

สรุปผลการวิจัย อภิปรายผล และข้อเสนอแนะ

การวิจัยครั้งนี้มีวัตถุประสงค์ เพื่อศึกษาเปรียบเทียบความสอดคล้องในการให้คะแนนของผู้ตรวจ 2 คน ที่ตรวจแบบสอบถามความเรียงโดยใช้วิธีเดียวกัน และความสอดคล้องในการให้คะแนนระหว่างวิธีการตรวจที่แตกต่างกัน 4 วิธี คือ การตรวจแบบวิธีวิเคราะห์ให้เสร็จเป็นรายชื่อ การตรวจแบบวิธีวิเคราะห์ให้เสร็จเป็นรายคน การตรวจแบบวิธีประเมินรวมให้เสร็จเป็นรายชื่อ การตรวจแบบวิธีประเมินรวมให้เสร็จเป็นรายคน กลุ่มตัวอย่างที่ใช้ในการวิจัยประกอบด้วย กลุ่มผู้ตอบและกลุ่มผู้ตรวจ โดยกลุ่มตัวอย่างผู้ตอบ เป็นนักเรียนระดับชั้นประถมศึกษาปีที่ 6 ปีการศึกษา 2535 สังกัดสำนักงานการประถมศึกษาจังหวัดสตูล ในเขตอำเภอทุ่งหว้า จำนวน 6 โรงเรียน นักเรียน 30 คนซึ่งได้จากการสุ่มกลุ่มตัวอย่างโดยวิธีการสุ่มหลายชั้น (Multi-stage Sampling) กลุ่มตัวอย่างผู้ตรวจเป็นครูผู้สอนในกลุ่มวิชาสร้างเสริมประสบการณ์ชีวิต ระดับชั้นประถมศึกษาปีที่ 5 - 6 สังกัดสำนักงานการประถมศึกษาจังหวัดสตูลในเขตอำเภอทุ่งหว้า จำนวน 8 คน ได้มาโดยการสุ่มแบบเจาะจง (Purposive Sampling) มีคุณสมบัติ ดังนี้ เป็นครูที่มีประสบการณ์การสอน ในวิชาสร้างเสริมประสบการณ์ชีวิตไม่น้อยกว่า 5 ปี และสำเร็จการศึกษาระดับปริญญาตรีทางการศึกษา

เก็บรวบรวมข้อมูล โดยใช้แบบสอบถามความเรียงที่ผู้วิจัยสร้างขึ้น จำนวน 6 ข้อ พร้อมทั้งเกณฑ์การตรวจ นำแบบสอบไปสอบกับนักเรียนที่เป็นกลุ่มตัวอย่าง แล้วนำกระดาษคำตอบจากกลุ่มตัวอย่างจำนวน 30 คน ไปถ่ายเอกสารจำนวน 8 ชุด สำหรับผู้ตรวจ 8 คน แล้วจัดสถานที่ นัดหมายวัน เวลา ให้ผู้ตรวจทั้ง 8 คนตรวจกระดาษคำตอบพร้อมกัน ตามเกณฑ์การตรวจและวิธีการตรวจที่กำหนด ผู้ตรวจสุ่มวิธีการตรวจโดยการจับสลาก ในแต่ละวิธีมีผู้ตรวจ 2 คน ผู้ตรวจ 1 คน ตรวจ 1 วิธี เท่านั้น นำข้อมูลที่ได้มาหาค่าสถิติพื้นฐาน และหาความสัมพันธ์ของการตรวจให้คะแนน โดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์ของเพียร์สัน ในการตรวจสอบความสอดคล้องของการให้คะแนนระหว่างผู้ตรวจและระหว่างวิธีการตรวจ วิเคราะห์ค่าสถิติต่าง ๆ โดยใช้โปรแกรมสำเร็จรูป

SPSS^x ของสถาบันคอมพิวเตอร์ จุฬาลงกรณ์มหาวิทยาลัย

ผลการวิจัย

1. ค่าสถิติพื้นฐานในการให้คะแนนของผู้ตรวจ 8 คน ภายใต้วิธีการตรวจ 4 วิธี

การให้คะแนนของผู้ตรวจทั้ง 8 คนแตกต่างกันไปสรุปได้ ดังนี้ ผู้ตรวจ ทั้ง 8 คน ตรวจให้คะแนนสูงสุดอยู่ในช่วง 33-48 คะแนนต่ำสุดอยู่ในช่วง 9-21 ค่าเฉลี่ย (Mean) อยู่ใน ช่วง 22.13-35.6 คะแนน ส่วนเบี่ยงเบนมาตรฐาน (S.D.) อยู่ในช่วง 4.42-8.33 และเมื่อ เฉลี่ยคะแนนผู้ตรวจ 2 คนที่ตรวจวิธีเดียวกันเป็นคะแนนการตรวจในวิธีนั้นการตรวจ 4 วิธี มีค่าสถิติ ดังนี้ การให้คะแนนสูงสุดอยู่ในช่วง 33-46 คะแนนต่ำสุดอยู่ในช่วง 11-15 ค่าเฉลี่ย (Mean) อยู่ ในช่วง 22.57-31.17 คะแนน ส่วนเบี่ยงเบนมาตรฐาน (S.D.) อยู่ในช่วง 4.39 - 6.69

2. การหาความสัมพันธ์ในการตรวจให้คะแนนระหว่างผู้ตรวจของแต่ละวิธี

การหาความสัมพันธ์ในการให้คะแนนระหว่างผู้ตรวจที่ตรวจวิธีเดียวกันใช้สัมประสิทธิ์ สหสัมพันธ์แบบเพียร์สัน เป็นตัวบอกขนาดและรูปแบบ พบว่าในแต่ละวิธี ผู้ตรวจที่ตรวจวิธีเดียวกันมี ความสัมพันธ์สอดคล้องกัน ในทิศทางเดียวกัน มากน้อย ดังนี้ การตรวจวิธีวิเคราะห์ให้เสร็จเป็น รายข้อ การตรวจแบบประเมินรวมให้เสร็จเป็นรายข้อ การตรวจวิธีวิเคราะห์ให้เสร็จเป็นรายคน และการตรวจแบบวิธีประเมินรวมให้เสร็จเป็นรายคน เท่ากับ 0.8408 0.7500 0.7362 และ 0.6835 ตามลำดับ

3. ความสัมพันธ์ในการตรวจให้คะแนน 4 วิธี

การหาความสัมพันธ์ในการให้คะแนนระหว่างวิธีการตรวจเมื่อใช้คะแนนเฉลี่ยของ ผู้ตรวจ 2 คนเป็นคะแนนการตรวจของวิธีนั้น ใช้สัมประสิทธิ์สหสัมพันธ์แบบเพียร์สัน เป็นตัวบอก ความสัมพันธ์ในการให้คะแนนระหว่างวิธีการตรวจ 4 วิธี พบว่า แต่ละวิธีการให้คะแนนมีความสอด คล้องกัน ในทิศทางเดียวกัน มีค่าตั้งแต่ 0.8116 - 0.6299 สัมประสิทธิ์สหสัมพันธ์ของการตรวจ แบบวิธีวิเคราะห์ให้เสร็จเป็นรายข้อ กับการตรวจแบบวิธีประเมินรวมให้เสร็จเป็นรายข้อมีค่าสูงสุด คือ .8116 และที่มีค่าต่ำสุด คือ การตรวจแบบวิธีวิเคราะห์ให้เสร็จเป็นรายคนกับการตรวจแบบวิธี ประเมินรวมให้เสร็จเป็นรายข้อ คือ 0.6299 แต่ละวิธีมีความสัมพันธ์กับวิธีอื่น ๆ ในทิศทาง เดียวกัน

อภิปรายผลการวิจัย

ในการศึกษาผลการตรวจให้คะแนนแบบสอบความเรียงในลักษณะวัดความรู้ การนำหลักวิชาจากบทเรียนมาใช้แก้ปัญหา ในกลุ่มวิชาสร้างเสริมประสบการณ์ชีวิตระดับชั้นประถมศึกษาปีที่ 6 เรื่อง สิ่งที่อยู่รอบตัวเรา ผู้วิจัยได้สร้างและพัฒนาแบบสอบขึ้นโดยมีแนวเฉลยคำตอบที่มีขอบเขตของคำตอบ และมีเกณฑ์การตรวจตามวิธีการตรวจที่กำหนดให้ นำไปใช้สอบกับกลุ่มตัวอย่าง แล้วนำกระดาษคำตอบมาศึกษา ถึงการตรวจให้คะแนน เนื่องจากการวัดผลด้วยแบบสอบความเรียงนั้น สิ่งจำเป็นที่ต้องควบคุม คือ ข้อสอบและผู้ตรวจ เพราะทั้ง 2 ตัวนี้ มีผลเกี่ยวข้องซึ่งกันและกัน De Gruijter (1980:221) กล่าวว่าข้อสอบความเรียงที่ไม่ดี ไม่สามารถยกระดับคุณภาพให้ดีขึ้นโดยอาศัยวิธีการใด ๆ แม้วิธีการตรวจนั้นจะกำหนดเกณฑ์การตรวจไว้เด่นชัดเพียงใดก็ตาม ในทางกลับกัน ข้อสอบที่สร้างอย่างดีแต่ใช้วิธีการตรวจที่ไม่เหมาะสม ข้อสอบนั้นก็จะมีคุณค่าเช่นกัน ดังนั้น แบบสอบความเรียงที่ใช้ศึกษานี้จึงได้สร้างและพัฒนา ตามหลักการสร้างจนมีคุณภาพที่เชื่อถือได้ซึ่งพิจารณาได้จากการที่แบบสอบมีความตรงตามเนื้อหารายชื่อ โดยผู้เชี่ยวชาญมีความคิดเห็นสอดคล้องกันว่า ข้อสอบ มีความสอดคล้องกับจุดประสงค์และเนื้อหา ทั้ง 6 ข้อ อยู่ในช่วง 0.67-0.92 มีค่าความตรงเชิงเกณฑ์สัมพัทธ์เท่ากับ .54 และ .44 เมื่อคำนวณจาก คะแนนการตรวจวิธีวิเคราะห์ และวิธีประเมินรวม ตามลำดับ มีความเที่ยงแบบความสอดคล้องภายใน ตามสูตร อัลฟา ครอนบาค เท่ากับ 0.65 และ 0.59 เมื่อคำนวณจากคะแนนการตรวจวิธีวิเคราะห์และวิธีประเมินรวมตามลำดับ ส่วนผู้ตรวจ พบว่าเป็นแหล่งความแปรปรวนที่มีอิทธิพลต่อการอ้างอิง (ความเที่ยง) ในการประเมินความเรียงมากที่สุด (ไนรัตน์ วงษ์นาม 2532:112) ดังนั้นเพื่อลดความแตกต่างของผู้ตรวจ จึงได้กำหนดเกณฑ์คุณสมบัติของครูผู้ตรวจ คือ เป็นผู้สำเร็จการศึกษาระดับปริญญาตรีทางการศึกษา ซึ่งนับเป็นผู้ที่มีความรู้ทางด้านการศึกษา และประเมินผลการศึกษา และเป็นผู้ที่ทำการสอนในกลุ่มวิชาสร้างเสริมประสบการณ์ชีวิตไม่น้อยกว่า 5 ปี และก่อนการตรวจได้มีการฝึก และทดลองตรวจ ถือได้ว่าครูผู้ตรวจ มีพื้นฐานความสามารถในการตรวจแบบสอบความเรียงครั้งนี้ใกล้เคียงกัน

ความสอดคล้องในการตรวจให้คะแนนระหว่างผู้ตรวจที่ตรวจวิธีเดียวกัน

จากผลการวิจัย เมื่อพิจารณาค่าสถิติพื้นฐานในการตรวจของผู้ตรวจแต่ละคน (ตารางที่ 5) จะเห็นว่า ค่าสถิติพื้นฐานที่ได้มีค่าแตกต่างกันไป บางคนให้คะแนนใกล้เคียงกับผู้ตรวจคนอื่น ๆ บางคนก็ให้คะแนนแตกต่างไปจากคนอื่น ๆ มาก ความแตกต่างที่เกิดขึ้นนี้ อาจเนื่องมาจากความแตกต่างของผู้ตรวจ หรืออาจเป็นเพราะวิธีการตรวจ เพราะในการตรวจแบบสอบถามเรียงนั้น คำตอบที่ถูกต้องของข้อสอบ ไม่สามารถทำได้โดยสมบูรณ์เหมือนข้อสอบปรนัย ธรรมชาติของการวัดแบบนี้ต้องใช้วิจารณญาณส่วนตัวของผู้ตรวจ จึงทำให้เกิดความคลาดเคลื่อนขึ้นได้ จากความแตกต่างในการให้คะแนนของผู้ตรวจ 8 คน เป็นไปตามที่ Coffman (1971:277) กล่าวว่า เกิดจากสาเหตุสำคัญเช่น ผู้ตรวจแต่ละคนมีความเข้มงวดต่างกัน บางคนชอบให้คะแนนมาก แต่บางคนชอบให้คะแนนน้อย ๆ การกระจายของคะแนน หรือ พิสัยของคะแนนจึงแตกต่างกัน บางคนให้คะแนนใกล้เคียง ๆ กับค่าเฉลี่ย บางคนให้คะแนนมีพิสัยกว้าง คือเกือบเต็ม และเกือบศูนย์ ถ้าการตรวจนั้นให้อิสระแก่ผู้ตรวจในการพิจารณาให้คะแนนก็จะทำให้เกิดความแตกต่างของคะแนนที่เนื่องมาจากผู้ตรวจมากขึ้น

เมื่อพิจารณาผู้ตรวจ 2 คน ที่ตรวจวิธีเดียวกัน จะเห็นว่าผู้ตรวจในวิธีวิเคราะห์ให้เสร็จเป็นรายข้อมีความสอดคล้องในการให้คะแนนมากที่สุด จากตารางที่ 9 จะเห็นได้ว่าค่าเฉลี่ย (Mean) ในการตรวจใกล้เคียงกันมากที่สุดคือมีค่าแตกต่างกัน เท่ากับ 0.87 การกระจายของการให้คะแนนพิจารณาจากค่า S.D. ใกล้เคียงกันมากที่สุด คือมีค่าแตกต่างกัน เท่ากับ 0.26 เมื่อเปรียบเทียบกับความใกล้เคียงกันของค่าเฉลี่ย (Mean) และ S.D. จากผู้ตรวจ 2 คน ในวิธีอื่น ๆ และมีค่าสหสัมพันธ์แบบเพียร์สัน ซึ่งแสดงถึงความสอดคล้องในการให้คะแนนมีค่าสูงที่สุด คือ $r = 0.8408$ ซึ่งสอดคล้องกับสมมติฐานการวิจัยข้อที่ 1 ที่ว่า ความสอดคล้องในการให้คะแนนแบบสอบถามเรียงระหว่างผู้ตรวจ 2 คนในแต่ละวิธี การตรวจวิธีวิเคราะห์ให้เสร็จเป็นรายข้อ น่าจะมีความสอดคล้องในการให้คะแนน มากกว่าวิธีอื่น ๆ ทั้งนี้ เพราะการตรวจวิธีวิเคราะห์ มีเกณฑ์ หรือประเด็นการตรวจพร้อมทั้งนำหนักคะแนนที่ค่อนข้างชัดเจน นอกจากนี้ การตรวจให้เสร็จเป็นรายข้อ ก็ยังสามารถลดความแปรปรวนในการใช้เกณฑ์การตรวจ ทำให้การตรวจวินิจฉัยคำตอบถูกต้องมากขึ้น ผู้ตรวจให้คะแนนได้ตรงกันมากกว่า ส่วนการตรวจวิธีอื่น ๆ เมื่อเปรียบเทียบกับ การตรวจแบบประเมินรวมให้เสร็จเป็นรายข้อมีความสอดคล้องเป็นลำดับที่ 2 เมื่อเปรียบเทียบกับ การตรวจแบบวิเคราะห์ให้เสร็จเป็นรายคน คือมีค่าสหสัมพันธ์ $r = 7500$ ซึ่งสูงกว่า และมีการกระจายในการตรวจให้คะแนน

ใกล้เคียงกันมากกว่า ($S.D.$ ต่างกัน = .69) ส่วนการตรวจแบบประเมินรวมให้เสร็จเป็นรายคน มีความสอดคล้องในการตรวจให้คะแนนระหว่างผู้ตรวจ น้อยที่สุด เพราะ มีค่าเฉลี่ย (Mean) และ $S.D.$ ของคะแนนแตกต่างกันมากที่สุด ค่าสหสัมพันธ์ระหว่างผู้ตรวจในวิธีนี้น้อยที่สุด

เมื่อพิจารณาค่าเฉลี่ยในการตรวจให้คะแนนพบว่าผู้ตรวจทุกคนตรวจให้คะแนนอยู่ในเกณฑ์ที่นักเรียนได้คะแนนน้อย เมื่อเทียบกับคะแนนเต็ม คือมีคะแนนเฉลี่ย (Mean) อยู่ในช่วง 22.13 - 35.6 จากคะแนนเต็ม 60 หรือ เมื่อเฉลี่ยคะแนนระหว่างผู้ตรวจ 2 คน ที่ตรวจวิธีเดียวกัน (ตารางที่ 6) จะเห็นว่า คะแนนเฉลี่ย (Mean) ของการให้คะแนน 4 วิธี อยู่ในช่วง 22.57 - 31.17 คะแนน จากคะแนนเต็ม 60 ถ้าพิจารณาค่าเฉลี่ยแล้ว แสดงว่าคำตอบของนักเรียนไม่ค่อยถูก หรือไม่ครอบคลุมคำตอบที่สมบูรณ์ ทั้งนี้อาจเนื่องมาจาก นักเรียนไม่คุ้นเคย กับการทำแบบสอบ ความเรียง ไม่ได้ฝึกฝนการเขียนบรรยายความคิด ความเข้าใจออกมา บางคนการอ่านและเขียนยังไม่คล่องเท่าที่ควร มักเขียนตอบสั้น ๆ ประเด็นคำตอบไม่ครอบคลุม สมบูรณ์ ซึ่งอาจเป็นอิทธิพลของการทำข้อสอบปรนัยอยู่เป็นประจำ โดยการตอบคำถามข้อหนึ่ง ๆ จะมีคำตอบถูกเพียงข้อเดียว และจากการพูดคุย ชักถาม นักเรียนในช่วงเก็บข้อมูล คือ หลังจากสอบเสร็จแต่ละห้อง พบว่า นักเรียนจะแสดงความพอใจและคิดว่าตนเองทำข้อสอบได้ เมื่อรู้ว่าประเด็นเดียวหรือสองประเด็นที่ตอบไปนั้นถูก แม้จะอธิบายว่าคำตอบถูกมีหลายประเด็น นักเรียนต้องตอบหลาย ๆ ประเด็นที่ถูกจึงถือว่ามีความรู้สมบูรณ์จะได้คะแนนมาก นักเรียนก็ยังแสดงว่าตั้งใจที่ตอบถูกประเด็นหนึ่ง บางคนใช้ภาษาวกวน เขียนคำตอบไม่ชัดเจน ทำให้การตรวจตามวิธีวิเคราะห์ให้คะแนนน้อยกว่าการตรวจตามวิธีประเมินรวม

ความสอดคล้องในการตรวจให้คะแนนระหว่างวิธีการตรวจ 4 วิธี

จากตารางที่ 5 ค่าสถิติพื้นฐานของการให้คะแนนของผู้ตรวจ 8 คน ใน 4 วิธี เมื่อพิจารณาคะแนนจากการตรวจเป็นรายข้อ กับรายคน จะเห็นว่า ค่าเฉลี่ย (Mean) ของการให้คะแนนเป็นรายข้อ คะแนนเฉลี่ยอยู่ในช่วง 22.13 - 28.5 ค่า $S.D.$ อยู่ในช่วง 4.45 - 6.63 ส่วนการให้คะแนนเป็นรายคน ค่าเฉลี่ย (Mean) อยู่ในช่วง 25.4 - 35.6 และค่า $S.D.$ อยู่ในช่วง 4.42 - 8.33 นั่นคือ การตรวจให้คะแนนแบบรายข้อ จะให้ระดับคะแนนใกล้เคียงกันมากกว่าการตรวจให้คะแนนเป็นรายคนและเมื่อใช้ค่าเฉลี่ยของคะแนนจากผู้ตรวจ 2 คน ที่ตรวจวิธีเดียวกัน เป็นคะแนนการตรวจของวิธีนั้น ๆ (ตารางที่ 6) พบว่า ในการตรวจให้คะแนน

เป็นรายชื่อ มีค่าเฉลี่ย (Mean) อยู่ในช่วง 22.57 - 26.06 และค่า S.D. อยู่ในช่วง 4.39 - 5.81 ส่วนการตรวจให้คะแนนเป็นรายคน มีค่าเฉลี่ย (Mean) อยู่ในช่วง 26.5 - 31.17 และค่า S.D. อยู่ในช่วง 5.64 - 6.69 เมื่อพิจารณาค่าสหสัมพันธ์แบบเพียร์สันระหว่างผู้ตรวจ (ตารางที่ 7) พบว่าในการตรวจให้เสร็จเป็นรายชื่อมีความสอดคล้องในการตรวจให้คะแนน สูงกว่า การตรวจแบบให้เสร็จเป็นรายคน ทั้งในวิธีตรวจแบบวิเคราะห์ และ แบบประเมินรวม และเมื่อใช้ ค่าเฉลี่ยของคะแนนจากผู้ตรวจ 2 คน ที่ตรวจวิธีเดียวกัน เป็นคะแนนการตรวจของวิธีนั้น ๆ แล้วหา ความสัมพันธ์ระหว่างวิธีการตรวจ โดยหาค่าสหสัมพันธ์แบบเพียร์สัน (ตารางที่ 8) ความสอดคล้อง ระหว่างวิธีที่ตรวจให้เสร็จเป็นรายชื่อ คือ วิธีวิเคราะห์รายชื่อ และวิธีประเมินรวมรายชื่อ มีค่า r เท่ากับ .8116 ซึ่งสูงกว่า ความสอดคล้องระหว่างวิธีที่ตรวจให้เสร็จเป็นรายคน คือ วิธีวิเคราะห์ รายคน และวิธีประเมินรวมรายคน มีค่า r เท่ากับ .7377 จากผลการวิจัยที่กล่าวมาแล้วนี้เป็นไป ตามสมมุติฐานข้อที่ 2 ที่ว่า ในการตรวจ 4 วิธี การตรวจให้เสร็จเป็นรายชื่อน่าจะมีความสอดคล้อง ในการตรวจให้คะแนนมากกว่า การตรวจให้เสร็จเป็นรายคน ซึ่งสอดคล้องกับนักวัดผลหลายท่าน ที่ได้เสนอแนะการตรวจแบบสอบถามเรียงไว้เช่น เยาวดี วิบูลย์ศรี (2528:145) พวงรัตน์ ทวีรัตน์ (2530:120) อุทุมพร จามรมาน (2535:26) ซึ่งเสนอแนะให้อ่านคำตอบทีละข้อของทุกคน โดยการตรวจเป็นรายชื่อ จะทำให้คะแนนมีความถูกต้อง เชื่อถือได้มากยิ่งขึ้น Bergman (1981:130 - 133) แนะนำว่า การตรวจทีละข้อจนครบทุกคนทำให้ครูกำหนดเกณฑ์ที่เหมาะสมในแต่ละข้อได้ และสามารถจัดการความคลาดเคลื่อนแบบ halo effects ได้ Cochran and Weidemann (1937 cited by Hopkins and Stanley 1972: 213) เสนอไว้ว่าการตรวจแต่ละข้อจนครบทุกคนทำให้เกิดผลดี 2 ประการ คือ (1) ได้คะแนนจากการเปรียบเทียบที่มีความถูกต้องและยุติธรรม (2) ทำให้ผู้ตรวจมีตัวแบบในการตรวจเพียงตัวแบบเดียว ทำให้ประหยัดเวลาในการตรวจและมีความ ถูกต้องแม่นยำมากขึ้น ผู้วิจัยคิดว่า เนื่องจากในการตรวจเป็นรายชื่อ มีการใช้เกณฑ์ได้คงที่มากกว่า จากการที่มีเกณฑ์หรือตัวแบบเดียวในการตรวจข้อหนึ่ง ๆ จนเสร็จทุกคน ส่วนการตรวจเป็นรายคนนั้น ต้องเปลี่ยนข้อในการตรวจทุกครั้ง ข้อสอบแต่ละข้อ มีสาระของคำตอบแตกต่างกันไป เกณฑ์การให้ คะแนนแต่ละข้อก็จะต่างกันไปตามสาระของคำตอบที่กำหนดไว้ เมื่อมีการเปลี่ยนข้อ เปลี่ยนคำตอบ ก็เปลี่ยนเกณฑ์การให้คะแนนไปด้วย เมื่อตรวจจนครบทุกข้อของผู้ตอบ 1 คน จึงมาเริ่มใช้เกณฑ์การ ให้คะแนนแต่ละข้อซ้ำอีก การเปลี่ยนเกณฑ์ในลักษณะนี้ทำให้การใช้เกณฑ์การตรวจข้อหนึ่ง ๆ ไม่ต่อเนื่อง การใช้เกณฑ์การตรวจจึงคลาดเคลื่อนได้มากขึ้น ซึ่งก็เป็นไปตามที่นักวัดผลหลายท่านที่กล่าว มาแล้วแนะนำไว้ ว่าควรตรวจเป็นรายชื่อ

เมื่อพิจารณาการตรวจวิธีวิเคราะห์และวิธีประเมินรวม จะเห็นว่า ส่วนใหญ่ในการตรวจวิธีวิเคราะห์ ทั้งการตรวจแบบรายชื่อและการตรวจแบบรายคน การตรวจให้คะแนนมีค่าเฉลี่ย (Mean) และส่วนเบี่ยงเบนมาตรฐาน (S.D.) ต่ำกว่าการตรวจแบบประเมินรวมทั้งแบบการตรวจเป็นรายชื่อ และการตรวจเป็นรายคน ดังจะเห็นได้จากตารางที่ 5 การตรวจวิธีวิเคราะห์ มีค่า \bar{x} อยู่ในช่วง 23-27.6 S.D. อยู่ในช่วง 4.45-7.62 ส่วนการตรวจวิธีประเมินรวม มีค่า \bar{x} อยู่ในช่วง 23.7-35.6 ค่า S.D. อยู่ในช่วง 5.94-8.33 นั่นคือการตรวจแบบวิธีวิเคราะห์ มีการให้คะแนน และการกระจายของคะแนน ใกล้เคียงกันมากกว่าการตรวจแบบประเมินรวมทั้งนี้อาจเนื่องมาจากในการตรวจวิธีวิเคราะห์นั้นมีเกณฑ์การให้คะแนนที่ชัดเจนกว่า คำตอบที่ไม่แน่ใจว่าตอบได้ถูกต้องตรงประเด็นหรือไม่ ก็จะถูกตัดสินหักคะแนนไปเลย ส่วนการตรวจแบบวิธีประเมินรวมนั้นอาศัยความคิดเห็นและการตัดสินใจของผู้ตรวจมากกว่าวิธีวิเคราะห์ การให้คะแนนตามวิธีประเมินรวมจึงให้ระดับคะแนนและการกระจายคะแนนได้มากกว่า และจากตารางที่ 8 จะเห็นว่าการตรวจแบบวิธีวิเคราะห์ให้เสร็จเป็นรายชื่อและวิธีประเมินรวมให้เสร็จเป็นรายชื่อ มีค่า r เท่ากับ .8116 ซึ่งมีความสอดคล้องในการให้คะแนนสูงกว่าวิธีอื่น ๆ

จากวิธีการตรวจแบบสอบถามความเรียงทั้ง 4 วิธี เมื่อใช้ค่าเฉลี่ยคะแนนระหว่างผู้ตรวจ 2 คนที่ตรวจวิธีเดียวกันแทนคะแนนการตรวจของวิธีนั้น ๆ เปรียบเทียบกัน พบว่าการตรวจวิธีประเมินรวมให้เสร็จเป็นรายคน มีความแตกต่างในการตรวจให้คะแนนกับวิธีอื่น ๆ ทุกวิธี ดังจะเห็นได้จากมีค่าเฉลี่ย (Mean) ส่วนเบี่ยงเบนมาตรฐาน (S.D.) และค่าสถิติพื้นฐานตัวอื่น ๆ ของการตรวจให้คะแนนในวิธีนี้ แตกต่างจากการตรวจในวิธีอื่น ๆ มาก (พิจารณาในตารางที่ 6) ทั้งนี้เนื่องจากวิธีนี้มีความคลาดเคลื่อนในการให้คะแนนมากกว่าวิธีอื่น ๆ ซึ่งมาจากความมีอิสระของผู้ตรวจในการพิจารณาให้คะแนนตามวิธีการตรวจแบบประเมินรวม และการใช้เกณฑ์การตรวจแต่ละข้อไม่ต่อเนื่องจากการตรวจให้เสร็จเป็นรายคนนั่นเอง

ข้อเสนอแนะ

จากผลการวิจัยครั้งนี้ ทำให้ได้แนวความคิดซึ่งอาจนำไปใช้ให้เกิดประโยชน์ ดังนี้

ข้อเสนอแนะในการนำไปใช้

1. ในการตรวจให้คะแนนแบบสอบความเรียง ควรมีเฉลยหรือเกณฑ์การตรวจให้คะแนนให้ชัดเจน
2. ควรเลือกการตรวจวิธีวิเคราะห์ให้เสร็จเป็นรายข้อ แต่ถ้ามีผู้ตรวจหลายคนแล้วใช้ค่าเฉลี่ยของคะแนนจากผู้ตรวจ แทนคะแนนผลสัมฤทธิ์ของนักเรียน อาจเลือกวิธีการตรวจแบบวิธีวิเคราะห์ให้เสร็จเป็นรายข้อ หรือการตรวจวิธีประเมินรวมให้เสร็จเป็นรายข้อก็ได้ โดยถ้าเป็นการตรวจแบบประเมินรวม ควรมีการจัดระดับชั้นคุณภาพคำตอบ 5 ระดับ เพราะการจัดระดับชั้นคุณภาพน้อยเกินไป จะแยกคุณภาพคำตอบได้แตกต่างกันน้อย และควรมีตัวแบบคำตอบที่สมบูรณ์ไว้เป็นตัวเกณฑ์เทียบ
3. ในการตรวจให้คะแนนแบบสอบความเรียงหากมีผู้ตรวจหลายคน ต้องแน่ใจว่าผู้ตรวจทุกคนมีความรู้ในเรื่องที่ตรวจอย่างดี ควรมีการชี้แจงถึงวิธีและเกณฑ์การตรวจให้เข้าใจตรงกัน

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. ควรศึกษาในลักษณะเดิม ในกลุ่มประชากรอื่น ๆ เช่น ในระดับมัธยม หรืออุดมศึกษา ซึ่งผู้ตอบมีความสามารถในการเขียนบรรยายได้มากขึ้น ทักษะการเขียนดีขึ้น ผู้ตรวจเปลี่ยนไปใช้ผู้ตรวจที่มาจากการสุ่มตามปกติ และตรวจสอบความสามารถที่แท้จริงของผู้ตอบด้วย
2. ควรศึกษาในวิธีการตรวจแบบเดิม แต่หาความสอดคล้องในการตรวจแบบการตรวจซ้ำ
3. ควรศึกษาเกี่ยวกับลำดับความสามารถของนักเรียน จากการตรวจแบบสอบความเรียงในลักษณะแบ่งกันตรวจตามจำนวนคน และแบ่งกันตรวจตามจำนวนข้อ