

การเปรียบเทียบประสิทธิภาพของแบบวัดความสามารถในการสร้างคำอธิบาย
ทางวิทยาศาสตร์ ระหว่างแบบอัตร้อยจำกัดคำตอบและไม่จำกัดคำตอบ
สำหรับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4

COMPARING THE EFFICIENCY BETWEEN ABILITY IN CONSTRUCTING
SCIENTIFIC EXPLANATION RESTRICTED- AND EXTENDED-RESPONSE
SUBJECTIVE TESTS FOR MATHAYOM SUKSA 4 STUDENTS

ปิริยะ วรณไทย^{1*} และ ชนินันท์ พฤกษ์ประมูล²

Piriya Wannathai^{1*} and Chaninan Pruekpramool²

มหาวิทยาลัยศรีนครินทรวิโรฒ 114 ซอยสุขุมวิท 23 เขตวัฒนา กรุงเทพมหานคร 10110^{1,2}
Srinakharinwirot University, 114 Soi Sukhumvit 23, Khlong Toei Nuea, Watthana, Bangkok 10110^{1,2}

*Corresponding author E-mail: piriya.wannathai@g.swu.ac.th

(Received: 22 Jul, 2021; Revised: 17 Dec, 2021; Accepted: 21 Dec, 2021)

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาและเปรียบเทียบประสิทธิภาพของแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์แบบอัตร้อยจำกัดคำตอบและไม่จำกัดคำตอบ สำหรับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 กลุ่มตัวอย่างคือ นักเรียนชั้นมัธยมศึกษาปีที่ 4 จำนวน 120 คน จาก 6 โรงเรียนในจังหวัดเพชรบุรี ได้มาด้วยวิธีการเลือกแบบเจาะจง วิเคราะห์ข้อมูลโดยหาค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าสัมประสิทธิ์สหสัมพันธ์ภายใน (ICC) ค่าความยาก ค่าอำนาจจำแนก ค่าสัมประสิทธิ์แอลฟาของครอนบาค และการทดสอบค่าทีแบบสองกลุ่มที่ไม่อิสระจากกัน

ผลการวิจัย พบว่า 1) แบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ที่พัฒนาขึ้น ประกอบด้วยแบบอัตร้อยจำกัดคำตอบที่กำหนดให้เขียนคำตอบแยกตามองค์ประกอบของการสร้างคำอธิบายทางวิทยาศาสตร์ และแบบอัตร้อยไม่จำกัดคำตอบที่สามารถเขียนคำตอบได้อย่างอิสระ แต่ละแบบวัดมีข้อคำถาม จำนวน 10 ข้อ ไม่จำกัดกลุ่มเนื้อหา และมีการให้ข้อมูลทางวิทยาศาสตร์ที่เพียงพอต่อการสรุปหลักการทางวิทยาศาสตร์มาใช้ในการตอบคำถาม และเกณฑ์การให้คะแนนแบบแยกองค์ประกอบเป็นรายข้อ มีค่าความยาก 0.26 – 0.47 และค่าอำนาจจำแนก 0.51 – 0.93 และ 2) แบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์แบบอัตร้อยไม่จำกัดคำตอบมีประสิทธิภาพสูงกว่าแบบจำกัดคำตอบ นักเรียนมีคะแนนเฉลี่ยความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์จากแบบอัตร้อยไม่จำกัดคำตอบสูงกว่าแบบจำกัดคำตอบ อย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($t = 6.027$, $p = .000$)

คำสำคัญ: แบบวัดความสามารถ, การสร้างคำอธิบายทางวิทยาศาสตร์, แบบวัดอัตร้อยจำกัดคำตอบ, แบบวัดอัตร้อยไม่จำกัดคำตอบ

¹ นิสิตหลักสูตรการศึกษาดุษฎีบัณฑิต สาขาวิทยาศาสตร์ศึกษา ศูนย์วิทยาศาสตร์ศึกษา คณะวิทยาศาสตร์

² ผู้ช่วยศาสตราจารย์ ดร. สาขาวิทยาศาสตร์ศึกษา ศูนย์วิทยาศาสตร์ศึกษา คณะวิทยาศาสตร์

ABSTRACT

The purposes of this research were to develop and compare the efficiency between ability in constructing scientific explanation tests, restricted- and extended-response subjective tests for Mathayom Suksa 4 students. The samples were 120 Mathayom Suksa 4 students from 6 schools in Phetchaburi province, selected by purposive sampling. The data were analyzed by mean, standard deviation, Intraclass Correlation Coefficient (ICC), difficulty, discrimination, Cronbach's alpha coefficient, and t-test for dependent samples.

The results found that 1) the ability in constructing scientific explanation tests composed of the restricted-response subjective test, which required to write the answers separately according to the components of the scientific explanation and the extended-response subjective test, which allowed to write the answer freely. Each form of the tests consisted of 10 questions with nonspecific science contents and provided sufficient scientific information for summarizing scientific principles to answer the questions and specific scoring rubric for each question. The difficulty indices of each item were 0.26 – 0.47 and the discrimination indices of each item were 0.51 – 0.93. And 2) the efficiency of ability in constructing scientific explanation extended-response subjective test was higher than the restricted-response subjective test by comparing the students' mean scores from the tests. The students' mean score from the extended-response subjective test was higher than the restricted-response subjective test at a statistical significance level of .05 ($t = 6.027$, $p = .000$).

KEYWORDS: Ability Test, Constructing Scientific Explanation, Restricted-response Subjective Test, Extended-response Subjective Test

บทนำ

การสร้างคำอธิบายทางวิทยาศาสตร์ เป็นกระบวนการที่ใช้ในการอธิบายปรากฏการณ์ที่เกิดขึ้น โดยใช้พื้นฐานความรู้ทางวิทยาศาสตร์ (McNeill & Krajcik, 2008b) คำอธิบายทางวิทยาศาสตร์ ประกอบด้วย 3 ส่วน คือ 1) ข้อกล่าวอ้าง เป็นคำตอบของปรากฏการณ์ทางวิทยาศาสตร์ตามความเข้าใจของนักเรียน 2) หลักฐาน เป็นข้อมูลทางวิทยาศาสตร์ที่ได้จากการค้นคว้าหรือลงข้อสรุปจากแหล่งข้อมูลที่เชื่อถือเพื่อใช้สนับสนุนข้อกล่าวอ้าง และ 3) การให้เหตุผล เป็นหลักการทางวิทยาศาสตร์ที่ใช้การอธิบายความเชื่อมโยงระหว่างข้อกล่าวอ้างและหลักฐานที่นำมาสนับสนุน (McNeill & Krajcik, 2008b) โดยกระบวนการได้มาซึ่งหลักฐานที่มีความเหมาะสมต่อการนำมาสนับสนุนคำอธิบายทางวิทยาศาสตร์ของนักเรียน ต้องมีการคิดวิเคราะห์ ไตร่ตรอง และประเมินข้อมูลก่อนตัดสินใจเชื่อ ซึ่งเป็นส่วนหนึ่งของทักษะการคิดอย่างมีวิจารณญาณ (วิจิตร คุสมภ์ และสุลี ทองวิเชียร, 2560) การวัดประเมินผลความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ จำเป็นต้องอาศัยเครื่องมือที่มีคุณภาพ เพื่อสะท้อนให้เห็นว่า นักเรียนมีความสามารถนี้อยู่ในระดับใด และมีปัญหาเกี่ยวกับความสามารถนี้อย่างไร ผลที่ได้จากการนำเครื่องมือไปวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ของนักเรียน สามารถนำมาใช้เป็นข้อมูลพื้นฐานในการออกแบบการจัดการเรียนรู้เพื่อส่งเสริมให้นักเรียนมีความสามารถนี้ในระดับที่สูงขึ้นต่อไป

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องกับการวัดและประเมินความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ของนักเรียน พบว่า ความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์สามารถวัดได้จากการเขียน (Yao & Guo, 2018) หรือการพูดอธิบายปรากฏการณ์ทางวิทยาศาสตร์นั้น ๆ (Andrade, Shwartz, Freire, & Baptista, 2021) นักวิจัยและนักการศึกษาส่วนใหญ่วัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ผ่านการเขียน โดยให้นักเรียนทำแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ ที่มีข้อคำถามจำนวน 4 – 5 ข้อ มีระยะเวลาในการทำประมาณ 60 นาที และมีเกณฑ์การให้คะแนนครอบคลุมทุกองค์ประกอบของคำอธิบายทาง

วิทยาศาสตร์ ได้แก่ ข้อกล่าวอ้าง หลักฐาน และการให้เหตุผล (McNeill & Krajcik, 2008b)

แบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ที่ผ่านมา มีข้อจำกัดอยู่ 2 ประการ ประการแรก คือ แบบวัดที่สร้างขึ้นมักเป็นแบบวัดที่อิงเนื้อหาวิทยาศาสตร์ตามสิ่งที่ผู้เรียนได้เรียนรู้ในรายวิชา ทำให้ไม่สามารถนำไปใช้ในการวัดและประเมินผลในเนื้อหาวิชาทางวิทยาศาสตร์อื่น ๆ ได้ (McNeill & Krajcik, 2008a; พัฒนิตา มีลา และร่มเกล้า อาจเดช, 2560; สุทธิชาติ เปรมกมล และสกลรัตน์ แก้วดี, 2560) ประการที่ 2 คือ แบบวัดที่สร้างขึ้นมักมีข้อคำถามที่ผู้เรียนต้องมีความรู้หรือมีประสบการณ์ที่เคยเรียนมาก่อนแล้ว หากนักเรียนคนใดที่ไม่มีความรู้ความเข้าใจหรือไม่ประสบการณ์ในเนื้อหาตามข้อคำถามมาก่อน จะไม่สามารถตอบคำถามนั้นได้ (สุทธิชาติ เปรมกมล และสกลรัตน์ แก้วดี, 2560) จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง พบว่า แบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์เป็นแบบวัดแบบอัตนัยที่มีแนวทางการเขียนคำตอบแตกต่างกันอยู่สองรูปแบบ คือ 1) แบบอัตนัยจำกัดคำตอบ ที่กำหนดพื้นที่ในการเขียนคำอธิบายแยกตามองค์ประกอบของการสร้างคำอธิบายทางวิทยาศาสตร์ (พัฒนิตา มีลา และร่มเกล้า อาจเดช, 2560) และ 2) แบบอัตนัยไม่จำกัดคำตอบ ที่กำหนดพื้นที่ในการเขียนคำอธิบายทางวิทยาศาสตร์แบบปลายเปิด โดยไม่ต้องแยกตามองค์ประกอบที่กำหนดให้ (สุทธิชาติ เปรมกมล และสกลรัตน์ แก้วดี, 2560; Yao, Guo & Neumann, 2016) ซึ่งแบบวัดแต่ละรูปแบบนั้นยังไม่ปรากฏผลการศึกษายืนยันว่ารูปแบบใดเหมาะสมต่อการนำไปใช้ในการวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ของนักเรียน

ด้วยข้อจำกัดของแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ที่ผ่านมา จึงเห็นควรมีการพัฒนาและเปรียบเทียบประสิทธิภาพของแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์แบบอัตนัยจำกัดคำตอบและไม่จำกัดคำตอบ โดยมีลักษณะเป็นแบบวัดที่ไม่จำกัดกลุ่มเนื้อหา เพื่อให้สามารถนำไปใช้ได้กับทุกรายวิชาทางวิทยาศาสตร์ มีการให้ข้อมูลทางวิทยาศาสตร์ที่เพียงพอสำหรับให้ผู้เรียนที่ไม่มีความรู้มาก่อน สามารถ

สรุปหลักการและหลักฐานทางวิทยาศาสตร์มาใช้ในการตอบคำถามได้ และมีการสร้างเกณฑ์การให้คะแนนเป็นรายข้อ ที่มีการแสดงตัวอย่างคำตอบ เพื่อให้ง่ายต่อการนำไปใช้และสร้างความเข้าใจที่ตรงกันในกรณีที่มีผู้ตรวจหลายคน จนเป็นแนวทางให้กับผู้ที่สนใจศึกษาเกี่ยวกับความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์นำไปประยุกต์ใช้วัดความสามารถของผู้เรียนได้อย่างเหมาะสม สะท้อนความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ที่แท้จริงของนักเรียน และนำผลการประเมินที่ได้มาใช้เป็นแนวทางในการออกแบบเทคนิค วิธีการ หรือรูปแบบจัดการเรียนรู้ เพื่อส่งเสริมความสามารถนี้ของผู้เรียนให้สูงขึ้นต่อไป

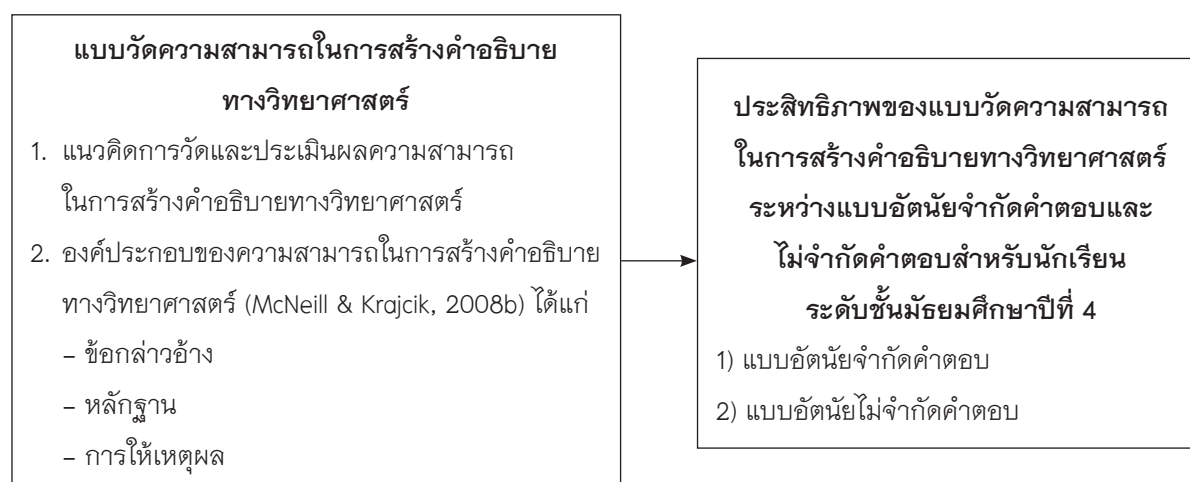
วัตถุประสงค์ของการวิจัย

1. เพื่อพัฒนาแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์แบบอัตนัยจำกัดคำตอบและไม่จำกัดคำตอบ สำหรับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4

2. เพื่อเปรียบเทียบประสิทธิภาพของแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์แบบอัตนัยจำกัดคำตอบและไม่จำกัดคำตอบ สำหรับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4

กรอบแนวคิดการวิจัย

จากการศึกษาเอกสารที่เกี่ยวข้องกับการวัดและประเมินผลความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ของนักเรียน ตลอดจนองค์ประกอบของความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ เพื่อเป็นแนวทางในการพัฒนาแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 4 นำมากำหนดเป็นกรอบแนวคิดการวิจัย ดังภาพที่ 1



ภาพที่ 1 กรอบแนวคิดการวิจัย

วิธีดำเนินการวิจัย

งานวิจัยครั้งนี้ เป็นงานวิจัยและพัฒนา (Research and Development) มีรายละเอียดดังนี้

ประชากร

นักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 ที่ศึกษาในภาคเรียนที่ 1 ปีการศึกษา 2562 สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเพชรบุรี

กลุ่มตัวอย่าง

นักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 ที่ศึกษาในภาคเรียนที่ 1 ปีการศึกษา 2562 จาก 6 โรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษาเพชรบุรี ประกอบด้วย โรงเรียนขนาดเล็ก กลาง และใหญ่ ขนาดละ 2 โรงเรียน โรงเรียนละ 20 คน รวมเป็น 120 คน ที่ได้มาด้วยวิธีการเลือกแบบเจาะจง โดยเป็นโรงเรียนที่สมัครใจตอบรับให้นักเรียนเข้าร่วมเป็นกลุ่มตัวอย่าง

ผู้วิจัยดำเนินการวิจัยและพัฒนาแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ 4 ขั้นตอน ตามแนวทางของ Johnson and Foa (1996) ดังนี้

ขั้นตอนที่ 1 การวิจัย (R_1): กำหนดจุดมุ่งหมายในการพัฒนาเครื่องมือ

1. สังเคราะห์ความหมาย ความสำคัญ องค์ประกอบ วิธีการวัดและประเมินผลความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ จากเอกสารและงานวิจัยที่เกี่ยวข้อง เพื่อให้ได้แนวทางการพัฒนาแบบวัด

2. วิเคราะห์การสร้างและพัฒนาแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ที่ผ่านมา ทั้งลักษณะของข้อมูลที่กำหนดให้ ลักษณะของข้อคำถาม ตลอดจนข้อจำกัดของเครื่องมือต่าง ๆ ที่มีการพัฒนาขึ้น เพื่อนำไปสู่การพัฒนาแบบวัดที่ลดข้อจำกัดต่าง ๆ ที่เกิดขึ้น

ขั้นตอนที่ 2 การพัฒนา (D_1): ออกแบบและพัฒนาเครื่องมือ

1. กำหนดนิยามและองค์ประกอบของความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ เพื่อเป็นกรอบในการพัฒนาแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์

2. สร้างแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ที่มีลักษณะเป็นข้อสอบแบบอัตนัยตามแนวคิดของ McNeill and Krajcik (2008a) จำนวน 2 รูปแบบ รูปแบบละ 10 ข้อคำถาม โดยมีข้อคำถามที่เหมือนกันทั้งสองรูปแบบ ข้อคำถามที่ใช้มีลักษณะที่ไม่อิงเนื้อหาตามรายวิชา มีการให้ข้อมูลทางวิทยาศาสตร์ที่เพียงพอสำหรับให้นักเรียนที่ไม่มีความรู้มาก่อน สามารถนำมาใช้ในการสร้างคำอธิบายทางวิทยาศาสตร์ได้ ทั้งนี้แบบอัตนัยจำกัดคำตอบ กำหนดให้เขียนคำตอบแยกตามองค์ประกอบของคำอธิบายทางวิทยาศาสตร์ ได้แก่ 1) ข้อกล่าวอ้าง โดยถามว่า “คำตอบของคำถามคืออะไร” 2) หลักฐาน โดยถามว่า “ข้อมูลหรือหลักฐานที่นักเรียนใช้ในการสนับสนุนคำตอบคืออะไร พร้อมอธิบาย” และ 3) การให้เหตุผล โดยถามว่า “เพราะเหตุใด นักเรียนจึงตอบเช่นนั้น” แสดงตัวอย่างดังภาพที่ 2 ส่วนแบบอัตนัยไม่จำกัดคำตอบเปิดให้นักเรียนได้เขียนคำอธิบายทางวิทยาศาสตร์อย่างอิสระ ซึ่งจำเป็นต้องใช้ข้อความเพิ่มเติมเพื่อชี้ให้นักเรียนเขียนให้ครบถ้วนตามองค์ประกอบว่า “นักเรียนพิจารณาจากสิ่งใด พร้อมให้เหตุผลประกอบ” แสดงตัวอย่างดังภาพที่ 3

ให้นักเรียนเขียนคำอธิบายทางวิทยาศาสตร์ที่แสดงให้เห็นว่า
“กิจกรรมการล่ามทะเล ส่งผลให้จำนวนประชากรปูทะเลยักษ์ในปี 1975
มีการเปลี่ยนแปลงไป จากปี 1972 อย่างไร”

1) คำตอบของคำถาม คืออะไร

2) ข้อมูลหรือหลักฐานที่นักเรียนใช้ในการสนับสนุนคำตอบ คืออะไร พร้อมอธิบาย

3) เพราะเหตุใด นักเรียนจึงตอบเช่นนั้น

ภาพที่ 2 ตัวอย่างลักษณะคำถามและพื้นที่เขียนคำตอบในแบบอัตนัยจำกัดคำตอบ

ให้นักเรียนเขียนคำอธิบายทางวิทยาศาสตร์ที่แสดงให้เห็นว่า
 “กิจกรรมการล่านากทะเล ส่งผลให้จำนวนประชากรปูทะเลยักษ์ในปี 1975
 มีการเปลี่ยนแปลงไป จากปี 1972 อย่างไร นักเรียนพิจารณาจากสิ่งใด
 พร้อมให้เหตุผลประกอบ”

คำอธิบายทางวิทยาศาสตร์

.....

.....

.....

.....

.....

ภาพที่ 3 ตัวอย่างลักษณะคำถามและพื้นที่เขียนคำตอบในแบบอัตโนมัติไม่จำกัดคำตอบ

3. สร้างเกณฑ์การให้คะแนนแบบวัดความสามารถในการสร้างคำอธิบายวิทยาศาสตร์เป็นรายข้อ ตามแนวคิดของ McNeill and Krajcik (2008a) ที่มีการประเมินแบบแยกองค์ประกอบของคำอธิบายทางวิทยาศาสตร์ คือ ข้อกล่าวอ้าง หลักฐาน และการให้เหตุผล องค์ประกอบละ

3 ระดับคะแนน คือ 0 1 และ 2 รวมเป็นคะแนนเต็มข้อละ 6 คะแนน พร้อมระบุตัวอย่างของคำตอบที่ถูกต้องในแต่ละระดับคะแนนของแต่ละองค์ประกอบ เพื่อเป็นแนวทางในการตรวจให้คะแนน แสดงตัวอย่างดังตารางที่ 1

ตารางที่ 1 ตัวอย่างเกณฑ์การให้คะแนนแบบวัดความสามารถในการสร้างคำอธิบายวิทยาศาสตร์แบบแยกองค์ประกอบเป็นรายข้อ คำถามเกี่ยวกับผลกระทบจากกิจกรรมการล่านากทะเล

คะแนน	เกณฑ์การให้คะแนน		
	ข้อกล่าวอ้าง	หลักฐาน	การให้เหตุผล
2	ระบุได้ถูกต้องว่า “ประชากรปูทะเลยักษ์มีจำนวนลดลง”	แสดงหลักฐานถูกต้องและครบทั้ง 2 อย่าง คือ “ประชากรเมนทะเลมีแนวโน้มเพิ่มขึ้น” และ “สาหร่ายเคลป์มีแนวโน้มลดลง”	มีการให้เหตุผลอย่างถูกต้องว่า “เมื่อสาหร่ายเคลป์ที่เป็นแหล่งอาหารภายในสายใยอาหารลดลง จะทำให้ประชากรของปูทะเลยักษ์ที่บริโภคอาหารนั้นลดลง”
1	ระบุได้ถูกต้อง แต่ไม่ชัดเจน เช่น “เป็นไปในทิศทางเดียวกับประชากรของนากทะเล” หรือ “เป็นไปในทิศทางเดียวกับประชากรของสาหร่ายเคลป์”	แสดงหลักฐานถูกต้อง เพียงอย่างใดอย่างหนึ่ง เช่น “ประชากรเมนทะเลมีแนวโน้มเพิ่มขึ้น” หรือ “สาหร่ายเคลป์มีแนวโน้มลดลง”	ให้เหตุผล โดยกล่าวถึงหลักฐานซ้ำ เช่น “ประชากรนากทะเลที่ลดลง จะทำให้ประชากรเมนทะเลมีแนวโน้มเพิ่มขึ้น และทำให้สาหร่ายเคลป์มีแนวโน้มลดลง”
0	ไม่ระบุข้อกล่าวอ้าง หรือระบุไม่ถูกต้อง เช่น “ประชากรปูทะเลยักษ์มีจำนวนเพิ่มขึ้น” หรือ “ไม่มีการเปลี่ยนแปลงจำนวน”	ไม่แสดงหลักฐาน หรือ แสดงหลักฐานไม่ถูกต้องบางส่วนหรือทั้งหมด เช่น “ประชากรเมนทะเลมีแนวโน้มลดลง” และ/หรือ “สาหร่ายเคลป์มีแนวโน้มเพิ่มขึ้น”	ไม่มีการให้เหตุผล หรือ ให้เหตุผลโดยใช้หลักการทางวิทยาศาสตร์ไม่ถูกต้อง เช่น “เมื่อแหล่งอาหารภายในสายใยอาหารลดลง จะทำให้ประชากรของสิ่งมีชีวิตที่บริโภคอาหารนั้นเพิ่มขึ้น”

4. นำแบบวัดความสามารถ และเกณฑ์การให้คะแนน ไปให้ผู้เชี่ยวชาญด้านวิทยาศาสตร์ศึกษา 3 ท่าน พิจารณาความตรงเชิงเนื้อหา (Content Validity) ความเหมาะสมของข้อคำถามและเกณฑ์การให้คะแนนต่อการนำไปใช้ พร้อมให้ข้อเสนอแนะ จากนั้นนำผลมาวิเคราะห์หาค่าดัชนีความสอดคล้อง (Index of Item Objective Congruence: IOC) ระหว่างข้อคำถามกับนิยามความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ โดยข้อคำถามที่มีความตรงเชิงเนื้อหาและนำไปใช้ได้ต้องมีค่า IOC มากกว่า 0.5 (วิเชียร อินทรสมพันธ์ และคนอื่น ๆ, 2563)

การวิเคราะห์ความเหมาะสมของเกณฑ์การให้คะแนนต่อการนำไปใช้ โดยให้ผู้เชี่ยวชาญพิจารณาแบ่งเป็น 5 ระดับ (Rating Scale) จากนั้นนำมาวิเคราะห์หาค่าเฉลี่ย และเปรียบเทียบกับเกณฑ์ในการแปลความหมายระดับความเหมาะสม 5 ช่วง คือ 4.51 – 5.00 หมายถึง เหมาะสมมากที่สุด 3.51 – 4.50 หมายถึง เหมาะสมมาก 2.51 – 3.50 หมายถึง เหมาะสมปานกลาง 1.51 – 2.50 หมายถึง เหมาะสมน้อย และ 1.00 – 1.50 หมายถึง เหมาะสมน้อยที่สุด (บุญชม ศรีสะอาด, 2560)

5. ปรับปรุงแก้ไขแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์และเกณฑ์การให้คะแนนตามข้อเสนอแนะของผู้เชี่ยวชาญ

ขั้นตอนที่ 3 การวิจัย (R_2): ทดลองใช้

1. นำแบบวัดความสามารถ ที่สร้างขึ้น ทั้งสองรูปแบบ ไปทดลองใช้กับกลุ่มตัวอย่าง จำนวน 120 คน และเพื่อลดข้อจำกัดในการทำแบบวัดแต่ละรูปแบบก่อนหรือหลังที่จะมีผลต่อการวิจัยจึงแบ่งกลุ่มตัวอย่างเป็น 2 กลุ่ม กลุ่มละ 60 คน แต่ละคนจะได้ทำแบบวัดทั้งสองรูปแบบ โดยการวัดครั้งแรก กลุ่มที่ 1 ทำแบบวัดจำกัดคำตอบ ส่วนกลุ่มที่ 2 ทำแบบวัดไม่จำกัดคำตอบ และการวัดครั้งที่สอง ให้สลับกลุ่มที่ 1 ทำแบบวัดไม่จำกัดคำตอบ ส่วนกลุ่มที่ 2 ทำแบบวัดจำกัดคำตอบ ผู้วิจัยดำเนินการวัดครั้งแรกและครั้งที่สองห่างกัน 4 สัปดาห์ เพื่อไม่ให้กลุ่มตัวอย่างสามารถจดจำข้อคำถามของแบบวัดได้

2. วิเคราะห์ค่าความเชื่อมั่นระหว่างผู้ประเมิน (Inter-rater Reliability: IRR) เพื่อพิจารณาความเป็นปรนัย

ของเกณฑ์การให้คะแนน โดยผู้ประเมิน 3 ท่าน ประกอบด้วย ผู้วิจัยและผู้ตรวจอีก 2 ท่าน ในฐานะผู้ร่วมวิจัย ตรวจให้คะแนนตามเกณฑ์การให้คะแนน แล้วคำนวณโดยหาค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (Intraclass Correlation Coefficient: ICC) (Secolsky & Denison, 2017)

3. วิเคราะห์ค่าความยาก (Difficulty) และค่าอำนาจจำแนก (Discrimination) ของข้อคำถามเป็นรายข้อ ด้วยเทคนิค ร้อยละ 25 โดยเรียงลำดับคะแนนของกลุ่มตัวอย่างจากมากไปน้อย แล้วแบ่งกลุ่มตัวอย่างออกเป็นกลุ่มสูงและกลุ่มต่ำ จำนวนร้อยละ 25 ของกลุ่มตัวอย่างทั้งหมด จากนั้นวิเคราะห์หาค่าความยากและอำนาจจำแนกโดยใช้ข้อมูลจากกลุ่มสูงและกลุ่มต่ำ ทั้งนี้แต่ละข้อคำถามควรมีค่าความยากอยู่ระหว่าง 0.20 – 0.80 และมีค่าอำนาจจำแนก 0.20 ขึ้นไป (วิเชียร อินทรสมพันธ์ และคนอื่น ๆ, 2563)

4. วิเคราะห์ค่าความเชื่อมั่น (Reliability) ทั้งฉบับของแบบวัดทั้งสองรูปแบบ โดยนำคะแนนที่นักเรียนทำได้ในแต่ละแบบวัด มาวิเคราะห์ค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha Coefficient) ทั้งนี้ ค่าความเชื่อมั่นของเครื่องมือในระดับที่นำไปใช้ได้ควรมีค่าตั้งแต่ 0.70 ขึ้นไป (วิเชียร อินทรสมพันธ์ และคนอื่น ๆ, 2563)

ขั้นตอนที่ 4 การพัฒนา (D_2): เปรียบเทียบประสิทธิภาพและปรับปรุง

1. การเปรียบเทียบประสิทธิภาพของแบบวัดทั้งสองรูปแบบ โดยพิจารณาจากความแตกต่างของคะแนนเฉลี่ยความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์จากแบบวัดนี้ไม่จำกัดคำตอบและแบบจำกัดคำตอบ โดยวิเคราะห์จากการทดสอบค่าที่แบบสองกลุ่มที่ไม่อิสระจากกัน (t-test for Dependent Samples) ที่ระดับนัยสำคัญทางสถิติ .05

2. ปรับปรุงแบบวัดทั้งสองรูปแบบภายหลังการทดลองใช้

การวิเคราะห์ข้อมูล

วิเคราะห์ข้อมูลโดยหาค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน ค่าความยาก ค่าอำนาจจำแนก ค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น (ICC) ค่าสัมประสิทธิ์แอลฟาของครอนบาค และการทดสอบค่าที่แบบสองกลุ่มที่ไม่อิสระจากกัน

ผลการวิจัย

จากการพัฒนาและเปรียบเทียบประสิทธิภาพของแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 4 นำเสนอผลการวิจัยได้ดังนี้

1. การพัฒนาแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์และเกณฑ์การให้คะแนน

1.1 จากการวิเคราะห์ความตรงเชิงเนื้อหาของแบบวัดความสามารถ จำนวน 10 ข้อ โดยให้ผู้เชี่ยวชาญ 3 ท่าน พิจารณาความสอดคล้องระหว่างข้อคำถามกับนิยามความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์และความเหมาะสมของเกณฑ์การให้คะแนนต่อการนำไปใช้พบว่า ทุกข้อคำถามมีค่า IOC เท่ากับ 1.00 และเกณฑ์

ที่สร้างขึ้นมีความเหมาะสมต่อการนำไปใช้อยู่ในระดับมากที่สุด ($\bar{x} = 5.00$, S.D. = 0.00)

1.2 จากการพิจารณาความเป็นปรนัยของเกณฑ์การให้คะแนน โดยวิเคราะห์ค่าความเชื่อมั่นระหว่างผู้ประเมิน (IRR) 3 ท่าน ประกอบด้วย ผู้วิจัยและผู้ตรวจอีก 2 ท่าน ในฐานะผู้ร่วมวิจัย แล้วนำคะแนนที่ตรวจได้มาคำนวณค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น ที่ความเชื่อมั่นร้อยละ 95 พบว่า เกณฑ์การให้คะแนนแบบอัตนัยจำกัดคำตอบ มีค่าความเชื่อมั่นระหว่างผู้ประเมินในภาพรวมอยู่ระหว่าง 0.73 – 1.00 ส่วนเกณฑ์การให้คะแนนแบบอัตนัยไม่จำกัดคำตอบ มีค่าระหว่าง 0.70 – 1.00 แสดงให้เห็นว่าเกณฑ์การให้คะแนนที่สร้างขึ้นมีความเป็นปรนัยสูง ดังตารางที่ 2

ตารางที่ 2 การวิเคราะห์ค่าความเชื่อมั่นระหว่างผู้ประเมิน (IRR) เพื่อพิจารณาความเป็นปรนัยเกณฑ์การให้คะแนนแบบอัตนัยจำกัดคำตอบและไม่จำกัดคำตอบ

ข้อ	แบบอัตนัยจำกัดคำตอบ			แบบอัตนัยไม่จำกัดคำตอบ		
	ข้อกล่าวอ้าง	หลักฐาน	การให้เหตุผล	ข้อกล่าวอ้าง	หลักฐาน	การให้เหตุผล
1	1.00	0.87	0.79	1.00	0.96	0.90
2	1.00	0.95	0.96	1.00	0.70	0.73
3	1.00	0.85	0.88	1.00	0.85	0.81
4	1.00	0.88	0.75	1.00	0.74	1.00
5	1.00	0.73	0.87	1.00	0.10	0.91
6	1.00	1.00	0.87	1.00	0.86	0.70
7	1.00	0.93	0.96	1.00	0.75	0.81
8	1.00	0.85	0.92	1.00	0.70	0.86
9	1.00	0.79	0.74	1.00	0.74	1.00
10	1.00	0.76	0.89	1.00	0.77	0.86

1.3 จากการวิเคราะห์ค่าความยาก และค่าอำนาจจำแนกของแบบอัตนัยจำกัดคำตอบและแบบไม่จำกัดคำตอบ ซึ่งมีข้อคำถามเหมือนกัน โดยแบ่งคะแนนของกลุ่มตัวอย่างออกเป็นกลุ่มสูงและกลุ่มต่ำโดยใช้เทคนิค 25% พบว่า ข้อคำถามในแบบวัดทั้งสองรูปแบบ มีความยากอยู่ในช่วงปานกลาง – ค่อนข้างยาก และมีอำนาจจำแนกดีมาก โดยแบบอัตนัยไม่จำกัดคำตอบ มีค่าความยากอยู่ในช่วง

0.26 – 0.47 และมีค่าอำนาจจำแนกอยู่ในช่วง 0.52 – 0.93 ส่วนแบบอัตนัยจำกัดคำตอบ มีค่าความยากอยู่ในช่วง 0.26 – 0.42 และมีค่าอำนาจจำแนกอยู่ในช่วง 0.51 – 0.83 และเมื่อพิจารณาข้อคำถามเดียวกัน พบว่า มีค่าความยากและค่าอำนาจจำแนกที่ใกล้เคียงกัน โดยภาพรวมแล้วข้อคำถามส่วนใหญ่ค่อนข้างยาก และทุกข้อสามารถจำแนกนักเรียนในกลุ่มเก่งและกลุ่มอ่อนได้ดีมาก ดังตารางที่ 3

ตารางที่ 3 ค่าความยากและค่าอำนาจจำแนกของข้อคำถามแบบอัตนัยจำกัดคำตอบและไม่จำกัดคำตอบ

ข้อ	แบบอัตนัยไม่จำกัดคำตอบ				แบบอัตนัยจำกัดคำตอบ			
	ความยาก	แปลความหมาย	อำนาจจำแนก	แปลความหมาย	ความยาก	แปลความหมาย	อำนาจจำแนก	แปลความหมาย
1	0.47	ปานกลาง	0.93	จำแนกดีมาก	0.42	ปานกลาง	0.83	จำแนกดีมาก
2	0.38	ค่อนข้างยาก	0.76	จำแนกดีมาก	0.33	ค่อนข้างยาก	0.66	จำแนกดีมาก
3	0.38	ค่อนข้างยาก	0.74	จำแนกดีมาก	0.36	ค่อนข้างยาก	0.71	จำแนกดีมาก
4	0.33	ค่อนข้างยาก	0.66	จำแนกดีมาก	0.33	ค่อนข้างยาก	0.66	จำแนกดีมาก
5	0.41	ปานกลาง	0.82	จำแนกดีมาก	0.34	ค่อนข้างยาก	0.67	จำแนกดีมาก
6	0.42	ปานกลาง	0.83	จำแนกดีมาก	0.38	ค่อนข้างยาก	0.75	จำแนกดีมาก
7	0.36	ค่อนข้างยาก	0.71	จำแนกดีมาก	0.35	ค่อนข้างยาก	0.69	จำแนกดีมาก
8	0.39	ค่อนข้างยาก	0.72	จำแนกดีมาก	0.36	ค่อนข้างยาก	0.73	จำแนกดีมาก
9	0.26	ค่อนข้างยาก	0.52	จำแนกดีมาก	0.26	ค่อนข้างยาก	0.51	จำแนกดีมาก
10	0.44	ปานกลาง	0.88	จำแนกดีมาก	0.40	ปานกลาง	0.81	จำแนกดีมาก

1.4 จากการวิเคราะห์ความเชื่อมั่นทั้งฉบับของแบบอัตนัยจำกัดคำตอบและแบบไม่จำกัดคำตอบ โดยนำคะแนนที่กลุ่มตัวอย่างทำได้ในแต่ละแบบวัดมาคำนวณค่าสัมประสิทธิ์แอลฟาของครอนบาค พบว่า

แบบวัดทั้งสองรูปแบบมีความเชื่อมั่นสูง โดยแบบอัตนัยไม่จำกัดคำตอบ มีค่าความเชื่อมั่นเท่ากับ 0.87 ซึ่งสูงกว่าแบบอัตนัยจำกัดคำตอบที่มีค่าความเชื่อมั่นเท่ากับ 0.85 ดังตารางที่ 4

ตารางที่ 4 การวิเคราะห์ค่าความเชื่อมั่นทั้งฉบับของแบบอัตนัยจำกัดคำตอบและไม่จำกัดคำตอบ

รูปแบบของแบบวัด	ค่าความเชื่อมั่นทั้งฉบับ	การแปลความหมาย
อัตนัยจำกัดคำตอบ	0.85	ความเชื่อมั่นสูง
อัตนัยไม่จำกัดคำตอบ	0.87	ความเชื่อมั่นสูง

2. จากการเปรียบเทียบประสิทธิภาพของแบบอัตนัยแบบจำกัดคำตอบและไม่จำกัดคำตอบโดยพิจารณาจากคะแนนเฉลี่ยของกลุ่มตัวอย่าง จำนวน 120 คน ที่ทำแบบวัดทั้งสองรูปแบบ คะแนนเต็ม 60 คะแนน พบว่า แบบอัตนัยจำกัดคำตอบ มีคะแนนเฉลี่ย 18.52 คะแนน และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10.45 ส่วนแบบอัตนัยไม่จำกัดคำตอบ มีคะแนนเฉลี่ย 22.13 คะแนน และส่วนเบี่ยงเบน

มาตรฐานเท่ากับ 11.43 เมื่อเปรียบเทียบคะแนนเฉลี่ยจากแบบวัดทั้งสองรูปแบบ ด้วยการทดสอบค่าทีแบบสองกลุ่มที่ไม่อิสระจากกัน พบว่า คะแนนเฉลี่ยของนักเรียนที่ทำแบบอัตนัยไม่จำกัดคำตอบสูงกว่าคะแนนเฉลี่ยแบบอัตนัยจำกัดคำตอบ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ($t = 6.027, p = .000$) ดังตารางที่ 5

ตารางที่ 5 ผลการเปรียบเทียบคะแนนเฉลี่ยระหว่างแบบอัตนัยจำกัดคำตอบและไม่จำกัดคำตอบ

รูปแบบของแบบวัด	n	คะแนนเต็ม	$\bar{x}$	S.D.	t	p
อัตนัยจำกัดคำตอบ	120	60	18.52	10.45	6.027*	.000
อัตนัยไม่จำกัดคำตอบ	120	60	22.13	11.43		

* $p < .05$

สรุปผลการวิจัย

1. แบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ที่พัฒนาขึ้น มี 2 รูปแบบ คือ 1) แบบอัตนัยจำกัดคำตอบกำหนดให้เขียนคำตอบแยกตามองค์ประกอบของการสร้างคำอธิบายทางวิทยาศาสตร์ และ 2) แบบอัตนัยไม่จำกัดคำตอบที่สามารถเขียนคำตอบได้อย่างอิสระโดยใช้คำถามชี้แนะให้นักเรียนเขียนคำตอบครบตามองค์ประกอบ แต่ละรูปแบบมีข้อความจำนวน 10 ข้อ ไม่จำกัดกลุ่มเนื้อหา และมีการให้ข้อมูลทางวิทยาศาสตร์ที่เพียงพอต่อการสรุปหลักการทางวิทยาศาสตร์มาใช้ในการตอบคำถาม แบบวัดทั้ง 2 รูปแบบ มีความเชื่อมั่นสูง ข้อคำถามในแบบวัดฯ มีความยากในช่วงปานกลาง – ยาก และมีอำนาจจำแนกดีมาก

2. แบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์แบบอัตนัยไม่จำกัดคำตอบ มีประสิทธิภาพดีกว่าแบบอัตนัยจำกัดคำตอบ โดยนักเรียนมีคะแนนเฉลี่ยแบบอัตนัยไม่จำกัดคำตอบที่สูงกว่าแบบจำกัดคำตอบแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

การอภิปรายผลการวิจัย

การศึกษา เรื่อง การเปรียบเทียบประสิทธิภาพของแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์แบบอัตนัยจำกัดคำตอบและไม่จำกัดคำตอบ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 4 มีประเด็นในการอภิปรายผลการวิจัย ดังนี้

1. การพัฒนาแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์แบบอัตนัยจำกัดคำตอบและไม่จำกัดคำตอบ สำหรับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4

1.1 แบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ที่พัฒนาขึ้น มีลักษณะเป็นแบบวัด

แบบอัตนัยที่มี 2 รูปแบบ คือ แบบอัตนัยจำกัดคำตอบที่กำหนดให้เขียนคำตอบแยกตามองค์ประกอบของการสร้างคำอธิบายทางวิทยาศาสตร์ และแบบอัตนัยไม่จำกัดคำตอบที่สามารถเขียนคำตอบได้อย่างอิสระ ซึ่งสามารถวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ของนักเรียนได้ ทั้งนี้อาจเป็นเพราะว่าเป็นแบบวัดที่สร้างขึ้นไม่จำกัดกลุ่มเนื้อหา และให้ข้อมูลทางวิทยาศาสตร์ เพื่อให้ผู้เรียนที่ไม่มีความรู้ในเรื่องนั้น ๆ มาก่อน สามารถนำมาใช้ในการสร้างคำอธิบายทางวิทยาศาสตร์ได้ ซึ่งผู้เรียนที่มีความรู้ความเข้าใจในเนื้อหาทางวิทยาศาสตร์ที่ศึกษามาสามารถสร้างคำอธิบายทางวิทยาศาสตร์ได้ดีขึ้น (Oktavianti, Handayanto, Wartono, & Saniso, 2018) อย่างไรก็ตามพบว่า นักเรียนทำแบบวัดได้คะแนนไม่มาก อาจเป็นเพราะว่าขาดการวิเคราะห์และเชื่อมโยงคำถามไปสู่องค์ความรู้ทางวิทยาศาสตร์ที่มีมาก่อน นักเรียนบางคนมีความเข้าใจเนื้อหาที่เรียนไม่ดีเท่าที่ควร ทำให้การเขียนเหตุผลประกอบคำอธิบายที่สร้างขึ้นไม่ถูกต้อง (กรกนก เลิศเดชาภัทร และปริญดา ลิ้มพานนท์ พรหมรัตน์, 2561) ดังนั้นการเพิ่มข้อมูลทางวิทยาศาสตร์ขึ้นมาจึงช่วยให้นักเรียนมีองค์ความรู้เพียงพอในการสร้างคำอธิบายทางวิทยาศาสตร์และช่วยลดข้อจำกัดของการให้เหตุผลประกอบทางวิทยาศาสตร์ สอดคล้องกับกฤตกร สภาสันติกุล (2559) ที่ศึกษาพบว่า แบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์แบบอัตนัยประเมินความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ของผู้เรียนได้อย่างมีประสิทธิภาพ (พรธมนภา อนิวรรณวงศ์ และร่มเกล้า จันทราธิ, 2562; Yao & Guo, 2018)

1.2 คุณภาพของแบบวัดความสามารถ พบว่า มีความตรงเชิงเนื้อหา มีความยากอยู่ในช่วงปานกลาง – ค่อนข้างยาก และทุกข้อคำถามมีอำนาจจำแนกดีมาก มีค่าความเชื่อมั่นมากกว่า 0.70 ทั้งนี้เป็นเพราะว่า

แบบวัดความสามารถ ที่พัฒนาขึ้นตามแนวทางการสร้าง คำอธิบายทางวิทยาศาสตร์ เน้นการประเมินความสามารถ ของผู้เรียนในการระบุข้อกล่าวอ้าง หลักฐาน และการให้ เหตุผล ตามนิยามของความสามารถในการสร้างคำอธิบาย ทางวิทยาศาสตร์ที่กำหนดขึ้น ทำให้มีค่าความสอดคล้อง ระหว่างข้อคำถามกับนิยามของความสามารถในการสร้าง คำอธิบายทางวิทยาศาสตร์ที่ประเมินโดยผู้เชี่ยวชาญมีค่าสูง สะท้อนให้เห็นว่า แบบวัดความสามารถ ที่พัฒนาขึ้นมี ความตรงเชิงเนื้อหา สามารถวัดได้ตรงตามสิ่งที่ต้องการวัด (อารยา องค์เอี่ยม และพงศ์ธรา วิจิตเวชไพศาล, 2561; วิเชียร อินทรสมพันธ์ และคนอื่น ๆ, 2563) ข้อคำถาม แต่ละข้อมีการให้ข้อมูลทางวิทยาศาสตร์ที่เพียงพอต่อ การสร้างคำอธิบายทางวิทยาศาสตร์ ทำให้ผู้เรียนที่ไม่มีความรู้ในเรื่องนั้นมาก่อนสามารถสรุปหลักการเพื่อใช้ในการตอบคำถามได้ จึงมีค่าความยากอยู่ในช่วงที่เหมาะสม สอดคล้องกับผลการพัฒนาแบบวัดความสามารถในการ สร้างคำอธิบายทางวิทยาศาสตร์ของ Ningsi, Suhandi, Kaniawati and Samsudin (2019) ที่พบว่า แบบวัดความสามารถ ที่พัฒนาขึ้นจะมีความยากมาก หากผู้เรียนขาด องค์ความรู้เกี่ยวกับข้อคำถามนั้น ๆ ในขณะเดียวกัน ทุกข้อ คำถามยังมีค่าอำนาจจำแนกดีมาก สะท้อนให้เห็นว่าข้อคำถาม แต่ละข้อสามารถจำแนกผู้เรียนออกเป็นกลุ่มเก่งและ กลุ่มอ่อนได้ (อารยา องค์เอี่ยม และพงศ์ธรา วิจิตเวชไพศาล, 2561) นอกจากนี้แบบวัดทั้งสองรูปแบบมีค่าความเชื่อมั่นสูง สะท้อนให้เห็นถึงความสม่ำเสมอของคะแนนที่ได้จากการ ทำแบบวัดหลายครั้ง เป็นเครื่องมือที่มีคุณภาพ และ เหมาะสมต่อการนำไปใช้ (Sener & Tas, 2017; Ningsi et al., 2019)

อย่างไรก็ตามแบบวัดความสามารถทั้งสองรูปแบบ จำเป็นต้องมีคำถามชี้้นำให้ผู้เรียนเขียนคำตอบได้ครบ ตามองค์ประกอบ เพื่อลดข้อจำกัดสำหรับนักเรียนที่ไม่ ได้รับการฝึกการสร้างคำอธิบายทางวิทยาศาสตร์มาก่อน ให้สามารถเขียนคำอธิบายทางวิทยาศาสตร์ได้ครบ ทุกองค์ประกอบ นอกจากนี้แบบวัดความสามารถ ทั้งสอง รูปแบบ มีลักษณะการเขียนคำตอบและความยากง่าย ในการตรวจคำตอบที่แตกต่างกัน โดยแบบอัตนัยจำกัด คำตอบผู้เรียนต้องเขียนคำตอบแยกองค์ประกอบของ

คำอธิบายทางวิทยาศาสตร์ทำให้สามารถตรวจให้คะแนน ตามเกณฑ์การให้คะแนนง่าย แต่หากเขียนคำตอบ ผิดตำแหน่งหรือเขียนลงในช่องใดช่องหนึ่ง เช่น เขียนทุก องค์ประกอบลงในช่องข้อกล่าวอ้างเพียงช่องเดียว จะทำให้ ไม่มีคะแนนในองค์ประกอบที่เว้นว่างไว้ ขณะที่แบบอัตนัย ไม่จำกัดคำตอบ ผู้เรียนสามารถเขียนองค์ประกอบใดก่อน และหลังได้อย่างอิสระ แต่การตรวจให้คะแนนจะยากกว่า แบบอัตนัยจำกัดคำตอบ เนื่องจากต้องพิจารณาว่าคำตอบ ที่ผู้เรียนเขียนในแต่ละตำแหน่งนั้นเป็นองค์ประกอบใด สอดคล้องกับกรรณก เลิศเดชภัทร (2559) ที่ศึกษาพบว่าการ ใช้แบบวัดความสามารถในการสร้างคำอธิบายทาง วิทยาศาสตร์ในลักษณะแบบอัตนัยจำกัดคำตอบสามารถ วิเคราะห์และประเมินความสามารถของผู้เรียนในแต่ละ องค์ประกอบของคำอธิบายทางวิทยาศาสตร์ได้อย่างมีประสิทธิภาพ (พรรณนภา อนิวรรณวงศ์ และร่วมเกล้า จันทราธิ, 2562; ฉลองวุฒิ จันทรหอม และสมเกียรติ พรพิสุทธิมาศ, 2563)

1.3 เกณฑ์การให้คะแนนคำอธิบายทาง วิทยาศาสตร์ของแบบวัดทั้งสองรูปแบบมีลักษณะเป็น เกณฑ์การให้คะแนนแบบรายข้อ แยกการให้คะแนนในแต่ละ องค์ประกอบของคำอธิบายทางวิทยาศาสตร์ พร้อมตัวอย่าง คำตอบที่ถูกต้องในแต่ละระดับคะแนน โดยเป็นเกณฑ์ที่มีความเหมาะสมต่อการนำไปใช้ ทั้งนี้เป็นเพราะว่า เกณฑ์การ ให้คะแนนที่พัฒนาขึ้น แต่ละองค์ประกอบมีการระบุพฤติกรรม และแสดงตัวอย่างคำตอบที่ผู้เรียนสามารถปฏิบัติได้ในแต่ละ ระดับคะแนนอย่างชัดเจน สะท้อนความสามารถในการสร้าง คำอธิบายทางวิทยาศาสตร์ของนักเรียนครอบคลุม ทุกองค์ประกอบของคำอธิบายทางวิทยาศาสตร์ได้ รวมทั้งผ่านการตรวจสอบความเหมาะสมต่อการนำไปใช้ จากผู้เชี่ยวชาญ 3 ท่าน และการตรวจสอบความเป็นปรนัย โดยวิเคราะห์ค่าความเชื่อมั่นระหว่างผู้ประเมิน 3 ท่าน โดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์ภายในซึ่งความเชื่อมั่น 95% พบว่า ในภาพรวมมีค่าระหว่าง 0.70 – 1.00 ซึ่งถือว่า ผู้ประเมินแต่ละท่านมีการประเมินที่สอดคล้องกันสูง (Bajpai, Bajpai & Chaturvedi, 2015) สะท้อนให้เห็นถึงความเป็นปรนัย ของเกณฑ์การให้คะแนนที่พัฒนาขึ้น สอดคล้องกับ Fort-vanmeerhaeghe, Montalvo, Lloyd, Read and Myer

(2017) ที่วิเคราะห์ค่าความเชื่อมั่นระหว่างผู้ประเมินโดยใช้ค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้น เพื่อพิจารณาความเป็นปรนัยของเกณฑ์ที่พัฒนาขึ้น โดยค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้นที่มีค่าสูง จะสะท้อนถึงความเป็นปรนัยสูง (Lucas et al., 2017; Murphy et al., 2019; Villa, Sprunger, Walton, Costello, & Isaacs, 2020)

2. การเปรียบเทียบประสิทธิภาพของแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์แบบอัตโนมัติจำกัดคำตอบและไม่จำกัดคำตอบ สำหรับนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 พบว่า แบบวัดความสามารถแบบอัตโนมัติไม่จำกัดคำตอบ มีประสิทธิภาพสูงกว่า ผู้เรียนมีคะแนนเฉลี่ยจากการทำแบบอัตโนมัติไม่จำกัดคำตอบสูงกว่าแบบจำกัดคำตอบ แตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 ทั้งนี้อาจเป็นเพราะว่าแบบอัตโนมัติจำกัดคำตอบกำหนดให้เขียนคำอธิบายทางวิทยาศาสตร์โดยแยกเป็นองค์ประกอบของข้อกล่าวอ้าง หลักฐาน และการให้เหตุผล เพื่อกระตุ้นให้ผู้เรียนสามารถเขียนคำอธิบายทางวิทยาศาสตร์ได้ครบทุกองค์ประกอบ แต่การกำหนดให้ผู้เรียนเขียนแยกตามแต่ละองค์ประกอบนั้นเปรียบเสมือนกรอบที่จำกัดความคิด ทำให้ไม่มีอิสระในการเขียนคำอธิบายทางวิทยาศาสตร์ และใช้ระยะเวลาทำความเข้าใจว่าแต่ละองค์ประกอบต้องเขียนคำอธิบายในลักษณะใด ขณะที่แบบอัตโนมัติไม่จำกัดคำตอบผู้เรียนสามารถเขียนคำอธิบายทางวิทยาศาสตร์ได้อย่างอิสระ โดยไม่จำกัดว่าคำอธิบายทางวิทยาศาสตร์ที่สร้างขึ้นต้องเขียนลงในองค์ประกอบใด สอดคล้องกับ Sumarni, Supardi and Widiarti (2017) ได้กล่าวไว้ว่า การใช้แบบวัดอัตโนมัติไม่จำกัดคำตอบผู้เรียนสามารถเขียนแสดงความคิดเห็นที่สอดคล้องกับทักษะหรือความสามารถที่มีอยู่และให้คำตอบประเด็นที่ศึกษาได้อย่างอิสระ (พัชรินทร์ ชมภูวิเศษ, 2559; Shafack, Alemnge & Eta, 2020) ซึ่งสะท้อนถึงความรู้และประสบการณ์ที่ผู้เรียนมีอยู่ได้เป็นอย่างดี (Ningsi et al., 2019; Traut, 2017; Yao and Guo, 2018; สุทธิชาติ เปรมมงคล และสกลรัตน์ แก้วดี, 2560)

เมื่อพิจารณาคะแนนเฉลี่ยที่ผู้เรียนทำได้ในแบบวัดแต่ละรูปแบบมีคะแนนค่อนข้างน้อยเมื่อเทียบกับคะแนนเต็ม แม้ว่าจะมีการให้ข้อมูลทางวิทยาศาสตร์ที่เพียงพอต่อการสร้างคำอธิบายทางวิทยาศาสตร์ก็ตาม

สะท้อนให้เห็นว่านักเรียนมีความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์อยู่ในระดับที่ไม่สูงมาก ทั้งนี้อาจเนื่องมาจากนักเรียนยังไม่สามารถระบุคำตอบและเชื่อมโยงองค์ประกอบทั้งสามของคำอธิบายทางวิทยาศาสตร์ได้ดีเท่าที่ควร ทั้งในส่วนของ การแสดงหลักฐานและการให้เหตุผล โดยเฉพาะอย่างยิ่งในส่วนของ การให้เหตุผล ที่ผู้เรียนทำคะแนนได้น้อยกว่าองค์ประกอบอื่น ๆ ซึ่งการขาดความรู้และความเข้าใจหลักการทางวิทยาศาสตร์ที่เกี่ยวข้องกับเนื้อหา จะทำให้ผู้เรียนไม่สามารถหาเหตุผลมาสนับสนุนคำอธิบายได้อย่างถูกต้อง (ฉลองวุฒิ จันทรหอม และสมเกียรติ พรพิสุทธิมาศ, 2563) สอดคล้องกับ Traut (2017) และ Oktavianti et al. (2018) ที่ศึกษาพบว่า ข้อกล่าวอ้างเป็นองค์ประกอบที่ผู้เรียนทำคะแนนได้มากที่สุด ส่วนองค์ประกอบที่ผู้เรียนมีปัญหามากที่สุดและทำคะแนนได้น้อยที่สุดคือ การให้เหตุผล เนื่องจากองค์ประกอบข้อกล่าวอ้างนั้นผู้เรียนอาจนำความรู้ที่ได้จากการสรุปข้อมูลที่กำหนดให้หรือประสบการณ์เดิมมาใช้ในการตอบคำถาม ทำให้ผู้เรียนทำคะแนนออกมาได้ดีกว่าองค์ประกอบอื่น ขณะที่การให้เหตุผลนั้นผู้เรียนต้องนำหลักการทางวิทยาศาสตร์ที่เกี่ยวข้องกับสถานการณ์ที่กำหนดมาเพื่อแสดงให้เห็นถึงความเชื่อมโยงระหว่างข้อกล่าวอ้างที่สร้างขึ้นกับหลักฐานที่นำมาสนับสนุน หากผู้เรียนขาดองค์ความรู้ในหลักการทางวิทยาศาสตร์ที่เกี่ยวข้อง หรือไม่สามารถสรุปหลักการทางวิทยาศาสตร์จากข้อมูลกำหนดให้ได้ ผู้เรียนจะไม่สามารถให้เหตุผลประกอบคำอธิบายทางวิทยาศาสตร์ได้ (Traut, 2017; Oktavianti et al., 2018; กฤตกร สภาสันติกุล, 2559)

หากพิจารณาค่าความเชื่อมั่นของแบบวัดความสามารถ ทั้งสองรูปแบบพร้อมด้วยแล้ว พบว่าค่าความเชื่อมั่นของแบบอัตโนมัติไม่จำกัดคำตอบสูงกว่าแบบอัตโนมัติจำกัดคำตอบ ซึ่งค่าความเชื่อมั่นที่สูงขึ้นสะท้อนประสิทธิภาพของแบบวัดว่าสามารถวัดผลได้อย่างแม่นยำและลดข้อผิดพลาดหรือความไม่น่าเชื่อถือของข้อมูลที่วัดได้มากขึ้น (Sinadia & Jatmika, 2020) ดังนั้นแบบวัดความสามารถ แบบอัตโนมัติไม่จำกัดคำตอบจึงเหมาะสมต่อการนำไปใช้วัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ของนักเรียนได้ดีกว่าแบบอัตโนมัติจำกัดคำตอบ

ข้อเสนอแนะ

ข้อเสนอแนะเพื่อการนำผลการวิจัยไปใช้

1. แบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ทั้งสองรูปแบบ มีลักษณะการเขียนคำตอบและความยากง่ายในการตรวจคำตอบที่แตกต่างกัน การนำแบบวัดความสามารถไปใช้ ผู้สอนควรพิจารณาประยุกต์ใช้ให้เหมาะสมกับบริบทของผู้เรียน

2. การนำแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ไปใช้ แม้จะยกตัวอย่างการเขียนคำอธิบายทางวิทยาศาสตร์ทั้งในภาพรวมและแยกองค์ประกอบ ให้ผู้เรียนศึกษาการทำแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์แล้วก็ตาม แต่พบว่าผู้เรียนส่วนใหญ่ยังไม่เข้าใจการเขียนข้อกล่าวอ้าง หลักฐาน และการให้เหตุผล เนื่องจากไม่มีประสบการณ์ในการสร้างคำอธิบายทางวิทยาศาสตร์มาก่อน ดังนั้นผู้สอนควรแสดงตัวอย่างคำอธิบายทางวิทยาศาสตร์ มากกว่า 1 ตัวอย่าง เพื่อเป็นแนวทางการเขียนคำอธิบายทางวิทยาศาสตร์ และอธิบายให้ผู้เรียนได้เข้าใจอย่างละเอียดว่าตำแหน่งใดของคำอธิบายทางวิทยาศาสตร์เรียกว่า ข้อกล่าวอ้าง หลักฐาน หรือการให้เหตุผล และลดความคลาดเคลื่อนในการตรวจให้คะแนน

3. ข้อคำถามแต่ละข้อในแบบวัดความสามารถ มีความซับซ้อนและข้อมูลที่กำหนดให้แตกต่างกัน คำถามบางข้อต้องอ่านและพิจารณาอย่างละเอียดจึงจะสามารถตอบ

คำถามได้ แต่บางข้อพิจารณาแล้วสามารถตอบได้ทันที ทำให้ผู้เรียนใช้เวลาสร้างคำอธิบายทางวิทยาศาสตร์แต่ละข้อแตกต่างกัน ดังนั้นควรกำหนดจำนวนคำถามที่ใช้วัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ ต้องพิจารณาให้เหมาะสมกับระยะเวลาที่ใช้ในการทดสอบ โดยคำถาม 4 – 5 ข้อ ควรใช้เวลาสำหรับการทำแบบวัด 60 นาที

ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. ควรมีการพัฒนาแบบวัดความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 4 ด้วยเครื่องมืออื่น เช่น แบบสังเกตพฤติกรรม แบบสัมภาษณ์ เป็นต้น เพื่อประเมินความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ของนักเรียนร่วมกัน และสะท้อนผลการวัดและประเมินความสามารถของผู้เรียนในหลายมิติทั้งในเชิงปริมาณและเชิงคุณภาพ

2. ควรมีการศึกษาความสัมพันธ์ระหว่างตัวแปรหรือปัจจัยที่ส่งผลต่อความสามารถในการสร้างคำอธิบายทางวิทยาศาสตร์ของผู้เรียน เช่น ความเข้าใจแนวคิดวิทยาศาสตร์ เจตคติต่อวิทยาศาสตร์ เพื่อพัฒนาการจัดการเรียนรู้การสร้างคำอธิบายทางวิทยาศาสตร์

เอกสารอ้างอิง

- กรกนก เลิศเดชภักทร. (2559). ผลของการสืบสอบแบบร่วมมือมีผลต่อความสามารถในการสร้างคำอธิบายเชิงวิทยาศาสตร์ ของนักเรียนมัธยมศึกษาตอนต้น. (วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาการศึกษา วิทยาศาสตร์ คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย).
- กรกนก เลิศเดชภักทร และปริณดา ลิมปานนท์ พรหมรัตน์. (2561). ผลของการสืบสอบแบบร่วมมือมีผลต่อความสามารถในการสร้างคำอธิบายเชิงวิทยาศาสตร์ของนักเรียนมัธยมศึกษาตอนต้น. *วารสารครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย*, 46(12), 1–20.
- กฤตกร สภาสันติกุล. (2559). ผลการสืบสอบแบบร่วมมือมีผลต่อความสามารถในการสร้างคำอธิบายเชิงวิทยาศาสตร์และความสามารถในการแก้ปัญหาแบบร่วมมือมีผลต่อความสามารถของนักเรียนมัธยมศึกษาตอนต้น. (วิทยานิพนธ์ครุศาสตรมหาบัณฑิต, สาขาวิชาการศึกษาวิทยาศาสตร์ คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย).
- ฉลอมวุฒิ จันทรหอม และสมเกียรติ พรพิสุทธิมาศ. (2563). การสำรวจความสามารถในการสร้างคำอธิบายเชิงวิทยาศาสตร์ เรื่อง ระบบย่อยอาหาร ของนักเรียนชั้นมัธยมศึกษาตอนปลาย. ใน *การประชุมวิชาการเสนอผลงานวิจัยระดับบัณฑิตศึกษาแห่งชาติ ครั้งที่ 21*, วันที่ 27 มีนาคม 2563 ณ มหาวิทยาลัยขอนแก่น. (หน้า 644–654). ขอนแก่น: มหาวิทยาลัยขอนแก่น.
- บุญชม ศรีสะอาด. (2560). *การวิจัยเบื้องต้น*. (พิมพ์ครั้งที่ 10). กรุงเทพฯ: สุวีริยาสาส์น.
- พรรณณา อนิวรรณวงศ์ และร่มเกล้า จันทราธิ. (2562). การประเมินผลของการสืบเสาะหาความรู้โดยใช้แบบจำลองเป็นฐานร่วมกับการใช้การเชื่อมโยงหลักฐานและแบบจำลองที่มีต่อการอธิบายทางวิทยาศาสตร์ เรื่อง สารละลาย อิเล็กโทรไลต์และนอนอิเล็กโทรไลต์ของนักเรียนชั้นมัธยมศึกษาปีที่ 5. *วารสารนวัตกรรมการเรียนรู้มหาวิทยาลัยวลัยลักษณ์*, 5(1), 65–83.
- พัชรินทร์ ชมภูวิเศษ. (2559). *การวัดและประเมินผลการศึกษา*. กรุงเทพฯ: คณะครุศาสตร์ มหาวิทยาลัยราชภัฏอุดรธานี.
- พนิดา มีลา และร่มเกล้า อาจเดช. (2560). การสืบเสาะหาความรู้โดยใช้แบบจำลองเป็นฐานและการอธิบายทางวิทยาศาสตร์: การส่งเสริมการสร้างความหมายในชั้นเรียน. *วารสารศึกษาศาสตร์ มหาวิทยาลัยนเรศวร*, 19(3), 1–15.
- วิจิตรา กุสมภ์ และสุลี ทองวิเชียร. (2560). การพัฒนาทักษะการคิดอย่างมีวิจารณญาณของนักศึกษาพยาบาล. *วารสารคณะพยาบาลศาสตร์ มหาวิทยาลัยบูรพา*, 25(3), 1–8.
- วิเชียร อินทรสมพันธ์, เพชราวดี จงประดับเกียรติ, ทวีศักดิ์ จงประดับเกียรติ, สุภาพร ศรีหามิ, อัศเดช เกตฉ่า, เพ็ญพร ทองคำสูง, ..., ทิตยา ลิทธิโสภาสกุล. (2563). *การวัดและประเมินผลการเรียนรู้*. กรุงเทพฯ: วรานนท์ เอ็นเตอร์ไพรส์.
- สุทธิชาติ เปรมกมล และสกลรัชต์ แก้วดี. (2560). ผลของการใช้การสืบสอบเน้นแบบจำลองเป็นฐานที่มีต่อความสามารถในการสร้างคำอธิบายเชิงวิทยาศาสตร์และการให้เหตุผลของนักเรียนมัธยมศึกษาตอนต้น. *วารสารอิเล็กทรอนิกส์ทางการศึกษา*, 12(1), 259–274.
- อารยา องค์เอี่ยม และพงศ์ธรา วิจิตเวชไพศาล. (2561). การตรวจสอบคุณภาพเครื่องมือวิจัย. *วิสัยทัศน์สาร*, 44(1), 36–42.
- Andrade, V., Shwartz, Y., Freire, S., and Baptista, M. (2021). Students' mechanistic reasoning in practice: Enabling functions of drawing, gestures and talk. *Science Education*, 106(1), 1–27.

- Bajpai, S., Bajpai, R. C., and Chaturvedi, H. K. (2015). Evaluation of inter-rater agreement and inter-rater reliability for observational data: an overview of concepts and methods. *Journal of the Indian Academy of Applied Psychology*, 41(3), 20–27. Retrieved from https://www.researchgate.net/publication/273451591_Evaluation_of_Inter-Rater_Agreement_and_Inter-Rater_Reliability_for_Observational_Data_An_Overview_of_Concepts_and_Methods
- Fort–Vanmeerhaeghe, A., Montalvo, A. M., Lloyd, R. S., Read, P., and Myer, G. D. (2017). Intra–and inter–rater reliability of the modified tuck jump assessment. *Journal of Sports Science & Medicine*, 16(1), 117.
- Johnson, K. A., and Foa, L. J. (1996). *Instructional design: New alternatives for effective education and training*. Phoenix: Oryx Press.
- Lucas, C., Bosnic–Anticevich, S., Schneider, C. R., Bartimote–Aufflick, K., McEntee, M., and Smith, L. (2017). Inter–rater reliability of a reflective rubric to assess pharmacy students’ reflective thinking. *Currents in Pharmacy Teaching and Learning*, 9(6), 989–995.
- McNeill, K. L., and Krajcik, J. (2008a). Assessing middle school students’ content knowledge and reasoning through written scientific explanations. *Assessing science learning: Perspectives from research and practice*, 101–116.
- McNeill, K. L., and Krajcik, J. (2008b). Scientific explanations: Characterizing and evaluating the effects of teachers’ instructional practices on student learning. *Journal of Research in Science Teaching*, 45(1), 53–78.
- Murphy, A. R., Ingram, H. E., Nelson, J. T., Bohm, M. R., Linsey, J. S., and Nagel, R. L. (2019). An update to a functional modeling scoring rubric with overall and question–level inter–rater reliability. *Journal of Mechanical Design*, 141(8), 084501–1–10. Retrieved from <https://asmedigitalcollection.asme.org/mechanicaldesign/article-abstract/141/8/084501/727183/An-Update-to-a-Functional-Modeling-Scoring-Rubric?redirectedFrom=fulltext>
- Ningsi, S., Suhandi, A., Kaniawati, I., and Samsudin, A. (2019). *KTG–SESC: Development of scientific explanation skills test instrument*. In *Journal of Physics: Conference Series*, August 28–29, 2018 Gedung Graha Cakrawala (p. 1–6). Indonesia: Universitas Negeri Malang.
- Oktavianti, E., Handayanto, S. K., Wartono, W., and Saniso, E. (2018). Students’ scientific explanation in blended physics learning with e–scaffolding. *Jurnal Pendidikan IPA Indonesia*, 7(2), 181–186. Retrieved from <https://journal.unnes.ac.id/nju/index.php/jpii/article/view/14232/7901>
- Secolsky, C., and Denison, D. B. (Eds.). (2017). *Handbook on measurement, assessment, and evaluation in higher education*. New York: Routledge.
- Sener, N., and Tas, E. (2017). Developing achievement test: A research for assessment of 5th grade biology subject. *Journal of Education and Learning*, 6(2), 254–271.
- Shafack, R. M., Alemnge, F. L., and Eta, E. R. (2020). Classroom assessment of essay writing, and students’ learning of English language in secondary schools in the south west region of Cameroon. *International Journal of Science Academic Research*, 1(8), 591–601.
- Sinadia, A. R., and Jatmika, S. (2020). Development of the academic achievement test for undergraduate students. *Advances in Social Science, Education and Humanities Research*, 438, 191–195.

- Sumarni, W., Supardi, K. I., and Widiarti, N. (2017). Development of assessment instruments to measure critical thinking skills. *IOP Conference Series: Materials Science and Engineering*, 349, 1–11.
- Traut, J. (2017). *Forming explanations from evidence using the claim–evidence–reasoning framework*. (Master thesis, Science Education, Montana State University).
- Villa, K. R., Sprunger, T. L., Walton, A. M., Costello, T. J., and Isaacs, A. N. (2020). Inter–rater reliability of a clinical documentation rubric within pharmacotherapy problem–based learning courses. *American Journal of Pharmaceutical Education*, 84(7), 886–891.
- Yao, J. X., and Guo, Y. Y. (2018). Validity evidence for a learning progression of scientific explanation. *Journal of Research in Science Teaching*, 55(2), 299–317.
- Yao, J. X., Guo, Y. Y., and Neumann, K. (2016). Towards a hypothetical learning progression of scientific explanation. *Asia–Pacific science education*, 2(1), 1–17.

