.42	0.45	0.11	0.24	-0.13	-0.79	คงที่
98	0.23	0.58	0.59	0.19	0.28	-0.08	-0.60	คงที่	0.25	0.69	0.67	-0.06	0.55	-0.61	-2.05*	ไม่คงที่	0.31	0.48	0.47	0.30	0.32	-0.02	-0.13	คงที่
99	-0.28	1.64	1.61	-0.47	-0.08	-0.38	-2.90*	ไม่คงที่	0.43	0.69	0.69	0.39	0.46	-0.07	-0.25	คงที่	-0.56	1.13	1.12	-0.75	-0.37	-0.38	-2.44*	ไม่คงที่
100	-0.72	1.95	1.94	-0.70	-0.74	0.04	0.33	คงที่	-0.79	1.05	1.05	-0.86	-0.73	-0.13	-0.47	คงที่	-0.58	1.61	1.61	-0.37	-0.79	0.43	2.77*	ไม่คงที่

ตาราง 32 (ต่อ)

ID	แยกองค์ประกอบ								รวมองค์ประกอบ								ผสมผสาน							
	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล
			t	วัน 1	วัน 2	ต่าง						วัน 1	วัน 2	ต่าง						วัน 1	วัน 2	ต่าง		
101	-0.09	1.09	1.07	-0.29	0.11	-0.40	-2.93*	ไม่คงที่	1.09	0.84	0.81	1.47	0.74	0.73	2.32*	ไม่คงที่	0.69	0.87	0.85	0.99	0.41	0.57	3.38*	ไม่คงที่
102	0.10	0.95	0.94	0.01	0.19	-0.18	-1.30	คงที่	-0.67	1.08	1.07	-1.42	0.11	-1.53	-5.32*	ไม่คงที่	-0.62	1.06	1.06	-1.10	-0.12	-0.99	-6.41*	ไม่คงที่
103	0.31	0.83	0.83	-0.20	0.88	-1.07	-7.51*	ไม่คงที่	0.89	1.17	1.20	-0.01	1.89	-1.90	-6.05*	ไม่คงที่	0.75	1.37	1.36	0.11	1.47	-1.35	-7.84*	ไม่คงที่
104	-0.27	1.03	1.00	0.12	-0.63	0.75	5.64*	ไม่คงที่	0.09	1.47	1.45	-0.01	0.20	-0.21	-0.72	คงที่	-0.16	1.11	1.10	-0.59	0.29	-0.89	-5.60*	ไม่คงที่
105	0.51	0.64	0.66	0.02	1.05	-1.04	-7.08*	ไม่คงที่	0.40	0.63	0.68	-0.01	0.83	-0.84	-2.80*	ไม่คงที่	0.18	0.59	0.64	0.39	-0.02	0.41	2.52*	ไม่คงที่
106	-0.38	1.12	1.13	-0.87	0.15	-1.02	-7.71*	ไม่คงที่	-1.22	0.91	0.92	-2.23	-0.19	-2.04	-7.08*	ไม่คงที่	-0.38	1.32	1.33	-1.13	0.43	-1.55	-9.85*	ไม่คงที่
107	0.20	0.82	0.85	0.32	0.08	0.24	1.71	คงที่	0.45	1.31	1.31	1.26	-0.27	1.53	5.01*	ไม่คงที่	0.57	0.88	0.90	0.73	0.43	0.31	1.83	คงที่
108	0.48	0.88	0.89	0.19	0.79	-0.60	-4.11*	ไม่คงที่	0.05	0.79	0.78	-0.40	0.51	-0.91	-3.08*	ไม่คงที่	0.71	0.91	0.94	0.86	0.56	0.30	1.79	คงที่
109	0.48	0.66	0.67	0.61	0.35	0.25	1.74	คงที่	0.05	0.57	0.59	-0.10	0.20	-0.30	-1.01	คงที่	-0.08	0.58	0.60	-0.32	0.16	-0.48	-3.02*	ไม่คงที่
110	-0.64	1.08	1.06	-0.64	-0.63	-0.01	-0.10	คงที่	-0.29	0.95	0.93	-0.40	-0.19	-0.21	-0.73	คงที่	-0.53	1.09	1.10	-0.56	-0.50	-0.06	-0.37	คงที่
111	0.74	1.17	1.21	0.88	0.60	0.28	1.85	คงที่	-0.21	0.70	0.71	-0.10	-0.31	0.21	0.73	คงที่	0.67	1.01	1.05	0.70	0.63	0.07	0.42	คงที่
112	0.14	0.92	0.94	0.70	-0.35	1.04	7.42*	ไม่คงที่	-0.38	0.71	0.70	-0.36	-0.40	0.04	0.13	คงที่	-0.31	0.77	0.79	0.16	-0.75	0.91	5.80*	ไม่คงที่
113	-1.86	0.80	0.81	-1.59	-2.14	0.55	4.43*	ไม่คงที่	-1.42	1.06	1.07	-0.82	-2.01	1.19	4.17*	ไม่คงที่	-1.93	0.67	0.67	-1.88	-1.98	0.10	0.67	คงที่
114	-0.63	0.63	0.66	-0.81	-0.46	-0.35	-2.73*	ไม่คงที่	-0.69	0.89	0.87	-1.02	-0.36	-0.67	-2.32*	ไม่คงที่	-0.33	0.89	0.92	-0.29	-0.36	0.07	0.43	คงที่
115	1.09	1.79	1.63	0.81	1.39	-0.58	-3.71*	ไม่คงที่	1.29	1.83	1.68	0.76	1.83	-1.07	-3.34*	ไม่คงที่	1.51	1.42	1.25	1.49	1.52	-0.02	-0.13	คงที่
116	-1.51	1.13	1.17	-1.64	-1.39	-0.25	-2.07*	ไม่คงที่	-0.63	1.33	1.34	-0.49	-0.77	0.28	0.97	คงที่	-1.62	1.19	1.21	-1.94	-1.31	-0.62	-4.15*	ไม่คงที่
117	0.52	1.45	1.42	0.59	0.46	0.13	0.89	คงที่	0.99	1.61	1.62	0.67	1.31	-0.65	-2.07*	ไม่คงที่	0.19	1.08	1.08	0.10	0.28	-0.18	-1.11	คงที่
118	0.12	0.65	0.65	0.07	0.17	-0.10	-0.75	คงที่	0.31	0.63	0.62	0.39	0.24	0.15	0.49	คงที่	-0.05	0.64	0.64	-0.26	0.16	-0.42	-2.63*	ไม่คงที่
119	0.08	1.38	1.36	-0.09	0.26	-0.35	-2.54*	ไม่คงที่	-0.77	0.79	0.78	-0.78	-0.77	-0.01	-0.04	คงที่	0.07	1.43	1.39	0.03	0.10	-0.06	-0.40	คงที่
120	0.27	0.63	0.65	0.13	0.42	-0.28	-2.00*	ไม่คงที่	-0.02	0.57	0.58	-0.27	0.24	-0.52	-1.75	คงที่	-0.30	0.70	0.72	-0.70	0.11	-0.81	-5.18*	ไม่คงที่
121	-0.75	0.61	0.61	-0.69	-0.81	0.11	0.89	คงที่	-0.83	0.60	0.60	-0.82	-0.85	0.03	0.10	คงที่	-0.98	0.52	0.54	-1.06	-0.91	-0.15	-1.00	คงที่
122	-0.18	1.30	1.28	-0.32	-0.03	-0.29	-2.18*	ไม่คงที่	0.20	1.32	1.32	-0.65	1.12	-1.77	-5.89*	ไม่คงที่	-0.41	0.97	0.97	-0.57	-0.25	-0.32	-2.06*	ไม่คงที่
123	-0.58	1.16	1.17	-0.86	-0.28	-0.58	-4.53*	ไม่คงที่	-0.42	0.97	0.96	0.12	-0.93	1.05	3.59*	ไม่คงที่	-0.34	1.20	1.19	-0.28	-0.39	0.11	0.74	คงที่
124	0.00	0.86	0.84	-0.31	0.32	-0.64	-4.64*	ไม่คงที่	0.36	0.63	0.64	0.21	0.51	-0.30	-1.00	คงที่	0.07	0.57	0.56	-0.02	0.16	-0.18	-1.12	คงที่
125	-0.40	1.51	1.51	-0.35	-0.45	0.10	0.78	คงที่	0.20	1.08	1.07	0.39	0.02	0.36	1.22	คงที่	-0.42	1.17	1.18	-0.41	-0.42	0.00	0.03	คงที่

ตาราง 32 (ต่อ)

ID	แยกองค์ประกอบ								รวมองค์ประกอบ								ผสมผสาน							
	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล	Logit	Infit	Outfit	Logit	Logit	ความ	Z	ผล
			t	วัน 1	วัน 2	ต่าง						วัน 1	วัน 2	ต่าง						วัน 1	วัน 2	ต่าง		
126	0.66	0.78	0.77	0.62	0.71	-0.09	-0.62	คงที่	0.56	0.49	0.47	0.72	0.42	0.30	0.98	คงที่	0.74	0.53	0.52	0.86	0.62	0.25	1.46	คงที่
127	-0.15	2.19	2.16	-0.40	0.10	-0.50	-3.72*	ไม่คงที่	0.05	1.64	1.62	-0.06	0.15	-0.21	-0.72	คงที่	-0.39	1.93	1.93	-1.11	0.40	-1.51	-9.62*	ไม่คงที่
128	0.04	1.28	1.27	0.04	0.04	0.00	0.00	คงที่	0.45	1.57	1.50	0.67	0.24	0.43	1.41	คงที่	0.31	1.38	1.33	0.52	0.11	0.41	2.48*	ไม่คงที่
129	-0.30	1.46	1.45	-0.47	-0.13	-0.34	-2.57*	ไม่คงที่	-0.27	1.03	1.03	-0.40	-0.15	-0.26	-0.88	คงที่	-0.15	1.53	1.51	0.00	-0.29	0.28	1.79	คงที่
130	-0.20	1.06	1.07	-0.42	0.04	-0.46	-3.44*	ไม่คงที่	-0.87	1.43	1.48	-1.02	-0.73	-0.30	-1.04	คงที่	-0.15	1.01	1.00	-0.09	-0.20	0.11	0.69	คงที่

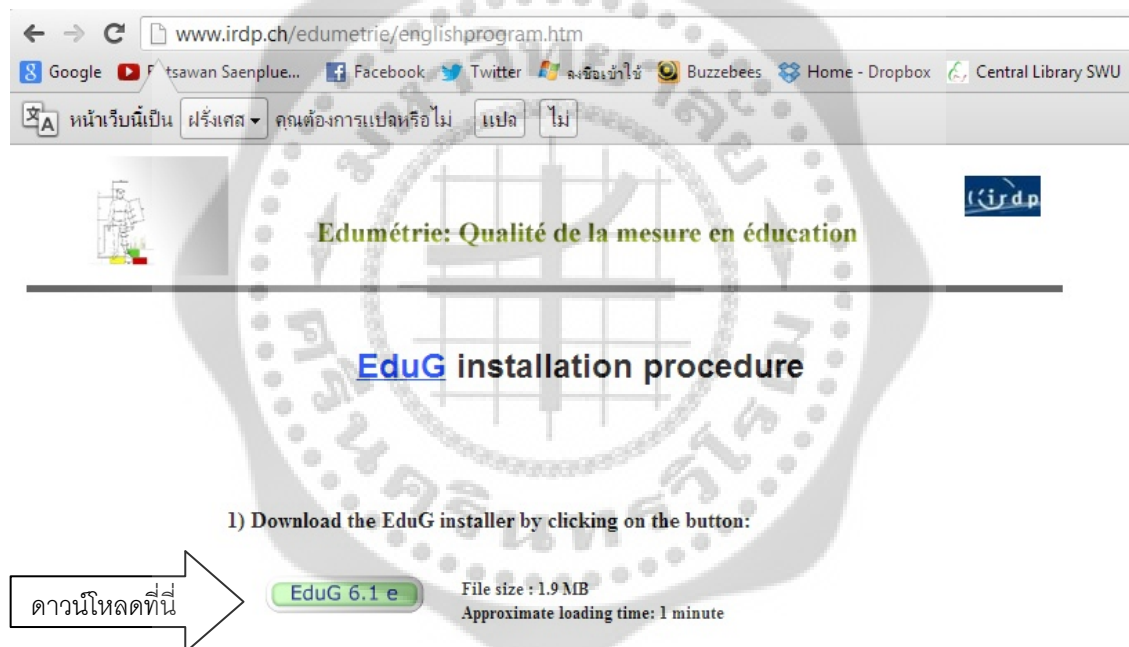


ภาคผนวก ค
คู่มือการใช้โปรแกรม EDUG และ FACETS

คู่มือการใช้โปรแกรม EDUG 6.1-e

ผู้วิจัยนำผลคะแนนที่ได้มาวิเคราะห์โดยใช้โมเดลการสรุปอ้างอิง เพื่อศึกษาความแปรปรวนจากแหล่งต่างๆ ของเครื่องมือที่ใช้ในการวิจัยโดยใช้การออกแบบเอกภพสำหรับสององค์ประกอบ (Two-Facet Design) แบบ Two-Facet Crossed Design โดยใช้โปรแกรม EDUG version 6.1-e กำหนดให้ Population คือ นักเรียน (P) วิธีการตรวจให้คะแนน (I) และ Facet 2 ผู้ให้คะแนน (R) Design คือ $p \times i \times r$ design G-Study คือ กลุ่มตัวอย่าง 50 คน ถูกให้คะแนนโดยใช้วิธีการตรวจให้คะแนน 3 วิธี โดยมีผู้ตรวจให้คะแนน 3 คน ดั้งชั้นตอนต่อไปนี่

ขั้นที่ 1 ดาวน์โหลดโปรแกรมจาก <http://www.irdp.ch/edumetrie/englishprogram.htm> โดยไม่มีค่าใช้จ่าย



The screenshot shows a web browser window with the URL www.irdp.ch/edumetrie/englishprogram.htm. The page content includes the text "Edumétrie: Qualité de la mesure en éducation" and "EduG installation procedure". Below this, there is a step instruction: "1) Download the EduG installer by clicking on the button:". An arrow points to a button labeled "EduG 6.1 e" with the following details: "File size : 1.9 MB" and "Approximate loading time: 1 minute".

1) Download the EduG installer by clicking on the button:



EduG 6.1 e File size : 1.9 MB
Approximate loading time: 1 minute

2) Unzip (decompress) the downloaded file "InstallerEduG61e.zip" by double-clicking its icon. This should create a file named "InstallerEduG61e.exe".

If you do not have a decompression program, you may download and install [Unzip Me](#). (It is free).

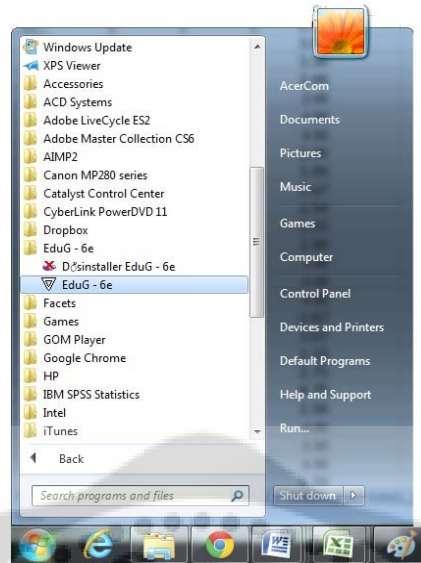
Some anti-virus security systems prevent the transmission of all executable programs. If you meet this problem then send an e-mail to: francois.ducrey@etat.ge.ch.

3) Double click the file name to launch the installer.

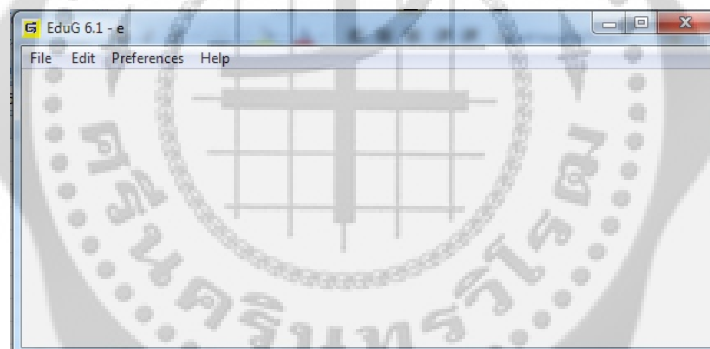
For the users of Windows Vista or Seven, it is imperative to read the following announcement :



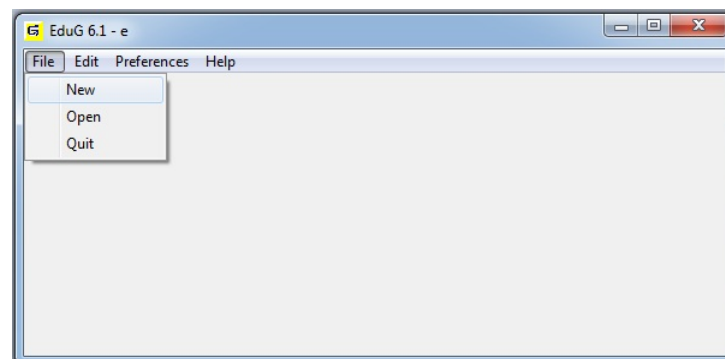
ขั้นที่ 2 เปิดโปรแกรมจาก Start



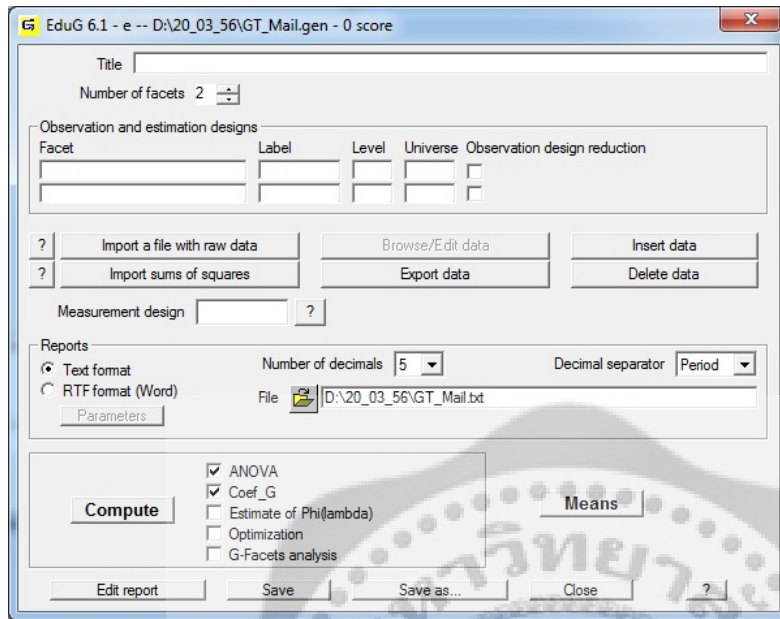
จะพบหน้าจอตั้งนี้



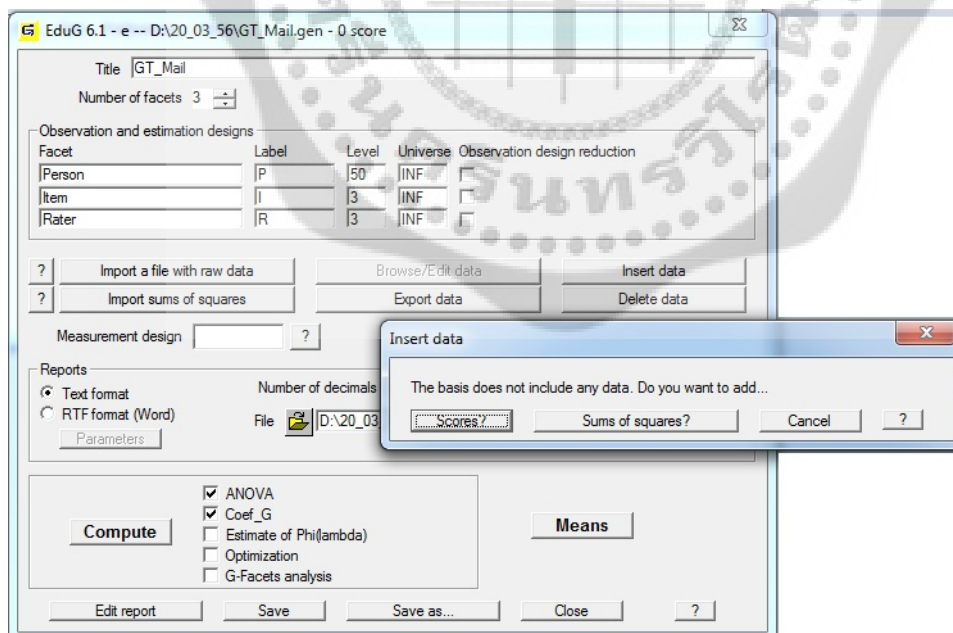
ขั้นที่ 3 ตั้งชื่อไฟล์สำหรับ Output โดยเลือก File>New>ตั้งชื่อ "GT_Mail" จากนั้นเลือก Open



จะพบหน้าจอนี้



ขั้นที่ 4 กรอกข้อมูล จากนั้นนำข้อมูลโดยเลือก Insert data โดยเราสามารถเลือกได้ว่าต้องการใช้ข้อมูลที่เป็นข้อมูลดิบ Scores หรือ Sums of Squares



ในที่นี้ผู้วิจัยใช้ข้อมูลดิบซึ่งสามารถคัดลอกได้จาก Excel โดยที่โปรแกรมได้ตั้งค่าข้อมูลให้แล้วว่าคอลัมน์หรือแถวคืออะไร

	P	I	R	Data
1	1	1	1	3.50
2	1	1	2	3.50
3	1	1	3	2.25
4	1	2	1	2.00
5	1	2	2	4.00
6	1	2	3	2.00
7	1	3	1	2.67
8	1	3	2	2.00
9	1	3	3	2.33
10	2	1	1	1.50
11	2	1	2	3.25
12	2	1	3	1.50
13	2	2	1	1.00
14	2	2	2	4.00
15	2	2	3	1.00
16	2	3	1	1.67
17	2	3	2	3.00
18	2	3	3	2.00

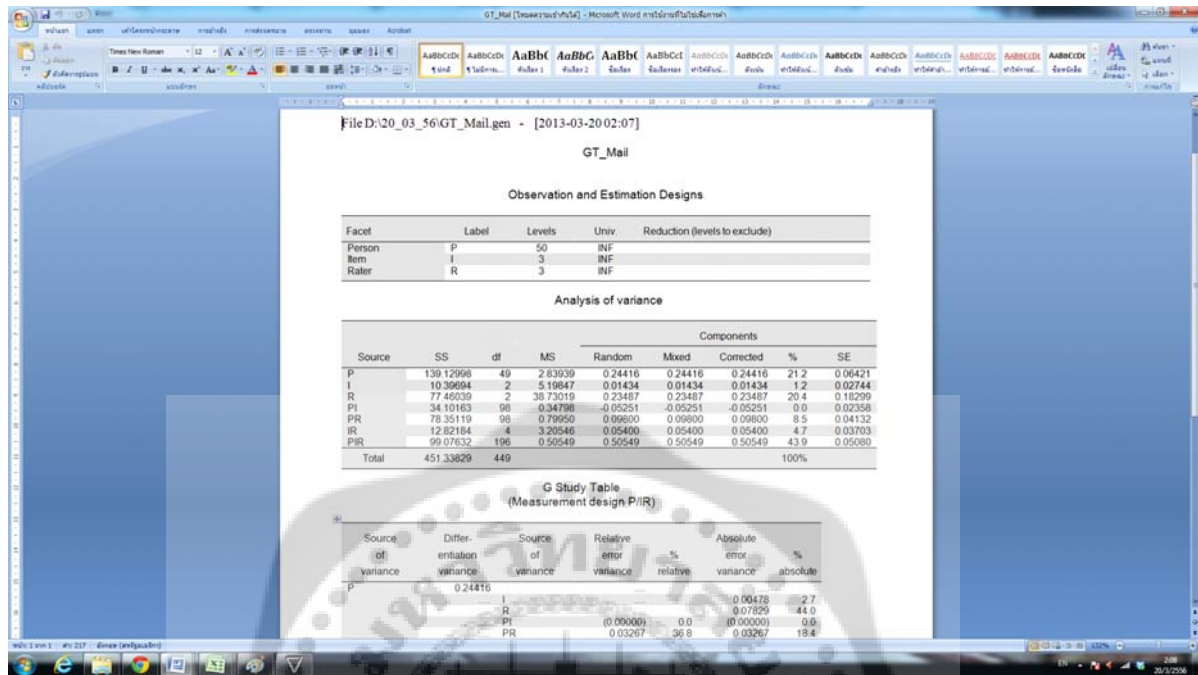
ขั้นที่ 5 เพื่อความสะดวกให้เลือก Reports > RTF format (Word) เพื่อให้Output แสดงใน MicroSoft Office Word จากนั้นสั่งให้เครื่องคำนวณเลือก Compute ด้านล่างซ้าย

The screenshot shows the EduG 6.1 software window with the following settings:

- Title: GT_Mail
- Number of facets: 3
- Observation and estimation designs table:

Facet	Label	Level	Universe	Observation design reduction
Person	P	50	IINF	<input type="checkbox"/>
Item	I	3	IINF	<input type="checkbox"/>
Rater	R	3	IINF	<input type="checkbox"/>
- Measurement design: P/R
- Reports: RTF format (Word) selected
- Number of decimals: 5
- Decimal separator: Period
- File: D:\20_03_56\GT_Mail.rtf
- Compute button is highlighted
- Other options: ANOVA, Coef_G, Estimate of Phi(lambda), Cut Score=, Optimization, G-Facets analysis

จะได้ Output ในรูปแบบ MicroSoft Office Word



ตัวอย่าง Output จากโปรแกรม EDUG 6.1-e

File D:\20_03_56\GT_Mail.gen - [2013-03-20 02:07]

GT_Mail

Observation and Estimation Designs

Facet	Label	Levels	Univ.	Reduction (levels to exclude)
Person	P	50	INF	
Item	I	3	INF	
Rater	R	3	INF	

Analysis of variance

Source	SS	df	MS	Components					SE
				Random	Mixed	Corrected	%		
P	139.12998	49	2.83939	0.24416	0.24416	0.24416	21.2	0.06421	
I	10.39694	2	5.19847	0.01434	0.01434	0.01434	1.2	0.02744	
R	77.46039	2	38.73019	0.23487	0.23487	0.23487	20.4	0.18299	
PI	34.10163	98	0.34798	-0.05251	-0.05251	-0.05251	0.0	0.02358	
PR	78.35119	98	0.79950	0.09800	0.09800	0.09800	8.5	0.04132	
IR	12.82184	4	3.20546	0.05400	0.05400	0.05400	4.7	0.03703	
PIR	99.07632	196	0.50549	0.50549	0.50549	0.50549	43.9	0.05080	
Total	451.33829	449					100%		

G Study Table

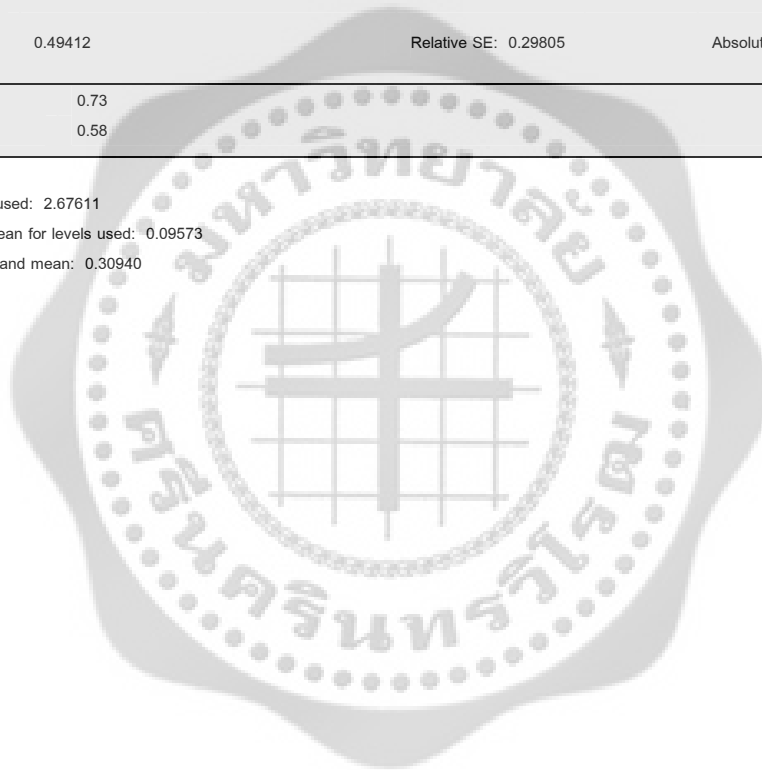
(Measurement design P/IR)

Source of variance	Differ-entiation variance	Source of variance	Relative error variance	% relative	Absolute error variance	% absolute
P	0.24416		
	I		0.00478	2.7
	R		0.07829	44.0
	PI	(0.00000)	0.0	(0.00000)	0.0
	PR	0.03267	36.8	0.03267	18.4
	IR		0.00600	3.4
	PIR	0.05617	63.2	0.05617	31.6
Sum of variances	0.24416		0.08883	100%	0.17790	100%
Standard deviation	0.49412		Relative SE: 0.29805		Absolute SE: 0.42179	
Coef_G relative	0.73					
Coef_G absolute	0.58					

Grand mean for levels used: 2.67611

Variance error of the mean for levels used: 0.09573

Standard error of the grand mean: 0.30940



คู่มือการใช้โปรแกรม FACETS 3.70.1

ผู้วิจัยใช้โปรแกรมเพื่อนำเสนอ พารามิเตอร์ของนักเรียนเพื่อแสดงให้เห็นค่าเฉลี่ยของความสามารถของนักเรียน จำนวน 50 คน พารามิเตอร์ของข้อสอบเรียงความ เพื่อแสดงให้เห็นค่าความยากในการเปลี่ยนขั้นของการใช้เกณฑ์ทั้ง 3 รูปแบบว่าเป็นไปตามลำดับและมีความเที่ยงตรงในการเปลี่ยนขั้นหรือไม่ และพารามิเตอร์ผู้ตรวจให้คะแนน เพื่อแสดงให้เห็นค่าเฉลี่ยของความเข้มงวดโดยรวมของผู้ให้คะแนน ทั้ง 130 คน และเพื่อศึกษาคุณลักษณะของผู้ให้คะแนน คือ ความเข้มงวด/ใจดีของผู้ให้คะแนน ความลำเอียงด้านเพศของผู้ให้คะแนน และการทำหน้าที่ต่างกันของผู้ให้คะแนนเมื่อเวลาผ่านไป

ขั้นที่ 1 เตรียมข้อมูลนำเข้าให้อยู่ในรูป "," (comma: *.csv) ใน Excel โดยเรียงลำดับตาม Facet ที่ต้องการคำนวณ เช่น นักเรียน ครู ข้อสอบ แล้วจึงเป็นคะแนน

ตัวอย่างการ key ข้อมูลแบบแยกองค์ประกอบ

rater	examinee	item	1	2	3	4
1	1	1-4	2	3	2	1
1	2	1-4	2	1	2	2
1	3	1-4	2	2	2	1
1	4	1-4	1	1	1	2
1	5	1-4	2	2	3	1
1	6	1-4	2	2	1	1
1	7	1-4	1	2	3	1
1	8	1-4	1	1	2	2
1	9	1-4	2	1	2	1
1	10	1-4	2	1	2	2

ตัวอย่างการ key ข้อมูลแบบรวมองค์ประกอบ

rater	examinee	rated
1	1	2
1	2	3
1	3	2
1	4	2
1	5	3
1	6	2
1	7	3
1	8	1
1	9	3
1	10	2

ตัวอย่างการ key ข้อมูลแบบผสมผสาน

rater	examinee	item	1	2	3
1	1	1-3	2	1	1
1	2	1-3	2	1	1
1	3	1-3	2	2	2
1	4	1-3	1	1	1
1	5	1-3	2	2	1
1	6	1-3	2	2	1
1	7	1-3	3	2	1
1	8	1-3	1	1	2
1	9	1-3	2	2	1
1	10	1-3	2	2	1

จากนั้น save ข้อมูลจาก *.xlsx เป็น *.csv แล้วเปิดโดยใช้ Notepad

ขั้นที่ 2 สร้างคำสั่งการคำนวณใน Notepad ดังนี้

ตัวอย่างคำสั่งให้คำนวณความเข้มงวด/ใจดี

โดยใช้เกณฑ์การให้คะแนนแบบแยกองค์ประกอบ

Title = Analytic Scoring Rubrics for SL -->ตั้งชื่อ

facet = 3 -->จำนวน Facets ที่ใช้คำนวณ

Unexpected = 2 -->ให้รายงานเมื่อพบค่า Standardized Residual มีค่ามากกว่า 2

Pt-biserial = measure -->ให้คำนวณค่า Point Biserial

Inter-rater = 2 --> บอกว่า ให้คำนวณ Inter-rater ซึ่งอยู่บน Facet ที่ 2

Model = ?,?,?,Essay --> บอกว่าให้รายงานค่าทุกโมเดล

Rating scale = Essay,R5 -->ตั้งชื่อสเกลกำหนดค่าเพื่อให้เข้าใจง่ายขึ้นในตารางรายงานผล PCM

1 = lowest

3 = middle

5 = highest

* -->ทุกครั้งที่จะเริ่มคำสั่งใหม่ให้ใส่ * เพื่อบอกว่าจบคำสั่ง

Labels= --> บอกว่าจะตั้งชื่อให้ตัวแปร

1,Student --> Facet ที่ 1 คือ นักเรียน

1=Korkiat

2=Kunakorn

3=Jatapat

4=Jittakorn

5=Chalerm

...

48=Aranya

49=Angsuma

50=Artittaya

* --> จบ

2,Teacher --> Facet ที่ 2 คือ ครู

1=Kathawut

2=Chrin

3=Trin

4=Bandasak

5=Boonkerd

....

128=Saranya

129=Sasirat

130=Sureporn

* --> จบ

3,Traits --> Facet ที่ 3 คือ ข้อสอบ (แบบแยกองค์ประกอบมี 4 ประเด็น)

1=Topic

2=Pattern

3=Presentation

4=Word

* --> จบ

Data= --> ข้อมูล

1,1,1-4,2,3,2,1

2,1,1-4,2,1,2,2

3,1,1-4,2,2,2,1

4,1,1-4,1,1,1,2

5,1,1-4,2,2,3,1

6,1,1-4,2,2,1,1

...

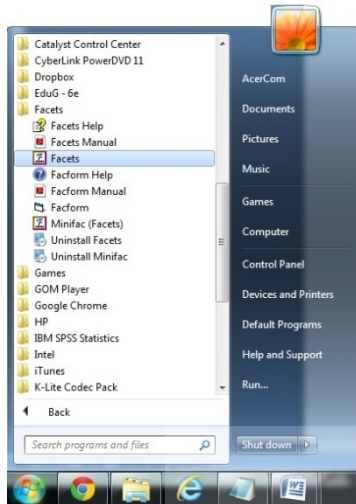
48,130,1-4,4,4,4,4

49,130,1-4,4,4,3,2

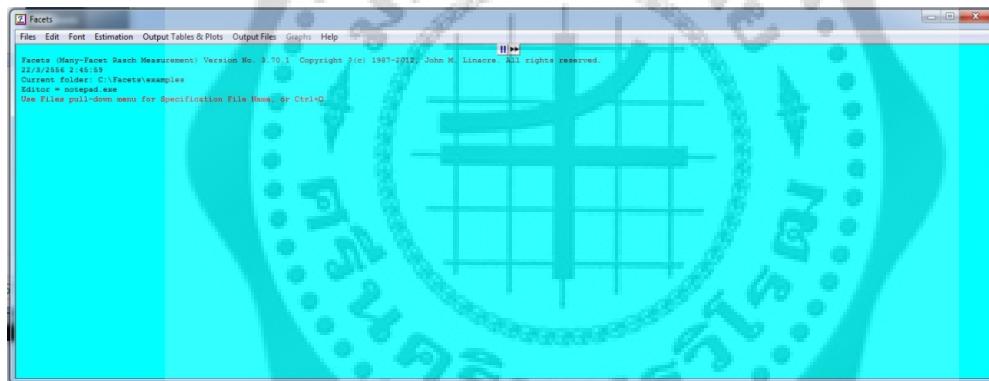
50,130,1-4,2,2,3,2

หลังข้อมูลชุดสุดท้ายไม่ต้องใส่ * ปล่อยให้ว่างไว้เฉยๆ จากนั้นบันทึกข้อมูลเป็นภาษาอังกฤษเท่านั้น เช่น analytic_SL.txt (*.txt อัตโนมัติ)

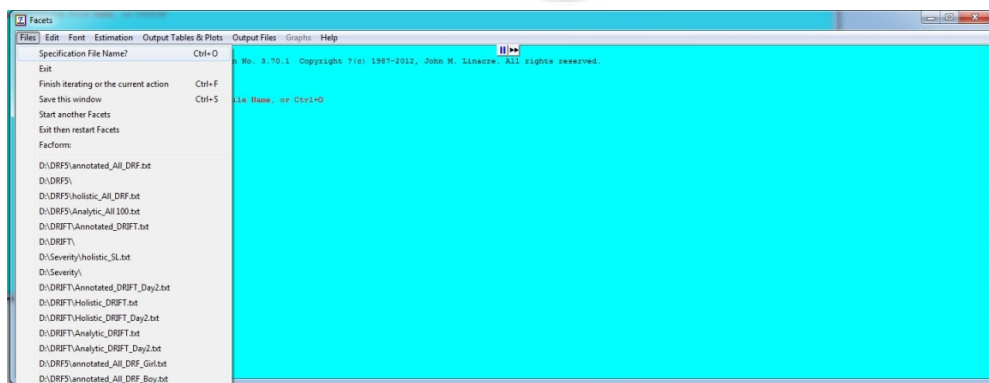
ขั้นที่ 2 เปิดโปรแกรม FACETS 3.70.1 จาก Start



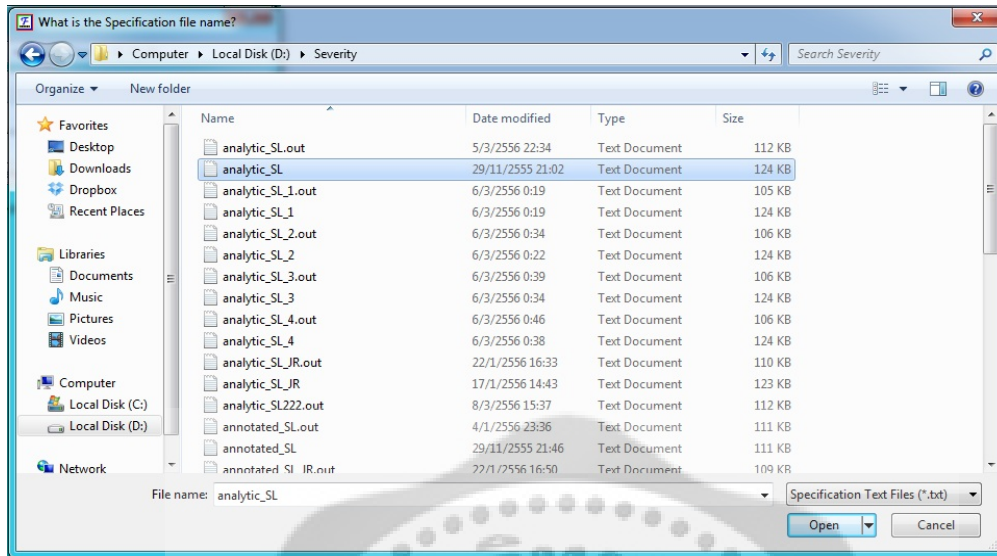
จะพบหน้าจอดังนี้



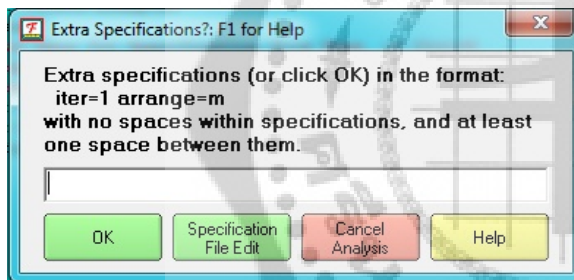
ขั้นที่ 3 เปิดคำสั่งที่เขียนไว้ เลือก File>Specification File Name?



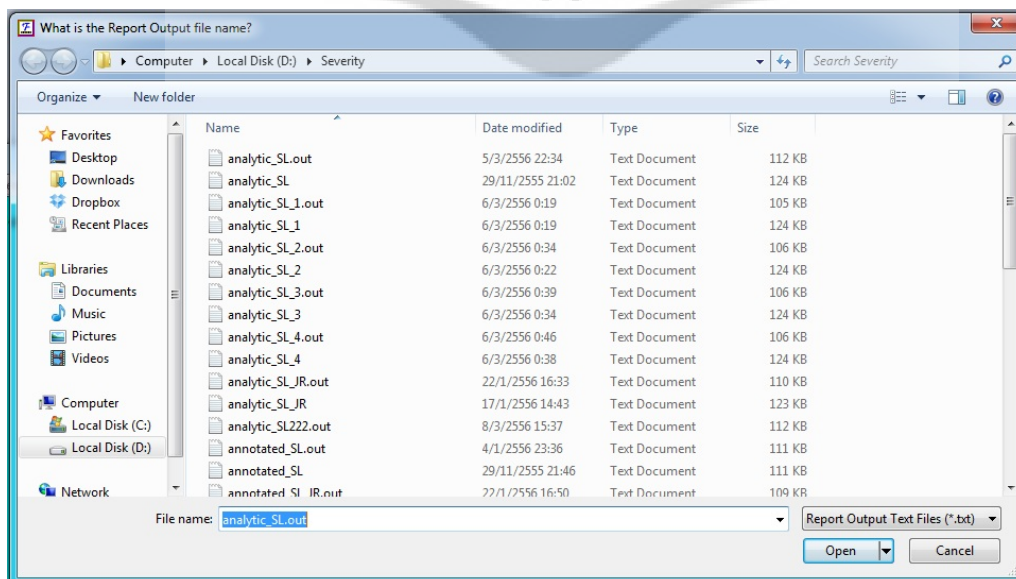
เลือกคำสั่ง analytic_SL ที่บันทึกไว้



จะพบหน้าต่างที่ถามว่ามีอะไรจะเพิ่มหรือไม่ ถ้าไม่ ให้เลือก OK



ขั้นที่ 4 โปรแกรมจะตั้งชื่อ Output ให้ แต่ถ้าไม่พอใจสามารถเปลี่ยนได้ จากนั้นเลือก Open



ขั้นที่ 5 โปรแกรมจะรันข้อมูลเมื่อเสร็จแล้วจะบอกเวลาที่ใช้คำนวณ

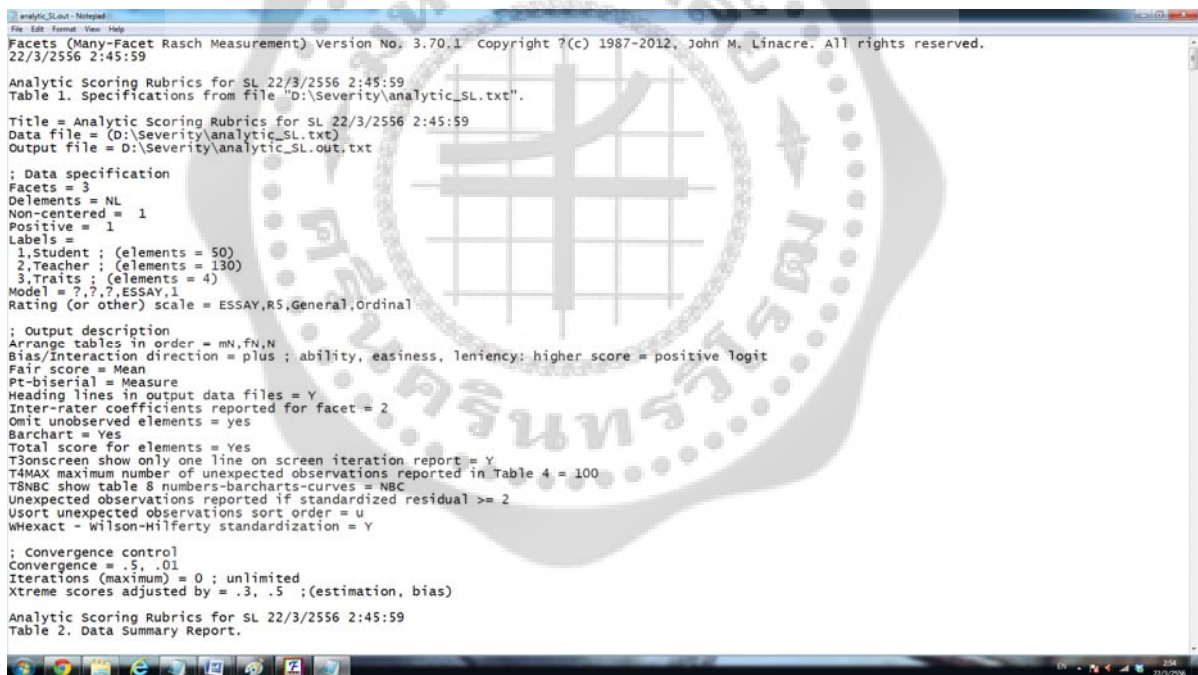


```

analytic_SL.txt
File Edit Font Estimation Output Tables & Plots Output Files Graphs Help
-----
INLG 51 -0.0295 0 1.4988 -0.014 -0.001
INLG 52 -1.7939 0 1.4361 -0.014 -0.001
INLG 53 -1.5809 0 1.3278 -0.014 -0.001
INLG 54 -1.4101 0 1.3769 -0.018 -0.001
INLG 55 -1.2500 0 1.0480 -0.002 -0.001
INLG 56 -1.1089 0 0.828 -0.002 -0.001
INLG 57 -0.9019 0 0.609 -0.002 -0.001
INLG 58 -0.8711 0 0.284 -0.002 -0.001
INLG 59 -0.7726 0 0.445 -0.002 -0.001
INLG 60 -0.6851 0 0.713 -0.011 -0.001
INLG 61 -0.6089 0 0.668 -0.001 -0.001
INLG 62 -0.5381 0 0.482 -0.001 -0.001
INLG 63 -0.4776 0 0.984 -0.001 -0.001

Calculating fit statistics
Computing inter-rater agreement
Table 4 Unexpected Responses - appears after Table 8
Table 5 Measurability Data Summary
Computing fit summary statistics
Table 0 0 All Facets' Vertical "Pulsars"
Table 1.1 Student Measurement Report (arranged by M)
Table 1.2 Student Measurement Report (arranged by W)
Table 1.3 Student Measurement Report (arranged by W)
Table 2.1 Teacher Measurement Report (arranged by M)
Table 2.2 Teacher Measurement Report (arranged by W)
Table 3.1 Traits Measurement Report (arranged by M)
Table 3.2 Traits Measurement Report (arranged by W)
Table 3.3 Traits Measurement Report (arranged by W)
Table 4 Category Statistics
Table 4.1 Unexpected Responses (100 residuals sorted by abs
Sorted in view of 0.10
Subset connection O.K
Output to: D:\Severity\analytic_SL.out.txt
  
```

จากนั้นโปรแกรมจะเปิด Output ออกมาให้สามารถใช้งานได้ทันที



```

analytic_SL.out - Notepad
File Edit Format View Help
Facets (Many-Facet Rasch Measurement) Version No. 3.70.1 Copyright ?(c) 1987-2012, John M. Linacre. All rights reserved.
22/3/2556 2:45:59

Analytic Scoring Rubrics for SL 22/3/2556 2:45:59
Table 1. Specifications from file "D:\Severity\analytic_SL.txt".

Title = Analytic Scoring Rubrics for SL 22/3/2556 2:45:59
Data file = (D:\Severity\analytic_SL.txt)
Output file = D:\Severity\analytic_SL.out.txt

; Data specification
Facets = 3
Delements = NL
Non-centered = 1
Positive = 1
Labels =
1,Student ; (elements = 50)
2,Teacher ; (elements = 130)
3,Traits ; (elements = 4)
Model = ?,.?.?.ESSAY,1
Rating (or other) scale = ESSAY,R5,General,Ordinal

; Output description
Arrange tables in order = mN,fN,N
Bias/Interaction direction = plus ; ability, easiness, leniency: higher score = positive logit
Fair score = Mean
PT-biserial = Measure
Heading lines in output data files = Y
Inter-rater coefficients reported for facet = 2
Omit unobserved elements = yes
Barchart = Yes
Total score for elements = Yes
T5onscreen show only one line on screen iteration report = Y
T4MAX maximum number of unexpected observations reported in Table 4 = 100
T8NBC show table 8 numbers-barcharts-curves = NBC
Unexpected observations reported if standardized residual >= 2
Usort unexpected observations sort order = u
WHexact - Wilson-Hilferty standardization = Y

; Convergence control
Convergence = .5, .01
Iterations (maximum) = 0 ; unlimited
Xtreme scores adjusted by = .3, .5 ;(estimation, bias)

Analytic Scoring Rubrics for SL 22/3/2556 2:45:59
Table 2. Data Summary Report.
  
```



