



บทที่ 5

สรุป อภิปรายผล และ ข้อเสนอแนะ

สรุปผลการวิจัย

การวิจัยเรื่องนี้เป็นการประยุกต์ทฤษฎีการอ้างอิงสรุปในการอธิบายความแม่นยำหรือความเที่ยงของแบบสอบความเรียง มีวัตถุประสงค์หลัก 3 ประการคือ (1) เพื่อประมาณค่าความแปรปรวนของตัวประกอบ 3 ตัวประกอบ ที่มีผลต่อค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง ได้แก่ ผู้ตอบ ข้อสอบ และ ผู้ตรวจ (2) เพื่อศึกษาและเปรียบเทียบค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง 6 ค่า ที่ประมาณค่าจากการตรวจ 2 วิธี คือ วิธีประเมินรวม และ วิธีวิเคราะห์ ภายใต้สถานการณ์การตรวจ 3 สถานการณ์ คือ สถานการณ์ที่ 1 ผู้ตรวจ ตรวจโดยใช้ประสบการณ์เดิม สถานการณ์ที่ 2 ผู้ตรวจ ตรวจตามได้รับการชี้แจงจากผู้วิจัยและไม่รู้ผลการเรียนของผู้ตอบ สถานการณ์ที่ 3 ผู้ตรวจ ตรวจตามได้รับการชี้แจงจากผู้วิจัยและรู้ผลการเรียนของผู้ตอบ (3) เพื่อเลือกวิธีตรวจ จำนวนผู้ตรวจ และจำนวนข้อสอบ ที่ให้ค่าสัมประสิทธิ์การอ้างอิง 0.50 เป็นอย่างต่ำ

การวิจัยนี้สนใจศึกษาตัวประกอบสำคัญที่เกี่ยวกับการสอบแบบความเรียง (Essay Examination) 3 ตัวประกอบคือ ผู้ตอบ ข้อสอบ และผู้ตรวจ กำหนดความสัมพันธ์ระหว่างตัวประกอบเป็นแบบ $p \times i \times r$ คือผู้ตอบแต่ละคนตอบข้อความเหมือนกันทุกข้อ ผู้ตรวจแต่ละคนตรวจคำตอบของผู้ตอบทุกคนและทุกข้อ เอกภพของการสังเกตที่ยอมรับได้ประกอบด้วย ผู้ตอบ ซึ่งเป็นนักเรียนชั้นประถมศึกษาปีที่ 6 และ ผู้ตรวจ ซึ่งเป็นครูที่สอนกลุ่มวิชาสร้างเสริมประสบการณ์ชีวิต ชั้นประถมศึกษาปีที่ 6 ของโรงเรียนสังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ (สพช.) ในอำเภอโนนสังจังหวัดอุดรธานี ปีการศึกษา 2531 และข้อสอบวัดความสามารถในการแสดงความคิดเห็นเกี่ยวกับข่าวและเหตุการณ์ กลุ่มตัวอย่างหรือเงื่อนไขของตัวประกอบที่ใช้ในการศึกษา เป็นตัวอย่างสุ่มจากเอกภพการสังเกต ประกอบด้วย ข้อสอบแบบความเรียง 5 ข้อ ผู้ตรวจ 20 คน และ ผู้ตอบ 30 คน

เครื่องมือที่ใช้ในการวิจัยประกอบด้วย ข้อสอบแบบความเรียงวัดความสามารถเกี่ยวกับข่าวและเหตุการณ์ที่ผู้วิจัยพัฒนาขึ้น จำนวน 5 ข้อ แต่ละข้อให้คะแนน 0 ถึง 10 คะแนน และคู่มือ

การตรวจที่ผู้วิจัยพัฒนาขึ้น เพื่อใช้เป็นเกณฑ์ประกอบในการพิจารณาให้คะแนน

ผลการวิจัยสามารถสรุปได้ดังนี้

1. ในการประเมินความสามารถในการแสดงความคิดเห็นเกี่ยวกับข่าวและเหตุการณ์ โดยใช้ข้อสอบแบบความเรียง ความแปรปรวนที่มีอิทธิพลต่อค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง ได้แก่ ความแปรปรวนของผู้ตรวจ ผู้ตอบ ผลร่วมระหว่างข้อสอบกับผู้ตรวจ ผู้ตอบ กับข้อสอบ ผู้ตอบกับผู้ตรวจ แต่ความแปรปรวนของข้อสอบมีผลน้อย จนสามารถกล่าวได้ว่า ข้อสอบไม่ใช่แหล่งความคลาดเคลื่อนในการอ้างอิงสรุป

2. การตรวจที่ผู้ตรวจ ตรวจสอบได้รับการชี้แจงจากผู้วิจัยและไม่รู้ผลการเรียนของผู้ตอบ ทำให้ความแปรปรวนของผู้ตรวจ ลดลงจากการตรวจโดยใช้ประสบการณ์เดิมถึงร้อยละ 50 สำหรับวิธีตรวจประเมินรวม และ ร้อยละ 25 สำหรับวิธีตรวจวิเคราะห์

3. การตรวจที่ผู้ตรวจ ตรวจสอบได้รับการชี้แจงจากผู้วิจัยและรู้ผลการเรียนของผู้ตอบ ทำให้ความแปรปรวนของผู้ตรวจเพิ่มขึ้นพอ ๆ กับการตรวจโดยใช้ประสบการณ์เดิม แต่ผลร่วมระหว่างผู้ตอบกับข้อสอบ มีค่ามากกว่าเมื่อตรวจโดยใช้ประสบการณ์เดิม

4. ค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง ที่อ้างอิงสรุปไปยัง เอกภพของผู้ตรวจอย่าง เดียวมีค่าสูงกว่า ที่อ้างอิงสรุปไปยัง เอกภพของข้อสอบและผู้ตรวจพร้อมกัน

5. ค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง ที่อ้างอิงไปยัง เอกภพของข้อสอบและผู้ตรวจ เมื่อการสอบนั้นใช้ ข้อสอบ 5 ข้อใด ๆ และผู้ตรวจ 10 คนใด ๆ สำหรับการตรวจแต่ละวิธี ในแต่ละสถานการณ์การตรวจมีค่าดังนี้

5.1 การตรวจวิธีประเมินรวม ที่ผู้ตรวจ ตรวจสอบโดยใช้ประสบการณ์เดิม มีค่า 0.3888

5.2 การตรวจวิธีวิเคราะห์ ที่ผู้ตรวจ ตรวจสอบ โดยใช้ประสบการณ์เดิม มีค่า 0.5655

5.3 การตรวจวิธีประเมินรวม ที่ผู้ตรวจ ตรวจสอบได้รับการชี้แจงจากผู้วิจัย และไม่รู้ผลการเรียนของผู้ตอบ มีค่า 0.4782

5.4 การตรวจวิธีวิเคราะห์ ที่ผู้ตรวจ ตรวจสอบได้รับการชี้แจงจากผู้วิจัย และ

ไม่รู้ผลการเรียนของผู้ตอบ มีค่า 0.5895

5.5 การตรวจวิธีประเมินรวม ที่ผู้ตรวจ ตรวจสอบได้รับการชี้แจงจากผู้วิจัย และรู้ผลการเรียนของผู้ตอบ มีค่า 0.3328

5.6 การตรวจวิธีวิเคราะห์ ที่ผู้ตรวจ ตรวจสอบได้รับการชี้แจงจากผู้วิจัยและรู้ ผลการเรียนของผู้ตอบ มีค่า 0.3348

6. ค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบถามเรียงที่อ้างอิงสรุปไปยัง เอกภพ ของผู้ตรวจอย่างเดี่ยว เมื่อการสอบแต่ละครั้งใช้ข้อสอบ 5 ข้อชุดเดิม แต่ใช้ผู้ตรวจ 10 คนใด ๆ สำหรับการตรวจแต่ละวิธี ในแต่ละสถานการณ์การตรวจ มีค่าดังนี้

6.1 การตรวจวิธีประเมินรวม ที่ผู้ตรวจ ตรวจโดยใช้ประสบการณ์เดิมมีค่า 0.4743

6.2 การตรวจวิธีวิเคราะห์ ที่ผู้ตรวจ ตรวจโดยใช้ประสบการณ์เดิมมีค่า 0.6023

6.3 การตรวจวิธีประเมินรวม ที่ผู้ตรวจ ตรวจสอบได้รับการชี้แจงจากผู้วิจัย และไม่รู้ผลการเรียนของผู้ตอบ มีค่า 0.6865

6.4 การตรวจวิธีวิเคราะห์ ที่ผู้ตรวจ ตรวจสอบได้รับการชี้แจงจากผู้วิจัย และไม่รู้ผลการเรียนของผู้ตอบ มีค่า 0.7761

6.5 การตรวจวิธีประเมินรวม ที่ผู้ตรวจ ตรวจสอบได้รับการชี้แจงจากผู้วิจัย และรู้ผลการเรียนของผู้ตอบ มีค่า 0.5268

6.6 การตรวจวิธีวิเคราะห์ ที่ผู้ตรวจ ตรวจสอบได้รับการชี้แจงจากผู้วิจัย และรู้ผลการเรียนของผู้ตอบ มีค่า 0.5985

7. ผลการทดสอบความแตกต่างทางสถิติของค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบ สอบความเรียง ที่อ้างอิงสรุปไปยัง เอกภพของข้อสอบและผู้ตรวจ (M_1) ทั้ง 6 ค่า จากการตรวจ 2 วิธี ภายใต้สถานการณ์ทดลอง 3 สถานการณ์ ได้แก่ ผู้ตรวจ ตรวจโดยใช้ประสบการณ์เดิม ผู้ ตรวจ ตรวจสอบได้รับการชี้แจงจากผู้วิจัยและ ไม่รู้ผลการเรียนของผู้ตอบ และ ผู้ตรวจ ตรวจสอบ ที่ได้รับการชี้แจงจากผู้วิจัยและรู้ผลการเรียนของผู้ตอบ เมื่อการสอบนั้นใช้ข้อสอบ 5 ข้อใด ๆ และ ผู้ตรวจ 10 คนใด ๆ มีค่าไม่แตกต่างกัน ชัดแย้งกับสมมุติฐานข้อที่ 1

8. ผลการทดสอบความแตกต่างทางสถิติ ของค่าสัมประสิทธิ์การอ้างอิงสรุป ของแบบ
สอบความเรียง ที่อ้างอิงไปยัง เอกภพของผู้ตรวจอย่างเดียว (M₂) ปรากฏผลดังนี้

8.1 ค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง 6 คำ จากการ
ตรวจ 2 วิธี คือ วิธีประเมินรวม และ วิธีวิเคราะห์ ภายใต้อาคารการตรวจ 3 สถานการณ์
ได้แก่ ผู้ตรวจตรวจโดยใช้ประสบการณ์เดิม ผู้ตรวจ ตรวจตามที่ได้รับคำสั่งแจงจากผู้วิจัยและ ไม่รู้
ผลการเรียนของผู้ตอบ และ ผู้ตรวจ ตรวจตามที่ได้รับคำสั่งแจงจากผู้วิจัยและรู้ผลการเรียนของผู้
ตอบ เมื่อการสอบใช้ข้อสอบ 5 ข้อชุดเดิม แต่ตรวจโดยผู้ตรวจ 10 คนใด ๆ มีค่าแตกต่างกัน
อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สอดคล้องกับสมมติฐานการวิจัย ข้อที่ 1

8.2 ค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง ที่ตรวจโดยวิธี
วิเคราะห์ กับที่ตรวจโดยวิธีประเมินรวม ในแต่ละสถานการณ์ และรวมทั้ง 3 สถานการณ์ มีค่าไม่
แตกต่างกัน ชัดแย้งกับสมมติฐานการวิจัย ข้อที่ 2, 3, 4 และ 5

8.3 ค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง ภายใต้อาคารการ
การตรวจ ที่ผู้ตรวจตรวจโดยใช้ประสบการณ์เดิม มีค่าไม่แตกต่างจากการตรวจ ที่ผู้ตรวจ ตรวจ
ตามที่ได้รับคำสั่งแจงจากผู้วิจัย ชัดแย้งกับสมมติฐานการวิจัย ข้อที่ 6

8.4 ค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง ภายใต้อาคารการ
การตรวจ ที่ผู้ตรวจ ตรวจโดยใช้ประสบการณ์เดิม มีค่าต่ำกว่าการตรวจ ที่ผู้ตรวจ ตรวจตามที่ได้
รับคำสั่งแจงจากผู้วิจัยและ ไม่รู้ผลการเรียนของผู้ตอบ อย่างมีนัยสำคัญทางสถิติ ที่ระดับ .05 สอด
คล้องกับสมมติฐานการวิจัย ข้อที่ 7

8.5 ค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง ภายใต้อาคารการ
การตรวจ ที่ผู้ตรวจตรวจโดยใช้ประสบการณ์เดิม กับการตรวจ ที่ผู้ตรวจ ตรวจตามที่ได้รับคำสั่ง
แจงจากผู้วิจัยและรู้ผลการเรียนของผู้ตอบ มีค่าไม่แตกต่างกัน สอดคล้องกับสมมติฐานการวิจัย
ข้อที่ 8

8.6 ค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง ภายใต้อาคารการ
การตรวจ ที่ผู้ตรวจตรวจตามที่ได้รับคำสั่งแจงจากผู้วิจัยและ ไม่รู้ผลการเรียนของผู้ตอบมีค่าสูงกว่า
การตรวจ ที่ผู้ตรวจตรวจตามที่ได้รับคำสั่งแจงจากผู้วิจัยและรู้ผลของผู้ตอบ อย่างมีนัยสำคัญทาง
สถิติที่ระดับ .05 สอดคล้องกับสมมติฐานการวิจัย ข้อที่ 9

8.7 ค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง ภายใต้อาคารการ

การตรวจ ที่ผู้ตรวจไม่รู้ผลการเรียนของผู้ตอบ มีค่าสูงกว่าการตรวจที่ผู้ตรวจรู้ผลการเรียนของผู้ตอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 สอดคล้องกับสมมติฐานการวิจัยข้อที่ 10

9. ผลการเลือกวิธีตรวจ จำนวนผู้ตรวจ และ จำนวนข้อสอบ ที่ให้ค่าสัมประสิทธิ์การอ้างอิงสรุป 0.50 เป็นอย่างต่ำ

9.1 รูปแบบการวัดที่อ้างอิงไปยังเอกภพของข้อสอบและผู้ตรวจ วิธีตรวจวิเคราะห์ ที่ผู้ตรวจตรวจตามที่ได้รับการชี้แจงจากผู้วิจัยและ ไม่รู้ผลการเรียนของผู้ตอบ เป็นวิธีที่ให้ค่า สัมประสิทธิ์การอ้างอิงสรุปสูงที่สุด จำนวนผู้ตรวจ และจำนวนข้อสอบ ที่เหมาะสม สามารถนำไปปฏิบัติได้ และ ให้ค่าสัมประสิทธิ์การอ้างอิงสรุป 0.50 เป็นอย่างต่ำ คือ ใช้ผู้ตรวจอย่างน้อย 5 คน ข้อสอบอย่างน้อย 6 ข้อ รองลงไปคือ วิธีตรวจวิเคราะห์โดยใช้ประสบการณ์ของผู้ตรวจ โดยใช้ผู้ตรวจอย่างน้อย 8 คน ข้อสอบ 6 ข้อ

9.2 รูปแบบการวัดที่อ้างอิงไปยังเอกภพของข้อสอบอย่างเดียว เมื่อกำหนดว่าในการสอบครั้งต่อไปจะใช้ข้อสอบ 5 ข้อชุดเดิม วิธีตรวจที่ผู้ตรวจ ตรวจตามได้รับการชี้แจงจากผู้วิจัย โดยไม่รู้ผลการเรียนของผู้ตอบ ทั้งวิธีประเมินรวมและวิธีวิเคราะห์ มีค่าสัมประสิทธิ์การอ้างอิงสรุปสูงกว่าการตรวจในสถานการณ์อื่น ๆ แต่วิธีตรวจวิเคราะห์ให้ค่าสูงกว่า จำนวนผู้ตรวจที่เหมาะสม สามารถนำไปปฏิบัติได้สะดวก และ ให้ค่าสัมประสิทธิ์การอ้างอิงสรุปไม่ต่ำกว่า 0.50 คือ ผู้ตรวจอย่างน้อย 3 คน ส่วนวิธีประเมินรวม ควรใช้ผู้ตรวจอย่างน้อย 5 คน

การอภิปรายผล

1. ค่าประมาณความแปรปรวน

จากการศึกษาอิทธิพลของพาเซ็คต่าง ๆ ที่เกี่ยวข้องกับประเมินความเรียง 3 พาเซ็ค กำหนดความสัมพันธ์ระหว่างพาเซ็ค เป็น pxixr ในขั้นการวิเคราะห์ความแปรปรวน จะพบว่า ในทุก ๆ สถานการณ์ ไม่ว่าจะเป็นการตรวจตามประสบการณ์เดิมของครู การตรวจตามที่ได้รับการชี้แจงจากผู้วิจัย โดยผู้ตรวจไม่รู้ผลการเรียนของผู้ตอบ และตรวจตามที่ได้รับการชี้แจงจากผู้วิจัยรู้ผลการเรียนของผู้ตอบก็ตาม ในบรรดาแหล่งความแปรปรวนเชิงระบบที่ระบุแหล่งได้ชัดเจน ผู้ตรวจเป็นแหล่งความแปรปรวนที่มีอิทธิพลต่อการอ้างอิงมากที่สุด ทั้งนี้จะเป็นเพราะธรรมชาติของการสอบแบบความเรียง Stalnaker (1951 cited by Coffman 1971:

271) กล่าวว่าข้อสอบแบบความเรียง เป็นข้อสอบที่ไม่ใช่มีคำตอบที่ถูกต้องเพียงคำตอบเดียว ในการตัดสินคุณภาพของคำตอบต้องอาศัยผู้เชี่ยวชาญ แต่ผู้เชี่ยวชาญดังกล่าวก็ไม่อาจชี้ถูกหรือผิดได้ โดยเด็ดขาด หากเพียงแต่สามารถบอกระดับคุณภาพของคำตอบได้เท่านั้น การที่ไม่สามารถจะชี้ถูกหรือผิดได้อย่างสมบูรณ์นี้เองทำให้ความแตกต่างระหว่างผู้ตรวจ มีอิทธิพลต่อความแตกต่างของคะแนนสังเกต ส่วนความแปรปรวนของข้อสอบพบว่ามีค่าน้อยมาก นั่นแสดงว่าข้อสอบมีความเป็นเอกพันธ์ (homogeneity) กันมาก จึงถือได้ว่า ข้อสอบไม่ใช่แหล่งความคลาดเคลื่อนในการอ้างอิงสรุป ผลการวิจัยนี้สอดคล้องกับการศึกษาของ Ibramim (1984) แดง กลางท่าไค้ (2531) ที่พบว่าในการประเมินโดยใช้ผู้เชี่ยวชาญนั้น ผู้เชี่ยวชาญเป็นแหล่งความแปรปรวนที่มีอิทธิพลต่อค่าสัมประสิทธิ์การอ้างอิงสรุปมากที่สุด

อย่างไรก็ตาม แม้ว่าจะไม่สามารถขจัดความแปรปรวนของผู้ตรวจได้โดยสิ้นเชิง แต่การตรวจภายใต้สถานการณ์ที่ต่างกัน มีผลทำให้ความแปรปรวนของแหล่งต่าง ๆ มีค่าต่างกันด้วย กล่าวคือ การตรวจวิธีประเมินรวมตามประสบการณ์เดิมของผู้ตรวจ พบความแปรปรวนของผู้ตรวจสูงถึงร้อยละ 59 ในขณะที่การตรวจวิธีวิเคราะห์ ที่ผู้ตรวจ ตรวจตามได้รับการชี้แจงจากผู้วิจัย และไม่รู้ผลการเรียนของผู้ตอบ พบความแปรปรวนเพียงร้อยละ 29 ซึ่งลดลงจากวิธีแรกถึงร้อยละ 50 ทั้งนี้ น่าจะเป็นเพราะการตรวจวิธีแรก อาศัยความประทับใจจากการอ่านที่ค่อนข้างรวดเร็ว ไม่มีเกณฑ์ให้ยึดถือแต่อย่างใด ผู้ตรวจจึงมีอิสระในการให้คะแนนมาก ยังผลให้เกิดความแตกต่างกันมาก สอดคล้องกับความคิดของ Coffman (1971: 277) ที่ว่า การตรวจความเรียง ปัญหาเรื่องความแตกต่างระหว่างผู้ตรวจเป็นเรื่องซับซ้อน เนื่องจากสาเหตุที่ไม่สามารถเลี่ยงได้ 3 ประการคือ (1) ผู้ตรวจต่างกันมีแนวโน้มจะให้คะแนนต่างกัน (2) ผู้ตรวจคนเดียวก็มีแนวโน้มจะให้คะแนนต่างกันตามโอกาสที่ตรวจ และ (3) ความไม่สอดคล้องจะยังมีมากเมื่อข้อสอบให้อิสระในการตอบมากขึ้น ทั้งนี้ เนื่องมาจาก ผู้ตรวจมีความเข้มงวดแตกต่างกันบางคนชอบให้คะแนนสูง ๆ บางคนชอบให้คะแนนต่ำ ๆ บางคนชอบให้คะแนนกระจายมาก บางคนชอบให้คะแนนใกล้ค่าเฉลี่ย ส่วนการตรวจวิธีหลัง เป็นวิธีที่นำวัดผลการศึกษาได้ให้คำแนะนำ ผู้ตรวจจะมีเกณฑ์ในการตรวจ พร้อมทั้งแนวปฏิบัติในการให้คะแนนชัดเจน จึงทำให้มีความคิดเห็นสอดคล้อง ยังผลให้ความแตกต่างระหว่างผู้ตรวจ มีอิทธิพลต่อความแตกต่างของคะแนนสังเกตน้อยลง

อิทธิพลของความคลาดเคลื่อนจากการประทับใจที่เกิดจากการรู้ผลการเรียนของผู้ตอบเป็นอีกตัวแปรหนึ่งที่ทำให้ ความแปรปรวนของผู้ตรวจมีค่าสูงขึ้น ความคลาดเคลื่อนในลักษณะ

นี้รู้จักกันโดยทั่วไปว่า halo effect จากการวิจัยครั้งนี้พบว่า ในสถานการณ์ที่ผู้ตรวจได้รับการชี้แจงจากผู้วิจัยแต่ตรวจคำตอบโดยรู้ผลการเรียนของผู้ตอบต่างกัน ความแปรปรวนของผู้ตรวจจะเพิ่มขึ้นจากเดิมที่ตรวจโดยไม่รู้ผลการเรียน กล่าวคือจากเดิม ร้อยละ 37 เป็นร้อยละ 46 สำหรับการตรวจวิธีประเมินรวม และจากร้อยละ 29 เป็นร้อยละ 38 สำหรับการตรวจวิธีวิเคราะห์ ค่าความแปรปรวนที่เพิ่มขึ้นนี้ เป็นผลจากความคาดหวังจากการรู้ผลการเรียนของผู้ตอบ ผู้วิจัยได้ตรวจสอบผลของความคาดหวังพบว่า ผู้ตอบกลุ่มที่มีผลการเรียนต่างกัน ได้คะแนนแตกต่างกัน กล่าวคือ กลุ่มที่มีผลการเรียนสูง ได้คะแนนสูงกว่ากลุ่มที่มีผลการเรียนต่ำ อย่างมีนัยสำคัญทางสถิติ .01 (รายละเอียดแสดงในภาคผนวก ค) ผลการวิจัยนี้สอดคล้องกับผลการศึกษาของ Chase (1979: 39-42) ซึ่งพบว่า ผู้ตอบ ที่ได้รับความคาดหวังสูงจะ ได้คะแนนสูงกว่าผู้ตอบที่ได้รับความคาดหวังต่ำ และยัง ได้ผลตรงกับการศึกษาของ Hughes, et al. 1983: 65-70)

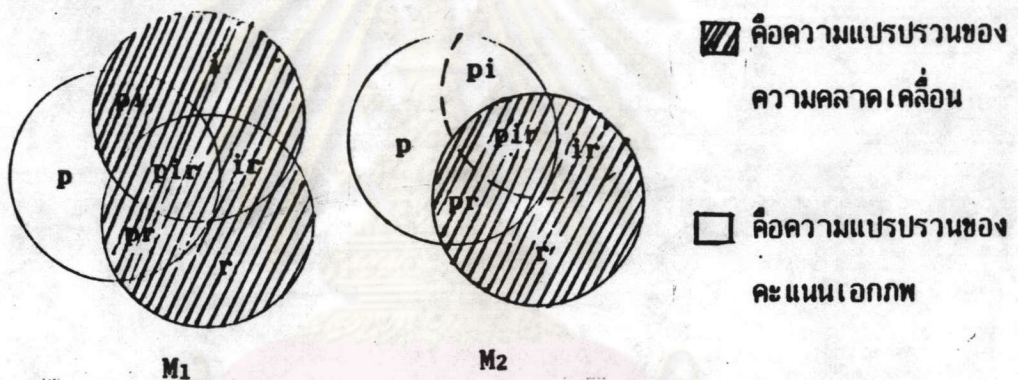
จากผลการวิจัยชี้ให้เห็นว่าการชี้แจงผู้ตรวจให้เข้าใจหลักการตรวจ รู้จักใช้เกณฑ์การตรวจ สามารถลดความแตกต่างระหว่างผู้ตรวจได้เฉพาะกรณีที่ผู้ตรวจไม่รู้จักระบวนการเรียนของผู้ตอบเท่านั้น หากผู้ตรวจรู้ผลผลการเรียนของผู้ตอบแล้ว ความคลาดเคลื่อนแบบ halo effect จะยังมีอิทธิพลต่อความแม่นยำในการอ้างอิงสรุป

2. ค่าประมาณสัมประสิทธิ์การอ้างอิงสรุป

2.1 ความสัมพันธ์ระหว่างรูปแบบการวัดกับค่าสัมประสิทธิ์การอ้างอิงสรุป ค่าสัมประสิทธิ์การอ้างอิงสรุป ซึ่งคำนวณจากอัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพ ต่อความแปรปรวนของคะแนนสังเกต ขึ้นอยู่กับรูปแบบการวัดและการสุ่มจำนวนข้อสอบหรือจำนวนผู้ตรวจที่สอดคล้องกับรูปแบบการวัดนั้น ๆ

รูปแบบ $M_1 (P/-/-, /I, R^*)$ ซึ่งเป็นรูปแบบการวัดที่ต้องการอ้างอิงไปยังเอกภพของข้อสอบและผู้ตรวจพร้อมกัน ค่าสัมประสิทธิ์การอ้างอิงสรุปจะมีค่าต่ำกว่า $M_2 (p/-/I/R^*)$ ซึ่งอ้างอิงไปยัง เอกภพของผู้ตรวจเพียงอย่างเดียว เป็นเพราะว่าตามรูปแบบการวัด M_1 นั้น เอกภพของข้อสอบมีขนาดใหญ่ การสุ่มข้อสอบในระดับต่าง ๆ เพื่อใช้วัดความแตกต่างของผู้ตอบนั้น จะมีความแปรปรวนของความคลาดเคลื่อนอันเนื่องมาจากการสุ่มเกิดขึ้น เช่น ความแตกต่างเนื่องจากความยากของข้อสอบ และผลร่วระหว่างข้อสอบกับตัวประกอบอื่น แต่ในรูปแบบการวัด M_2 จำกัดเอกภพของข้อสอบให้มีเพียง 5 ข้อ แล้วนำมาวัดความสามารถของผู้ตอบทุกข้อ เมื่อ

อ้างอิงไปยัง เอกภพของข้อสอบ จึงไม่มีความคลาดเคลื่อนอันเนื่องมาจากการสุ่ม เพราะสิ่งที่ต้องการรู้หรืออ้างอิงถึง ถูกนำมาศึกษาทั้งหมด อีกทั้งความแปรปรวนของผลร่วมระหว่างตัวประกอบผู้ตอบกับข้อสอบก็ไม่ถูกนับเข้าเป็นแหล่งความแปรปรวนของความคลาดเคลื่อนอีก แต่ได้กลายเป็นส่วนหนึ่งของความแปรปรวนของคะแนนเอกภพ (Cardinet and Allal 1983: 28; Cronbach, et al. 1971: 119) ด้วยเหตุนี้เอง M_1 จึงมีค่าความแปรปรวนของความคลาดเคลื่อนสูงกว่า M_2 นอกจากนี้ ยังมีความแปรปรวนของคะแนนเอกภพน้อยกว่า M_2 อีก ลักษณะ เช่นนี้ทำให้ค่าสัมประสิทธิ์การอ้างอิงสรุปของรูปแบบการวัด M_2 มีค่าสูงกว่า M_1 ในทุก ๆ สถานการณ์ที่ทำการทดลอง ดังรูปภาพที่ 4 ดังนี้



รูปภาพที่ 4 แผนภูมิ Venn แสดงแหล่งความแปรปรวนที่มีผลต่อค่าสัมประสิทธิ์การอ้างอิงสรุปของรูปแบบ M_1 และ M_2

ผลการวิจัยยืนยันข้อความรู้ที่ว่า ค่าสัมประสิทธิ์การอ้างอิงสรุปไปยัง เอกภพ เดียวมีค่าสูงกว่า เมื่ออ้างอิงไปยังเอกภพต่าง ๆ ตั้งแต่สองเอกภพพร้อม ๆ กัน ตามที่มึนนักวิจัยคนอื่น ๆ ได้ศึกษาไว้ (Kane, Gillmore and Crooks 1976; Smith 1979; Brennan and Kane 1979; แดง กลางท่าไค่ 2531) ข้อค้นพบนี้ตรงกับที่ Cardinet, et al. (1981) กล่าวไว้ในขั้นการปรับปรุงการวัด (optimization) ว่า เราสามารถเพิ่มความแม่นยำในการวัดโดยการเปลี่ยนพหุเซต I^R เป็น I^F (Cardinet and Allal 1981:199) นั่นคือเราสามารถเพิ่มค่าสัมประสิทธิ์การสรุปอ้างอิง โดยการจำกัดขอบเขตของเอกภพที่ต้องการอ้างอิง โดยกำหนดให้พหุเซตเครื่องมืออย่างใดอย่างหนึ่ง เป็นพหุเซตคงที่ ทั้งนี้ต้องให้มีพหุเซตเครื่องมือที่เป็นพหุเซตสุ่มอยู่อย่างน้อย หนึ่งพหุเซตด้วย และนอกจากนี้รูปแบบการวัดที่ปรับใหม่ จะต้องมีความ

หมายต่อการวัดผลด้วย จากการวิจัยนี้ ค่าสัมประสิทธิ์การอ้างอิงสรุปตามรูปแบบการวัด M 2 มีความหมายเช่นเดียวกับค่าความเที่ยงแบบ interrater reliability ของทฤษฎีการวัดแบบดั้งเดิม ซึ่งถือเป็นดัชนีความสอดคล้องกันระหว่างผู้ตรวจ

2.2 ความสัมพันธ์ระหว่างวิธีตรวจในแต่ละสถานการณ์การทดลอง จำนวนข้อสอบ จำนวนผู้ตรวจ และค่าสัมประสิทธิ์ การอ้างอิงสรุป

ตามรูปแบบการวัด M₁ ซึ่งอ้างอิงสรุปไปยังเอกภพของข้อสอบและผู้ตรวจพร้อม ๆ กันนั้น เมื่อทดสอบความแตกต่างระหว่างค่าสัมประสิทธิ์การอ้างอิงสรุปจากการสุ่มข้อสอบ 5 ข้อ และสุ่มผู้ตรวจ 10 คน พบว่า ค่าสัมประสิทธิ์การอ้างอิงสรุป ทั้ง 6 ค่า จากการตรวจ 2 วิธี ภายใต้สถานการณ์การตรวจ 3 สถานการณ์ ไม่พบความแตกต่างทางสถิติ แสดงให้เห็นว่าตามรูปแบบการวัด M₁ นี้ วิธีตรวจประเมินรวม หรือวิเคราะห์ก็ไม่ให้ผลแตกต่างกัน นอกจากนั้น การที่ผู้ตรวจได้รับการชี้แจงจากผู้วิจัยหรือไม่ได้รับการชี้แจง รับรู้หรือไม่รู้ผลการเรียนก็ตาม ไม่ทำให้ค่าสัมประสิทธิ์การอ้างอิงสรุปแตกต่างกันแต่อย่างใด ผลการวิจัยนี้ขัดแย้งกับสมมุติฐานที่ผู้วิจัยคาดหมายเอาไว้ ที่เป็นเช่นนั้นน่าจะเป็นเพราะว่า ในการประมาณค่าสัมประสิทธิ์การอ้างอิงสรุปในการวิจัยครั้งนี้ ใช้การตัดสินใจแบบสมบูรณ์ ซึ่งถือว่าแหล่งความแปรปรวนอื่น ๆ นอกเหนือจากความแปรปรวนของสิ่งที่ถูกวัด ซึ่งในการวิจัยครั้งนี้คือความแปรปรวนของผู้ตอบ $\sigma^2(p)$ เป็นความแปรปรวนของความคลาดเคลื่อนทั้งสิ้น ประกอบด้วย $\sigma^2(i)$, $\sigma^2(r)$, $\sigma^2(pi)$, $\sigma^2(pr)$, $\sigma^2(ir)$ และ $\sigma^2(pir)$ ความแปรปรวนของทั้ง 6 แหล่งนี้ถ้าพิจารณาจากผลการวิเคราะห์ความแปรปรวนจะพบว่า สถานการณ์การตรวจไม่ทำให้ $\sigma^2(i)$ เปลี่ยนแปลงมากนัก คือไม่ว่าจะใช้ประสบการณ์เดิม หรือตรวจตามวิธีที่ได้รับการชี้แจงจากผู้วิจัย โดยที่รู้หรือไม่รู้ผลการเรียนก็ตาม $\sigma^2(i)$ จะมีค่าไม่เกินร้อยละ 1 ของความแปรปรวนรวมทั้งหมด ดังนั้นค่าสัมประสิทธิ์การอ้างอิงสรุปจึงขึ้นอยู่กับ ความแปรปรวนของ 5 แหล่งที่เหลือ ซึ่งจะเห็นว่า การตรวจโดยใช้ประสบการณ์เดิม แม้ว่า $\sigma^2(r)$ จะค่อนข้างสูง แต่ความแปรปรวนของผลรวมอื่น ๆ ที่เหลือนั้นมีค่าน้อยมาก แสดงว่าแม้ผู้ตรวจมีความเห็นแตกต่างกันก็จริง แต่มีความสอดคล้องกันในการตรวจผู้ตอบแต่ละคน ข้อสอบแต่ละข้อ จึงทำให้ผลรวมของผู้ตอบกับข้อสอบมีค่าน้อย ซึ่งหมายความว่า ผู้ตอบได้คะแนนในแต่ละข้อสอดคล้องกัน ส่วนการตรวจที่ผู้ตรวจ ตรวจตามได้รับการชี้แจงจากผู้วิจัยโดยไม่รู้ผลการเรียน แม้ว่าค่า $\sigma^2(r)$ จะลดลงจากการที่ผู้ตรวจมีเกณฑ์การตรวจที่ชัดเจนก็จริง แต่กลับปรากฏว่าค่า $\sigma^2(pi)$, $\sigma^2(pr)$, $\sigma^2(ir)$ มีค่าเพิ่มขึ้น ทั้งนี้เนื่องจาก วิธีการตรวจที่ผู้ตรวจ

ตรวจตามได้รับการชี้แจงจากผู้วิจัย เน้นให้ตรวจทีละข้อจนครบทุกคนจึงเริ่มตรวจข้ออื่น และตรวจ โดยไม่ต้องย้อนกลับไปดูคะแนนข้อที่ตรวจผ่านมา วิธีปฏิบัติเช่นนี้ทำให้อิทธิพลของผลรวมที่มีข้อสอบ รวมอยู่ ทั้ง pi และ ir มีค่าสูงขึ้น ซึ่งสะท้อนให้เห็นว่าผู้ตอบที่ได้คะแนนสูงในข้อหนึ่งแต่ได้ คะแนนในอีกข้อหนึ่งต่ำ (ดูตาราง 8) ด้วยเหตุนี้ การตรวจโดยใช้ประสบการณ์เดิมหรือการ ตรวจที่ผู้ตรวจ ตรวจตามได้รับการชี้แจงจากผู้วิจัย จึงไม่สามารถลดความคลาดเคลื่อนลงได้โดย เด็ดขาด วิธีการตรวจและสถานการณ์ในการตรวจที่ต่างกัน จึงมีผลเพียงการเพิ่มและลดความ คลาดเคลื่อนต่างแหล่งเท่านั้นเอง ทำให้อัตราส่วนความแปรปรวน หรือค่าสัมประสิทธิ์การอ้างอิง สรุปรูปไม่แตกต่างกัน

อย่างไรก็ตามการที่ค่าสัมประสิทธิ์การอ้างอิงสรุปรูปของรูปแบบการวัด $M_1(p/-/-/I,R^*)$ ของวิธีการตรวจในแต่ละสถานการณ์ไม่แตกต่างกันดังที่เสนอมานั้น เป็นเพียงการสอบ ที่ลุ่มจำนวน ข้อสอบ 5 ข้อ และผู้ตรวจ 10 คนเท่านั้น ตามทฤษฎีการอ้างอิงสรุปรูป ค่าสัมประสิทธิ์สหสัมพันธ์ของ รูปแบบการวัด M_1 จะขึ้นอยู่กับจำนวนเงื่อนไขการวัดของพาเซตข้อสอบและผู้ตรวจ ซึ่งไม่สามารถ จะทำนายความสัมพันธ์ระหว่างค่าสัมประสิทธิ์การอ้างอิงสรุปรูปกับการเพิ่มขนาดของข้อสอบและผู้ตรวจ โดยสูตรของ Spearman-Brown ได้ (Cronbach, et al. 1970: 171) กรณี $pxIxR$ จะ เห็นว่าการเพิ่มข้อสอบจะลดความแปรปรวนของ I, pI, IR และ PIR เท่านั้นแต่ไม่ลดความ แปรปรวนของ R และ pR เช่นเดียวกัน การเพิ่มผู้ตรวจก็จะลดความแปรปรวนของ R, IR, pR และ PIR แต่จะไม่ลดความแปรปรวนของ I และ PI ดังนั้นการลุ่มข้อสอบและผู้ตรวจที่ระดับต่าง ๆ จะให้ค่าสัมประสิทธิ์การอ้างอิงสรุปรูปเปลี่ยนแปลงไป ตัวอย่าง จากตารางที่ 15 จะพบว่าเมื่อลุ่ม ข้อสอบ 6 ข้อ ผู้ตรวจ 1 คน ค่าสัมประสิทธิ์การอ้างอิงสรุปรูปของการตรวจวิธีวิเคราะห์ ที่ผู้ตรวจ ตรวจ ตรวจตามที่ได้รับการชี้แจงจากผู้วิจัยและ ไม่รู้ผลการเรียนของผู้ตอบ (A2) และ การตรวจ วิธีวิเคราะห์ ที่ผู้ตรวจ ตรวจโดยใช้ประสบการณ์เดิม (A1) E_{ρ^2} มีค่า 0.2823 และ 0.1238 ตามลำดับ ถ้าเราเพิ่มผู้ตรวจอย่างเดียวเป็น 20 คน และใช้สูตร Spearman-Brown จะพบว่า ค่าความเที่ยงใหม่ของการตรวจแต่ละวิธีดังนี้

$$\begin{aligned} \text{วิธีตรวจ A2 } r_{tt} &= (20 \times 0.2823) / \{1 + (19 \times 0.2823)\} \\ &= 0.8872 \end{aligned}$$

$$\begin{aligned} \text{วิธีตรวจ A1 } r_{tt} &= (20 \times 0.1238) / \{1 + (19 \times 0.1238)\} \\ &= 0.7386 \end{aligned}$$

แต่ E_p^2 ตามทฤษฎีการอ้างอิงสรุป (จากตารางที่ 15) ของวิธีตรวจ A2 และ A1 มีค่าเป็น 0.6955 และ 0.7228 ซึ่งค่าความของทฤษฎีทั้งสองให้ค่าแตกต่างกัน ด้วยเหตุนี้ผลการเปรียบเทียบค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบถามเรียงที่สุ่มข้อสอบ 5 ข้อ และผู้ตรวจ 10 คน ภายใต้สถานการณ์การทดลองทั้งสามสถานการณ์ ซึ่งพบว่ามีความไม่แตกต่างกันอย่างมีนัยสำคัญทางสถิติ จึงไม่ได้หมายความว่าเมื่อสุ่มข้อสอบและผู้ตรวจที่ระดับอื่น ๆ จะไม่แตกต่างกันด้วย จากกราฟแสดงความสัมพันธ์ระหว่าง จำนวนข้อสอบ จำนวนผู้ตรวจ และค่าสัมประสิทธิ์การอ้างอิงสรุป (แสดงในภาคผนวก ค) เมื่อกำหนดให้ผลคูณของจำนวนข้อสอบกับจำนวนผู้ตรวจเป็น 50 จะเห็นว่า ผลการเปรียบเทียบค่าสัมประสิทธิ์การอ้างอิงสรุปของการตรวจ ทั้ง 6 ภายใต้สถานการณ์การตรวจ 3 สถานการณ์มีค่าแปรเปลี่ยนตามจำนวนข้อสอบ (n_i) และจำนวนผู้ตรวจ (n_r) เมื่อการสอบใช้ผู้ตรวจ 50 คน และใช้ข้อสอบเพียง 1 ข้อ การตรวจวิธีวิเคราะห์ ที่ผู้ตรวจ ตรวจโดยใช้ประสบการณ์เดิม (A1) จะมีความแม่นยำในการอ้างอิงสรุปสูงสุด และการตรวจวิธีวิเคราะห์ ที่ผู้ตรวจ ตรวจตามที่ได้รับแจ้งจากผู้วิจัยและรู้ผลการเรียนของผู้ตอบ (A3) มีความแม่นยำต่ำที่สุด และที่จำนวนผู้ตรวจและข้อสอบระดับนี้ เป็นตำแหน่งที่มีความแตกต่างของค่าสัมประสิทธิ์การอ้างอิงสรุปสูงสุด เมื่อลดจำนวนผู้ตรวจเป็น 25 คนและเพิ่มข้อสอบเป็น 2 ข้อ วิธีตรวจ A1 ยังคงมีความแม่นยำในการอ้างอิงสรุปสูงสุด และ วิธีตรวจ A3 จะยังมีความแม่นยำน้อยที่สุดเช่นเดิม แต่ช่วงความแตกต่างจะแคบเข้า เมื่อลดจำนวนผู้ตรวจตั้งแต่ 10 คนลงไป และเพิ่มข้อสอบตั้งแต่ 5 ข้อขึ้นไป วิธีตรวจวิเคราะห์ ที่ผู้ตรวจ ตรวจตามที่ได้รับแจ้งจากผู้วิจัยและไม่รู้ผลการเรียนของผู้ตอบ (A2) จะมีความแม่นยำในการวัดเหนือกว่าวิธีอื่น ๆ แสดงว่าความแตกต่างของค่าสัมประสิทธิ์การอ้างอิงสรุปของการสุ่มข้อสอบและผู้ตรวจในแต่ละคู่ มีรูปแบบไม่คงที่ ทั้งนี้เป็นเพราะความแปรปรวนของความคลาดเคลื่อน อันเนื่องมาจากข้อสอบและผู้ตรวจ ของการตรวจแต่ละสถานการณ์ มีค่าแตกต่างกัน วิธีตรวจ A1 และ H1 มีความแปรปรวนของผู้ตรวจสูงกว่าวิธีตรวจอื่น ๆ ดังนั้นการสุ่มผู้ตรวจมาก ๆ จึงเพิ่มความแม่นยำในการสรุปอ้างอิงสูงกว่าวิธีอื่น วิธี A2 และ H2 เน้นการตรวจทีละข้อทำให้ผลของข้อสอบ ทั้งผลหลักและผลร่วมกับตัวประกอบอื่น ๆ มีค่าสูงกว่าการตรวจวิธีอื่น ๆ ดังนั้นการเพิ่มข้อสอบแต่ละผู้ตรวจ จึงทำให้วิธีการทั้งสองมีค่าความแม่นยำในการอ้างอิงสรุปสูงกว่าการตรวจตามสถานการณ์อื่น ๆ Cardinet, et al. (1981: 119) กล่าวว่า การสุ่มเงื่อนไขการวัดของพาเซ็คการอ้างอิงสรุปที่มีความแปรปรวนสูง ๆ จะทำให้ค่าสัมประสิทธิ์การอ้างอิงสรุปสูงขึ้นโดยอัตโนมัติ

เนื่องจากรูปแบบการวัดที่อ้างอิงไปยัง เอกภพองฟาเซ็คส์ตั้งแต่ 2 ฟาเซ็คส์เช่นนี้ การ
 สุ่มขนาดของข้อสอบและผู้ตรวจ สามารถจะพิจารณาได้มากมายหลาย combination ครูหรือผู้
 เกี่ยวข้องกับการประเมินจะต้องพิจารณาคู่ที่สามารถปฏิบัติได้ สะดวกและประหยัด โดยทั่วไปใน
 การสอบแบบความเรียง เราสามารถเพิ่มข้อสอบได้มากกว่าเพิ่มผู้ตรวจ โดยเฉพาะการใช้ครู
 มากกว่า 5 คนเป็นสิ่งที่ทำได้ยากสำหรับการประเมินในโรงเรียน แต่การเพิ่มข้อสอบมากกว่า 5
 ข้อสามารถทำได้ง่าย วิธีตรวจวิเคราะห์ตามที่ได้รับการชี้แจงจากผู้วิจัยและตรวจโดยหลีกเลี่ยง
 การรู้ผลการเรียนของผู้ตอบ จึงเป็นวิธีที่เหมาะสมทั้งในแง่การปฏิบัติและความแม่นยำในการวัด
 แต่ถ้าวัดที่สามารถจัดหาผู้ตรวจได้มากกว่า 10 คน โดยที่ใช้ข้อสอบไม่เกิน 5 ข้อ วิธี
 ตรวจวิเคราะห์ที่ผู้ตรวจแต่ละคน กำหนดเกณฑ์การให้คะแนนของตนเอง จะให้ค่าสัมประสิทธิ์การ
 อ้างอิงสรุปเหนือกว่าวิธีอื่น

ผลการเปรียบเทียบค่าสัมประสิทธิ์การอ้างอิงสรุปของรูปแบบการวัด M₂ ซึ่งอ้างอิงไป
 ยัง เอกภพองฟาเซ็คส์เพียงอย่างเดียว เป็นรูปแบบที่ดีความหมายค่าสัมประสิทธิ์การอ้างอิงสรุปเช่น
 เดียวกับการหาความเที่ยงระหว่างผู้ตรวจ (interrater reliability) ของทฤษฎีการวัดแบบ
 ดั้งเดิม ผลการวิเคราะห์สามารถอภิปรายได้ดังนี้

ก. การเปรียบเทียบระหว่างวิธีตรวจประเมินรวมกับวิธีตรวจวิเคราะห์

ผลการเปรียบเทียบระหว่างวิธีตรวจประเมินรวมกับวิธีตรวจวิเคราะห์ พบว่า
 การตรวจทั้งสองวิธี ให้ค่าสัมประสิทธิ์การอ้างอิงสรุปไม่แตกต่างกัน ชัดแย้งกับสมมุติฐานการวิจัย
 ที่คาดหมายไว้แต่ต้นว่า ค่าสัมประสิทธิ์การอ้างอิงสรุปของการตรวจวิธีวิเคราะห์ น่าจะสูงกว่าการ
 ตรวจวิธีประเมินรวม ทั้งนี้ น่าจะมาจากความยืดหยุ่นของผลการเรียนที่วัด มีลักษณะก้ำกึ่งกัน
 ระหว่างความรู้ตามเนื้อหา กับความสามารถในการเขียน โดยปกตินักวัดผลชี้ว่า ถ้าสิ่งที่ถามเป็น
 เนื้อหาวิชาที่มีกฎเกณฑ์ความถูกต้องตายตัว เช่น วิชาประวัติศาสตร์ เคมี การตรวจด้วยวิธี
 วิเคราะห์จะให้ความเที่ยงสูง (Coffman 1971 :291) เช่น ผลการศึกษาของ Swineford
 (cited by Coffman 1971 :293) ที่ศึกษาผลการตรวจวิชาประวัติศาสตร์อเมริกา พบว่า
 ความเที่ยงเป็น 0.95 วิชาเคมี มีค่าเป็น 0.98 เป็นต้น แต่ถ้าเป็นการวัดความสามารถด้าน
 การเขียนซึ่งวัดความสามารถด้านความคิดเห็นและความสามารถเชิงภาษา วิธีตรวจประเมินรวม
 จะให้ความเที่ยงสูงดังเช่น การศึกษาผลการตรวจเรียงความ 3 เรื่องของ Gosling (1966
 cited by Coffman 1971:279) ซึ่งพบความเที่ยงสูงถึง 0.96-0.98 ถ้าพิจารณาผลการ

เรียนที่มุ่งวัดในการวิจัยครั้งนี้ซึ่ง ได้แก่ ความสามารถในการแสดงความคิดเห็นเกี่ยวกับข่าวและเหตุการณ์" เมื่อนำเสนอต่อคนอื่นในรูปการเขียนจะมีธรรมชาติคล้ายกับการเขียนเรียงความทั่วไป ทำให้ข้อดีของความสามารถในการเขียนเข้ามามีส่วนในการให้คะแนนมากขึ้น ยิ่งผู้ตอบมีอิสระในการแสดงความคิดเห็นได้หลากหลาย ก็ยิ่งทำให้วิธีตรวจวิเคราะห์มีผลใกล้เคียงกับการตรวจวิธีประเมินรวม Mehrens and Lehmann (1972: 231) กล่าวว่า แม้ว่าวิธีการตรวจวิธีวิเคราะห์จะเป็นวิธีที่พยายามควบคุมตัวแปรภายนอกที่ไม่เกี่ยวข้องกับผลการวัดก็ตาม แต่ในทางปฏิบัติ การควบคุมความแตกต่างของมนุษย์ให้มีความเห็นตรงกันเป็นเรื่องยากที่จะทำได้สำเร็จ ผลการวิจัยนี้ยังสอดคล้องกับผลการศึกษาของ (Coffman and Kurfman 1968) ที่พบว่า การตรวจวิธีวิเคราะห์และวิธีประเมินรวมมีความเที่ยงไม่แตกต่างกัน นอกจากนี้ยังสอดคล้องกับผลการศึกษาของ Bauer (1982 cited by Purves and Takala 1982: 309) ที่พบว่าวิธีการตรวจวิเคราะห์ หรือ วิธีประเมินรวมมีความเที่ยงพอ ๆ กัน เพียงแต่การตรวจแต่ละวิธีให้สารสนเทศในรายละเอียดแตกต่างกันเท่านั้น

อย่างไรก็ตาม หากพิจารณาผลการเปรียบเทียบในแง่ปฏิบัติ โดยการเปรียบเทียบจำนวนข้อสอบและ จำนวนผู้ตรวจที่ต้องการในแต่ละวิธี หากต้องการให้ค่าสัมประสิทธิ์การอ้างอิงสรุปมีค่าพอ ๆ กัน โดยใช้อัตราส่วน Signal - Noise (แสดงในภาคผนวก ค) จะเห็นว่า การตรวจวิธีวิเคราะห์ มีประสิทธิภาพสูงกว่าการตรวจวิธีประเมินรวม

ข. การเปรียบเทียบระหว่างการตรวจตามที่ได้รับการชี้แจงและผู้ตรวจไม่รู้ผลการเรียนของผู้ตอบกับการตรวจโดยใช้ประสบการณ์เดิม

ค่าสัมประสิทธิ์การอ้างอิงสรุปของการตรวจ ที่ผู้ตรวจ ตรวจตามได้รับการชี้แจง จากผู้วิจัยและ ไม่รู้ผลการเรียนของผู้ตอบ มีค่าสูงกว่าการตรวจตามประสบการณ์เดิม สอดคล้องกับสมมุติฐานการวิจัย รวมทั้งข้อเสนอแนะของนักวัดผลที่ว่า การฝึกอบรมและชี้แจงผู้ตรวจ การใช้เกณฑ์การตรวจที่กำหนดไว้ล่วงหน้าอย่างชัดเจน สามารถทำให้ผลการตรวจมีความเที่ยงสูงขึ้นได้ (Coffman 1971 : 285-296; Ebel and Frisbie 1981: 131; Bergman 1981: 130-131; Kubiszyn and Borich 1981 :100-101) และยังสอดคล้องของกับผลการวิจัยของ Gilbert (1981) ที่พบว่า การฝึกอบรมผู้ตรวจในการตรวจคำตอบความเรียง ประเภทวัดความสามารถในการเขียน ทำให้การตรวจมีความเที่ยงเพิ่มขึ้น อย่างไรก็ตาม มีผลการเปรียบเทียบอยู่หนึ่งการเปรียบเทียบที่พบว่า การตรวจตามที่ได้รับการชี้แจงกับการตรวจโดยใช้ประสบการณ์เดิม

มีค่าสัมประสิทธิ์การอ้างอิงสรุปไม่ต่างกัน ได้แก่ การตรวจวิธีประเมินรวมที่ผู้ตรวจ ตรวจสอบได้ รับการชี้แจงจากผู้วิจัย ก็กับการตรวจวิธีวิเคราะห์โดยใช้ประสบการณ์เดิม ที่เป็นเช่นนี้อาจเนื่องมา จาก วิธีการทั้งสองนั้นมีเกณฑ์ในการตรวจไม่ชัดเจนพอ ๆ กัน กล่าวคือ การตรวจประเมินรวม ที่ผู้ตรวจ ตรวจสอบได้รับการชี้แจงจากผู้วิจัย ใช้คำตอบของนักเรียนเป็นตัวแบบจึงไม่ใช่ตัวเฉลี่ยที่ สมบูรณ์ ประกอบกับประเภทของคำถามได้ให้อิสระในการตอบมาก ทำให้คำตอบนักเรียนแต่ละคน เบี่ยงเบนจากคำตอบตัวแบบ ทำให้ผู้ตรวจต้องใช้ความเห็นของตนเองมากขึ้น คำถามยิ่งให้อิสระ ในการตอบมากเท่าใดความเที่ยงในการตรวจจะยิ่งต่ำมากตามไปด้วย (Coffman 1971: 277) อีกสาเหตุหนึ่งน่าจะมาจาก การตรวจวิธีวิเคราะห์โดยใช้ประสบการณ์เดิมนั้น ผู้ตรวจใช้เกณฑ์ทาง ด้านภาษาซึ่งถึงแม้ไม่ใช่จุดประสงค์ที่ต้องการ แต่ผู้ตรวจก็ใช้เป็นเกณฑ์ตัดสินคุณภาพคำตอบอย่าง สอดคล้องกัน จากการให้ผู้ตรวจระบุเกณฑ์ที่ตนใช้พิจารณาให้คะแนน พบว่า มีความเห็นตรงกันใน เรื่องของ ความสะอาด การสะกดถูกต้อง ลายมืออ่านง่าย การที่ผู้ตรวจใช้เกณฑ์คล้ายกันนี้เองทำ ให้ดัชนีความสอดคล้องในการตรวจมีค่าสูงขึ้น

ค. การเปรียบเทียบระหว่างการตรวจที่ผู้ตรวจรู้ผลการเรียนกับการไม่รู้ผลการเรียน ของผู้ตอบ

ผลการวิจัยพบว่าวิธีตรวจ ที่ผู้ตรวจ ตรวจสอบที่ได้รับการชี้แจงจากผู้วิจัยจะให้ค่า สัมประสิทธิ์ในการอ้างอิงสรุป หรือความเที่ยงสูงกว่าการตรวจโดยใช้ประสบการณ์เดิม ก็ต่อเมื่อ เป็นการตรวจภายใต้สถานการณ์ ที่ผู้ตรวจไม่รู้ผลการเรียนของผู้ตอบเท่านั้น ถ้าผู้ตรวจที่ได้รับการ ชี้แจงจากผู้วิจัยรู้ผลการเรียนของผู้ตอบแล้ว ค่าสัมประสิทธิ์การอ้างอิงสรุปของการตรวจใน สถานการณ์ทั้งสองจะ ไม่แตกต่างกัน นอกจากนั้นค่าสัมประสิทธิ์การอ้างอิงสรุปของการตรวจใน สถานการณ์ที่ผู้ตรวจ ตรวจสอบที่ได้รับการชี้แจงจากผู้วิจัยและ ไม่รู้ผลการเรียนของผู้ตอบ มีค่าสูง กว่าการตรวจที่ได้รับการชี้แจงจากผู้วิจัยและรู้ผลการเรียนของผู้ตอบ สอดคล้องกับสมมุติฐานของ การวิจัย ที่เป็นเช่นนี้น่าจะเป็นผลมาจาก ความคลาดเคลื่อนประเภท halo effect จากความ คาดหวังที่รู้ผลการเรียนของผู้ตอบนั่นเอง สอดคล้องกับผลการศึกษาของ Chase (1979) Hughes, et al. (1983) ที่ได้ศึกษาโดยใช้เกณฑ์การตรวจที่กำหนดไว้ล่วงหน้าทั้งวิธีตรวจ ประเมินรวมและวิธีวิเคราะห์เพื่อขจัดอิทธิพลของความคาดหวัง แต่ก็ยังพบว่า อิทธิพลของความ คาดหวังมีผลต่อการตรวจคำตอบความเรียงอย่างมีนัยสำคัญทางสถิติอยู่นั่นเอง

ข้อเสนอแนะ

จากผลการวิจัยที่นำเสนอมาโดยลำดับ ผู้วิจัยใคร่เสนอแนะดังนี้

1. การนำผลการวิจัยไปใช้

1. ในการตรวจแบบความเรียง (Essay Examination) ควรจะมีการฝึกอบรมผู้ตรวจให้เข้าใจตรงกัน ทั้งในหลักการตรวจทั่วไป เนื้อหาที่มุ่งวัดการใช้คู่มือการให้คะแนน
2. ในการตรวจให้คะแนนแต่ละครั้งต้องหลีกเลี่ยงการรู้จักนักเรียนเจ้าของคำตอบ เพราะจะทำให้เกิดความคลาดเคลื่อนในการประเมินได้ มีวิธีหลีกเลี่ยงได้หลายวิธี เช่น ให้นักเรียนเขียนชื่อไว้ด้านหลังแผ่นสุดท้ายของกระดาษคำตอบ หรือให้เขียนชื่อในรูปของรหัสอื่น เป็นต้น
3. ในการอ้างอิงผลการวัดไปยัง เอกภพของข้อสอบและผู้ตรวจพร้อม ๆ กัน เมื่อใช้ผู้ตรวจไม่เกิน 5 คน ควรใช้ข้อสอบแบบความเรียงอย่างน้อย 6 ข้อ และใช้วิธีตรวจแบบวิเคราะห์ มีเกณฑ์การตรวจที่ชัดเจน เพราะการตรวจวิธีนี้จะให้ค่าความเที่ยงอย่างน้อย 0.50 โดยประมาณ แต่ถ้าสามารถหาผู้ตรวจได้เกิน 10 คน ในการประเมินแต่ละครั้ง สามารถใช้การตรวจวิธีวิเคราะห์โดยใช้ประสบการณ์เดิมแทนได้
5. ในการอ้างอิงผลการวัดไปยัง เอกภพของผู้ตรวจอย่างเดียว เมื่อการสอบแต่ละครั้งใช้ข้อสอบ 5 ข้อ ควรมีการชี้แจงหรือฝึกอบรมผู้ตรวจ พร้อมทั้งกำหนดเกณฑ์การตรวจไว้ล่วงหน้าให้ชัดเจน หากตรวจวิธีวิเคราะห์ ควรใช้ผู้ตรวจอย่างน้อย 3 คน ถ้าตรวจวิธีประเมินรวม ควรใช้ผู้ตรวจอย่างน้อย 5 คน แต่ถ้าเป็นการสอบวัดที่ต้องการความแม่นยำในการวัดสูง ๆ เช่น การสอบคัดเลือก เป็นต้น ควรใช้วิธีตรวจวิเคราะห์และใช้ผู้ตรวจอย่างน้อย 8 คน จึงจะให้ค่าสัมประสิทธิ์การอ้างอิงสูงกว่า .70
6. ในการอ้างอิงสรุปผลการวัดควรให้สัมพันธ์กับขนาดของเอกภพการอ้างอิง เช่น ถ้าเป็นการประเมินผลภายในกลุ่มโรงเรียน ครูที่อยู่ในเอกภพของการอ้างอิงควรได้แก่ครูที่สอนอยู่ภายในกลุ่มโรงเรียนเท่านั้น แต่ถ้าเป็นการประเมินผลเทียบกันภายในอำเภอ เอกภพของผู้ตรวจก็ควรจะหมายถึงครูที่สอนอยู่ภายในอำเภอนั้น ขนาดของเอกภพมีผลต่อการอ้างอิงสรุป ถ้าเอกภพมีขนาดเล็กค่าสัมประสิทธิ์การอ้างอิงสรุปจะมีค่าสูงกว่าเอกภพที่มีขนาดใหญ่
7. ในการเลือกรูปแบบการวัดว่าจะอ้างอิงไปยัง เอกภพของข้อสอบและผู้ตรวจ

พร้อม ๆ กัน หรือจะเลือกรูปแบบการวัดที่อ้างอิงไปยังผู้ตรวจเพียงอย่างเดียว ครูหรือผู้ประเมิน หรือผู้ใช้ผลวิจัยจะต้องพิจารณาว่า ในการสอบครั้งต่อ ๆ ไป กับนักเรียนในประชากรนั้น ๆ จะต้องพิจารณาว่า จะเลือกแบบสอบชุดเดิมหรือไม่ ถ้าใช้แบบสอบเดิมถือว่ารูปแบบที่เลือกเป็นแบบอ้างอิงไปยังเอกภพของผู้ตรวจเพียงอย่างเดียว แต่ถ้าจะใช้แบบสอบชุดใหม่ แสดงว่าเลือกรูปแบบการวัดที่ต้องการอ้างอิงไปยังทั้งสองเอกภพ


2. การวิจัยในครั้งต่อไป

1. การศึกษาในลักษณะ เดิมแต่ออกแบบการทดลองเป็นแบบ 2 X 3 Factorial Design คือการตรวจ 2 วิธี สถานการณ์การตรวจ 3 สถานการณ์ ให้ผู้ตรวจแต่ละคนรับ combination ของการทดลองเพียง combination เดียว เพื่อลดอิทธิพลของความคลาดเคลื่อนต่าง ๆ
2. การศึกษาค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียงประเภทวัดความสามารถในทางการเขียน หรือที่เรียกว่า เรียงความ หรือความเรียงวัดผลสัมฤทธิ์ตามเนื้อหาวิชาอื่น
3. การศึกษาในประชากรอื่น ๆ เช่น นักเรียนและครูในระดับมัธยม หรือระดับอุดมศึกษา
4. การวิจัยในลักษณะ เดิมแต่ขยายเอกภพของการสังเกตที่ยอมรับได้ให้ใหญ่ขึ้น เช่น ขยายเป็นระดับจังหวัด หรือระดับภาคภูมิศาสตร์ เป็นต้น
5. การจะศึกษาพาเซตอื่นที่มีผลต่อค่าสัมประสิทธิ์การอ้างอิงสรุป เช่น จำนวนครั้งในการตรวจ (O) โรงเรียน (S) ผู้ตรวจที่แบ่งอยู่ในโรงเรียน (R:S) วิธีดำเนินการสอบ (M) เป็นต้น
6. การวิจัยในลักษณะ เดิมแต่ผนวกวิธีการตรวจ (M) และสถานการณ์ (S) ไว้ในแบบแผนการทดลอง เป็น $pxix(r:m)$ หรือ $pxix(r:m:s)$ โดยไม่ต้องศึกษาเปรียบเทียบค่าสัมประสิทธิ์การอ้างอิงสรุปของวิธีตรวจ หรือ สถานการณ์การตรวจ
7. เปลี่ยนรูปแบบความสัมพันธ์ระหว่างพาเซตที่ศึกษาในขั้นการปรับปรุงรูปแบบการวัด (optimization หรือ ชั้น D-study) เป็นแบบ nested กล่าวคือ จาก $pxIxR$ เป็นแบบ $Px(I:R)$ ซึ่งเป็นรูปแบบการประเมินที่ผู้ตรวจแต่ละคนไม่จำเป็นต้องตรวจข้อสอบทุกข้อ

เหมือนกันหมด จะทำให้สามารถลดภาระและเวลาในการตรวจได้เป็นอย่างดี

8. ศึกษาเปรียบเทียบค่าสัมประสิทธิ์การอ้างอิงสรุปของการวัด ที่ใช้รูปแบบความสัมพันธ์ระหว่างพาเซ็ด แบบ crossed กับแบบ nested ในแง่จำนวนข้อสอบ จำนวนผู้ตรวจ

9. ศึกษาเปรียบเทียบค่าสัมประสิทธิ์การอ้างอิงสรุปในแง่ประสิทธิภาพในทางปฏิบัติ เปรียบเทียบจำนวนข้อสอบ หรือจำนวนผู้ตรวจโดยอาศัย Signal to Noise Ratio แทนการทดสอบความมีนัยสำคัญทางสถิติ



ศูนย์วิทยพัชร์พยากร
จุฬาลงกรณ์มหาวิทยาลัย