



การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนแบบทดสอบความเรียง
วิชาภาษาไทย ชั้นประถมศึกษาปีที่ 3 ที่มีจำนวนผู้ตรวจ วิธีการตรวจและรูปแบบการตรวจ
ต่างกัน : กรณีศึกษาสำนักงานเขตพื้นที่การศึกษาประถมศึกษากำแพงเพชร เขต 2
A Generalizability Coefficient Comparison of Thai Essay Test Scores for
Phathomsuksa 3 Students Using Different Numbers of Raters, Scoring
Methods and Scoring Patterns : A Case Study of Kamphaeng Phet Primary
Education Service Area Office 2

ณธีรรัตน์ เมฆสิงห์*

Nateerat Makesing

ปณณวิชญ์ ไบกุลลาบ**

Punnawit Baikularb

Received : September 24, 2021

Revised : April 21, 2022

Accepted : June 24, 2022

บทคัดย่อ

การวิจัยในครั้งนี้มีวัตถุประสงค์เพื่อศึกษาขนาดขององค์ประกอบความแปรปรวนเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนน และเปรียบเทียบค่าความแปรปรวนของความคลาดเคลื่อนของคะแนนแบบทดสอบความเรียงวิชาภาษาไทย ชั้นประถมศึกษาปีที่ 3 ที่มีจำนวนผู้ตรวจ วิธีการตรวจและรูปแบบการตรวจต่างกัน กลุ่มตัวอย่างที่ใช้ในการวิจัยในครั้งนี้ คือ นักเรียนชั้นประถมศึกษาปีที่ 3 ปีการศึกษา 2563 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษากำแพงเพชร เขต 2 จำนวน 182 โรงเรียน โดยการสุ่มโรงเรียนแบ่งตามขนาดของโรงเรียนทั้งหมด 3 ขนาด ซึ่งได้มาด้วยวิธีการสุ่มอย่างง่าย (Simple Random Sampling) ได้โรงเรียนทั้งหมด 7 โรงเรียน โดยกลุ่มตัวอย่างที่ใช้ในการวิจัยในครั้งนี้ คือนักเรียนจำนวน 60 คน เครื่องมือที่ใช้เป็นแบบทดสอบการเขียนความเรียงวิชาภาษาไทย จำนวน 6 ข้อ วิเคราะห์ข้อมูลโดยการหาค่าสัมประสิทธิ์การสรุปอ้างอิง โดยใช้โปรแกรม GENOVA ผลการวิจัยสรุปได้ว่า 1. ค่าความแปรปรวนในทุกเงื่อนไขจำนวนผู้ตรวจ รูปแบบการตรวจ และวิธีการตรวจให้คะแนน ความแปรปรวนที่มีค่ามากที่สุดคือผู้สอบ รองลงมาคือ

*นักศึกษาลัทธิศาสตรมหาบัณฑิต สาขาการวิจัยและประเมินทางการศึกษา คณะครุศาสตร์

มหาวิทยาลัยราชภัฏพิบูลสงคราม

Master of Education student, Department of Educational Research and Assessment, Faculty of Education, Pibulsongkram Rajabhat University.

**อาจารย์ประจำสาขาการวิจัยและประเมินทางการศึกษา คณะครุศาสตร์ มหาวิทยาลัยราชภัฏพิบูลสงคราม

Lecturer in Educational Research and Evaluation, Faculty of Education, Pibulsongkram Rajabhat University

ข้อสอบ และผู้ตรวจ 2. ค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนที่มีจำนวนผู้ตรวจ วิธีการตรวจและรูปแบบการตรวจต่างกัน พบว่า 2.1) ค่าสัมประสิทธิ์การสรุปอ้างอิงที่มีจำนวนผู้ตรวจต่างกัน เมื่อรูปแบบการตรวจและวิธีการตรวจให้คะแนนเหมือนกัน มีค่าสัมประสิทธิ์การสรุปอ้างอิงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นรูปแบบตรวจข้อสอบบางข้อของผู้สอบทุกคน 2.2) ค่าสัมประสิทธิ์การสรุปอ้างอิงที่มีวิธีการตรวจให้คะแนนต่างกัน ที่มีจำนวนผู้ตรวจเท่ากัน รูปแบบการตรวจเหมือนกัน มีค่าสัมประสิทธิ์การสรุปอ้างอิงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้น รูปแบบการตรวจข้อสอบบางข้อของผู้สอบทุกคน 2.3) ค่าสัมประสิทธิ์การสรุปอ้างอิงที่มีรูปแบบการตรวจต่างกัน เมื่อมีจำนวนผู้ตรวจเท่ากันโดยวิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ พบว่า มีค่าสัมประสิทธิ์ การสรุปอ้างอิงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้น จำนวนผู้ตรวจ 1 คน ส่วนวิธีการตรวจให้คะแนนแบบประเมินรวม มีค่าสัมประสิทธิ์การสรุปอ้างอิงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้น จำนวนผู้ตรวจ 3 คน 3. ค่าความแปรปรวนของความคลาดเคลื่อนของคะแนนข้อสอบที่มีจำนวนผู้ตรวจ และวิธีการตรวจให้คะแนนต่างกัน เมื่อเพิ่มจำนวนผู้ตรวจ ค่าความแปรปรวนของความคลาดเคลื่อนลดลง ทั้ง 2 วิธีการตรวจให้คะแนน ยกเว้น รูปแบบตรวจข้อสอบบางข้อของผู้สอบทุกคน

คำสำคัญ : ทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด / แบบทดสอบความเรียงวิชาภาษาไทย / ขนาดขององค์ประกอบความแปรปรวน / ชั้นประถมศึกษาปีที่ 3 สำนักงานเขตพื้นที่การศึกษาประถมศึกษากำแพงเพชร เขต 2

ABSTRACT

The purposes of this research were 1) to study the size of variance component compared with scores' Generalizability Coefficient and 2) to compare deviation's variance of Thai language essay test scores of grade 3 students with a different number of examiners, checking methods, and checking patterns. The research used the Multi-stage Sampling method to collect samples. The sample was 60 students in grade 3 of the academic year 2020 from schools under the Kamphaeng Phet Primary Educational Service Area Office 2. The tool of the research was a 6-item Thai essay writing test. Moreover, this research's data was analyzed by Generalizability Coefficient with GENOVA. The results of the research concluded that 1. The variance in all conditions, test takers was the most valuable, followed by test and examiners respectively. 2. Scores' Generalizability Coefficients with a different number of examiners, checking methods, and checking patterns were found that 2.1) Generalizability Coefficient with different numbers of examiners but same patterns and methods were found that the Generalizability Coefficient differed statistically significant at the 0.01 level except for checking some items of all test takers pattern. 2.2) Generalizability Coefficient with different checking

methods but the same number of examiners and checking pattern differed statistically significant at the 0.01 level except for checking some items of all test takers pattern.

2.3) Generalizability Coefficient with different checking patterns but the same number of examiners by discrete elements checking method was found that the Generalizability Coefficient differed statistically significant at the 0.01 level except for one examiner. In part of Generalizability Coefficient with the overall assessment method differed statistically significant at the 0.01 level except for 3 numbers of examiners. 3) The variance of test score's deviation with a different number of examiners, checking methods, and patterns resulting that the deviation's variance reduced when the number of examiners increased except for checking some items of all test takers pattern.

Keywords : Generalizability Theory / Thai Language Essay Test /

The Size of the Variance Components / Grade 3, Kamphaeng Phet Primary
Education Service Area Office 2

ความเป็นมาและความสำคัญของปัญหา

หลักสูตรการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ในกลุ่มสาระการเรียนรู้ภาษาไทยได้กำหนดมาตรฐานการเรียนรู้ช่วงชั้น ป.1-3 ไว้ว่าผู้เรียนต้องมีความรู้ความสามารถในการเขียนคำได้ถูกความหมายและสะกดการันต์ให้ถูกต้อง ใช้ความรู้และประสบการณ์เขียนประโยค ข้อความ และเรื่องราวแสดงความคิด ความรู้สึก ความต้องการ และจินตนาการ รวมทั้งใช้กระบวนการเขียนพัฒนางานเขียน นอกจากนี้ผู้เรียนต้องมีการยาททางการเขียนและนิสัยรักการเขียนและใช้ทักษะการเขียน จดบันทึกความรู้ประสบการณ์และเรื่องราวในชีวิตประจำวันได้ (Ministry of Education, 2008, p.10) ดังนั้นทักษะทางการเขียนจึงนับเป็นทักษะหนึ่งที่มีความสำคัญและจำเป็นอย่างยิ่งในชีวิตประจำวัน

สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน มีนโยบายในการประเมินคุณภาพผู้เรียน ที่สามารถสะท้อนถึงคุณลักษณะของผู้เรียนในศตวรรษที่ 21 ซึ่งการประเมินคุณภาพผู้เรียนในระดับชั้นประถมศึกษาปีที่ 3 มีการประเมินความสามารถด้านภาษาไทย และความสามารถด้านคณิตศาสตร์กับผู้เรียนจากทุกสังกัดทั่วประเทศ (Office of The Basic Education Commission, 2019) ในการสอบวัดความสามารถด้านภาษาไทย มีการตรวจข้อสอบความเรียง ซึ่งศูนย์สอบดำเนินการแต่งตั้งคณะกรรมการตรวจข้อสอบได้กำหนดแนวทางการตรวจโดยให้มีคณะกรรมการตรวจข้อสอบอย่างน้อย 2 คน ต่อหนึ่งฉบับ ทั้งนี้จำนวนผู้ตรวจข้อสอบให้อยู่ในดุลยพินิจของศูนย์สอบ (Office of The Basic Education Commission, 2019, p.31) ดังที่กล่าวข้างต้น ทำให้ทราบว่ายังไม่มีจำนวนผู้ตรวจที่สามารถเชื่อถือผลของคะแนนได้ จำนวนผู้ตรวจต่างกันจึงมีผลต่อความน่าเชื่อถือของคะแนน แบบทดสอบความเรียงเป็นแบบทดสอบที่มีประสิทธิภาพในการวัดพฤติกรรมการเรียนรู้ขั้นสูง แต่ผู้ตรวจข้อสอบประเภทนี้มักจะต้องเผชิญกับปัญหาหรือจุดอ่อนที่สำคัญ คือ จำนวนนักเรียนรู้สอบมีจำนวนมาก

ทำให้ผู้ตรวจไม่สามารถอ่านคำตอบของแบบทดสอบความเรียงได้อย่างรอบคอบ ทำให้มีจุดอ่อนในเรื่องของความยุติธรรมในการให้คะแนน หรือเกิดความลำเอียงในการตรวจให้คะแนนกับผู้สอบแต่ละคน ดังนั้นไม่ควรให้ผู้ตรวจเพียงคนเดียวเพราะจะทำให้ขาดความเที่ยงตรงของคะแนน (Uebersax, 2003) นอกจากนี้ยังพบว่าเมื่อเพิ่มจำนวนผู้ตรวจจะทำให้ค่าความน่าเชื่อถือของคะแนนเพิ่มสูงขึ้น (Swartz, et al. 1999; Sudweeks, Reeve & Bradshaw, 2005; Brennan Gao, & Colton, 1995) แต่จำนวนผู้ตรวจที่เหมาะสมนั้นยังไม่สามารถสรุปได้แน่นอน เนื่องจากผลของการศึกษายังพบแตกต่างกัน และเกณฑ์การตรวจให้คะแนนมีลักษณะเป็นเกณฑ์การให้คะแนนแบบประเมินรวม ซึ่งนักวัดผลการศึกษาและนักวิจัยเสนอว่าโดยทั่วไปแล้วการตรวจด้วยเกณฑ์การให้คะแนนแบบแยกองค์ประกอบ จะให้ค่าความน่าเชื่อถือของคะแนนสูงกว่า และการตรวจให้คะแนนของผู้สอบที่มีผู้สอบจำนวนมากนั้น การตรวจที่ให้คะแนนที่ยุติธรรมที่สุดนั้น คือให้ผู้ตรวจทุกคนตรวจแบบทดสอบทุกข้อของผู้สอบทุกคน แต่ต้องใช้เวลาในการตรวจค่อนข้างมากและมีค่าใช้จ่ายสูง หากสามารถวางแผนตรวจหรือหาวิธีการตรวจเพื่อลดภาระลดเวลาในการตรวจ โดยมีความเป็นไปได้ในทางปฏิบัติ คือสามารถนำไปใช้กับโรงเรียนที่มีครูผู้ตรวจให้คะแนนที่มีความเชี่ยวชาญในจำนวนจำกัด และมีภาระงานค่อนข้างมาก และผลของคะแนนต้องมีความน่าเชื่อถือและให้ค่าสถิติที่ยอมรับได้ จึงจะเป็นทางหนึ่งในการแก้ปัญหาที่เกิดขึ้นได้ (Linacre & Wright, 2002, pp.490-493)

ทฤษฎีการสรุปอ้างอิง (Generalizability Theory) จึงเป็นสถิติที่สามารถวิเคราะห์ความเชื่อถือได้ของผลการวัดพฤติกรรมในสถานการณ์การวัดผล หรือเงื่อนไขการทดสอบในลักษณะต่าง ๆ ได้ ดังนั้นทฤษฎีการสรุปอ้างอิงความน่าเชื่อถือของผลการวัด จึงมีความเหมาะสมในการวิเคราะห์ความแปรปรวนของคะแนนความคลาดเคลื่อนจากแหล่งต่าง ๆ ทำให้สามารถตรวจสอบถึงแหล่งความแปรปรวนที่ไม่พึงประสงค์ และสามารถควบคุมแหล่งความคลาดเคลื่อนได้ตรงประเด็น (Kanchanawasee, 2012)

ด้วยเหตุผลดังกล่าวข้างต้น ผู้วิจัยจึงสนใจศึกษาการเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงความน่าเชื่อถือของคะแนนข้อสอบความเรียงวิชาภาษาไทย ชั้นประถมศึกษาปีที่ 3 ที่มีจำนวนผู้ตรวจ วิธีการตรวจ และรูปแบบการตรวจต่างกัน เพื่อเป็นแนวทางในการเลือกใช้งานผู้ตรวจ วิธีการตรวจให้คะแนนที่ให้ค่าความน่าเชื่อถือของผลคะแนน และรูปแบบการตรวจที่เหมาะสม ประหยัดเวลา ค่าใช้จ่าย และสามารถเป็นสารสนเทศประกอบการตัดสินใจในการตรวจให้คะแนนแบบทดสอบความเรียงวิชาภาษาไทยต่อไป

วัตถุประสงค์ของการวิจัย

1. เพื่อศึกษาขนาดขององค์ประกอบความแปรปรวนของคะแนนแบบทดสอบความเรียงวิชาภาษาไทย ชั้นประถมศึกษาปีที่ 3 ที่มีจำนวนผู้ตรวจ วิธีการตรวจและรูปแบบการตรวจต่างกัน
2. เพื่อเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนแบบทดสอบความเรียงวิชาภาษาไทย ชั้นประถมศึกษาปีที่ 3 ที่มีจำนวนผู้ตรวจ วิธีการตรวจและรูปแบบการตรวจต่างกัน
3. เพื่อเปรียบเทียบค่าความแปรปรวนของความคลาดเคลื่อนของคะแนนแบบทดสอบความเรียงวิชาภาษาไทย ชั้นประถมศึกษาปีที่ 3 ที่มีจำนวนผู้ตรวจ และวิธีการตรวจต่างกัน

สมมติฐานของการวิจัย

1. ค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนข้อสอบความเรียงวิชาภาษาไทย ชั้นประถมศึกษาปีที่ 3 ที่มีจำนวนผู้ตรวจต่างกัน รูปแบบการตรวจ และวิธีการตรวจให้คะแนนเหมือนกัน มีค่าแตกต่างกัน
2. ค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนข้อสอบความเรียงวิชาภาษาไทย ชั้นประถมศึกษาปีที่ 3 ที่มีวิธีการตรวจให้คะแนนต่างกัน จำนวนผู้ตรวจเท่ากัน รูปแบบการตรวจเหมือนกัน มีค่าแตกต่างกัน
3. ค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนข้อสอบความเรียงวิชาภาษาไทย ชั้นประถมศึกษาปีที่ 3 ที่มีรูปแบบการตรวจต่างกัน จำนวนผู้ตรวจเท่ากัน และมีวิธีการตรวจให้คะแนนเหมือนกัน มีค่าแตกต่างกัน
4. ค่าความแปรปรวนของความคลาดเคลื่อนที่ได้จากคะแนนข้อสอบความเรียงวิชาภาษาไทย ชั้นประถมศึกษาปีที่ 3 ที่มีผู้ตรวจเพิ่มขึ้น มีค่าลดลง ในทุกวิธีการตรวจให้คะแนน

วิธีดำเนินการวิจัย

ประชากรและกลุ่มตัวอย่างที่ใช้ในการวิจัย

ประชากรที่ใช้ในการวิจัย คือ นักเรียนชั้นประถมศึกษาปีที่ 3 ของโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษากำแพงเพชร เขต 2 จำนวน 182 โรงเรียน กลุ่มตัวอย่างที่ใช้ในการวิจัย คือ นักเรียนชั้นประถมศึกษาปีที่ 3 ของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษากำแพงเพชร เขต 2 จำนวน 60 คน โดยวิธีการสุ่มแบบหลายขั้นตอน คือ ได้โรงเรียนทั้งหมด 7 โรงเรียน จากนั้นสุ่มนักเรียนจากห้องเรียนที่ได้จากโรงเรียนต่าง ๆ วิธีการจับฉลาก และได้จำนวนนักเรียนที่เป็นกลุ่มตัวอย่างทั้งหมด 60 คน

ในการกำหนดขนาดของกลุ่มตัวอย่างนั้น ผู้วิจัยใช้แนวคิดเกี่ยวกับทฤษฎีการสรุปอ้างอิง ดังนั้นเพื่อให้การประมาณค่าสัมประสิทธิ์การสรุปอ้างอิงมีความแม่นยำ ขนาดของตัวอย่างขั้นต่ำที่ใช้ในการศึกษาควรทำให้เกิดค่าสังเกต (observation) อย่างน้อย 1,000 ค่า (Smith, 1978, p.319-346) ผู้วิจัยได้ศึกษา 2 องค์ประกอบในรูปแบบ $p \times i \times r$ โดยมีจำนวนนักเรียน 60 คน จำนวนข้อสอบ 6 ข้อ และจำนวนผู้ตรวจ 3 คน ดังนั้น $p \times i \times r = 60 \times 6 \times 3$ มีค่าเท่ากับ 1,080 ซึ่งค่าสังเกตมีค่ามากกว่าหน่วยขั้นต่ำที่สมมติเสนอ เป็นขนาดตัวอย่างที่เหมาะสมในการนำไปใช้ประมาณค่าการสรุปอ้างอิง

การสร้างเครื่องมือ

เครื่องมือที่ใช้ในการวิจัย คือ แบบทดสอบการเขียนความเรียง สารการเรียนรู้ ภาษาไทย ชั้นประถมศึกษาปีที่ 3 ในสาระการเรียนรู้ที่ 2 การเขียนเรื่องจากภาพตามจินตนาการ ที่ผู้วิจัยได้จัดทำขึ้นจำนวน 6 ฉบับ ๆ ละ 1 ข้อ และสร้างเกณฑ์การให้คะแนนแบบรูบริคแบบแยกองค์ประกอบ ส่วนเกณฑ์การให้คะแนนแบบประเมินรวม จะใช้เกณฑ์การให้คะแนนการประเมินคุณภาพผู้เรียน (NT) ที่สร้างขึ้นโดยสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน

การทดสอบคุณภาพของเครื่องมือ

นำแบบทดสอบการเขียนความเรียงและเกณฑ์การให้คะแนนตรวจสอบความเที่ยงตรงเชิงเนื้อหา พร้อมทั้งปรับปรุงแก้ไข โดยนำแบบทดสอบและเกณฑ์การให้คะแนนนำเสนอผู้เชี่ยวชาญจำนวน 3 ท่าน โดยเป็น

ผู้เชี่ยวชาญด้านการวัดและประเมินผลทางการศึกษา และผู้เชี่ยวชาญด้านภาษาไทย เพื่อพิจารณาตรวจสอบโดยใช้ดัชนีความสอดคล้อง (IOC) ปรากฏผลการคำนวณของค่าความเที่ยงตรงเชิงเนื้อหาของแบบทดสอบความเรียงมีค่าเท่ากับ 1.00 ทุกข้อ และเกณฑ์การตรวจให้คะแนนแบบแยกองค์ประกอบ และแบบประเมินรวมมีค่าเท่ากับ 1.00 แล้วนำแบบทดสอบการเขียนความเรียงวิชาภาษาไทยไปทดลองใช้กับนักเรียนชั้นประถมศึกษาปีที่ 3 ที่ไม่ใช่กลุ่มตัวอย่างในงานวิจัยจำนวน 10 คน นำผลที่ได้มาวิเคราะห์และปรับปรุงแก้ไขเกณฑ์การตรวจให้คะแนนแบบแยกองค์ประกอบเพิ่มเติม จากนั้นจัดพิมพ์แบบทดสอบฉบับสมบูรณ์ทั้ง 6 ข้อ และเกณฑ์การตรวจให้คะแนนสำหรับผู้ตรวจ

การเก็บรวบรวมข้อมูล

ผู้วิจัยดำเนินการเก็บรวบรวมข้อมูล ดังนี้ ดำเนินการสอบโดยนำแบบทดสอบความเรียงวิชาภาษาไทย ชั้นประถมศึกษาปีที่ 3 ไปทดสอบกับนักเรียนที่เป็นกลุ่มตัวอย่างจำนวน 60 คน โดยให้กลุ่มตัวอย่างทุกคนทำแบบทดสอบความเรียงทุกข้อ แต่ละข้อใช้เวลาทำข้อละ 1 ชั่วโมง รวมจำนวนทั้งหมด 6 วัน จากนั้นนำผลงานการเขียนความเรียงมาถ่ายเอกสาร และใส่เลขที่ที่ผลงานแต่ละฉบับ จากนั้นชี้แจงวิธีการตรวจและกฎเกณฑ์การให้คะแนนที่ผู้วิจัยได้สร้างขึ้นให้ชัดเจน เพื่อสร้างความเข้าใจที่ตรงกัน เมื่อผู้ตรวจเข้าใจตรงกันแล้ว จึงให้อิสระผู้ตรวจแต่ละคนตรวจข้อสอบทุกข้อของผู้สอบทุกคน การตรวจข้อสอบในแต่ละวิธีการตรวจให้คะแนนจะใช้ระยะเวลาต่างกัน 14 วัน โดยในการตรวจข้อสอบนั้นให้ผู้ตรวจที่มีวุฒิทางการศึกษาตั้งแต่ระดับปริญญาตรีขึ้นไปทางด้านวิชาเอกภาษาไทย และมีประสบการณ์ในการสอนวิชาภาษาไทย ไม่น้อยกว่า 3 ปี จำนวน 3 คน จากนั้นดำเนินการตรวจให้คะแนนลงในแบบบันทึก แล้วนำผลการตรวจให้คะแนนมาจัดกระทำข้อมูลตามเงื่อนไขที่ผู้วิจัยต้องการศึกษา คือตรวจให้คะแนนข้อสอบทุกข้อของผู้สอบทุกคน ($p \times i \times r$) โดยใช้วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบและแบบประเมินรวม และตรวจข้อสอบบางข้อของผู้สอบทุกคน [$p \times (i : r)$] โดยวิธีการตรวจให้คะแนนแบบแยกองค์ประกอบและแบบประเมินรวม

การวิเคราะห์ข้อมูล

นำผลที่ได้จากการตรวจให้คะแนนในแต่ละเงื่อนไขมาวิเคราะห์ข้อมูล ดังนี้

วิเคราะห์ค่าสถิติพื้นฐานของคะแนนสอบ โดยคำนวณค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าสูงสุด ค่าต่ำสุด ความเบ้ และความโด่ง โดยใช้โปรแกรมสำเร็จรูป วิเคราะห์องค์ประกอบของความแปรปรวน โดยใช้ทฤษฎีการสรุปอ้างอิงด้วยโปรแกรม GENOVA และประมาณค่าและเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงทดสอบความแตกต่างของค่าสัมประสิทธิ์การสรุปอ้างอิง โดยใช้สูตร UX_1 ของ Woodruff & Feldt (1986, pp.393-413)

สรุปผลการวิจัย

ผู้วิจัยได้ดำเนินการวิจัยตามวัตถุประสงค์ของการวิจัย ได้ผลสรุปตามลำดับ ดังนี้

1. ขนาดขององค์ประกอบความแปรปรวนของคะแนนแบบทดสอบความเรียงวิชาภาษาไทย ซึ่งพิจารณาแหล่งของความแปรปรวนหลัก คือ ผู้สอบ ข้อสอบ และผู้ตรวจ พบว่า รูปแบบตรวจข้อสอบทุกข้อของผู้สอบทุกคน และตรวจข้อสอบบางข้อของผู้สอบทุกคน โดยวิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ และ

วิธีการตรวจให้คะแนนแบบประเมินรวม ค่าความแปรปรวนมากที่สุด คือ ความแปรปรวนของผู้สอบ รองลงมา คือ ความแปรปรวนของข้อสอบ และค่าความแปรปรวนของผู้ตรวจ ตามลำดับ ปรากฏดังตารางที่ 1

ตารางที่ 1 ขนาดขององค์ประกอบความแปรปรวนของคะแนนแบบทดสอบความเรียงวิชาภาษาไทย
ชั้นประถมศึกษาปีที่ 3 ที่มีจำนวนผู้ตรวจ วิธีการตรวจและรูปแบบการตรวจต่างกัน

วิธีการตรวจ ให้คะแนน	รูปแบบ การตรวจ	แหล่งความ แปรปรวน	จำนวนผู้ตรวจ		
			1 คน	2 คน	3 คน
			ความแปรปรวนตามจำนวนผู้ตรวจ		
แบบแยก องค์ประกอบ	ตรวจข้อสอบทุกข้อ ของผู้สอบทุกคน	P	32.7443	33.7052	31.6240
		I	8.7190	0.0000	2.1034
		R	-	0.0000	1.7335
	ตรวจข้อสอบบางข้อ ของผู้สอบทุกคน	P	29.0867	30.7308	27.7105
		I	1.6763	-	-
		R	-	1.2436	7.3134
แบบประเมิน รวม	ตรวจข้อสอบทุกข้อ ของผู้สอบทุกคน	P	6.7612	12.2815	16.1925
		I	6.0046	6.8357	6.9396
		R	-	2.6746	1.9901
แบบประเมิน รวม	ตรวจข้อสอบบางข้อ ของผู้สอบทุกคน	P	21.9207	21.8312	22.5549
		I	7.1283	-	-
		R	-	3.3799	0.0000

2. ค่าสัมประสิทธิ์การสุร่อ้างอิงของคะแนนข้อสอบความเรียงวิชาภาษาไทย ที่มีจำนวนผู้ตรวจ วิธีการตรวจให้คะแนน และรูปแบบการตรวจต่างกัน พบว่า

2.1 ค่าสัมประสิทธิ์การสุร่อ้างอิงที่มีจำนวนผู้ตรวจต่างกัน เมื่อรูปแบบการตรวจเหมือนกัน คือ ตรวจข้อสอบทุกข้อของผู้สอบทุกคน โดยวิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ และแบบประเมินรวม พบว่า มีค่าสัมประสิทธิ์การสุร่อ้างอิงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 และมีค่าความแตกต่างรายคู่ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ทุกคู่ ยกเว้นรูปแบบการตรวจข้อสอบบางข้อของผู้สอบทุกคน ปรากฏดังตารางที่ 2

ตารางที่ 2 การทดสอบความแตกต่างรายคู่ของค่าสัมประสิทธิ์การสรุปอ้างอิงที่มีจำนวนผู้ตรวจต่างกัน เมื่อรูปแบบการตรวจเหมือนกัน โดยวิธีการตรวจให้คะแนนเหมือนกัน

วิธีการตรวจให้คะแนน	รูปแบบการตรวจ	จำนวนผู้ตรวจ	ค่าสัมประสิทธิ์การสรุปอ้างอิง	UX ₁
แบบแยกองค์ประกอบ	ตรวจข้อสอบทุกข้อของผู้สอบทุกคน	1	0.7705	13.0435**
		2	0.8309	
		1	0.7705	44.577**
		3	0.8680	
		2	0.8309	35.1874**
		3	0.8680	
แบบประเมินรวม	ตรวจข้อสอบทุกข้อของผู้สอบทุกคน	1	0.3174	8.3574**
		2	0.5512	
		1	0.3174	24.5326**
		3	0.6600	
		2	0.5512	75.5132**
		3	0.6600	

**p < 0.01

2.2 ค่าสัมประสิทธิ์การสรุปอ้างอิงที่มีวิธีการตรวจให้คะแนนต่างกัน ที่มีจำนวนผู้ตรวจเท่ากัน เมื่อใช้รูปแบบการตรวจเหมือนกัน พบว่า เมื่อใช้รูปแบบตรวจข้อสอบทุกข้อของผู้สอบทุกคน มีค่าสัมประสิทธิ์การสรุปอ้างอิงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ทุกจำนวนผู้ตรวจ ปรากฏดังตารางที่ 3 และเมื่อใช้รูปแบบตรวจข้อสอบบางข้อของผู้สอบทุกคน มีค่าไม่แตกต่างกัน

ตารางที่ 3 การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของวิธีการตรวจให้คะแนนต่างกัน ที่มีจำนวนผู้ตรวจเท่ากัน เมื่อใช้รูปแบบการตรวจเหมือนกัน

รูปแบบการตรวจให้คะแนน	จำนวนผู้ตรวจ	วิธีการตรวจ	ค่าสัมประสิทธิ์การสรุปอ้างอิง	UX ₁
ตรวจข้อสอบทุกข้อของผู้สอบทุกคน	1	แบบแยกองค์ประกอบ	0.7705	13.4728**
		แบบประเมินรวม	0.3174	
	2	แบบแยกองค์ประกอบ	0.8309	51.8057**
		แบบประเมินรวม	0.5512	
	3	แบบแยกองค์ประกอบ	0.8680	69.4344**
		แบบประเมินรวม	0.6600	

**p < 0.01

2.3 ค่าสัมประสิทธิ์การสรุปอ้างอิงที่มีรูปแบบการตรวจต่างกัน เมื่อมีจำนวนผู้ตรวจเท่ากัน และมีวิธีการตรวจให้คะแนนเหมือนกัน วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ พบว่า มีค่าสัมประสิทธิ์การสรุปอ้างอิงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ทุกจำนวนผู้ตรวจ ยกเว้น จำนวนผู้ตรวจ 1 คน ส่วนวิธีการตรวจให้คะแนนแบบประเมินรวม พบว่า มีค่าสัมประสิทธิ์การสรุปอ้างอิงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ทุกจำนวนผู้ตรวจ ยกเว้น จำนวนผู้ตรวจ 3 คน ปรากฏดังตารางที่ 4

ตารางที่ 4 การทดสอบความแตกต่างรายคู่ของค่าสัมประสิทธิ์การสรุปอ้างอิงที่มีรูปแบบการตรวจต่างกัน เมื่อมีจำนวนผู้ตรวจเท่ากัน และมีวิธีการตรวจให้คะแนนเหมือนกัน

วิธีการตรวจให้คะแนน	จำนวนผู้ตรวจ	รูปแบบการตรวจให้คะแนน	ค่าสัมประสิทธิ์การสรุปอ้างอิง	UX ₁
แบบแยกองค์ประกอบ	1 คน	ตรวจข้อสอบทุกข้อของผู้สอบทุกคน	0.7705	0.7262
		ตรวจข้อสอบบางข้อของผู้สอบทุกคน	0.7160	
	2 คน	ตรวจข้อสอบทุกข้อของผู้สอบทุกคน	0.8309	44.2591**
		ตรวจข้อสอบบางข้อของผู้สอบทุกคน	0.7417	

ตารางที่ 4 (ต่อ)

วิธีการตรวจ ให้คะแนน	จำนวน ผู้ตรวจ	รูปแบบการตรวจให้คะแนน	ค่าสัมประสิทธิ์ การสุปร้องอิง	UX ₁
	3 คน	ตรวจข้อสอบทุกข้อของผู้สอบ ทุกคน	0.8680	50.9219**
		ตรวจข้อสอบบางข้อของผู้สอบ ทุกคน	0.7515	
แบบ ประเมินรวม	1 คน	ตรวจข้อสอบทุกข้อของผู้สอบ ทุกคน	0.3174	7.0332**
		ตรวจข้อสอบบางข้อของผู้สอบ ทุกคน	0.6496	
	2 คน	ตรวจข้อสอบทุกข้อของผู้สอบ ทุกคน	0.5512	9.1831**
		ตรวจข้อสอบบางข้อของผู้สอบ ทุกคน	0.6532	
	3 คน	ตรวจข้อสอบทุกข้อของผู้สอบ ทุกคน	0.6600	0.0005
		ตรวจข้อสอบบางข้อของผู้สอบ ทุกคน	0.6605	

**p < 0.01

3. ค่าความแปรปรวนของความคลาดเคลื่อนของคะแนนแบบทดสอบความเรียงวิชาภาษาไทย
ชั้นประถมศึกษาปีที่ 3 ที่มีจำนวนผู้ตรวจ วิธีการตรวจ ต่างกัน พบว่า วิธีการตรวจให้คะแนนแบบแยก
องค์ประกอบและแบบประเมินรวม ของรูปแบบการตรวจข้อสอบทุกข้อของผู้สอบทุกคน เมื่อเพิ่มจำนวนผู้ตรวจ
1 คน 2 คน และ 3 คน ค่าความแปรปรวนของความคลาดเคลื่อนลดลง ยกเว้น รูปแบบตรวจข้อสอบบางข้อของ
ผู้สอบทุกคน

ตารางที่ 5 ค่าความแปรปรวนของความคลาดเคลื่อนของคะแนนแบบทดสอบความเรียงวิชาภาษาไทย
ชั้นประถมศึกษาปีที่ 3 ที่มีจำนวนผู้ตรวจ และวิธีการตรวจต่างกัน

วิธีการตรวจ ให้คะแนน	รูปแบบการตรวจ ให้คะแนน	ความแปรปรวนของความคลาดเคลื่อน และ ค่าสัมประสิทธิ์การสุรุปร่าง	จำนวนผู้ตรวจ		
			1	2	3
แบบแยก องค์ประกอบ	ตรวจข้อสอบทุกข้อ ของผู้สอบทุกคน	ความแปรปรวนของความคลาดเคลื่อน	0.1579	0.1119	0.0663
		ค่าสัมประสิทธิ์การสุรุปร่าง	0.7705	0.8309	0.8680
	ตรวจข้อสอบบางข้อ ของผู้สอบทุกคน	ความแปรปรวนของความคลาดเคลื่อน	0.0998	0.1669	0.1422
		ค่าสัมประสิทธิ์การสุรุปร่าง	0.7160	0.7417	0.7515
แบบ ประเมินรวม	ตรวจข้อสอบทุกข้อ ของผู้สอบทุกคน	ความแปรปรวนของความคลาดเคลื่อน	0.4574	0.3518	0.2918
		ค่าสัมประสิทธิ์การสุรุปร่าง	0.3174	0.5512	0.6600
	ตรวจข้อสอบบางข้อ ของผู้สอบทุกคน	ความแปรปรวนของความคลาดเคลื่อน	0.4061	0.4324	0.3974
		ค่าสัมประสิทธิ์การสุรุปร่าง	0.6496	0.6532	0.6605

อภิปรายผลการวิจัย

1. การศึกษาขนาดขององค์ประกอบความแปรปรวนของคะแนนแบบทดสอบความเรียงวิชาภาษาไทย
ชั้นประถมศึกษาปีที่ 3 ที่มีจำนวนผู้ตรวจ วิธีการตรวจและรูปแบบการตรวจต่างกัน พบว่า ความแปรปรวนของ
ผู้สอบเป็นแหล่งความแปรปรวนที่มีค่ามากที่สุด แสดงว่า ผู้สอบแต่ละคนมีความรู้ในเนื้อหาต่างกัน และมีทักษะ
การเขียนความเรียงภาษาไทย ที่แตกต่างกันตามประสบการณ์พื้นฐานในแต่ละข้อนั่นเอง ประกอบกับ
กลุ่มตัวอย่าง ที่ผู้วิจัยได้มาโดยการสุ่ม ทำให้เป็นตัวแทนที่ดีของประชากร ทำให้ได้กลุ่มตัวอย่างที่มีทักษะการ
เขียนความเรียงภาษาไทยหลากหลายตั้งแต่ต้นน้อยถึงมาก จึงทำให้ผู้สอบมีทักษะแตกต่างกัน (Khwanja, 2013,
p.88) ความแปรปรวนรองลงมา คือ แหล่งความแปรปรวนของข้อสอบ แสดงว่ามีความแตกต่างกันของผู้สอบใน
การตอบข้อสอบไม่มากนัก ซึ่งทำให้ค่าเฉลี่ยของคะแนนในแต่ละข้อต่างกัน แสดงถึงความยากของข้อสอบที่
ต่างกัน อาจเนื่องมาจากธรรมชาติของข้อสอบแต่ละข้อ ความแตกต่างของระดับความรู้ของผู้สอบ หรือความ
สนใจของผู้สอบในประเด็นหรือเงื่อนไขสิ่งที่โจทย์กำหนดไว้ (Sudweeks, Reeve & Bradshaw, 2005,
p.248) และแหล่งความแปรปรวนที่มีค่าน้อยมาก คือ ความแปรปรวนของผู้ตรวจ แสดงว่าผู้ตรวจแต่ละคนมีการ
ตรวจให้คะแนนที่มีความสอดคล้องกัน ความแตกต่างในการตรวจให้คะแนนมีค่อนข้างน้อย ทั้งนี้เนื่องจาก
ผู้ตรวจได้รับการประชุมชี้แจง แนวปฏิบัติในการตรวจให้คะแนนที่ชัดเจน ซึ่งช่วยให้ผู้ตรวจมีความเข้าใจแนว
ปฏิบัติในการตรวจให้คะแนนที่ตรงกัน (Intanate, 2011, p.185)

2. การเปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของคะแนนแบบทดสอบความเรียงวิชาภาษาไทย ชั้นประถมศึกษาปีที่ 3 ที่มีจำนวนผู้ตรวจ วิธีการตรวจและรูปแบบการตรวจต่างกัน พบว่า

2.1 ค่าสัมประสิทธิ์การสรุปอ้างอิงที่มีจำนวนผู้ตรวจต่างกัน เมื่อใช้รูปแบบตรวจข้อสอบทุกข้อของผู้สอบทุกคน โดยวิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ และแบบประเมินรวม พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงที่มีจำนวนผู้ตรวจต่างกัน เมื่อรูปแบบการตรวจ และวิธีการตรวจให้คะแนนเหมือนกัน มีค่าสัมประสิทธิ์การสรุปอ้างอิงแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 ยกเว้นรูปแบบตรวจข้อสอบบางข้อของผู้สอบทุกคน แสดงว่า จำนวนผู้ตรวจต่างกันมีผลทำให้ค่าสัมประสิทธิ์การสรุปอ้างอิงแตกต่างกัน และเมื่อมีจำนวนผู้ตรวจเพิ่มมากขึ้นค่าสัมประสิทธิ์การสรุปอ้างอิงก็จะมีแนวโน้มเพิ่มขึ้น เช่นเดียวกับ Kanchanawasee (2007, p.37) ที่กล่าวว่า ไม่ว่าจะใช้รูปแบบการตรวจให้คะแนนแบบใด ถ้าเป็นไปได้ควรให้ผู้ตรวจหลายๆ คน ในการตรวจให้คะแนนผู้สอบแต่ละคนแล้วหาคะแนนเฉลี่ยหรือคะแนนรวมแทนความสามารถของผู้สอบ จะช่วยลดความแปรปรวนอันเนื่องมาจากผู้ตรวจ และความคลาดเคลื่อนที่หลีกเลี่ยงได้ หากให้ผู้ตรวจคนเดียว ผู้ตรวจอาจไม่ใส่ใจหรืออาจจะไม่เข้าใจในคุณลักษณะที่ประเมินอย่างแท้จริงก็เป็นไปได้ (Baikularb, 2007, p.281) ซึ่งส่งผลต่อคะแนนที่ได้อาจไม่น่าเชื่อถือเท่ากับการใช้ผู้ตรวจหลายๆ คน ในส่วนของรูปแบบการตรวจข้อสอบบางข้อของผู้สอบทุกคน ในแต่ละวิธีการตรวจให้คะแนนของจำนวนผู้ตรวจต่างกัน มีค่าไม่แตกต่างกันนั้น แสดงว่า ผู้สอบที่ผู้วิจัยได้ทำการสุ่มตามเงื่อนไขที่ศึกษามีความสามารถในการทำแบบทดสอบที่หลากหลาย จึงทำให้คะแนนของผลการทำแบบทดสอบมีความแตกต่างกัน เกิดการกระจายของคะแนนที่แตกต่างกัน และพฤติกรรมกรรมการตรวจให้คะแนนของผู้ตรวจบางคนอาจมีความเข้มงวดต่างกันในการตรวจให้คะแนนกับผู้สอบ (Taoto, 2016, p.113)

2.2 ค่าสัมประสิทธิ์การสรุปอ้างอิงของวิธีการตรวจให้คะแนนต่างกัน ที่มีจำนวนผู้ตรวจเท่ากัน เมื่อใช้รูปแบบการตรวจข้อสอบทุกข้อของผู้สอบทุกคน พบว่า ค่าสัมประสิทธิ์การสรุปอ้างอิงแตกต่างกันทุกคู่ ซึ่งสอดคล้องกับสมมติฐานที่ตั้งไว้ ยกเว้น เมื่อใช้รูปแบบการตรวจข้อสอบบางข้อของผู้สอบทุกคน มีค่าไม่แตกต่างกัน เนื่องจากในวิธีการตรวจให้คะแนนแบบแยกองค์ประกอบและแบบประเมินรวม มีการให้ระดับคะแนนที่แตกต่างกัน วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบมีระดับคะแนนที่ละเอียดและชัดเจนมากกว่าวิธีการให้คะแนนแบบประเมินรวมที่มีเพียงระดับคะแนน 2 ระดับ แม้จะมีเกณฑ์การให้คะแนนแยกเป็นแต่ละระดับเช่นเดียวกัน โดยที่จะให้อิสระแก่ผู้ตรวจในการพิจารณาคำตอบ จึงทำให้เกิดความคลาดเคลื่อนของคะแนนได้มาก และเมื่อมีความคลาดเคลื่อนมาก จึงส่งผลต่อค่าสัมประสิทธิ์สรุปอ้างอิงที่ได้มีค่าน้อย (Peartit, 2002)

2.3 ค่าสัมประสิทธิ์การสรุปอ้างอิงที่มีรูปแบบการตรวจต่างกัน เมื่อมีจำนวนผู้ตรวจเท่ากัน โดยวิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ พบว่า มีค่าสัมประสิทธิ์สรุปอ้างอิงแตกต่างกันทุกจำนวนผู้ตรวจ ยกเว้น จำนวนผู้ตรวจให้คะแนน 1 คน และรูปแบบการตรวจต่างกัน โดยวิธีการตรวจให้คะแนนแบบประเมินรวม มีค่าสัมประสิทธิ์การสรุปอ้างอิงแตกต่างกัน ทุกจำนวนผู้ตรวจ ยกเว้น จำนวนผู้ตรวจ 3 คน ซึ่งไม่สอดคล้องกับสมมติฐานที่ตั้งไว้ ทั้งนี้เนื่องจากรูปแบบการตรวจให้คะแนนในแต่ละรูปแบบ มีจำนวนผู้สอบและ

ข้อสอบแตกต่างกันไปตามจำนวนผู้ตรวจและรูปแบบการตรวจให้คะแนนที่ต่างกัน โดยผู้วิจัยนำผลการตรวจให้คะแนนมาจัดกระทำข้อมูลตามรูปแบบการตรวจที่ต่างกันตามเงื่อนไขที่ศึกษา โดยใช้วิธีการสุ่มอย่างง่าย จึงอาจสุ่มได้ผู้สอบที่มีความสามารถทางภาษาไทยที่ใกล้เคียงกันหรือแตกต่างกันมากอยู่ในกลุ่มเดียวกัน ถ้าความสามารถของผู้สอบในกลุ่มมีความแตกต่างกันมาก การกระจายของคะแนนจะมาก ทำให้ค่าสัมประสิทธิ์การสุบอ้างอิงสูง แต่ถ้าความสามารถของกลุ่มผู้เข้าสอบใกล้เคียงกันมาก การกระจายของคะแนนจะน้อยทำให้ค่าสัมประสิทธิ์การสุบอ้างอิงต่ำ (Gronlund, 1976, p.117)

3. ค่าความแปรปรวนของความคลาดเคลื่อนของคะแนนแบบทดสอบความเรียงวิชาภาษาไทย ชั้นประถมศึกษาปีที่ 3 ที่มีจำนวนผู้ตรวจ รูปแบบการตรวจ และวิธีการตรวจต่างกัน พบว่า วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบและแบบประเมินรวม เมื่อเพิ่มจำนวนผู้ตรวจ 1 คน 2 คน และ 3 คน ในรูปแบบการตรวจให้คะแนนตรวจข้อสอบทุกข้อของผู้สอบทุกคน ค่าความแปรปรวนของความคลาดเคลื่อนลดลง ซึ่งสอดคล้องกับสมมติฐาน ยกเว้นรูปแบบการตรวจตรวจข้อสอบบางข้อของผู้สอบทุกคน เนื่องจากในรูปแบบการตรวจให้คะแนนตรวจข้อสอบทุกข้อของผู้สอบทุกคน ผู้วิจัยตรวจแบบทดสอบที่มีจำนวนข้อมากขึ้น จาก 6 ข้อ 12 ข้อ และ 18 ข้อ สอดคล้องกับงานวิจัยของ Sriwaleerat (2015, p.99) พบว่าค่าความแปรปรวนของความคลาดเคลื่อนที่ได้จากคะแนนสอบที่มีจำนวนข้อเพิ่มขึ้น เมื่อกำหนดให้จำนวนผู้ตรวจคงที่ ค่าความแปรปรวนของความคลาดเคลื่อนลดลงอย่างมีนัยสำคัญทางสถิติที่ระดับ 0.01 และพบว่าเมื่อเพิ่มจำนวนคนตรวจมากขึ้น ค่าสัมประสิทธิ์การสุบอ้างอิงก็จะมีแนวโน้มเพิ่มขึ้น

ข้อเสนอแนะ

ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1. เมื่อใช้จำนวนผู้ตรวจที่ต่างกันในทุกรูปแบบการตรวจที่เหมือนกัน วิธีการตรวจให้คะแนนเหมือนกัน ควรเลือกใช้จำนวนผู้ตรวจ 3 คน ซึ่งเป็นจำนวนที่เหมาะสมและมีความน่าเชื่อถือของผลคะแนน

2. วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ มีค่าสัมประสิทธิ์การสุบอ้างอิงสูงกว่าวิธีการตรวจให้คะแนนแบบประเมินรวม ดังนั้นควรเลือกใช้วิธีการตรวจให้คะแนนแบบแยกองค์ประกอบในการตรวจแบบทดสอบความเรียง เพราะมีระบบการให้คะแนนที่ชัดเจน ทำให้มีความเป็นปรนัย

3. ในรูปแบบการตรวจต่างกัน โดยวิธีการตรวจให้คะแนนแบบแยกองค์ประกอบ ของจำนวนผู้ตรวจ 1 คน และวิธีการตรวจให้คะแนนแบบประเมินรวม ของจำนวนผู้ตรวจ 3 คน ค่าสัมประสิทธิ์การสุบอ้างอิงทั้งสองรูปแบบไม่แตกต่างกัน ดังนั้น จึงควรเลือกใช้รูปแบบการตรวจข้อสอบบางข้อของผู้สอบทุกคน ซึ่งมีความสะดวก ประหยัดเวลา ค่าใช้จ่าย ลดภาระงานในการตรวจ

4. หากต้องการให้ผลการวัดมีค่าสัมประสิทธิ์การสุบอ้างอิงสูงตามกำหนด ในกรณีที่แบบทดสอบความเรียงมีจำนวนน้อยข้อ ต้องใช้ผู้ตรวจที่มีจำนวนมากขึ้น แต่ถ้ามีผู้ตรวจมีจำนวนน้อย ต้องใช้แบบทดสอบความเรียงที่มีจำนวนข้อมากขึ้น

ข้อจำกัด

1. หากจะนำผลการวิจัยไปประยุกต์ใช้นั้น การตรวจข้อสอบในรูปแบบการตรวจข้อสอบบางข้อของผู้สอบทุกคน ซึ่งมีความสะดวก ประหยัดเวลา ค่าใช้จ่าย ลดภาระงานในการตรวจนั้น ควรนำไปใช้กับการตรวจข้อสอบที่มีความต้องการความน่าเชื่อถือของผลคะแนนในการแข่งขันคัดเลือกผู้ที่มีความรู้เฉพาะด้านหรือคัดเลือกผู้สอบในการศึกษาต่อ และใช้กับการคัดเลือกผู้สอบที่มีจำนวนมาก

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. ควรมีการศึกษาความสัมพันธ์การสรุปอ้างอิงที่มีจำนวนผู้ตรวจ 1 คน 2 คน และ 3 คน ในวิธีการตรวจให้คะแนนแบบอื่น เช่น วิธีการให้คะแนนแบบตรวจสอบรายการ (Checklist)
2. ควรมีการศึกษาความสัมพันธ์การสรุปอ้างอิงจากเกณฑ์การตรวจให้คะแนนแบบแยกองค์ประกอบที่มีเกณฑ์การให้คะแนนที่เป็น 4 ระดับ หรือ 5 ระดับ ในรายวิชาต่างๆ
3. ควรมีการศึกษาความสัมพันธ์การสรุปอ้างอิงจากลักษณะของผู้ตรวจที่มีความเข้มงวดการให้คะแนนที่แตกต่างกัน

References

- Baikularb, A. (2007). **An Investigation of Rater Agreement Indexes for Categorized Data**. Doctor of Education Degree Srinakharinwirot University, Graduate School, Testing and Measurement.
- Brennan, R.L., Gao, X. & Colton, D.A. (1995, April). Generalizability Analysis Work Key Listening and Writing Testing. **Educational and Psychological Measurement**, 55(2), 157-176.
- Gronlund, Norman E. (1976). **Measurement and Evaluation in Teaching**. (3 rd ed.). New York : Macmillan.
- Intanate, N. (2011). **Characteristics of the Open-Ended Mathematics Test Scores for Different Numbers of Raters and Scoring Patterns Using Generalizability Model and Many-Facet Rasch Model**. Doctor of Education Degree in Testing and Measurement Graduate School, Srinakharinwirot University.
- Kanjanawasee, S. (2007). **Modern Test Theories**. Bangkok : Chulalongkorn University Press.
- Kwanja, N. (2013). **The Comparison of Generalizability Coefficient of Science Process Skill Test Mathayomsuksa 4 with Different Scoring Pattern**. Master of Education Degree, Mahasarakham University, Faculty of Education, Educational Measurement.
- Linacre, John M. & Wright, Benjamin D. (2002). Construction of Measures from Manyfacet Data. **Journal of Applied Measurement**, 3(4), 486-512.
- Ministry of Education. (2001). **Basic Education Curriculum 2001**. Bangkok : Kurusapa Ladprao Publishing.

- Office of the Basic Education Commission, Bureau of Educational Testing. (2019). **Student Quality Evaluation Manual of Phatumsuksa 3 Students Academic Year 2019**. [Online]. Available : <https://bet.obec.go.th/New2020> [2020, March 20].
- Peeartit, P. (2002). **A study of the optimal number of raters and number of writing tasks using different scoring rubrics**. Master of Education Thesis Measurement and Evaluation in Education Graduate School Chulalongkorn University.
- Smith, P.L. (1978). Sampling errors of variance components in small sample multifaceted generalizability studies. **Journal of Educational-Statistics**, 3(4), 319-346.
- Sriwaleerat, W. (2015). **A Generalization Coefficient Comparison of Thai Essay Test for Phatumsuksa 6 Students between different Number of Items and Number of Raters**. Master of Education Thesis Educational Research and Evaluation, Kanchanaburi Rajabhat University.
- Sudweeks, R.R., Reeve, S. & Bradshaw, W.S. (2005). A Comparison of Generalizability Theory and Many-Facet Rasch Measurement in an Analysis of College Sophomore Writing. **Assessment Writing**, 9(3), 239-261.
- Swartz, C.W., Hooper, S.R., Montgomery, J.W., Wakely, M.B., De Kruif, R.E.L., Reed, M., Brown, T.T., Levine, M.D., & White, K.P. (1999, June). Using Generalizability Theory to Estimate the Reliability of Writing Scores Derived from Holistic and Analytical Scoring Methods. **Education and Psychological Measurement**, 59(3), 492-506.
- Taoto, J. (2014). **A Study of the Reliability of Mathematics Essay Test Score for Matayomsuksa 2 Students: The Different Number of Raters and Scoring Patterns Using Generalizability Theory**. Master of Education Thesis Educational Research and Evaluation Pibulsongkram Rajabhat University.
- Woodruff, D.J. & Feldt, L.S. (1986, September). Test for Equality of Several Alpha Coefficients When Their Sample Estimates are Dependent. **Psychometrika**, 51(2), 393-413.