

บทความวิจัย (Research Article)

การศึกษาคุณลักษณะของผู้ตรวจให้คะแนนความสามารถในการเขียนความเรียง ของนักเรียนระดับชั้นมัธยมศึกษาตอนปลาย

THE STUDY OF RATERS' CHARACTERISTICS ON LANGUAGE WRITING ABILITY BY THE SECONDARY SCHOOLING

Received: December 25, 2017

Revised: February 14, 2018

Accepted: April 22, 2018

ณัฐณี ประวัง^{1*} และน้ำผึ้ง อินทะเนตร²

Nattinee Prawang^{1*} and Nampuang Intanate²

^{1,2}คณะศึกษาศาสตร์ มหาวิทยาลัยเชียงใหม่

^{1,2}Faculty of Education, Chiang Mai University, Chiang Mai 50200, Thailand

*Corresponding Author, E-mail: prawang@gmail.com

บทคัดย่อ

งานวิจัยนี้มีจุดประสงค์เพื่อพัฒนาเกณฑ์ในการตรวจให้คะแนนเรียงความ 3 รูปแบบ และศึกษาคุณลักษณะ ความเข้มงวด ใจดี และเป็นกลางของผู้ตรวจให้คะแนนเมื่อมีประสบการณ์ต่างกันและใช้เกณฑ์การตรวจให้คะแนนต่างกัน เครื่องมือที่ใช้ในการวิจัย คือ ข้อสอบการเขียนภาษาอังกฤษ เกณฑ์การตรวจให้คะแนน 3 รูปแบบ และคู่มือการใช้เกณฑ์ การประเมินที่ผู้วิจัยสร้างขึ้น กลุ่มตัวอย่าง คือ ครูผู้สอนวิชาภาษาอังกฤษที่มีประสบการณ์ในการสอนภาษาอังกฤษ ต่ำกว่า 5 ปี และกลุ่ม 10 ปีขึ้นไป จำนวน 8 คน และนักเรียนชั้นมัธยมศึกษาตอนปลาย จำนวน 120 คน ได้มาจากการสุ่มอย่างง่าย ผลการวิจัย พบว่า เกณฑ์การตรวจให้คะแนน 3 รูปแบบที่ผู้วิจัยสร้างขึ้น มีค่าความเชื่อมั่นอยู่ ระหว่าง .82 - .83 และ คุณลักษณะของผู้ตรวจให้คะแนน พบว่า 1) เกณฑ์แยกองค์ประกอบมีความเข้มงวด 2 คน ใจดี 1 คน และเป็นกลาง 5 คน 2) เกณฑ์รวมองค์ประกอบ มีความเข้มงวด 2 คน ใจดี 3 คน และเป็นกลาง 3 คน 3) เกณฑ์แบบตรวจสอบรายการมีความเข้มงวด 2 คน ใจดี 4 คน และเป็นกลาง 2 คน ซึ่งพบว่า มีผู้ตรวจที่มีคุณลักษณะ คงเดิม 2 แบบ คือ เข้มงวดและใจดีเพียง 3 คน (ร้อยละ 37.50) เท่านั้น และผู้ตรวจอีก 5 คน เมื่อใช้เกณฑ์ต่างรูปแบบกัน จะมีคุณลักษณะที่แตกต่างกัน โดยผู้ที่มีคุณลักษณะคงเดิม ได้แก่ ผู้ตรวจที่มีประสบการณ์ต่ำกว่า 5 ปี 1 คน และสูงกว่า 10 ปี 2 คน และไม่มีผู้ตรวจให้คะแนนคนใดที่แปรเปลี่ยนคุณลักษณะแบบสุดขั้ว คือ จากใจดีเป็นเข้มงวดหรือจากเข้มงวด เป็นใจดี ดังนั้น จึงยังไม่สามารถสรุปได้ว่าประสบการณ์ไม่มีผลต่อคุณลักษณะของผู้ตรวจ

คำสำคัญ : เกณฑ์การตรวจให้คะแนน ผู้ตรวจให้คะแนน ความเข้มงวด ใจดี

Abstract

The purposes of this research were to develop essays rubrics such as three criterion scoring methods and to examine the generalized ability characteristics of the raters with different experiences and using different rubrics. The research instrument was the criteria for scoring. The sample consisted of English teachers with less than 5 years of teaching experience and more than 10 years of teaching experience and 120 high school students. Three types of scoring methods were used. The reliability scores ranged from .825 to .833 and the characteristics of the inspectors were as follows: 1) the raters' characteristic using analytic rubrics was 2 severities, 1 leniency and 5 neutral. 2) There were 2 strict, 3 lenient and 3 neutral in holistic rubric using. 3) The criterion for checklists was strict 2 people, 4 lenient and 2 neutral. When compared with the experience group, different experienced raters did not make the feature score differently. There were 3 respondents (37.5%) having consistent characteristics when using any criteria. And no one who evaluates the extremes transport characteristic from severe to lenient or lenient to severe.

Keywords: Essays Rubrics, Raters, Severity, Leniency

บทนำ

การวัดและประเมินผลเป็นกิจกรรมส่วนหนึ่งของกระบวนการเรียนการสอน ซึ่งเป็นหน้าที่โดยตรงของครูผู้สอนทุกระดับและทุกประเภท ที่ต้องดำเนินการวัดและประเมินผลผู้เรียนทุกคนเพื่อตรวจสอบความสามารถในการเรียนรู้ Aiamsaard et al. (2016, pp. 184-192) กล่าวว่า องค์ประกอบที่เป็นคุณลักษณะสำคัญของครูภาษาอังกฤษนอกจากความรู้ความสามารถในการสอนวิชาเอกของตนเองแล้ว ยังต้องมีความรู้ความสามารถในด้านการวัดและประเมินผล ทั้งในส่วนของสร้างเครื่องมือในการวัดประเมินผล เนื่องจากในการดำเนินการวัดและประเมินผล การเรียนรู้ของผู้เรียนนั้น นอกจากจะต้องดำเนินการประเมินของแต่ละหน่วยงานและสถาบันแล้ว ยังต้องคำนึงถึงความถูกต้อง ความเป็นธรรม และดำเนินการอย่างเป็นระบบ เพื่อให้ผลการวัดและประเมินสามารถสะท้อนศักยภาพของผู้เรียนได้ตรงตามความเป็นจริงมากที่สุด และเกิดข้อผิดพลาดน้อยที่สุด (Wongpanich, 2000)

การวัดประเมินผลทักษะการเขียนถือเป็นอีกทักษะหนึ่งที่มีความละเอียดอ่อนในการประเมินที่สุด เนื่องจากมีคำตอบได้หลากหลายแนวทาง มิใช่มีคำตอบที่ตายตัวเพียงอย่างเดียวอย่างใดอย่างหนึ่งเท่านั้น ดังนั้น การตรวจให้คะแนนงานเขียนนั้น ต้องอาศัยการตัดสินใจของผู้ประเมิน เพื่อให้มีความเป็นปรนัย เครื่องมือในการตรวจให้คะแนนงานเขียน เพื่อใช้ประกอบสำหรับการตัดสินใจในการให้คะแนน ให้การประเมินการเรียนรู้ของผู้เรียนเป็นไปอย่างคงเส้นคงวามีความยุติธรรม ซึ่งสามารถเป็นได้ทั้งเกณฑ์แบบรูบริค และเกณฑ์แบบแยกองค์ประกอบ รวมถึงเกณฑ์รวมองค์ประกอบ (Mertler, 2001) ซึ่งนักวิจัยต่างก็พยายามที่จะพัฒนาวิธีที่จะสามารถเพิ่มความเที่ยงตรง และความน่าเชื่อถือของการประเมินงานเขียนด้วยการพัฒนา เปรียบเทียบรูปแบบเครื่องมือที่เหมาะสม (Meier et al.; Razaee & Lovorn cited

in Dusmair et al., 2015) เนื่องจากอิทธิพลของความคลาดเคลื่อนในการประเมินส่วนหนึ่งมาจากรูปแบบเกณฑ์การประเมิน จึงมีงานวิจัยหลายงานที่ได้ทดลองเปรียบเทียบความเชื่อมั่นของเกณฑ์ ความเชื่อมั่นระหว่างผู้ตรวจให้คะแนน ความน่าเชื่อถือของผลการประเมิน แหล่งความคลาดเคลื่อนจากการประเมิน และความเหมาะสมในการตรวจให้คะแนน ซึ่งนอกจากกระบวนการวิเคราะห์โดยพิจารณาองค์ประกอบเงื่อนไขการวัดแล้ว นักวิจัยส่วนหนึ่งยังมุ่งอธิบายคุณลักษณะของผู้ตรวจร่วมด้วย เนื่องจากการให้คะแนนขึ้นอยู่กับ การตัดสินใจของผู้ประเมิน ซึ่งสามารถกล่าวได้ว่า แหล่งความคลาดเคลื่อนส่วนใหญ่มาจากผู้ตรวจให้คะแนน ดังนั้น การอธิบายถึงคุณลักษณะของผู้ประเมินจะเป็นการช่วยอธิบายคะแนน และพฤติกรรมการตรวจให้คะแนนของผู้ตรวจด้วยการวิเคราะห์โมเดลหลายองค์ประกอบของราส์ช (Multi-Facet Rasch Model: MFRM) ว่าผู้ตรวจมีความเข้มงวด (Severity) หรือใจดี (Leniency) ในการให้คะแนนอย่างไร เพื่ออธิบายว่าผู้ตรวจมีแนวโน้มการให้คะแนนด้วยความเข้มงวดหรือใจดีเหมือนกันหรือไม่ ทำให้ตรวจสอบอิทธิพลของผู้ตรวจในการวัดประเมินได้ ซึ่งมีการนำไปประยุกต์ในการศึกษาคูณลักษณะของผู้ประเมินอย่างหลากหลาย (Intanate, 2011; Saenplue, 2013; Schaefer, 2008; Prieto & Nieto, 2014) หากศึกษารูปแบบเกณฑ์ที่เหมาะสมต่อผู้ตรวจในแต่ละคุณลักษณะ จะสามารถช่วยลดความคลาดเคลื่อนในการวัดประเมิน หรือช่วยผู้ตรวจในการตัดสินใจเลือกเครื่องมือในการตรวจให้คะแนนให้เหมาะสมกับงาน และเป็นแหล่งข้อมูลสารสนเทศเพื่อเป็นประโยชน์ต่อการศึกษากระบวนการวัดประเมินงานเขียนสำหรับหน่วยงาน หรือผู้ที่เกี่ยวข้อง ดังนั้นผู้วิจัยจึงสนใจที่จะพัฒนาเกณฑ์ในการตรวจให้คะแนนเรียงความ 3 รูปแบบ และศึกษาคูณลักษณะความเข้มงวด ใจดี และเป็นกลางของผู้ตรวจให้คะแนนเมื่อมีประสบการณ์ต่างกันและใช้เกณฑ์การตรวจให้คะแนนต่างกัน

วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาเกณฑ์ในการตรวจให้คะแนนเรียงความ 3 รูปแบบ
2. ศึกษาคุณลักษณะความเข้มงวด ใจดี และเป็นกลางของผู้ตรวจให้คะแนนเมื่อมีประสบการณ์ต่างกันและ

ใช้เกณฑ์การตรวจให้คะแนนต่างกัน

วิธีดำเนินการวิจัย

1. ประชากรและกลุ่มตัวอย่าง/กลุ่มเป้าหมายที่ใช้ในการวิจัย กลุ่มเป้าหมายที่ใช้ในการวิจัยมี 2 กลุ่ม ได้แก่ นักเรียนชั้นมัธยมศึกษาปีที่ 4-6 ปีการศึกษา 2559 จำนวน 120 คน โดยได้มาจากการสุ่มอย่างง่าย เพื่อใช้เป็นผู้สอบในการสร้างและพัฒนาคุณภาพของเกณฑ์ในการตรวจให้คะแนน และครูภาษาอังกฤษจำนวน 8 คน ได้มาจากการสุ่มจากครูภาษาอังกฤษในภาคเหนือ และคัดเลือกเจาะจงจากบุคคลที่มีความพร้อมและมีความเต็มใจในการให้ความร่วมมือตลอดขั้นตอนการให้คะแนนงานเขียนทั้ง 3 รอบ โดยเป็นผู้ที่มีประสบการณ์ในการสอนต่ำกว่า 5 ปี (ผู้ตรวจหมายเลข 1-4) และ มีประสบการณ์มากกว่า 10 ปี (ผู้ตรวจหมายเลข 5-8) หมายความว่า ผู้ตรวจต้องเคยมีประสบการณ์ในการตรวจงานเขียนของนักเรียนมาแล้ว

2. เครื่องมือที่ใช้ในการวิจัย ได้แก่ ข้อสอบการเขียนภาษาอังกฤษ เกณฑ์การตรวจให้คะแนน 3 รูปแบบ และคู่มือการใช้เกณฑ์การประเมินที่ผู้วิจัยสร้างขึ้น โดยแบ่งเป็น

2.1 ข้อสอบการเขียนภาษาอังกฤษ มีลักษณะเป็นคำถามปลายเปิด สร้างขึ้นโดยอาศัยจากเกณฑ์ การกำหนดคุณภาพภาษาที่ผู้ใช้ภาษาในระดับ B1 จะสามารถทำได้ อาศัยประเด็นการเขียนเพื่อการสื่อสาร (แสดง ความคิดเห็น) โดยหาคุณภาพจากการให้ผู้เชี่ยวชาญพิจารณาความสอดคล้องเหมาะสมกับประเด็นที่ต้องการวัด และ ระดับของผู้ที่ถนัด ซึ่งเน้นการให้ข้อเสนอแนะและคำแนะนำเชิงคุณภาพจากผู้เชี่ยวชาญด้านการสอนภาษาอังกฤษ จำนวน 3 ท่าน และพิจารณาความสอดคล้องเหมาะสมรายประเด็น (ความถูกต้องของภาษา ความเหมาะสมของภาษา ความเหมาะสมของระดับการวัด ความเป็นปรนัย และการเปิดโอกาสในการตอบที่อิสระและหลากหลาย) ด้วยมาตรวัด แบบมาตราส่วนประมาณค่า 5 ระดับ พบว่า ข้อสอบมีความสอดคล้องเหมาะสมมากที่สุด (4.60)

2.2 เกณฑ์การให้คะแนนพัฒนาขึ้นจากการประเมินระดับความสามารถใช้ภาษาของ Cambridge และ TWE เพื่อนำมากำหนดองค์ประกอบหลักในการให้คะแนน โดยแบ่งออกเป็น 3 รูปแบบ ดังนี้ 1) แบบแยก องค์ประกอบ ประกอบด้วย 4 ประเด็นในการพิจารณาคือ เนื้อหา (Content) การบรรลุจุดประสงค์การสื่อสาร (Communicative Achievement) โครงสร้าง (Organization) ภาษา (Language) โดยให้คะแนนเป็นมาตราประมาณค่า 3 ระดับ โดยให้ผู้เชี่ยวชาญพิจารณาความสอดคล้องของประเด็นในการวัด (IOC) เท่ากับ 1.00 และปรับปรุงตาม ข้อเสนอแนะของผู้เชี่ยวชาญ 2) แบบรวมองค์ประกอบ อาศัยจากประเด็นหลักตามเกณฑ์แยกองค์ประกอบ เพียงแต่ มีการตรวจให้คะแนนแบบรวมเพียงหนึ่งหัวข้อ และมีการพิจารณาคุณภาพตามระดับอยู่ 5 ระดับใหญ่ๆ 3) แบบตรวจสอบรายการ สร้างโดยการนำรายละเอียดของงานที่ต้องการมาแยกเป็นประเด็นในการพิจารณาให้คะแนน เป็นรายชื่อหลักๆ 10 ข้อ

2.3 คู่มือการใช้เกณฑ์การให้คะแนนที่ผู้วิจัยสร้างขึ้นมีลักษณะเป็นรูปเล่มซึ่งประกอบด้วยจุดประสงค์ ตัวอย่างงานเขียน เกณฑ์ และการให้คะแนนด้วยเกณฑ์ประเภทต่างๆ โดยระบุเหตุผลและคำอธิบายในการให้คะแนนทั้ง 3 รูปแบบ หาคุณภาพเบื้องต้นจากการให้ผู้เชี่ยวชาญ จำนวน 5 คน ทดลองใช้เกณฑ์การให้คะแนน และประเมินคู่มือ ด้วยแบบสอบถามแบบมาตราส่วนประมาณค่า 5 ระดับ โดยมีประเด็นในการพิจารณา ดังนี้ ความชัดเจนของจุดมุ่งหมาย และกระบวนการให้คะแนนในภาพรวม สามารถใช้เกณฑ์ได้ทั้ง 3 รูปแบบหลังจากการศึกษาคู่มือ รายละเอียดครบถ้วน ทุกประเด็น ภาษาที่ใช้ และรูปแบบคู่มือ พบว่า กลุ่มผู้เชี่ยวชาญเห็นว่าคู่มือมีความเหมาะสมมากที่สุด (4.86)

3. การเก็บรวบรวมข้อมูล ผู้วิจัยดำเนินการเก็บข้อมูลการสอบเขียนเรียงความภาษาอังกฤษของนักเรียน จำนวน 120 คน แล้วนำมาคัดกรองเบื้องต้นเพื่อคัดคำตอบชุดที่ไม่สมบูรณ์ออกไป ให้ได้จำนวน 50 ฉบับ เพื่อให้เหมาะสม ต่อการตรวจให้คะแนน โดยดำเนินการ ดังนี้

3.1 สร้างเกณฑ์การตรวจให้คะแนน 3 รูปแบบ และนำไปหาความสอดคล้องเชิงประจักษ์จาก การพิจารณาของผู้เชี่ยวชาญด้านภาษาอังกฤษ และผู้เชี่ยวชาญด้านการประเมินผล จำนวน 3 ท่าน และปรับปรุงตาม คำแนะนำของผู้เชี่ยวชาญ

3.2 หาคุณภาพเบื้องต้นของเกณฑ์การตรวจให้คะแนนทั้ง 3 เกณฑ์ โดยพิจารณาเชิงคุณภาพเบื้องต้น และทดลองใช้ให้คะแนนเพื่อหาความสอดคล้องภายในเกณฑ์การให้คะแนนด้วยสัมประสิทธิ์แอลฟาครอนบาค โดยมีค่า อยู่ระหว่าง .825 - .833 กล่าวคือ มีความสอดคล้องภายในสูงและหาความสอดคล้องระหว่างเกณฑ์ทั้ง 3 ด้วยการพิจารณา สหสัมพันธ์ของเพียร์สัน ซึ่งมีค่าอยู่ระหว่าง .835 - .882 และมีนัยสำคัญทางสถิติที่ระดับ .01 แสดงดังตาราง

	1	2	3
1. เกณฑ์แยกองค์ประกอบ	-	0.877**	0.835**
2. เกณฑ์รวมองค์ประกอบ	-	-	0.882**
3. เกณฑ์แบบตรวจสอบรายการ	-	-	-

** ค่าสัมประสิทธิ์สหสัมพันธ์มีนัยสำคัญทางสถิติที่ระดับ .01

3.3 สร้างคู่มือการตรวจให้คะแนน และให้ผู้เชี่ยวชาญพิจารณาความเหมาะสม ชัดเจนของคู่มือ

3.4 นำงานเขียนของนักเรียนที่คัดเลือกไว้ทั้ง 50 ชุดมาทำสำเนาเพื่อส่งให้ผู้ตรวจคะแนนพิจารณาให้คะแนน พร้อมกับชุดคู่มือการตรวจให้คะแนน โดยผู้วิจัยนัดหมายวันเวลาเพื่อเข้าไปชี้แจงการให้คะแนนกับผู้ตรวจก่อนจะแบ่งรอบการตรวจให้คะแนนเป็นสามรอบตามรูปแบบเกณฑ์การให้คะแนน และเว้นช่วงในการตรวจให้คะแนนแต่ละรอบเป็นระยะเวลาสองสัปดาห์

3.5 วิเคราะห์คุณลักษณะผู้ตรวจให้คะแนนจากผลคะแนนงานเขียนโดยใช้เกณฑ์ทั้งสามรูปแบบเป็นรายบุคคลด้วยการวิเคราะห์ พารามิเตอร์ ด้วยโปรแกรม Minifac (Facets) Rasch

ผลการวิจัย

1. ผลการสร้างและพัฒนาเกณฑ์การตรวจให้คะแนน 3 รูปแบบ พบว่า ผลการพิจารณาความสอดคล้องของเกณฑ์ทั้งสามรูปแบบ พบว่า สัมประสิทธิ์สหสัมพันธ์มีค่าอยู่ระหว่าง .835 - .882 และมีนัยสำคัญทางสถิติที่ระดับ .01 กล่าวคือ เกณฑ์ทั้งสามมีความสอดคล้องสัมพันธ์กันสูง หากให้คะแนนด้วยเกณฑ์รูปแบบหนึ่งได้ผลคะแนนมาก อีกเกณฑ์ก็จะได้มากเช่นกัน และเมื่อพิจารณาความสอดคล้องภายในเกณฑ์การให้คะแนน พบว่า เกณฑ์แบบแยกองค์ประกอบมีค่าสัมประสิทธิ์แอลฟาครอนบาคเท่ากับ .833 เกณฑ์แบบรวมองค์ประกอบ .831 และเกณฑ์แบบตรวจสอบรายการ .825 ซึ่งถือว่าอยู่ในระดับดี (George & Mallory, 2003) มีความสอดคล้องภายในสูง ไม่ว่าจะใช้ผู้ประเมินใดในการประเมินค่อนข้างที่จะให้ผลใกล้เคียงกัน

2. การพิจารณาคูณลักษณะของผู้ตรวจให้คะแนน พบว่า การตรวจให้คะแนนด้วยเกณฑ์ทั้งสามรูปแบบ มีความสอดคล้องกลมกลืนกับโมเดล โดยพิจารณาจากค่าสถิติ Infit อยู่ในช่วง 0.5 – 1.5 หมายถึง สามารถวัดได้ตามโครงสร้าง (Linacre, 2012) แล้วจึงพิจารณาความเข้มงวด ใจดี และความเป็นกลางของผู้ตรวจให้คะแนนจากค่า Logit ที่ได้จากการวิเคราะห์ด้วยโปรแกรม Minifac (Facets) Rasch ซึ่งมีเกณฑ์พิจารณา ดังนี้

ค่า Logit มีค่ามากกว่า +1.00	แสดงถึงความเข้มงวดของผู้ประเมิน
ค่า Logit มีค่าน้อยกว่า -1.00	แสดงถึงความใจดี
ค่า Logit มีค่าอยู่ในช่วง -1 ถึง +1	แสดงว่าผู้ประเมินมีความเป็นกลาง

ตาราง 2 แสดงการระบุความเข้มงวด/ใจดีของผู้ประเมิน จำแนกตามเกณฑ์

เกณฑ์	คุณลักษณะผู้ประเมิน			
	ไม่สอดคล้องกับ โมเดล	เข้มงวด	เป็นกลาง	ใจดี
แยกองค์ประกอบ (Analytic)	-	R4, R6	R3, R1, R2, R7, R8	R5
รวมองค์ประกอบ (Holistic)	-	R4, R6	R2, R7, R8	R5, R3, R1
แบบตรวจสอบรายการ (Checklist)	-	R4, R6	R1, R3	R5, R2, R7, R8

ตาราง 3 แสดงการระบุความเข้มงวด/ใจดีของผู้ประเมิน จำแนกตามผู้ตรวจ

ผู้ตรวจให้คะแนน	เกณฑ์		
	แยกองค์ประกอบ (Analytic)	รวมองค์ประกอบ (Holistic)	แบบตรวจสอบรายการ (Checklist)
<i>ประสบการณ์ต่ำกว่า 5 ปี</i>			
R1	เป็นกลาง	ใจดี	เป็นกลาง
R2	เป็นกลาง	เป็นกลาง	ใจดี
R3	เป็นกลาง	ใจดี	เป็นกลาง
R4	เข้มงวด	เข้มงวด	เข้มงวด
<i>ประสบการณ์ 10 ปีขึ้นไป</i>			
R5	ใจดี	ใจดี	ใจดี
R6	เข้มงวด	เข้มงวด	เข้มงวด
R7	เป็นกลาง	เป็นกลาง	ใจดี
R8	เป็นกลาง	เป็นกลาง	ใจดี

จากตาราง พบว่า ผู้ตรวจทุกคนให้คะแนนเชิงประจักษ์สอดคล้องเหมาะสมกับโมเดลคิดเป็นร้อยละ 100 โดยคุณลักษณะของผู้ตรวจให้คะแนน ดังนี้ 1) เกณฑ์แบบแยกองค์ประกอบมีความเข้มงวด 2 คน ได้แก่ผู้ตรวจหมายเลข 4, 6 ใจดี 1 คน ได้แก่ผู้ตรวจหมายเลข 5 และเป็นกลาง 5 คน ได้แก่ผู้ตรวจหมายเลข 1, 2, 3, 7 และ 8 2) เกณฑ์แบบรวมองค์ประกอบ ผู้ตรวจหมายเลข 4, 6 (เข้มงวด) ผู้ตรวจหมายเลข 1, 3, 5 (ใจดี) และผู้ตรวจหมายเลข 2, 7, 8 เป็นกลาง 3) เกณฑ์แบบตรวจสอบรายการมีความเข้มงวด 2 คน ได้แก่ผู้ตรวจหมายเลข 4, 6 ใจดี 4 คน ได้แก่ผู้ตรวจหมายเลข 2, 5, 7, 8 และเป็นกลาง 2 คน ได้แก่ผู้ตรวจหมายเลข 1, 3 และเมื่อเปรียบเทียบรายกลุ่มประสบการณ์ จากตารางที่ 3 พบว่า ยังไม่สามารถสรุปได้ว่าประสบการณ์ของผู้ตรวจที่แตกต่างกันไม่ส่งผลต่อคุณลักษณะในการให้คะแนน (ผู้ตรวจคนที่ R1-4 คือคนที่มีประสบการณ์ต่ำกว่า 5 ปีและ คนที่ R5-8 คือ กลุ่มผู้มีประสบการณ์มากกว่า 10 ปี) และพบว่า มีผู้ตรวจให้คะแนนเข้มงวดทั้ง 3 รูปแบบจำนวน 2 คน คิดเป็นร้อยละ 25 ใจดีทั้ง 3 รูปแบบจำนวน 1 คน คิดเป็นร้อยละ 12.5 และ

ไม่ปรากฏผู้ตรวจที่แสดงความเป็นกลางทั้ง 3 รูปแบบ ซึ่งแสดงให้เห็นว่าเกณฑ์ทั้ง 3 ให้ผลการจำแนกคุณลักษณะของผู้ตรวจที่เป็นกลางแตกต่างกัน

การอภิปรายผลการวิจัย

1. เกณฑ์การประเมินทั้ง 3 รูปแบบที่ผู้วิจัยสร้างและพัฒนาขึ้น คือแบบแยกองค์ประกอบ แบบรวมองค์ประกอบ และแบบตรวจสอบรายการ มีคุณภาพที่เหมาะสมสำหรับการนำไปใช้โดยมีค่าความเชื่อมั่นที่เหมาะสม ทั้งนี้เป็นเพราะการใช้เกณฑ์มาเป็นเครื่องมือในการให้คะแนนช่วยเป็นบรรทัดฐานในการจัดลำดับคุณภาพของงาน โดยเป็นการเขียนอธิบายคุณลักษณะของงานที่ชัดเจน หรือมีการกำหนดลักษณะของงานที่มีคุณภาพไว้จึงทำให้ผลการประเมินที่ได้มีความน่าเชื่อถือ และมีความเชื่อมั่นที่เหมาะสมมากกว่าการประเมินหรือให้คะแนนโดยที่ไม่ใช้เกณฑ์ ซึ่งสอดคล้องกับงานวิจัยของ Suaysod (2011, pp. 107-114) ที่ได้ทำการเปรียบเทียบความเชื่อมั่นและความเป็นปรนัยของคะแนนการตรวจเรียงความภาษาไทยของนักเรียนชั้นประถมศึกษาปีที่ 5 จำนวน 30 ชิ้นงาน โดยผู้ตรวจ 3 คน ตรวจให้คะแนนสองวิธี คือ แบบใช้เกณฑ์และไม่ใช้เกณฑ์ พบว่า ค่าสัมประสิทธิ์ความเชื่อมั่น และดัชนีความสอดคล้องระหว่างผู้ตรวจ (Rater Agreement Index: RAI) ของการตรวจแบบไม่ใช้เกณฑ์มีค่าสูงกว่าการตรวจแบบใช้เกณฑ์ และงานวิจัยของ Saenplue (2013) ได้สร้างและพัฒนาเกณฑ์การตรวจให้คะแนนเรียงความ 3 รูปแบบที่มีคุณภาพที่เหมาะสมสำหรับการนำไปใช้ คือ แบบเกณฑ์รวม แบบเกณฑ์แยก และแบบผสมผสาน จึงสามารถกล่าวได้ว่าในการตรวจงานเขียนหรือเรียงความนั้น สามารถพัฒนาเกณฑ์ที่มีความสอดคล้องกันได้หลากหลายรูปแบบตามความเหมาะสมของวัตถุประสงค์การนำไปใช้ และความเหมาะสมกับสภาพการเรียนการสอน

2. จากการศึกษาคุณลักษณะของผู้ตรวจให้คะแนนโดยพิจารณาจากประสบการณ์ผู้ตรวจ พบว่ามีความเฉพาะของแต่ละบุคคล ไม่สามารถสรุปได้ว่ากลุ่มประสบการณ์น้อยจะมีความใจติกว่า หรือเข้มงวดกว่ากลุ่มประสบการณ์มากกว่า 10 ปี ซึ่งไม่สอดคล้องกับงานวิจัยของ Bowsiripon (2000) ที่ได้เปรียบเทียบค่าสัมประสิทธิ์การสรุปอ้างอิงของแบบทดสอบวิชาคณิตศาสตร์ที่เปรียบเทียบการให้คะแนนจากสองกลุ่มผู้ตรวจที่มีประสบการณ์ต่างกันแล้ว พบว่า ประสบการณ์ที่ต่างกันส่งผลให้คะแนนที่ได้มีค่าต่างกัน ซึ่งอาจเป็นเพราะการตรวจให้คะแนนงานเขียนภาษาต่างประเทศนั้นค่อนข้างขึ้นอยู่กับความสามารถเฉพาะบุคคลสูงกว่าเรื่องประสบการณ์ ซึ่งสอดคล้องกับงานวิจัยของ Leckie and Baird (2011) ที่ได้ทำการศึกษาผลกระทบที่เกิดจากผู้ตรวจในการให้คะแนนเรียงความโดยการใช้ Multilevel Models และการวิเคราะห์เพื่อศึกษาแนวโน้มและศึกษาความแตกต่างของการเปลี่ยนแปลงของความเข้มงวดของผู้ตรวจเมื่อช่วงเวลาผ่านไป (Severity Drift) และการประเมินแบบกลางหรือการไม่กล้าตัดสินใจของผู้ตรวจ (Central Tendency Effects) โดยพิจารณาจากประสบการณ์ของผู้ตรวจ ซึ่งพบว่า ผู้ตรวจที่มีประสบการณ์น้อยจะเข้มงวดกว่าผู้ตรวจที่มีประสบการณ์ แต่ไม่พบระดับนัยสำคัญ และพบว่า ความเข้มงวดของผู้ตรวจไม่คงที่เมื่อข้ามช่วงเวลาอย่างมีนัยสำคัญ

เมื่อพิจารณาการระบุคุณลักษณะของผู้ประเมินความเข้มงวด ใจติรายบุคคล เมื่อใช้เกณฑ์การตรวจให้คะแนนต่างกัน พบว่า มีผู้ตรวจให้คะแนนสอดคล้องกันทั้งสามรูปแบบ ร้อยละ 37.5 และไม่ผู้ตรวจให้คะแนนคนใดที่มีคุณลักษณะเปลี่ยนไปแบบสุดขั้ว คือเปลี่ยนจากเข้มงวดเป็นใจติ หรือใจติเป็นเข้มงวด ร้อยละ 100 แสดงว่าความเข้มงวด

ใจดีของผู้ตรวจไม่เปลี่ยนไปตามรูปแบบของเกณฑ์การตรวจที่ใช้ ซึ่งสอดคล้องกับงานวิจัยของ Saenplue (2013) ที่ได้ทำการศึกษาระบบความเข้มงวด/ใจดีของผู้ตรวจให้คะแนนเรียงความภาษาไทย โดยใช้เกณฑ์ 3 รูปแบบ ได้แก่ แบบแยกองค์ประกอบ แบบรวมองค์ประกอบ และแบบผสมผสาน ซึ่งพบว่า การระบุความเข้มงวดใจดีของผู้ตรวจให้คะแนนไม่แปรเปลี่ยนไปแม้ว่าจะเปลี่ยนรูปแบบการตรวจให้คะแนน ในงานวิจัยนี้ก็เช่นกัน ผู้ตรวจให้คะแนนจะมีคุณลักษณะค่อนข้างคงที่ไม่ว่าจะเปลี่ยนไปใช้วิธีการตรวจให้คะแนนแบบใดก็ตาม ทั้งนี้ อาจมาจากคุณลักษณะของผู้ตรวจให้คะแนนเป็นสิ่งที่ติดตัวมาก่อนแล้ว ซึ่งค่อนข้างที่จะคงที่ไม่แปรเปลี่ยนโดยง่าย และเกณฑ์ได้คุณภาพที่เหมาะสมสอดคล้องกันทั้ง 3 รูปแบบไม่ได้มีแบบใดแบบหนึ่งที่เอื้อต่อการให้คะแนนแบบเข้มงวดเกินไปหรือใจดีเกินไป คุณลักษณะของผู้ตรวจจึงค่อนข้างคงที่ และหากมีการเปลี่ยนแปลงก็จะเกิดขึ้นเพียงเข้มงวด-เป็นกลาง หรือใจดี-เป็นกลาง ไม่ปรากฏคุณลักษณะแบบใจดี-เข้มงวด

ข้อเสนอแนะจากการวิจัย

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

จากผลการวิจัย พบว่า เกณฑ์ที่สร้างขึ้นมีคุณภาพที่เหมาะสมและมีความสอดคล้องกันทั้ง 3 เกณฑ์ จึงสามารถนำรูปแบบเกณฑ์ดังกล่าวมาใช้ในการประเมินหรือตรวจให้คะแนนงานเขียนได้ โดยที่มีการทำความเข้าใจต่อเกณฑ์ และการอบรมหรือให้ความรู้ด้านการวัดและประเมินผล หรือการสร้างเกณฑ์การประเมินให้กับครูผู้เกี่ยวข้องก่อนการทำวิจัยในลักษณะดังกล่าวนี้จะทำให้ได้ผลการประเมินที่น่าเชื่อถือ และพบว่า ทุกเกณฑ์จำแนกผู้ตรวจที่มีคุณลักษณะเป็นกลางแตกต่างกัน และมีผู้ตรวจที่มีคุณลักษณะคงที่อยู่เพียง ร้อยละ 37.50 คือ เข้มงวดและใจดี จึงยังไม่สามารถสรุปได้ว่าการใช้เกณฑ์ในการประเมินที่แตกต่างกัน หรือประสบการณ์ที่แตกต่างกันไม่ได้ทำให้คุณลักษณะของผู้ตรวจให้คะแนนแปรเปลี่ยนไป

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

ในการทำวิจัยครั้งต่อไปควรมีการศึกษามุมหลังของผู้ตรวจให้คะแนนในฐานะแหล่งของความคลาดเคลื่อนในการให้คะแนน เช่น ความสามารถด้านภาษา และควรมีการศึกษาในเชิงคุณภาพต่อไปเนื่องจากยังมีผู้ทำการศึกษาเปรียบเทียบในลักษณะนี้ไม่มากนัก จึงควรมีการศึกษาเพิ่มเติมในลักษณะเดียวกัน หรืออาจศึกษาการใช้เกณฑ์แบบต่างๆ นี้ในบริบทที่แตกต่างกัน

กิตติกรรมประกาศ

งานวิจัยนี้ได้รับทุนอุดหนุนการวิจัยนักศึกษาระดับปริญญาโท ปีงบประมาณ 2559 มหาวิทยาลัยเชียงใหม่

References

- Aiamsaard, A., Pansri, O., Parnichparinchai, T. (2016). Factors analysis of English teacher characteristics in school under the Secondary Educational Service Area Office. *Journal of Educational Naresuan University*, 18(1), 184-192. [in Thai]
- Bowsiripon, U. (2000). *A comparison of generalizability coefficients of mathematics tests with different scoring methods, numbers of raters and experience of raters* (Master thesis). Bangkok: Srinakharinwirot University. [in Thai]
- Dunsmuir, S., Kyriacou, M., Batuwitige, S., Hinson, E., Ingram, V., & O'Sullivan, S. (2015). An evaluation of the writing assessment measure (WAM) for children's narrative writing. *Assessing Writing*, 23, 1 – 18.
- George, D., & Mallery, P. (2003). *SPSS for Windows step by step: A simple guide and reference 11.0 update* (4th ed.). Boston: Allyn & Bacon.
- 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* (Doctoral dissertation). Bangkok: Srinakharinwirot University. [in Thai]
- Leckie, G., & Baird, J. A. (2011). Rater effects on essay scoring: A multilevel analysis of severity drift, central tendency, and rater experience. *Journal of Educational Measurement*, 48, 399–418.
- Linacre, J. M. (2012). *Many-Facet Rasch Measurement: Facets Tutorial*.
<http://winsteps.com/tutorials.htm>
- Mertler, C. A. (2001). Designing scoring rubrics for your classroom. *Practical Assessment, Research & Evaluation*, 7, 1-10.
- Prieto, G., & Nieto, E. (2014). Analysis of rater severity on written expression exam using many faceted Rasch measurement. *Journal of Psicologica*, 35, 385-397.
- Saenplue, B. (2013). *The study of agreement of rater's severity/leniency and differential rater functioning over time on essay composition by the elementary schooling grade 3* (Doctoral dissertation). Bangkok: Srinakharinwirot University. [in Thai]
- Schaefer, E. (2008). Rater bias patterns in an EFL writing assessment. *Language Testing*, 25(4), 465-493.
- Suaysod, K. (2011). A comparison of reliability and objectivity of score that scored by using and not using scoring rubric. *Kasalongkham Research Journal*, 5(2), 107-114. [in Thai]
- Wongpanich, P. (2000). *Measurement and evaluation in undergraduate education*. Bangkok: Research and Development, office of the Higher Education Commission. [in Thai]