



บทที่ 1

บทนำ

ความเป็นมาและแนวคิดเชิงทฤษฎีของปัญหา

แบบสอบความเรียง เป็นเครื่องมือที่มีความสำคัญอย่างยิ่งต่อการวัดและประเมินผลการศึกษา เป็นแบบสอบที่ให้อิสระในการแสดงความคิด ส่งเสริมการจัดระเบียบความคิด การประยุกต์ความรู้กับสถานการณ์ใหม่ ความคิดสร้างสรรค์ สามารถวัดผลการเรียนรู้ตามจุดประสงค์ด้านพุทธิปริเขต (cognitive domain) ระดับสูงของ Bloom (1956) ได้แก่ การวิเคราะห์ การสังเคราะห์ และการประเมินค่าได้ดี (Thorndike and Hagen 1977: 263; Ebel 1972: 145; La Fave 1966: 65-69 cited by Stanley and Hopkins 1978: 198; Mehrens and Lehmann 1972: 213; เขียวดี วิบูลย์ศรี 2528: 122-132; บุญเชิด ภิญโญนันตพงษ์ 2519: 35; ไพศาล หวังพานิช 2528: 89-97) เป็นเครื่องมือที่มีความตรงแบบ face validity โดยเฉพาะอย่างยิ่ง การวัดความสามารถในการเขียน การใช้ภาษา (Hopkins and Stanley 1981: 225)

นอกจากคุณภาพในแง่การวัดผลการเรียนที่ซับซ้อนได้เหมาะสมแล้ว แบบสอบความเรียงยังส่งเสริมการเรียนการสอนให้มีประสิทธิภาพมากยิ่งขึ้น สามารถสร้างและดำเนินการสอบได้ง่าย ไม่สิ้นเปลืองวัสดุอุปกรณ์ในการจัดพิมพ์ ทดสอบได้ทุกเวลาที่ทำการสอน เพียงแต่เขียนข้อความบนกระดาษดำ นักเรียนทั้งห้องสามารถเข้าร่วมกันได้ทันที (Coffman 1971: 272; No11 1965: 133) การสอบด้วยแบบสอบความเรียงถือเป็นการเสริมแรงการเรียนอย่างหนึ่งแก่ผู้เรียน เนื่องจากผู้เรียนเรียนรู้เหตุการณ์ต่าง ๆ ในรูปผลรวม ดังนั้นการสอบความเรียงจึงเป็นการเสริมแรงเปิดโอกาสให้ผู้เรียนได้แสดงความคิดเห็น เรียบเรียง และประยุกต์ความรู้อย่างกว้างขวาง (Bergman 1981: 127) นอกจากนี้ ยังมีผู้พบว่าผู้เรียนที่เรียนโดยใช้แบบสอบความเรียงเป็นเครื่องมือในการวัดผล มีผลสัมฤทธิ์ทางการเรียนสูงกว่าผู้เรียนที่ใช้การวัดผลแบบปรนัย (Mayer 1934 cited by Bergman 1981: 127)

มีนักวัดผลและนักวิจัยหลายคนเสนอว่า แบบสอบปรนัยที่สร้างอย่างดี ตรงตามหลักวิชาการวัดผล สามารถวัดพฤติกรรมระดับสูงของ Bloom รวมทั้งความสามารถในการเขียนได้เช่นกัน (Bracht and Hopkins 1970: 359-364; Brenland & Gaynor 1979: 119-128; Coffman 1966: 151-156; ชาวล แพร์ตกุล 2507: 143-158; กิ่งกาญจน์ สิริสุนทร 2521: 24-40) จากข้อเสนอนี้มีได้หมายความว่าแบบสอบความเรียงไม่มีความสำคัญอีกต่อไป การสร้างแบบสอบปรนัยให้สามารถวัดพฤติกรรมในระดับสูง ไม่ใช่สิ่งที่ทำได้ง่ายนัก โดยเฉพาะครูในระดับประถมและมัธยมศึกษา น้อยคนที่มีความรู้ด้านการวัดผลอย่างดีพอ ทำให้แบบสอบปรนัยที่สร้างไม่สามารถวัดพฤติกรรมที่ต้องการได้ อีกประการหนึ่งการที่นักเรียนตอบข้อสอบปรนัยได้ถูกต้อง ยังมีปัญหาอีกกว่าเป็นเพราะมีความสามารถจริง หรือเพราะความบังเอิญ ปัญหาข้อสอบแบบความเรียงสามารถขจัดได้ (Green 1963: 5) ที่สำคัญที่สุด การใช้แบบสอบปรนัยเพียงอย่างเดียวในทุก ๆ สถานการณ์ อาจก่อให้เกิดผลกระทบทางลบต่อการจัดการศึกษาได้ (สมหวัง พิธิยานุวัฒน์ 2531: 69-70; อุทุมพร จามรมาน 2531: 71-74; เอกวิทย์ ณ กลาง 2529: 511; ประกอบคุปรัตน์ 2531: 1-6) Coffman (1971: 286) กล่าวว่าแบบสอบความเรียงควรใช้ในการประเมินผลระหว่างการเรียนการสอนให้มาก เพราะนอกจากจะเปิดโอกาสให้นักเรียนได้ฝึกฝนทักษะการเขียนและบูรณาการความรู้ที่ได้เรียนมาแล้ว สถานการณ์เช่นนี้มีผู้สอบไม่มาก การสร้างข้อสอบปรนัยที่ดีจะเสียเวลาพอ ๆ กับการตรวจความเรียงซึ่งใช้เวลาสร้างน้อยกว่า

แม้ว่าแบบสอบความเรียงจะมีจุดเด่นหลายประการดังกล่าว แต่แบบสอบความเรียงมีจุดอ่อนหรือปัญหาในด้านการวัดผลที่สำคัญอยู่ 2 ประการคือ ปัญหาความเที่ยงกับปัญหาการสิ้นเปลืองเวลาและแรงงานในการตรวจ แบบสอบความเรียงมักจะมีค่าความเที่ยงต่ำ จากการตรวจสอบเอกสารงานวิจัย Coffman (1971: 277) พบว่า งานวิจัยบางเรื่องมีค่าความเที่ยง .35 เท่านั้น Finlayson (1951: 126-134) ใช้ผู้ตรวจ 4 คน ตรวจคำตอบความเรียง 2 ชุด พบค่าความเที่ยง .687 และ .703 Godshalk, Swineford and Coffman (1966: 39-40) พบว่า ค่าความเที่ยงของข้อสอบความเรียง 1 ข้อ ใช้ผู้ตรวจ 5 คนเป็น .40 ถ้าใช้ 5 ข้อ โดยตรวจคนเดียวจะมีค่า .25 การที่แบบสอบความเรียงมีค่าต่ำ Ebel and Frisbie (1986: 129) ชี้ว่าเกิดจากสาเหตุสำคัญ 3 ประการ คือ (1) ค่าตามมีน้อยไม่ครอบคลุมเนื้อหา

(2) ความไม่จำเพาะเจาะจงของคำถาม และ (3) ความเป็นอัตนัยของการตรวจอันเนื่องมาจากตัวแปรต่าง ๆ ที่ไม่เกี่ยวข้องกับผลการเรียนที่มุ่งวัด มีผลงานวิจัยพบว่าคำตอบความเรียงประเภทวัดความสามารถในการเขียนหรือเรียงความที่มีคุณภาพปานกลาง ถ้าตรวจตามหลังเรียงความที่มีคุณภาพต่างกันจะได้คะแนนต่างกัน กล่าวคือเมื่อตรวจตามหลังเรียงความที่มีคุณภาพต่ำจะได้คะแนนสูงกว่าเมื่อตรวจตามหลังชุดที่มีคุณภาพสูง (Hales and Tokar 1975: 115-117; Hughes, Keeling & Tuck 1980:131-135, 1983: 65-70, 1984: 277-281) นอกจากนี้ ยังมีงานวิจัยอีกมากที่ชี้ให้เห็นถึงอิทธิพลของตัวแปรอื่น ๆ เช่น การที่ผู้ตรวจรู้จักประวัติการเรียนของผู้ตอบ ลายมือ ความเรียบร้อย การสะกดการันต์ ความบกพร่องในด้านไวยากรณ์ เป็นต้น (Chase 1979: 39-42, 1983: 293-297; Scannell and Marshall 1966: 125-130; Marshall 1967: 375-386; Marshall & Power 1969: 97-101)

ปัญหาความเที่ยงต่ำนี้ มีนักวัดผลการศึกษาได้ให้ข้อเสนอแนะวิธีปรับปรุงเพื่อให้ผลการวัดมีความแม่นยำไว้หลายประการ De Gruijter (1980: 221) กล่าวว่า การควบคุมแหล่งความคลาดเคลื่อนไว้ล่วงหน้าจะทำให้ผลการวัดมีความเชื่อถือได้มากขึ้น ในการวัดแบบความเรียงนั้น ตัวประกอบ (Facet) ที่สำคัญจำเป็นต้องควบคุม คือ ข้อสอบ และ ผู้ตรวจ ตัวประกอบทั้งสองนี้มีผลเกี่ยวข้องซึ่งกันและกัน ข้อสอบความเรียงที่ไม่ดี ไม่สามารถยกระดับคุณภาพให้ดีขึ้นโดยอาศัยวิธีตรวจใด ๆ แม้วิธีตรวจนั้นจะกำหนดเกณฑ์การตรวจไว้เด่นชัดเพียงใดก็ตาม ในทางกลับกันข้อสอบที่สร้างอย่างดี แต่ใช้วิธีตรวจไม่เหมาะสมข้อสอบนั้นก็จะได้คุณค่าเช่นกัน เพื่อให้การตรวจมีความเที่ยงเพิ่มขึ้น ผู้ตรวจควรปฏิบัติต่อไปนี้อย่างเคร่งครัด (1) ใช้วิธีการตรวจที่เหมาะสมสามารถจัดอคติได้ (2) การตรวจต้องมุ่งสนใจอยู่เฉพาะประเด็นที่เกี่ยวข้องกับผลการเรียนที่มุ่งวัดเท่านั้น (3) ระมัดระวังอย่าให้มีส่วนตัวของผู้ตรวจมีอิทธิพลต่อคะแนน (4) ต้องใช้เกณฑ์การตรวจกับนักเรียนทุก ๆ คนอย่างคงเส้นคงวา (Mehrens and Lehmann 1972: 228) นอกจากนี้ 4 ข้อดังกล่าว นักวัดผลคนอื่น ๆ ให้ข้อเสนอแนะคล้ายคลึงกันว่าในการตรวจควรตรวจทีละข้อจนครบทุก ๆ คน ปกปิดชื่อของนักเรียนเอาไว้อย่าให้รู้ว่าการตรวจคำตอบของใคร ถ้าทำได้ควรตรวจหลายครั้งหรือตรวจหลายคน (Multiple rating) คือ ให้ผู้ตรวจคนเดียวตรวจหลายครั้ง หรือในการตรวจครั้งหนึ่ง ใช้ผู้ตรวจหลายคน แล้วใช้คะแนนเฉลี่ยหรือคะแนนรวมแทนคะแนนความสามารถของผู้สอบ จะให้ผลการวัดมีความเที่ยงสูงขึ้นได้ (Bergman 1981: 130-131;

Gronlund 1977: 229-232; Ebel and Frisbie 1986: 134-135; Cochran and Weideman 1937 cited by Hopkins and Stanley 1981: 223; Kubiszyn and Borich 1981: 100-101; Linnvall and Nitko 1975: 51-52) นอกจากนั้น การฝึกอบรมผู้ตรวจเป็นอีกวิธีหนึ่งที่ทำให้ผลการวัดมีความเที่ยงเพิ่มขึ้นได้ (Gilbert 1981: 4270)

วิธีตรวจคำตอบความเรียงที่นักวัดผลกล่าวถึงมี 2 วิธี (Coffman 1971: 291-293; Ebel 1972: 49-50; Mehrens and Lehmann 1984: 113-116; Glover, Bruning and Filbeck 1982: 420) คือ ตรวจแยกทีละประเด็น ให้คะแนนสูงสุดในแต่ละประเด็นเท่ากับค่าน้ำหนักคะแนนที่กำหนดให้ เรียกว่า "การตรวจวิธีวิเคราะห์" (Analytic method หรือ Point method) และ วิธีตรวจให้คะแนนรวมเพียงองค์ประกอบเดียว อาศัยความประทับใจจากการอ่านอย่างรวดเร็ว เรียกว่า "การตรวจวิธีประเมินรวม" (Holistic method หรือ Global method) วิธีตรวจแต่ละวิธีเหมาะสมกับสถานการณ์การประเมินและเนื้อหาวิชาแตกต่างกัน วิธีแรก เหมาะกับการตรวจความเรียงประเภทวัดเนื้อหา ในสถานการณ์การประเมินผลการเรียนการสอนซึ่งมีผู้สอบไม่มากนัก จุดอ่อนของวิธีนี้คือต้องเสียเวลาตรวจมากเพราะต้องตรวจทีละประเด็น และผลการตรวจประเด็นต้น ๆ ก่อให้เกิดความประทับใจมีผลส่งถึงประเด็นหลัง ๆ เช่น ประเด็นแรกได้คะแนนสูงทำให้ประเด็นหลัง ๆ ได้คะแนนสูงด้วย ความคลาดเคลื่อนนี้เรียกว่า halo effect (Coffman 1971: 292-293) นอกจากนั้น การแยกประเด็นน้อยมาก ๆ อาจวัดคุณลักษณะที่ไม่ตรงกับจุดมุ่งหมายการเรียนการสอนที่ต้องการได้ (Wiseman 1949 cited by Coffman 1971: 293) การตรวจวิธีที่ 2 สามารถกระทำได้รวดเร็ว เหมาะกับการตรวจความเรียงประเภทวัดความสามารถในการเขียน และในสถานการณ์ที่มีผู้สอบมาก ๆ การตรวจวิธีประเมินรวมมักให้ค่าความเที่ยงต่ำกว่าการตรวจวิธีวิเคราะห์ (De Gruijter 1980: 247) อย่างไรก็ตาม การจัดกลุ่มคุณภาพของคำตอบ (sorting) หรือใช้คำตอบตัวแบบ (essay model) เข้าช่วยจะทำให้ความเที่ยงของการตรวจวิธีประเมินรวมมีค่าสูงขึ้นได้ (Ebel and Frisbie 1986: 134; Mehrens and Lehmann 1972: 232; Linnvall and Nitko (1975: 50) กล่าวว่า การตรวจวิธีที่ 2 เป็นการเปรียบเทียบผู้ตอบหรือนักเรียนกับกลุ่ม แต่วิธีแรกเปรียบเทียบกับเกณฑ์

การเพิ่มข้อสอบความเรียงหรือการเพิ่มผู้ตรวจ เป็นอีกปัญหาหนึ่งที่จะต้องพิจารณาให้รอบคอบทั้งในแง่ทฤษฎีและการปฏิบัติ การสอบแบบความเรียงผู้ตอบต้องใช้เวลาในการเขียนมาก ครูก็ต้องใช้เวลาในการตรวจมากเช่นกัน ในการเพิ่มข้อสอบ หรือผู้ตรวจเพื่อเพิ่มค่าความเที่ยงนั้น ทฤษฎีการวัดแบบดั้งเดิมอธิบายด้วยสูตรของ Spearman-Brown ซึ่งเหมาะกับการวัดที่มีข้อสอบเป็นตัวประกอบสำคัญเพียงตัวเดียว แต่การสอบแบบความเรียงมีตัวประกอบสำคัญเพิ่มมาอีกคือ ผู้ตรวจ การเพิ่มข้อสอบหรือผู้ตรวจจึงไม่สามารถจะใช้สูตร Spearman-Brown ได้ครบทุกรูปแบบของการวัด (Cronbach, et al. 1972: 172) ในทางปฏิบัติการเพิ่มข้อสอบความเรียงแม้เพียง 1 ข้อ ครูต้องสิ้นเปลืองเวลาและแรงงานในการตรวจมาก ทำให้กิจกรรมการวัดและประเมินผลเป็นเรื่องน่าเบื่อหน่าย ในบางครั้งจำนวนครุมน้อยไม่เพียงพอที่ระดับความแม่นยำที่ต้องการ

ปัญหาการตีความค่าความเที่ยง เป็นอีกปัญหาหนึ่งที่มีความสำคัญในการวัดและประเมินผลการศึกษา ทฤษฎีการวัดแบบดั้งเดิม นิยามความเที่ยงในความหมายของคำว่า "ความคงที่" (consistency) กล่าวคือ ความเที่ยงของแบบสอบ หมายถึงความคงที่ของคะแนน ที่ได้จากการสอบคนกลุ่มเดียวกันสองครั้งด้วยแบบสอบฉบับเดิม หรือ สอบคนกลุ่มเดิมด้วยแบบสอบที่มีคุณสมบัติเท่าเทียมกันสองฉบับ (Anatasi 1968: 28; Edell 1972: 409) จากนิยามนี้นำไปสู่ วิธีการประมาณค่าความเที่ยงโดยวิธีการต่าง ๆ เช่น วิธีวัดความคงที่ของคะแนนจากการสอบด้วยแบบสอบฉบับเดียว 2 ครั้ง (measures of stability) วิธีวัดความคงที่ของคะแนนจากการสอบครั้งเดียวแต่ใช้แบบสอบคู่ขนาน 2 ฉบับ (measures of equivalence) วิธีวัดความคงที่ของคะแนนจากการสอบด้วยแบบสอบคู่ขนาน 2 ฉบับ ในเวลาที่ต่างกัน (measures of equivalence and stability) วิธีวัดความคงที่ของคะแนนจากการสอบด้วยแบบสอบฉบับเดียวเพียงครั้งเดียว (measures of internal consistency) (Mehrens and Lehmann 1984: 271-277) วิธีการเหล่านี้จะให้ค่าประมาณที่ไม่ลาเอียงก็ต่อเมื่อแบบสอบ 2 ฉบับ หรือ ส่วนแบ่งของแบบสอบมีคุณสมบัติคู่ขนานกัน (Cronbach, Rajaratnam and Gleser 1963: 137-140) แบบสอบตั้งแต่ 2 ฉบับขึ้นไปจะมีคุณสมบัติความเท่าเทียมกันก็ต่อเมื่อมี ค่าเฉลี่ยเท่ากัน ค่าความแปรปรวนเท่ากัน และ ค่าสหสัมพันธ์ระหว่างแบบสอบเท่ากัน (Gulliksen 1950: 173; Magnusson 1967: 62) หรือในกรณีที่มีเพียง 2 ฉบับซึ่งมีค่าสหสัมพันธ์ระหว่างแบบสอบเพียงค่า

เดียวแบบสอบทั้งสองจะต้องมีค่าสหสัมพันธ์กับเกณฑ์ภายนอกเท่ากัน (Votaw cited by Gulliksen 1950: 173) ในการวัดและประเมินผลโดยทั่วไป มักไม่สามารถจัดหาหรือสร้างเครื่องให้ที่มีคุณสมบัติครบทั้ง 3 ได้อย่างสมบูรณ์ โดยเฉพาะอย่างยิ่ง การวัดผลที่ต้องใช้ความเห็นของผู้ประเมินหรือผู้ตรวจ อย่างเช่นการสอบแบบความเรียง (essay examination) ในสถานการณ์การสอบดังกล่าว ผู้ตรวจแต่ละคนคือเครื่องมือวัด ซึ่งแต่ละคนให้คะแนนแตกต่างกันทั้งค่าเฉลี่ย ความแปรปรวน และ ค่าสหสัมพันธ์ระหว่างกัน เมื่อเป็นเช่นนี้ การใช้ทฤษฎีความเที่ยงแบบดั้งเดิมกับการสอบดังกล่าวจึงไม่เหมาะสม

ทฤษฎีการอ้างอิงสรุป (Generalizability Theory) ซึ่งกลุ่มนักวิจัยโดย Cronbach (Cronbach, et al. 1972) ได้พัฒนาขึ้น และต่อมา Cardinet, Tourneur and Allal (1976, 1981, 1983) ได้ขยายทฤษฎีในบางประเด็น เพื่อให้สามารถใช้อธิบายผลการวัดได้สมบูรณ์มากยิ่งขึ้น ทฤษฎีนี้มุ่งขยายความทฤษฎีความเที่ยงแบบดั้งเดิมให้ชัดเจนยิ่งขึ้น Vann der Kamp (1976: 173-174) กล่าวว่า ทฤษฎีการอ้างอิงสรุปสามารถใช้ในการอธิบายความเที่ยงได้ทุกสถานการณ์ โดยเฉพาะอย่างยิ่งการสอบแบบความเรียง ทั้งนี้เนื่องจากทฤษฎีนี้ไม่กำหนดข้อตกลงเกี่ยวกับคุณสมบัติความเท่าเทียม ผู้ตรวจแต่ละคนไม่จำเป็นต้องมีคุณสมบัติความเท่าเทียม แต่กำหนดมโนทัศน์เกี่ยวกับเอกภพ (Universe or Domain) เป็นกรอบอ้างอิงให้ชัดเจน กล่าวคือ ผู้ตรวจทั้งหมดที่ผู้วัดต้องการจะอ้างอิงถึงหมายถึงใคร มีจำนวนเท่าใด ข้อสอบที่ต้องการนำมาสอบมีขอบเขตเนื้อหากว้างเพียงใด การที่ถามว่า การตรวจความเรียงมีค่าความเที่ยงหรือความคล่องกันระหว่างผู้ตรวจ (rater agreement) เพียงใด หมายถึงถามว่าผู้วัดหรือผู้ประเมินผลสามารถจะอ้างอิงสรุป (generalize) ผลการวัดจากคะแนนที่ได้ชุดหนึ่งไปยังคะแนนจากการตรวจชุดอื่น ๆ ได้ดีเพียงใดนั่นเอง ค่าการวัดแต่ละชุดก็เป็นตัวอย่างหนึ่งของเอกภพของค่าการวัดที่เป็นไปได้ทั้งหมด ค่าการวัดแต่ละชุดเหมาะกับเอกภพที่ต่างกัน ความเที่ยงตามทฤษฎีนี้ หมายถึงความแม่นยำในการอ้างอิงสรุปไปยังกลุ่ม (class) ค่าสังเกตกลุ่มอื่น ๆ ที่เป็นสมาชิกของเอกภพที่ต้องการโดยอาศัยกลุ่มค่าการวัดที่มีอยู่ในมือ ทฤษฎีนี้ถือว่า คะแนนจริงคือคะแนนเอกภพ (Universe Score) ซึ่งเป็นค่าเฉลี่ยของคะแนนสังเกตที่เป็นไปได้ทั้งหมดภายในเอกภพที่ผู้วัดผลสนใจ คะแนนเอกภพของผู้สอบแต่ละคนมิได้หลายค่า เช่นเดียวกับค่าสัมประสิทธิ์การอ้างอิงสรุป ทั้งนี้ขึ้นอยู่กับเอกภพที่ผู้วัดหรือประเมินผลสนใจ เช่น คะแนนเอกภพในการเขียน

เรียงความเรื่อง "โรงเรียนของฉัน" ของนักเรียนชั้นประถมศึกษาปีที่ 6 คนหนึ่งที่ได้รับการตรวจจากครูที่สอนภาษาไทย 5 คน หากผู้วัดผลสนใจและต้องการจะรู้คะแนนความสามารถของนักเรียนคนนี้เฉพาะการเขียนเรียงความเรื่องนี้จากการตรวจของครูเพียง 5 คนที่กล่าวมา คะแนนเอกภพของนักเรียนคนนี้จะแตกต่างจากคะแนนเอกภพที่ผู้วัดสนใจอยากรู้ความสามารถในการเขียนเรียงความเรื่องอื่น ๆ และให้ผู้ตรวจ 5 คนใดๆ ที่สอนภาษาไทยเป็นคนตรวจ คะแนนเอกภพเมื่อสนใจเฉพาะผลการสอบในสัปดาห์ที่ทำการสอบจะแตกต่างจากคะแนนเอกภพที่สนใจการสอบสัปดาห์ใด ๆ ในเดือนที่ทำการสอบ เป็นต้น แต่ทฤษฎีการวัดแบบดั้งเดิมถือว่าคะแนนจริงของผู้สอบแต่ละคนมีเพียงค่าเดียว และความเที่ยงของเครื่องมือที่ใช้วัดมีเพียงค่าเดียวเช่นกัน

การตัดสินใจว่าความเที่ยงของแบบสอบควรเป็นเท่าไรจึงจะสูงถึงระดับที่น่าพอใจ ขึ้นอยู่กับจุดประสงค์ของการใช้แบบสอบ นักวัดผลบางคนกล่าวว่า โดยทั่วไปแบบสอบควรมีค่าความเที่ยงอย่างน้อย 0.80 แต่บางคนแย้งว่าไม่จำเป็นต้องเสียเวลาและงบประมาณในการสร้างเครื่องมือให้มีความเที่ยงสูงขนาดนั้น (Nunnally 1978: 245) Coffman (1971:278) ให้ความเห็นว่า การตัดสินใจว่าความเที่ยงของแบบสอบความเรียงควรเป็นเท่าไร เป็นเรื่องที่ยากมาก ทั้งนี้เนื่องจากระดับความเที่ยงจะเปลี่ยนไปตามปัจจัยที่เกี่ยวข้อง เช่น ลักษณะของผู้สอบ ผู้ตรวจ เนื้อหาวิชาที่วัด เป็นต้น ในการประเมินผลการเรียน ความเที่ยงสูงอย่างเดียวไม่อาจให้สารสนเทศที่มีประโยชน์ต่อการตัดสินใจมากนักหากว่าเครื่องมือนั้นขาดความตรง (Cronbach 1970: 152) จากการตรวจสอบงานวิจัยเกี่ยวกับการสอบแบบความเรียงของ Coffman (1971: 278) พบค่าความเที่ยงอยู่ระหว่าง 0.35-0.98 และส่วนมากจะอยู่ระหว่าง 0.40-0.60 เมื่อพิจารณาจากแนวคิดและพิสัยความเที่ยงที่เป็นไปได้ ผู้วิจัยจึงเห็นว่าน่าจะอนุมานได้ว่า มาตรฐานความเที่ยงของแบบสอบความเรียงในการวิจัยครั้งนี้ ควรมีค่าอย่างต่ำเป็น 0.50 นอกเหนือจากเหตุผลดังกล่าวมาแล้ว ความเที่ยงระดับ 0.50 เป็นระดับความแม่นยำของการวัด ที่ยอมให้ค่าความแปรปรวนของคะแนนเอกภพมีค่าเท่ากับครึ่งหนึ่งของความแปรปรวนของคะแนนสังเกตพอดี Woodbury (cited by Cronbach and Gleser 1964: 467-479) ใช้ค่าความเที่ยง 0.50 เป็นเกณฑ์เปรียบเทียบระดับประสิทธิภาพมาตรฐาน (standard efficiency) ระหว่างแบบสอบเช่นกัน

จากที่กล่าวมาสรุปได้ว่า โดยทั่วไปแบบสอบถามเรียงมีความเที่ยงต่ำ ตัวประกอบที่มีผลต่อค่าสัมประสิทธิ์ความเที่ยงจากการสอบแบบความเรียง จำนวนเป็น แบบสอบ ผู้ตรวจ วิธีการตรวจ และ ทฤษฎีที่ใช้ในการประมาณค่าความเที่ยง จากข้อเสนอแนะของนักวัดผลการศึกษาชี้ให้เห็นว่า การควบคุมแหล่งความคลาดเคลื่อนของแบบสอบ ผู้ตรวจ และ วิธีการตรวจจะทำให้ผลการวัดมีความแม่นยำเพิ่มขึ้น การใช้ทฤษฎีประมาณค่าความเที่ยงที่มีข้อดกลงสอดคล้องกับธรรมชาติของการวัดจะทำให้สามารถอธิบายผลการวัดได้เหมาะสมยิ่งขึ้น จากข้อเรียกร้องของนักการศึกษาที่ต้องการให้มีการใช้แบบสอบถามเรียงในการประเมินผลการศึกษาให้มากขึ้น โดยเฉพาะอย่างยิ่ง การประเมินผลการเรียนการสอนในชั้นเรียน ซึ่งในสถานการณ์ดังกล่าวครุฑมักรู้จักผลการเรียนของนักเรียนอยู่ก่อนแล้ว ครุฑจึงเผชิญกับความคลาดเคลื่อนแบบ halo effect อย่างหลีกเลี่ยงไม่ได้ ผลการวิจัยพบว่า ผู้ที่มีประวัติการเรียนดีได้คะแนนสูงกว่าผู้มีประวัติการเรียนอ่อนทั้ง ๆ ที่เนื้อหาคำตอบเหมือนกันทุกประการ (Chase 1979: 39-42; Hughes, et al. 1983: 65-72) ปัญหาจึงมีว่า การควบคุมแหล่งความคลาดเคลื่อนของตัวประกอบที่เกี่ยวข้องดังกล่าวสามารถขจัดอิทธิพลของการรู้จักผลการเรียนของนักเรียนได้หรือไม่ เมื่อเทียบกับการควบคุมข้อสอบเพียงอย่างเดียวแต่ให้ผู้ตรวจตรวจโดยใช้ประสบการณ์เดิม จะมีผลแตกต่างกันหรือไม่ ผลการศึกษาอิทธิพลของความคลาดเคลื่อนแบบ halo effect ที่ผ่านมาส่วนใหญ่เปรียบเทียบความแตกต่างระหว่างค่าเฉลี่ยในกลุ่มที่รู้ผลการเรียนสูงกับกลุ่มที่รู้ผลการเรียนต่ำ มิใช่การเปรียบเทียบค่าความเที่ยง ซึ่งเป็นดัชนีแสดงระดับความแม่นยำของการวัดอย่างแท้จริง เพื่อให้เข้าใจในเรื่องนี้ ผู้วิจัยจึงสนใจที่จะศึกษา ผลของการควบคุมแหล่งความคลาดเคลื่อนของตัวประกอบต่อไปนี้ (1) แบบสอบ โดยการกำหนดโครงสร้างคำถาม ระบุความสามารถที่ต้องใช้ในการตอบ และระบุเกณฑ์การให้คะแนนในคำชี้แจงของคำถามให้ชัดเจน (2) ผู้ตรวจ โดยการชี้แจงผู้ตรวจให้เข้าใจหลักการตรวจ การใช้เกณฑ์การตรวจที่กำหนดไว้ล่วงหน้า (3) วิธีการตรวจ โดยกำหนดเกณฑ์ในแต่ละวิธีให้เหมาะสม เพื่อให้ได้สารสนเทศอันจะเป็นประโยชน์ต่อการประยุกต์การสอบแบบความเรียงครอบคลุมผลการตรวจในหลายสถานการณ์ ผู้วิจัยจึงศึกษาเปรียบเทียบผลการควบคุมตัวประกอบเป็น 3 ระดับ คือ สถานการณ์ที่ 1 ผู้ตรวจตรวจโดยใช้ประสบการณ์เดิม เป็นสถานการณ์ที่ควบคุมตัวประกอบแบบสอบเพียงอย่างเดียว แต่ให้ผู้ตรวจมีอิสระในการกำหนดเกณฑ์เอง สถานการณ์ที่ 2 ผู้ตรวจตรวจตามที่ได้รับคำชี้แจงจากผู้วิจัยและไม่รู้ผลการเรียนของผู้ตอบ เป็นสถานการณ์ที่ควบคุมการตรวจสอบข้อสอบ ผู้ตรวจ และ วิธีการตรวจมากที่สุด และสถานการณ์ที่

3 ผู้ตรวจติดตามที่ได้รับชี้แจงจากผู้วิจัยและรู้ผลการเรียนของผู้ตอบ นอกเหนือจากการเปรียบเทียบวิธีการตรวจในแต่ละสถานการณ์แล้ว โดยทฤษฎีการวัด การเพิ่มข้อสอบ และผู้ตรวจ ทำให้ค่าการวัดมีความแม่นยำสูงขึ้น ผู้วิจัยสนใจจะค้นหาว่าจำนวนข้อสอบและผู้ตรวจควรเป็นเท่าใด ภายใต้วิธีการตรวจในสถานการณ์ใด จึงจะให้ค่าความเที่ยง 0.50 เป็นอย่างต่ำ

เนื่องจากทฤษฎีการอ้างอิงสรุป (generalizability Theory) เป็นทฤษฎีที่อธิบายความสัมพันธ์ระหว่างความเที่ยงกับการเพิ่มข้อสอบหรือผู้ตรวจได้เหมาะสมมากกว่าสูตรของ Spearman-Brown เป็นทฤษฎีที่ไม่กำหนดข้อตกลงเกี่ยวกับคุณสมบัติความเป็นคู่ขนานกันของเครื่องมือ ผู้วิจัยจึงกำหนดทฤษฎีการอ้างอิงสรุปในการศึกษาในครั้งนี้

วัตถุประสงค์ของการวิจัย

1. เพื่อประมาณค่าความแปรปรวนของตัวประกอบ 3 ตัวประกอบ ที่มีผลต่อค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง ได้แก่ ผู้ตอบ ข้อสอบ และ ผู้ตรวจ
2. เพื่อศึกษาและ เปรียบเทียบค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง 6 ค่า ที่ประมาณค่าจากการตรวจ 2 วิธี คือ วิธีประเมินรวม และ วิธีวิเคราะห์ ภายใต้สถานการณ์การตรวจ 3 สถานการณ์ คือ
 - 2.1 ผู้ตรวจตรวจโดยใช้ประสบการณ์เดิม
 - 2.2 ผู้ตรวจติดตามที่ได้รับการชี้แจงจากผู้วิจัยและไม่รู้ผลการเรียนของผู้ตอบ
 - 2.3 ผู้ตรวจติดตามที่ได้รับการชี้แจงจากผู้วิจัยและรู้ผลการเรียนของผู้ตอบ
3. เพื่อเลือกวิธีตรวจ จำนวนผู้ตรวจ และจำนวนข้อสอบ ที่ให้ค่าสัมประสิทธิ์การอ้างอิงสรุป 0.50 เป็นอย่างต่ำ

ขอบเขตการวิจัย

ก. เอกภพของการวิจัย

1. ผู้ตอบ เป็นนักเรียนชั้นประถมศึกษาปีที่ 6 สังกัดสำนักงานคณะกรรมการ

การประเมินศึกษาแห่งชาติ ในอำเภอโนนสัง จังหวัดอุดรธานี ปีการศึกษา 2531 จำนวน 1448 คน

2. ผู้ตรวจ เป็นครูสอนกลุ่มวิชาสร้างเสริมประสบการณ์ชีวิต ในโรงเรียนประถมศึกษาศึกษาสังกัดสำนักงานคณะกรรมการการศึกษาแห่งชาติ อำเภอโนนสัง จังหวัดอุดรธานี ปีการศึกษา 2531 จำนวน 154 คน

3. ข้อสอบแบบความเรียง เป็นคำถามวัดความรู้และความสามารถตามจุดประสงค์การเรียนรู้ในกลุ่มวิชาสร้างเสริมประสบการณ์ชีวิตของชั้นประถมศึกษาปีที่ 6 เรื่อง ข้าวและเหตุการณ์สำคัญ มีจำนวนอนันต์ (∞)

ข. ทฤษฎีที่ใช้

ในการวิจัยครั้งนี้ ใช้ทฤษฎีการอ้างอิงสรุปตามแนวความคิดของ Cardinet, et al. (1976, 1981, 1983) ซึ่งนอกจากจะขยายทฤษฎีการอ้างอิงสรุปให้สามารถศึกษาสิ่งที่ถูกวัดได้ครอบคลุมแล้ว ยังใช้สัญลักษณ์ในการอธิบายรูปแบบการวัดได้ชัดเจน ตัวประกอบที่ใช้ในการศึกษา ได้แก่ นักเรียน (p) ผู้ตรวจ (r) และ ข้อสอบ (i) กำหนดความสัมพันธ์ของตัวประกอบที่ศึกษาเป็น $pxixr$

สมมติฐานในการวิจัย

โดยทั่วไปแบบสอบถามความเรียงมักจะมีความเที่ยงต่ำ และความเที่ยงส่วนใหญ่เป็นดัชนีความสอดคล้องระหว่างผู้ตรวจต่างคน หรือภายในผู้ตรวจคนเดียวกันที่ตรวจต่างโอกาส (Intra or Interrater reliability) ผู้ตรวจต่างคน หรือคนเดียวกันแต่ตรวจต่างโอกาสมีแนวโน้มที่จะให้คะแนนต่างกัน ความไม่สอดคล้องจะยิ่งมากขึ้นเมื่อข้อสอบให้อิสระในการตอบมากขึ้น (Coffman 1971: 277) การอ้างอิงสรุปจากกลุ่มตัวอย่างสุ่มไปยังประชากร ขึ้นอยู่กับธรรมชาติของประชากรเป็นสำคัญ กล่าวคือถ้าประชากรมีความเป็นเอกพันธ์ (Homogeneity) ตัวอย่างสุ่มจำนวนไม่มากนักก็เป็นตัวแทนที่ดีได้ (Chadwick, Bahr and Albrecht 1984: 67-68) นั้น แสดงว่า ถ้าผู้ตรวจมีความเห็นสอดคล้องกัน การอ้างอิงสรุปไปยังประชากรของผู้ตรวจ จะมีความแม่นยำสูง ซึ่งหมายถึงผลการตรวจมีความเที่ยงสูงนั่นเอง นักวัดผลการศึกษาและนักวิจัยเสนอว่าการชี้แจงให้ผู้ตรวจมีความรู้ ความเข้าใจในเรื่องที่ตรวจ รู้จักใช้แบบแผนหรือคู่มือการตรวจที่

เหมาะสม สามารถทำให้คะแนนมีความเที่ยงสูงขึ้นได้ (Hopkins and Stanley 1981: 222-223; Bergman 1981: 130-131; Kubiszyn and Borich 1981: 100-101; Wallapa Devahastin and Pateep Methakunavudht 1986: 24-25) วิธีการตรวจเป็นอีกปัจจัยหนึ่งที่ให้ค่าความเที่ยงแตกต่างกัน โดยทั่วไปการตรวจวิธีวิเคราะห์จะให้ค่าความเที่ยงสูงกว่าวิธีประเมินรวม (Hopkins and Stanley 1981: 210; De Gruijter 1980: 247; Lloyed-Jones 1977 cited by Wesdrop, Bauer and Purves 1982: 304) ผู้ตอบที่ได้รับความคาดหวังจากผู้ตรวจสูงจะได้คะแนนสูงกว่าผู้ตอบที่ได้รับความคาดหวังต่ำ (Chase 1979: 39-42; Hughes, et al. 1983: 65-72)

จากจุดประสงค์ของการวิจัยและการศึกษาทฤษฎีและงานวิจัยที่กล่าวมาแล้ว ผู้วิจัยกำหนดสมมุติฐานการวิจัยเกี่ยวกับการสอบแบบความเรียง ที่ใช้ข้อสอบและจำนวนผู้ตรวจเท่ากัน ดังต่อไปนี้

สมมุติฐานที่ 1 ค่าสัมประสิทธิ์การอ้างอิงสรุบทั้ง 6 ค่า ที่ประมาณค่าจากการตรวจ 2 วิธี คือ วิธีประเมินรวม และ วิธีวิเคราะห์ ภายใต้สถานการณ์การตรวจ 3 สถานการณ์ คือ ผู้ตรวจตรวจโดยใช้ประสบการณ์เดิม ผู้ตรวจตรวจตามที่ได้รับภาระชี้แจงจากผู้วิจัยและไม่รู้ผลการเรียนของผู้ตอบ และ ผู้ตรวจตรวจตามที่ได้รับภาระชี้แจงจากผู้วิจัยและรู้ผลการเรียนของผู้ตอบ มีค่าแตกต่างกัน

สมมุติฐานที่ 2 ภายใต้สถานการณ์การตรวจที่ผู้ตรวจตรวจโดยใช้ประสบการณ์เดิม สัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียงที่ตรวจโดยวิธีวิเคราะห์ มีค่าสูงกว่าที่ตรวจโดยวิธีประเมินรวม

สมมุติฐานที่ 3 ภายใต้สถานการณ์การตรวจที่ผู้ตรวจตรวจตามที่ได้รับภาระชี้แจงจากผู้วิจัยและไม่รู้ผลการเรียนของผู้ตอบ สัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียงที่ตรวจโดยวิธีวิเคราะห์ มีค่าสูงกว่าที่ตรวจโดยวิธีประเมินรวม

สมมุติฐานที่ 4 ภายใต้สถานการณ์การตรวจที่ผู้ตรวจตรวจตามที่ได้รับภาระชี้แจงจากผู้วิจัยและรู้ผลการเรียนของผู้ตอบ สัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียงที่ตรวจโดยวิธีวิเคราะห์ มีค่าสูงกว่าที่ตรวจโดยวิธีประเมินรวม

สมมุติฐานที่ 5 ภายใต้สถานการณ์การตรวจทั้ง 3 สถานการณ์ สัมประสิทธิ์การ

อ้างอิงสรุปของแบบสอบความเรียงที่ตรวจโดยวิธีวิเคราะห์ มีค่าสูงกว่าที่ตรวจโดยวิธีประเมินรวม

สมมติฐานที่ 6 สัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียงภายใต้สถานการณ์การตรวจที่ผู้ตรวจตรวจโดยใช้ประสบการณ์เดิมมีค่าต่ำกว่าการตรวจที่ผู้ตรวจตรวจตามที่ได้รับการชี้แจงจากผู้วิจัย

สมมติฐานที่ 7 สัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียงภายใต้สถานการณ์การตรวจที่ผู้ตรวจตรวจโดยใช้ประสบการณ์เดิมมีค่าต่ำกว่าการตรวจที่ผู้ตรวจตรวจตามที่ได้รับการชี้แจงจากผู้วิจัยและไม่รู้ผลการเรียนของผู้ตอบ

สมมติฐานที่ 8 สัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียงภายใต้สถานการณ์การตรวจที่ผู้ตรวจตรวจโดยใช้ประสบการณ์เดิมกับการตรวจที่ผู้ตรวจตรวจตามที่ได้รับการชี้แจงจากผู้วิจัยและรู้ผลการเรียนของผู้ตอบมีค่าไม่แตกต่างกัน

สมมติฐานที่ 9 สัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียงภายใต้สถานการณ์การตรวจที่ผู้ตรวจตรวจตามที่ได้รับการชี้แจงจากผู้วิจัยและไม่รู้ผลการเรียนของผู้ตอบมีค่าสูงกว่าการตรวจที่ผู้ตรวจตรวจตามที่ได้รับการชี้แจงจากผู้วิจัยและรู้ผลการเรียนของผู้ตอบ

สมมติฐานที่ 10 สัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียงภายใต้สถานการณ์การตรวจที่ผู้ตรวจไม่รู้ผลการเรียนของผู้ตอบมีค่าสูงกว่าการตรวจที่ผู้ตรวจรู้ผลการเรียนของผู้ตอบ

ตัวแปรในการวิจัย

ตัวแปรต้น ประกอบด้วยตัวแปรต่าง ๆ ดังนี้

1. วิธีการตรวจ จำแนกเป็น 2 ประเภท ได้แก่
 - 1.1 วิธีประเมินรวม
 - 1.2 วิธีวิเคราะห์
2. สถานการณ์การตรวจ จำแนกเป็น 3 ระดับ ได้แก่
 - 2.1 ผู้ตรวจตรวจโดยใช้ประสบการณ์เดิม
 - 2.2 ผู้ตรวจตรวจตามที่ได้รับการชี้แจงจากผู้วิจัยและไม่รู้ผลการเรียนของผู้ตอบ

2.3 ผู้ตรวจติดตามที่ได้รับการชี้แจงจากผู้วิจัยและรู้ผลการเรียน การเรียนของผู้ตอบ

ตัวแปรตาม ได้แก่

1. ค่าความแปรปรวน
2. ค่าสัมประสิทธิ์ความสามารถในการอ้างอิงสรุป

ข้อตกลงเบื้องต้น

ในการวิจัยครั้งนี้ผู้วิจัยตั้งข้อตกลงเบื้องต้นดังนี้

1. ครูที่เข้าร่วมในการวิจัย ได้ตรวจให้คะแนนคำตอบความเรียง ด้วยความตั้งใจ และเต็มใจ
2. เพศของผู้ตรวจไม่มีผลต่อค่าสัมประสิทธิ์การอ้างอิงสรุปของแบบสอบความเรียง

คำจำกัดความและนิยามศัพท์เฉพาะ

1. แบบสอบความเรียง หมายถึง แบบสอบที่ต้องการให้นักเรียนเขียนคำตอบแสดงความคิดเห็นเกี่ยวกับข่าวและเหตุการณ์ ตามแผนการสอนกลุ่มวิชาสร้างเสริมประสบการณ์ชีวิตชั้นประถมศึกษาปีที่ 6 ของกรมวิชาการ กระทรวงศึกษาธิการ พ.ศ. 2526
2. ผู้ตอบ หมายถึง นักเรียนชั้นประถมศึกษาปีที่ 6 โรงเรียนสังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ อำเภอโนนสัง จังหวัดอุดรธานี ในปีการศึกษา 2531
3. ผู้ตรวจ (Rater) หมายถึง ครูที่สอนกลุ่มวิชาสร้างเสริมประสบการณ์ชีวิต ในชั้นประถมศึกษาปีที่ 6 โรงเรียนสังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ อำเภอโนนสัง จังหวัดอุดรธานี ปีการศึกษา 2531
4. การชี้แจงผู้ตรวจ หมายถึง การอธิบายหลักการและวิธีการตรวจ รวมทั้งการใช้เกณฑ์การตรวจให้คะแนนคำตอบความเรียงตามที่ผู้วิจัยกำหนดให้
5. วิธีตรวจแบบวิเคราะห์ หมายถึง การตรวจให้คะแนนคำตอบความเรียง แยกเป็น 3 ประเด็น คือ (1) แสดงความคิดเห็นและยกเหตุผลได้ตรงประเด็น (2) ยกตัวอย่างหรือข้อ

มูลสนับสนุน (3) ความคิดเห็นมีคุณค่าเชิงจริยธรรม หรือถูกต้องตามหลักวิชา

6. วิธีตรวจแบบประเมินรวม หมายถึง การตรวจให้คะแนนคำตอบความเรียง โดยพิจารณาคุณภาพรวมเพียงหน่วยเดียวตามความประทับใจของผู้ตรวจ

7. การตรวจโดยใช้ประสบการณ์เดิม หมายถึง การให้คะแนนคำตอบความเรียงตามวิธีการและเกณฑ์ของผู้ตรวจแต่ละคน

8. การตรวจตามที่ได้รับคำสั่งจากผู้วิจัย หมายถึง การให้คะแนนคำตอบความเรียงตามวิธีการและเกณฑ์ที่ผู้วิจัยกำหนดให้

9. การรู้ผลการเรียนของผู้ตอบ หมายถึง การที่ผู้ตรวจรู้จักข้อมูลแสดงระดับผลการเรียนรายวิชาของผู้ตอบแต่ละคนก่อนลงมือตรวจ

10. เงื่อนไขการวัด (Condition) หมายถึง จำนวนข้อสอบ หรือจำนวนผู้ตรวจที่สามารถนำมาใช้ในการประเมินผลแบบความเรียง

11. พาเซต (Facets) หมายถึง ตัวประกอบ (Factor) หรือลักษณะร่วมกันของกลุ่มเงื่อนไขการวัดชนิดเดียวกัน ในการวิจัยครั้งนี้มี 3 พาเซต ได้แก่ นักเรียน ข้อสอบ และ ผู้ตรวจ

12. เอกภพ (Universe) ของพาเซต หมายถึง ประชากรของตัวประกอบที่ศึกษา ได้แก่ นักเรียนชั้นประถมศึกษาปีที่ 6 และครูที่สอนกลุ่มวิชาสร้างเสริมประสบการณ์ชีวิต ชั้นประถมศึกษาปีที่ 6 โรงเรียนสังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ ปีการศึกษา 2531 ทุกคน และ ข้อสอบที่วัดความคิดเห็นเกี่ยวกับข่าวและเหตุการณ์ทุกข้อ

เอกภพของการวิจัยครั้งนี้แบ่งเป็น 2 ชนิดตามแนวทฤษฎีการอ้างอิงสรุป ได้แก่

12.1 เอกภพของการสังเกตที่ยอมรับได้ (Universe of Admissible Observation) หมายถึง กลุ่มเงื่อนไขการวัดที่เป็นไปได้ทั้งหมด ได้แก่ นักเรียนทุกคน ข้อสอบทุกข้อ และผู้ตรวจทุกคนที่อยู่ในเอกภพของแต่ละพาเซต ผู้วิจัยหรือนักวิจัยคนอื่นสามารถทำการสังเกตหรือวัดค่าได้

12.2 เอกภพของการอ้างอิงสรุป (Universe of Generalization) หมายถึง กลุ่มเงื่อนไขของพาเซต ที่ผู้วิจัยต้องการอ้างอิงสรุปผลการวัด ในการวิจัยครั้งนี้ต้องการอ้างอิงไปยังเอกภพที่มีขอบเขตต่างกัน 2 ระดับ ได้แก่ (1) เอกภพของข้อสอบและ เอกภพของผู้ตรวจพร้อม ๆ กันโดยที่เอกภพของข้อสอบประกอบด้วยข้อสอบเป็นจำนวนอนันต์ (∞) และ เอกภพ

ของผู้ตรวจซึ่งประกอบด้วยครูที่สอนวิชาสร้างเสริมประสบการณ์ชีวิตชั้นประถมศึกษาปีที่ 6 ในโรงเรียนสังกัดสำนักงานคณะกรรมการการประถมศึกษาแห่งชาติ อำเภอโนนสัง จังหวัดอุดรธานี จำนวน 154 คน (2) เอกภพของผู้ตรวจเพียงอย่างเดียว

13. สัมประสิทธิ์การอ้างอิงสรุป (Generalizability Coefficient) หมายถึง ค่าเฉลี่ยของค่าสหสัมพันธ์ระหว่างชุดของเงื่อนไขการวัด ที่สุ่มมาจากเอกภพที่ต้องการอ้างอิงสรุป เช่น ถ้าค่าสัมประสิทธิ์การอ้างอิงสรุปของการสอบความเรียงของนักเรียนชั้นประถมศึกษาปีที่ 6 จากการใช้ข้อสอบ 5 ข้อเป็น .6 หมายความว่า ถ้าสอบนักเรียนชั้นประถมศึกษาปีที่ 6 กลุ่มใด ๆ ด้วยข้อสอบซึ่งสุ่มมาจากเอกภพของข้อสอบ (โคเมน) ครั้งละ 5 ข้อ นาคะแนนแต่ละชุดมาหาค่าสหสัมพันธ์รายคู่ แล้วหาค่าเฉลี่ยของค่าสหสัมพันธ์ที่ได้จะมีค่าเป็น .6 ค่าสัมประสิทธิ์การอ้างอิงสรุป แสดงถึงความสอดคล้องกันระหว่างชุดของข้อสอบในเอกภพของข้อสอบ มีค่าอยู่ระหว่าง 0 ถึง 1 คำนวณค่าจากอัตราส่วนระหว่างความแปรปรวนของคะแนนเอกภพ ต่อค่าคาดหวังของความแปรปรวนของคะแนนสังเกต

14. รูปแบบการวัด หมายถึงการกำหนดตัวประกอบที่ศึกษาให้เป็นสิ่งที่ถูกวัด กับ เป็นเครื่องมือ และกำหนดประเภทของแต่ละตัวประกอบให้เป็นตัวประกอบสุ่มหรือตัวประกอบคงที่ ในการศึกษาครั้งนี้กำหนดรูปแบบการวัดเป็น 2 รูปแบบ คือ รูปแบบที่ 1 ให้นักเรียนเป็นสิ่งที่ถูกวัด กำหนดให้เป็นตัวประกอบสุ่ม ให้ข้อสอบและผู้ตรวจเป็นเครื่องมือในการวัดและให้เป็นตัวประกอบสุ่ม รูปแบบที่ 2 ให้นักเรียนเป็นสิ่งที่ถูกวัด กำหนดให้เป็นตัวประกอบสุ่ม ให้ข้อสอบและผู้ตรวจเป็นเครื่องมือวัดโดยที่ข้อสอบเป็นตัวประกอบคงที่ แต่ผู้ตรวจเป็นตัวประกอบสุ่ม

ประโยชน์ที่คาดว่าจะได้รับ

จากผลการวิจัยนี้คาดว่าจะได้ประโยชน์อย่างน้อย 5 ประการ คือ

1. ทำให้นักวัดผล ครูและนักวิจัยได้ทราบขนาดความแปรปรวนของตัวประกอบสำคัญในการวัดที่มีอิทธิพลต่อความเที่ยงของแบบความเรียงวัดผลสัมฤทธิ์ทางการเรียน ซึ่ง ได้แก่ ข้อสอบ ผู้ตรวจ และผลรวมของข้อสอบและผู้ตรวจ รวมทั้งความคลาดเคลื่อนเชิงสุ่มที่ไม่สามารถระบุแหล่งได้ชัดเจน

2. ทำให้นักวัดผล ครู และนักวิจัยได้ทราบค่าสัมประสิทธิ์การอ้างอิงสรุปของการสอบ

แบบความเรียงที่ตรวจโดยวิธีประเมินรวม และวิธีวิเคราะห์ ในสถานการณ์การตรวจที่ต่างกัน กล่าวคือ ผู้ตรวจได้รับการชี้แจงกับไม่ได้รับการชี้แจงแนวการตรวจจากผู้วิจัย การตรวจโดยที่ผู้ตรวจรู้ผลการเรียนของผู้ตอบ

3. ทำให้นักวัดผล และครูสามารถเลือกรูปแบบการวัด วิธีการตรวจ จำนวนข้อสอบ จำนวนผู้ตรวจ ที่เหมาะสมและให้ค่าการวัดมีความแม่นยำสูง

4. เป็นการกระตุ้นครูให้สนใจและ เห็นคุณค่าของการสอบแบบความเรียงและหันมาใช้แบบสอบความเรียงในการวัดและประเมินผลการเรียนมากขึ้น อันจะเป็นประโยชน์ต่อการพัฒนาสมรรถภาพของนักเรียนด้านความรู้ความคิดในระดับสูง รู้จักคิดวิเคราะห์ สังเคราะห์และประเมินค่า ตลอดจนสามารถสื่อความคิดในรูปการเขียนได้อย่างมีประสิทธิภาพ

5. เป็นแนวทางและเป็นตัวอย่างสำหรับการประยุกต์ทฤษฎีการอ้างอิงสรุป ในการศึกษาปัญหาทางการวัดผล ในด้านอื่น ๆ เพื่อจะได้เข้าใจปัญหาได้หลายแง่มุม อันจะนำไปสู่การพัฒนาการวัดผลการศึกษาให้มีประสิทธิภาพมากยิ่งขึ้น

ศูนย์วิทยทรัพยากร
จุฬาลงกรณ์มหาวิทยาลัย