การพัฒนาเครื่องมือประเมินทักษะกระบวนการวิทยาศาสตร์ ด้วยวิธีสร้างโมเดลภาวะสันนิษฐาน

Development of a Measure for Assessment of Scientific Process Skills, Using Construct Modeling Approach

บุษกร สิงห์คู[1] และพัชรี จันทร์เพ็ง[2]
Bussakorn Singku[1] and Putcharee Junpeng[2]

(Received: July 12, 2019; Revised: September 19, 2019; Accepted: September 20, 2019)

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์หลักเพื่อพัฒนาเครื่องมือประเมินทักษะกระบวนการวิทยาศาสตร์แบบพหุมิติสำหรับเนื้อหาปริมาณสัมพันธ์ ระดับมัธยมศึกษาปีที่ 5 มีวัตถุประสงค์ย่อย คือ 1) เพื่อพัฒนาแผนที่ภาวะสันนิษฐานทักษะกระบวนทางวิทยาศาสตร์ และ 2) เพื่อสร้างและตรวจสอบคุณภาพของเครื่องมือประเมินทักษะดังกล่าว กลุ่มตัวอย่าง คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 5 ปีการศึกษา 2561 จำนวน 1,200 คน สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา ภาคตะวันออกเฉียงเหนือ เครื่องมือประเมินพัฒนาด้วยวิธีการสร้างโมเดลภาวะสันนิษฐาน มี 4 ขั้นตอน คือ 1) การพัฒนาแผนที่ภาวะสันนิษฐานทักษะกระบวนการทางวิทยาศาสตร์แบบพหุมิติ 2) การออกแบบข้อคำถาม 3) การให้คะแนนคำตอบ และ 4) การวิเคราะห์ข้อมูลตามโมเดลสัมประสิทธิ์สุ่มสำหรับข้อมูลหลายค่าที่มาจากการวัดหลายมิติ ด้วยโปรแกรม conquest ผลการวิจัยปรากฏดังนี้

1) แผนที่ภาวะสันนิษฐานทักษะกระบวนการทางวิทยาศาสตร์ มี 2 ด้าน คือ ด้านทักษะการพยากรณ์ มี 4 ระดับ ตั้งแต่ระดับต่ำสุด คือ นักเรียนรู้การเกิดปฏิกิริยาแต่ไม่สามารถตอบหรืออธิบายการเกิดปฏิกิริยาได้ถูกต้อง ไปยังระดับสูงสุด คือ ให้เหตุผลและบอกความแตกต่างของปฏิกิริยา และนำความรู้เรื่องอัตราการเกิดปฏิกิริยาไปใช้กับได้สอดคล้องกับบริบทจริง และด้านทักษะการจัดกระทำและสื่อความหมายของข้อมูล มี 4 ระดับ ตั้งแต่ระดับต่ำสุด คือ นักเรียนมีความรู้บอกได้ว่า อัตราการเกิดปฏิกิริยาเปลี่ยนตามพื้นที่ผิว แต่บอกไม่ได้ว่าสารตั้งต้นที่ทำปฏิกิริยากับหินปูนคือสารใด ไปจนถึงระดับสูงสุด คือ นักเรียนสามารถนำความรู้พื้นฐานทางวิทยาศาสตร์กระบวนการคิดมาประยุกต์ใช้เพื่อแสดงวิธีคำนวณหาปริมาณไฮโดรเจนไอออนที่ทำปฏิกิริยากับหินปูนได้และบอกอุณหภูมิที่มีผลต่อปฏิกิริยาได้

[1] นักศึกษาระดับมหาบัณฑิต สาขาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

[2] ผู้ช่วยศาสตราจารย์ สาขาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

[1] Master degree student in Educational Measurement and Evaluation Program, Faculty of Education Khon Kaen University

[2] Assistant professor, Department of Educational Measurement and Evaluation, Faculty of Education, Khon Kaen University

Corresponding author E-mail: bussakorntu30@gmail.com

2) เครื่องมือประเมินทักษะกระบวนการทางวิทยาศาสตร์ มีทั้งหมด 15 ข้อ เป็นข้อคำถามปลายเปิดและตรวจให้คะแนนคำตอบหลายค่า ตั้งแต่ 0 ถึง 3 คะแนน ตามระดับทักษะในแผนที่ภาวะสันนิษฐานที่พัฒนาขึ้น มีข้อคำถามด้านการพยากรณ์ จำนวน 8 ข้อ และด้านการจัดกระทำและสื่อความหมายของข้อมูล จำนวน 7 ข้อ

3) ผลการตรวจสอบความตรงของผลการประเมินทักษะกระบวนการทางวิทยาศาสตร์ ด้านหลักฐานความตรงของเนื้อหาของข้อคำถาม พบว่า ข้อคำถามมีค่าความยากครอบคลุมช่วงความสามารถของนักเรียน นั่นคือ เนื้อหาของข้อคำถามเป็นนิยามของทักษะกระบวนทางวิทยาศาสตร์ของนักเรียน ด้านหลักฐานความตรงเกี่ยวกับกระบวนการตอบสนอง พบว่า นักเรียนเข้าใจเนื้อหาของข้อคำถาม ตรงกับความตั้งใจของผู้สร้างข้อคำถาม และด้านหลักฐานเกี่ยวกับโครงสร้างภายใน พบว่า ข้อคำถามสามารถนำไปใช้วัดทักษะกระบวนการรายด้านได้ (Infit MNSQ ระหว่าง 0.81 ถึง 1.11) โดยโมเดลพหุมิติสอดคล้องกลมกลืนกับข้อมูลคำตอบอย่างมีนัยสำคัญทางสถิติ (χ^2=332.48, df =2, p < .01) และมีค่า AIC และ BIC ต่ำกว่าโมเดลเอกมิติ เครื่องมือประเมินด้านการพยากรณ์และด้านการจัดกระทำและสื่อความหมายของข้อมูล มีค่าความเที่ยงแบบ EAP/PV เท่ากับ 0.78 และ 0.70 ตามลำดับ ซึ่งอยู่ในเกณฑ์ที่ยอมรับได้ และมีความคลาดเคลื่อนมาตรฐานในการวัดต่ำ

คำสำคัญ การพัฒนาเครื่องมือ วิธีสร้างโมเดลภาวะสันนิษฐาน ทักษะกระบวนการทางวิทยาศาสตร์ การวัดแบบพหุมิติ วิทยาศาสตร์ศึกษา

Abstract

The main objective of this study was to develop a measure for assessment of multidimensional scientific process skills on the topic of Stoichiometry for grade 11. There were two specific objectives: (1) to develop construct maps of scientific process skills; and (2) to create and check the quality of the measure. The sample consisted of 1,200 grade 11 students in the academic year 2018 from schools under the Secondary Educational Service Area Offices in the Northeast, Thailand. The measure was developed according to the construct modeling approach which consisted of 4 steps: (1) development of the multidimensional scientific process skill construct maps; (2) designing the test items; (3) scoring the responses; and (4) analysis of the data according to the multidimensional random coefficients multinomial logit model, using ConQuest program. The research findings were as follows:

(1) The multidimensional scientific process skill construct maps had 2 dimensions: the predicting dimension that had 4 levels, from the lowest level that the student knew the reaction occurred but could not answer or explain the occurrence correctly, to the highest level that the student could give the reasons and tell the difference of the reactions, and could make use of the knowledge of the rate of chemical reaction consistently with the authentic context; and the dimension of organizing data and communication which had 4

levels, from the lowest level that the student knew and could tell that the rate of reaction changed according to the area of the surface, but could not tell which reactant reacted to the limestone, to the highest level that the student could apply basic scientific knowledge and the thinking process in showing how to calculate the amount of hydrogen ion that reacted with the limestone and could tell the temperature which affected the reaction.

(2) The measure for assessment of the science process skills consisted of 15 open-ended items and they were polytomous items to which a score from 0 to 3 could be given, according to the level of skill in the multidimensional scientific process skill construct maps that had been developed. There were 8 items on the predicting dimension and 7 items for the organizing data and communication dimension.

(3) The results of the assessment of the science process skills were validated. On the aspect of validity evidence of the items, it was found that the item difficulties covered the range of the student skill estimates. That means the content of the items could be the definition of the student scientific process skills. On the aspect of validity evidence of the response process, it was found that the students understood the content of the items as intended. And on the aspect of validity evidence related to the internal structure, it was found that the items could be used to measure the process skills dimensionally (Infit MNSQ between 0.81 and 1.11) as the multidimensional model fitted the response data significantly (χ^2=332.48, df =2, p < .01), and the AIC and BIC were lower than the unidimensional model. The measure showed the EAP/PV reliabilities of 0.78 for the predicting dimension, 0.70 for the organizing data and communication dimension, respectively, which were acceptable; and the standard errors of measurement was low.

Keywords: measure development, construct modeling approach, scientific process skills, multidimensional measurement, science education

บทนำ

ทักษะกระบวนการทางวิทยาศาสตร์เป็นหัวใจสำคัญของการเรียนรู้วิทยาศาสตร์ และนำไปสู่การเรียนรู้ตลอดชีวิต (Arantika et al., 2019) และเป็นรากฐานสำคัญต่อการพัฒนานักวิทยาศาสตร์ให้เป็นนวัตกรรม ที่คิดค้นสิ่งที่เป็นประโยชน์ต่อโลกปัจจุบันและอนาคต ทักษะกระบวนการทางวิทยาศาสตร์แบ่งออกเป็น ทักษะขั้นพื้นฐานและทักษะขั้นบูรณาการ (American Association for the Advancement of Science, 1993; Chiappetta & Koballa, 2002; สถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยี, 2556) ทักษะขั้นพื้นฐานเป็นทักษะอย่างง่าย ประกอบด้วย การสังเกต การวัด การสรุปรวบรวม การจำแนก การทำนาย

และการจัดกระทำและสื่อความหมายข้อมูล ในขณะที่ทักษะขั้นบูรณาการเป็นทักษะขั้นสูง ประกอบด้วย การสร้างสมมติฐาน การกำหนดและควบคุมตัวแปร การออกแบบการทดลอง และการตีความหมายข้อมูล

นักเรียนที่มีทักษะกระบวนการทางวิทยาศาสตร์ขั้นพื้นฐาน จะสามารถกำหนดทิศทางการเรียนรู้และสืบเสาะหาความรู้ได้ด้วยตนเอง (Ongowo & Indoshi, 2013; Rambuda & Fraser, 2004) ตัวอย่างเช่น การวัดปริมาณสารเคมีเพื่อเร่งปฏิกิริยาในห้องปฏิบัติการ จำเป็นต้องอาศัยการสังเกตปฏิกิริยาของสาร และจัดกระทำข้อมูล และสื่อความหมายของปฏิกิริยาเคมีนั้นในรูปแบบสัญลักษณ์ จากนั้นนำผลจากการสังเกต การวัดและการทดลอง มาจัดกระทำเพื่อสื่อสารกับผู้อื่นให้ชัดเจนและถูกต้อง ด้วยการพูด การเขียน การใช้แผนภูมิ แผนภาพ กราฟ ไดอะแกรม หรือสมการ เป็นต้น จากตัวอย่างข้างต้นจะเห็นว่า ทักษะขั้นพื้นฐานทำให้นักเรียนเกิดกระบวนการคิดขั้นสูง โดยเฉพาะการคิดแก้ปัญหา (Ozgelen, 2012; Padilla, 1990) ซึ่งเป็นหัวใจของการบูรณาการระหว่างการสร้างสรรค์ความรู้ทางวิทยาศาสตร์ การเรียนรู้วิทยาศาสตร์ และการพัฒนากระบวนการทางวิทยาศาสตร์

จากการวิเคราะห์เนื้อหา มาตรฐาน ตัวชี้วัดของสาระที่ 3 ปริมาณสัมพันธ์ ชั้นมัธยมศึกษาปีที่ 5 ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พ.ศ. 2551 (กระทรวงศึกษาธิการ, 2551) พบว่า นักเรียนจำเป็นต้องมีทักษะขั้นพื้นฐาน โดยเฉพาะอย่างยิ่ง ทักษะการทำนาย และการจัดกระทำและสื่อความหมายของข้อมูล เนื่องจากสาระการเรียนรู้ดังกล่าว มุ่งเน้นการวิเคราะห์และอธิบายการจัดเรียงอิเล็กตรอนในอะตอม ความสัมพันธ์ระหว่างอิเล็กตรอนในระดับพลังงานนอกสุดกับสมบัติของธาตุและการเกิดปฏิกิริยา การทดลอง การอธิบายและการเขียนสมการของปฏิกิริยาเคมีทั่วไปที่พบในชีวิตประจำวัน รวมทั้งการอธิบายผลของสารเคมีที่มีต่อสิ่งมีชีวิตและสิ่งแวดล้อม และการทดลองและอธิบายอัตราการเกิดปฏิกิริยาเคมี ปัจจัยที่มีผลต่ออัตราการเกิดปฏิกิริยาเคมี และการนำความรู้ไปใช้ประโยชน์ (สำนักวิชาการและมาตรฐานการศึกษา, 2551) ทำให้การพัฒนาเครื่องมือประเมินทักษะข้างต้นจึงเป็นสิ่งจำเป็นด้วย แต่อย่างไรก็ตาม ผลการทบทวนเอกสารและงานวิจัยที่เกี่ยวข้อง พบว่า เครื่องมือประเมินทักษะพื้นฐานส่วนใหญ่เป็นแบบทดสอบ ที่รายงานคะแนนดิบรวม (total raw score) หรือคะแนนรายทักษะ แล้วนำคะแนนเหล่านั้นเปรียบเทียบกับระดับคุณภาพ (Azizah et al., 2017) จุดอ่อนที่สำคัญของเครื่องมือเหล่านี้ คือ ขาดการตีความสารสนเทศเพื่อให้นักเรียนสามารถนำไปใช้กำหนดทิศทางการพัฒนาทักษะด้วยตนเอง และขาดสารสนเทศที่ผู้สอนสามารถนำไปใช้ปรับปรุงและออกแบบกิจกรรมการเรียนรู้ให้เหมาะสมกับนักเรียน นั่นหมายความว่า เครื่องมือประเมินเหล่านั้นไม่สามารถบรรยายได้ว่า นักเรียนมีทักษะพื้นฐานแต่ละด้านเป็นอย่างไร และอยู่ในระดับใด และนักเรียนจะพัฒนาทักษะพื้นฐานแต่ละด้านให้สูงขึ้นได้อย่างไร และสะท้อนให้เห็นว่า เครื่องมือส่วนใหญ่ไม่สามารถให้หลักฐานสนับสนุนความตรงและความเที่ยงของผลการประเมิน ตามมาตรฐานการทดสอบทางการศึกษาและจิตวิทยา ของประเทศสหรัฐอเมริกา (American Educational Research Association et al., 2014)

วิธีสร้างโมเดลภาวะสันนิษฐาน (Construct modeling approach: Wilson, 2005) เป็นวิธีพัฒนาเครื่องมือประเมิน ที่เน้นการสร้างโครงสร้างหรือระดับคุณลักษณะแฝง (latent trait) ที่มีอยู่ภายในบุคคล และใช้โครงสร้างดังกล่าวในการตีความผลการวัด และสรุปอ้างอิงคุณลักษณะของบุคคล (วีรภัทร์ สุขศิริ และ

ชนม์ชกรณ์ วรอินทร์, 2559) นอกจากนี้วิธีดังกล่าวยังสามารถเชื่อมโยงได้ดีกับเป้าหมายของหลักสูตรและการสอน ทำให้การประเมินสอดคล้องกับหลักสูตร กิจกรรมการเรียนการสอน และกิจกรรมการประเมินการเรียนรู้

การวัดทางจิตวิทยาและการศึกษาเกี่ยวข้องกับการกำหนดค่าเชิงปริมาณให้กับลักษณะที่สนใจ เช่น ค่า 1 หมายถึง ลักษณะ A, 2 หมายถึง ลักษณะ B และ 3 หมายถึง ลักษณะ C ซึ่งในความเป็นจริงแล้ว กระบวนการดังกล่าวเป็นการวัดที่ยังไม่สมบูรณ์ เพราะจะต้องประกอบด้วย การตรวจสอบความถูกต้องของผลการวัด และการนำผลการวัดไปใช้ประโยชน์ และรวมถึงการแปลความหมายของค่าที่ได้อย่างเป็นเหตุเป็นผลเพื่อสรุปอ้างอิงเกี่ยวกับลักษณะของสิ่งที่วัด (Wilson, 2005) อย่างไรก็ตาม เครื่องมือวัดนั้นจะต้องให้ผลการวัดที่ครอบคลุมถึงโครงสร้างของคุณลักษณะที่มุ่งวัด และคำตอบของนักเรียนจะต้องสัมพันธ์กับ ระดับต่าง ๆ ในโครงสร้างของคุณลักษณะที่พัฒนาไว้ก่อนหน้า การวัดตามแนวคิดของ Wilson (2005) หรือวิธีสร้างโมเดลภาวะสันนิษฐานสามารถแก้ไขข้อจำกัดข้างต้น โดยประกอบด้วย 4 ขั้นตอน ได้แก่ การสร้างแผนที่ภาวะสันนิษฐาน (construct map) การออกแบบข้อคำถาม (items design) การให้คะแนนคำตอบ (outcome space) และการประมาณค่าคุณลักษณะแฝงของบุคคลด้วยโมเดลการวัด (measurement model)

วิธีสร้างโมเดลภาวะสันนิษฐาน มีหลักการพื้นฐานอยู่บนการวัดแบบ Rasch (1980) จึงให้แผนที่ของ Wright (Wilson & Draney, 2002) เป็นผลผลิตที่สำคัญต่อการตีความผลการประมาณค่าคุณลักษณะแฝงของนักเรียน แผนที่ของ Wright นี้สามารถใช้เป็นหลักฐานเพื่อสนับสนุนความตรงของผลการวัด กล่าวคือ ใช้บ่งชี้ประสิทธิภาพของการออกแบบการประเมินว่า เหมาะสมกับนักเรียนที่มีความสามารถแตกต่างกัน และสภาพห้องเรียนที่มีบริบทแตกต่างกันหรือไม่ สารสนเทศที่ได้จาก Wright map จะแสดงถึงความก้าวหน้าในการเรียนรู้ของนักเรียน จากระดับพื้นฐานสู่ระดับซับซ้อน ทั้งแบบภาพรวม (unidimensional) และแบบรายด้าน (multidimensional) โดยมีโมเดลสัมประสิทธิ์สุ่มสำหรับข้อมูลหลายค่าที่มาจากการวัดหลายมิติ (Multidimensional random coefficients multinomial logit [MRCML] model: Adams et al., 1997) เป็นเครื่องมือสำคัญต่อการประมาณค่าพารามิเตอร์ต่าง ๆ และเชื่อมโยงสารสนเทศระหว่างแผนที่ของ Wright และแผนที่ภาวะสันนิษฐาน

ความมุ่งหมายของการวิจัย

การวิจัยนี้เล็งเห็นถึงความสำคัญของการประเมินทักษะกระบวนการทางวิทยาศาสตร์ และจุดเด่นของวิธีสร้างโมเดลภาวะสันนิษฐาน จึงได้พัฒนาเครื่องมือประเมินทักษะการพยากรณ์ และทักษะการจัดทำและสื่อความหมายข้อมูล ที่เกี่ยวข้องกับเนื้อหาเรื่องปริมาณสัมพันธ์ สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 5 ด้วยวิธีสร้างโมเดลภาวะสันนิษฐาน มีวัตถุประสงค์ย่อย คือ 1) เพื่อพัฒนาแผนที่โครงสร้างทักษะกระบวนทางวิทยาศาสตร์ 2) เพื่อสร้างและตรวจสอบคุณภาพของเครื่องมือประเมินทักษะกระบวนทางวิทยาศาสตร์ และ 3) เพื่อกำหนดจุดเปลี่ยนผ่านของผลประเมินทักษะกระบวนทางวิทยาศาสตร์

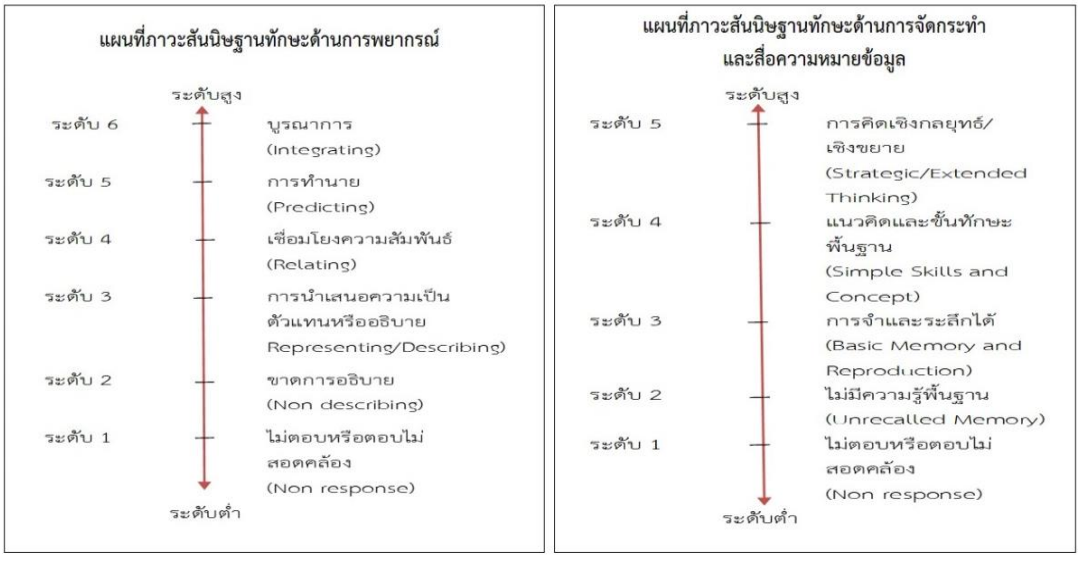
วิธีดำเนินการวิจัย

1. กลุ่มผู้สอบ

ผู้วิจัยกำหนดขนาดกลุ่มตัวอย่างเท่ากับ 1,200 คน ให้เพียงพอต่อการประมาณค่าพารามิเตอร์ด้วยโมเดล MRCML และให้ครอบคลุมระดับความสามารถของนักเรียนตามแผนที่ภาวะสันนิษฐาน แทนที่จะคำนึงถึงความเป็นตัวแทนประชากรเพื่อสรุปอ้างอิงนัยทั่วไป (Demars, 2010) จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง พบว่า ขนาดกลุ่มตัวอย่าง อย่างน้อย 200 คน เพียงพอสำหรับประมาณค่าพารามิเตอร์ ให้ความคลาดเคลื่อนต่ำ และเกิดความแปรเปลี่ยนต่ำระหว่างกลุ่มผู้สอบและข้อสอบ (Sass et al., 2008) และเมื่อเพิ่มขนาดตัวอย่างขั้นต่ำเป็นระหว่าง 250 – 500 คน พบว่า ค่าพารามิเตอร์แม่นยำมากยิ่งขึ้น (Linacre, 1994; Wright & Stone, 1979) นอกจากนี้ ผู้วิจัยได้กำหนดเกณฑ์การเลือกตัวอย่างนักเรียนชั้นมัธยมศึกษาปีที่ 5 กลุ่มโรงเรียนภาคตะวันออกเฉียงเหนือ โดยพิจารณาจาก 1) ผลการประเมินนักเรียนทางวิทยาศาสตร์ทั้งในระดับชาติและนานาชาติ 2) ขนาดโรงเรียน และ 3) ความพร้อมของโรงเรียน

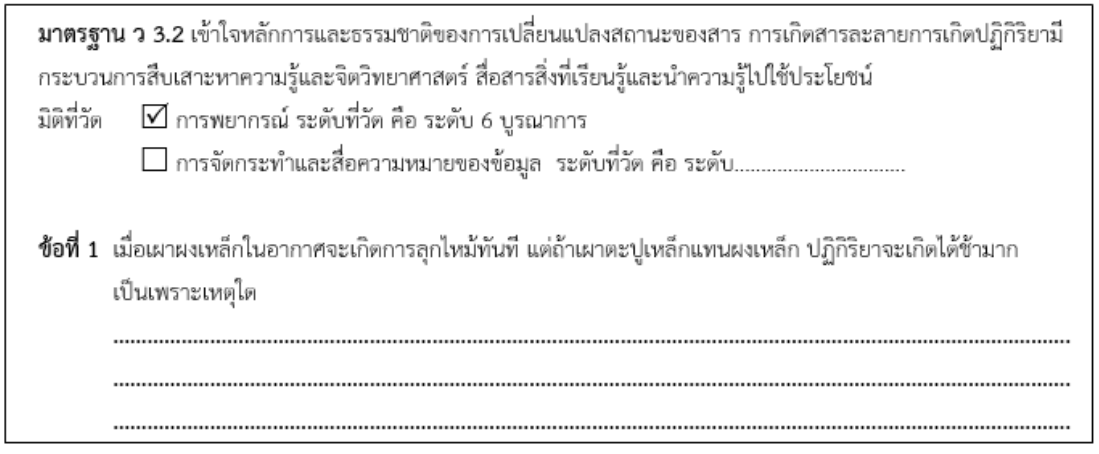
2. ระเบียบวิธีวิจัย

ขั้นตอนที่ 1 การพัฒนาแผนที่ภาวะสันนิษฐานทักษะกระบวนการทางวิทยาศาสตร์ โดยผู้วิจัยบรรยายระดับทักษะของนักเรียนด้านการพยากรณ์ และด้านการจัดกระทำและสื่อความหมายของข้อมูล ซึ่งได้จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง (Junpeng et al., 2018; Webb, 1997; พนิดา กระทุ่มนอก และ พัชรี จันทร์เพ็ง, 2560) จากการสังเกตในชั้นเรียนจริง และจากกระบวนการคิดเสียงดัง (think aloud) ของนักเรียน ทำให้ได้ต้นร่างแผนที่ภาวะสันนิษฐานทักษะด้านการพยากรณ์ ประกอบด้วย 6 ระดับ ตั้งแต่ระดับต่ำสุด คือ ขั้นไม่ตอบหรือตอบไม่สอดคล้อง ไปจนถึงระดับสูงสุด คือ ขั้นบูรณาการ และต้นร่างแผนที่ภาวะสันนิษฐานทักษะด้านการจัดกระทำและสื่อความหมายข้อมูล ประกอบด้วย 5 ระดับ ตั้งแต่ระดับต่ำสุด คือ ขั้นไม่ตอบหรือตอบไม่สอดคล้อง ไปจนถึงระดับสูงสุด คือการคิดเชิงกลยุทธ์/เชิงขยาย แสดงดังภาพประกอบ 1



ภาพประกอบ 1 แผนที่ภาวะสันนิษฐานทักษะกระบวนการทางวิทยาศาสตร์

ขั้นตอนที่ 2 การพัฒนาข้อคำถามตามแผนที่ภาวะสันนิษฐาน เป็นการออกแบบข้อคำถามเพื่อให้ได้ข้อมูลตามที่บรรยายไว้ในแผนที่ภาวะสันนิษฐาน โดยมีข้อคำถามด้านการพยากรณ์จำนวน 8 ข้อ และด้านการจัดกระทำและสื่อความหมายข้อมูลจำนวน 7 ข้อ แต่ละข้อตรวจให้คะแนนแบบหลายค่า โดยข้อคำถาม 1 ข้อวัดหนึ่งมิติ ตัวอย่างข้อคำถามดังภาพประกอบ 2



มาตรฐาน ว 3.2 เข้าใจหลักการและธรรมชาติของการเปลี่ยนแปลงสถานะของสาร การเกิดสารละลายการเกิดปฏิกิริยามีกระบวนการสืบเสาะหาความรู้และจิตวิทยาศาสตร์ สื่อสารสิ่งที่เรียนรู้และนำความรู้ไปใช้ประโยชน์

มิติที่วัด ☑ การพยากรณ์ ระดับที่วัด คือ ระดับ 6 บูรณาการ

☐ การจัดกระทำและสื่อความหมายของข้อมูล ระดับที่วัด คือ ระดับ.....

ข้อที่ 1 เมื่อเผาผงเหล็กในอากาศจะเกิดการลุกไหม้ทันที แต่ถ้าเผาตะปูเหล็กแทนผงเหล็ก ปฏิกิริยาจะเกิดได้ช้ามาก เป็นเพราะเหตุใด

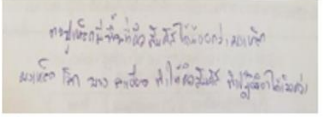
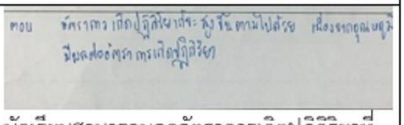
.....

.....

.....

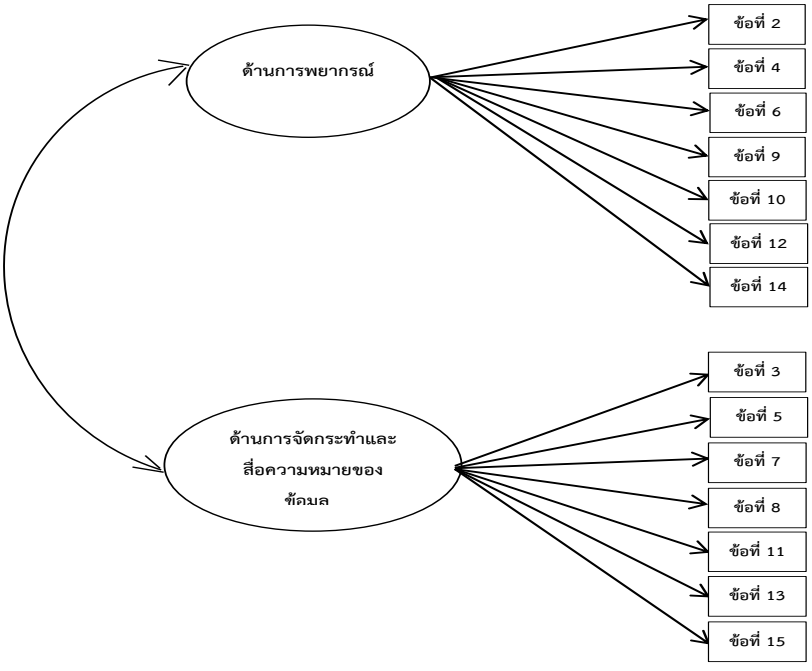
ภาพประกอบ 2 ตัวอย่างข้อคำถามตามแผนที่โครงสร้างด้านการพยากรณ์

ขั้นตอนที่ 3 การให้คะแนนคำตอบ เป็นการให้คะแนนคำตอบของนักเรียนที่มีต่อข้อคำถามที่สร้างขึ้น โดยเกณฑ์การให้คะแนนคำตอบจะสัมพันธ์กับระดับของทักษะกระบวนการทางวิทยาศาสตร์ ตามแผนที่ภาวะสันนิษฐาน คะแนนที่นักเรียนจะได้รับมีตั้งแต่ 0 ถึง 5 คะแนน สำหรับข้อคำถามด้านการพยากรณ์ และตั้งแต่ 0 ถึง 4 คะแนน สำหรับข้อคำถามด้านการจัดกระทำและสื่อความหมายข้อมูล ตัวอย่างการให้คะแนนคำตอบแสดงดังภาพประกอบ 3

ระดับ	คะแนน	ชื่อ	คำอธิบาย	ตัวอย่างผลการตอบ
ตัวอย่างเกณฑ์การให้คะแนนและตัวอย่างคำตอบของข้อคำถามด้านการพยากรณ์				
4	3	ขั้นเชื่อมโยงความสัมพันธ์ (Relating)	- นักเรียนสามารถบอกสถานะที่แตกต่างกันของสารตั้งต้นได้ - นักเรียนมีมโนมิติที่ถูกต้องต่ออัตราการเกิดปฏิกิริยาและสามารถเปรียบเทียบอัตราการเกิดปฏิกิริยาได้	 นักเรียนอธิบายการเกิดปฏิกิริยาของสารจากพื้นที่ผิวสัมผัสได้อย่างเหมาะสม
ตัวอย่างเกณฑ์การให้คะแนนและตัวอย่างคำตอบของข้อคำถามด้านการจัดกระทำและสื่อความหมาย				
4	3	แนวคิดและขั้นทักษะพื้นฐาน (Simple Skills and Concept)	- นักเรียนสามารถอธิบายได้ว่าปฏิกิริยาเคมีที่เป็นแก๊ส และอัตราการเกิดปฏิกิริยาแปรผันตามอุณหภูมิ	 นักเรียนสามารถบอกอัตราการเกิดปฏิกิริยาที่แปลผันตามอุณหภูมิ

ภาพประกอบ 3 ตัวอย่างเกณฑ์การให้คะแนนของการประเมินทักษะกระบวนการทางวิทยาศาสตร์

ขั้นตอนที่ 4 การวิเคราะห์ข้อมูลตามโมเดลการวัด เป็นขั้นตอนที่ผู้วิจัยนำคะแนนคำตอบมาประมาณค่าพารามิเตอร์ต่าง ๆ ตามโมเดล MRCML ดังภาพ 4 ด้วยโปรแกรม ConQuest 2.0 (Wu et al., 2007)



ภาพประกอบ 4 โมเดลการวัดทักษะกระบวนการทางวิทยาศาสตร์แบบพหุมิติ

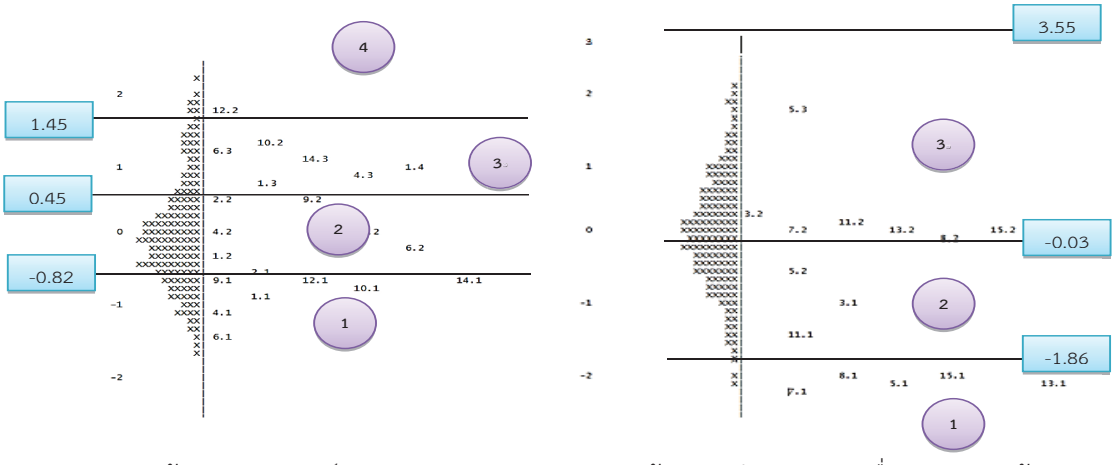
สารสนเทศที่ได้จากขั้นตอนที่ 1 ถึง 4 นำมาใช้เป็นหลักฐานความตรง และคำนวณหาจุดเปลี่ยนผ่านของระดับทักษะกระบวนการทางวิทยาศาสตร์ ถ้าจุดเปลี่ยนผ่านใกล้มีตำแหน่งบนแผนที่ของ Wright ใกล้กันและไม่สามารถจำแนกความแตกต่างของระดับทักษะได้ ผู้วิจัยจะยุบรวมระดับเหล่านั้นเข้าด้วยกัน จุดเปลี่ยนผ่านคำนวณจากค่าเฉลี่ยของค่าความยากขั้น (item threshold) ของแต่ละด้าน จากนั้นผู้วิจัยจึงแปลค่าประมาณทักษะกระบวนการของแต่ละด้านเป็นคะแนนสเกล $\theta^{*} = 50 + (10*\theta)$ เมื่อ θ คือ ค่าพารามิเตอร์ทักษะกระบวนการทางวิทยาศาสตร์ ดังนั้นนักเรียนแต่ละคนจะมีคะแนนสเกล 2 ค่าตามจำนวนด้านที่ทำการศึกษา (พัชรี จันทร์เพ็ง, 2561; วีรภัทร์ สุขศิริ และ ชนม์ชกรณ์ วรอินทร์, 2559)

ผลการวิจัย

1. ผลการพัฒนาแผนที่ภาวะสันนิษฐานทักษะกระบวนทางวิทยาศาสตร์

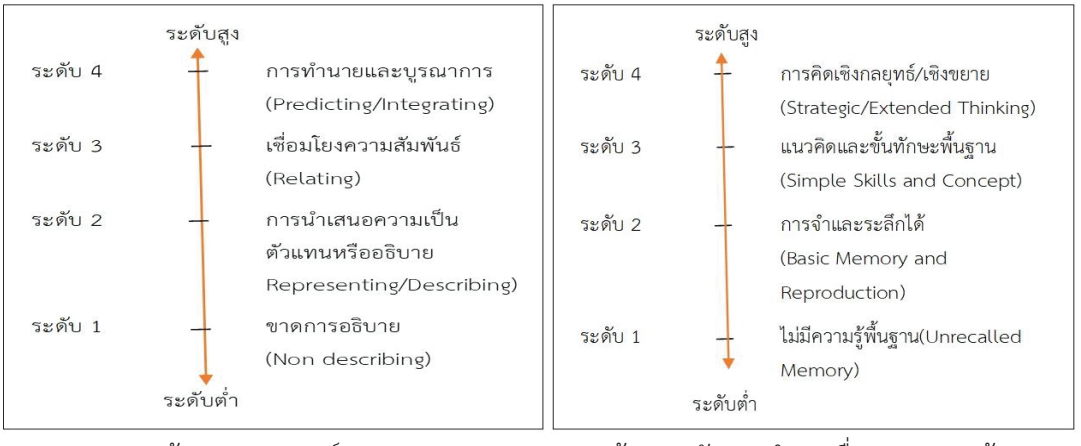
ผลการความตรงของแผนที่ภาวะสันนิษฐานทักษะกระบวนการทางวิทยาศาสตร์ จากแผนที่ของ Wright และค่าเฉลี่ยจุดเปลี่ยนผ่านของแต่ละระดับของด้านการพยากรณ์ พบว่า ผู้วิจัยรวมระดับ 1 และ 2 เข้าด้วยกัน เป็นขั้นการไม่อธิบาย และรวมระดับ 5 และ 6 เป็นขั้นการทำนายและบูรณาการ แผนที่ภาวะสันนิษฐานใหม่นี้จึงมี 4 ระดับ ดังภาพ 6a นอกจากนี้ผู้วิจัยกำหนดจุดเปลี่ยนผ่านจำนวน 3 จุด คือ จากระดับ 1 ไป 2, ระดับ 2 ไป 3 และระดับ 3 ไป 4 เท่ากับ -0.82 0.45 และ 1.45 ลอจิต ซึ่งเท่ากับคะแนนสเกล 54.50, 64.50 และ 79.00 ตามลำดับ ดังภาพ 5a

สำหรับด้านการจัดกระทำและสื่อความหมาย พบว่า ผู้วิจัยรวม มีระดับ 1 และ 2 เข้าด้วยกัน เป็นขั้นไม่มีความรู้พื้นฐาน แผนที่ภาวะสันนิษฐานใหม่นี้จึงมี 4 ระดับ ดังภาพประกอบ 6b และมีจุดเปลี่ยนผ่านจำนวน 3 จุด เท่ากับ -1.86, -0.03 และ 3.55 ลอจิต ตามลำดับ ซึ่งเท่ากับคะแนนสเกล 31.40, 49.70 และ 85.50 ตามลำดับ ดังภาพประกอบ 5b



(a) ด้านการพยากรณ์ (b) ด้านการจัดกระทำและสื่อความหมายข้อมูล

ภาพประกอบ 5 แผนที่ของ Wright และจุดเปลี่ยนผ่านระดับทักษะกระบวนการทางวิทยาศาสตร์



(a) ด้านการพยากรณ์ (b) ด้านการจัดกระทำและสื่อความหมายข้อมูล

ภาพประกอบ 6 แผนที่ภาวะสันนิษฐานเชิงประจักษ์ทักษะกระบวนการทางวิทยาศาสตร์

ผลการตรวจสอบคุณภาพของเครื่องมือเป็นไปตามเกณฑ์ที่ยอมรับได้ กล่าวคือ ข้อคำถามมีค่าความยากกระจายครอบคลุมช่วงค่าประมาณทักษะกระบวนการของนักเรียน ดังภาพประกอบ 7 นั่นหมายความว่า ข้อสอบที่พัฒนาขึ้นเป็นนิยามของทักษะกระบวนการทางวิทยาศาสตร์ ทั้งด้านการพยากรณ์ และด้านการกระทำและสื่อความหมายของข้อมูล นอกจากนี้ ข้อคำถามมีค่าความสอดคล้องกลมกลืนแบบถ่วงน้ำหนัก (infit MNSQ) ระหว่าง 0.81 ถึง 1.11 ซึ่งอยู่ในช่วงที่ยอมรับได้ (Adams & Khoo, 1996) นั่นคือ ข้อคำถามทั้ง 15 ข้อสอดคล้องกลมกลืนกับโมเดลการวัดแบบพหุมิติ และอธิบายความผันแปรของทักษะกระบวนการของนักเรียนได้

ภาพประกอบ 7 หลักฐานความตรงเกี่ยวกับเนื้อหาของข้อคำถาม

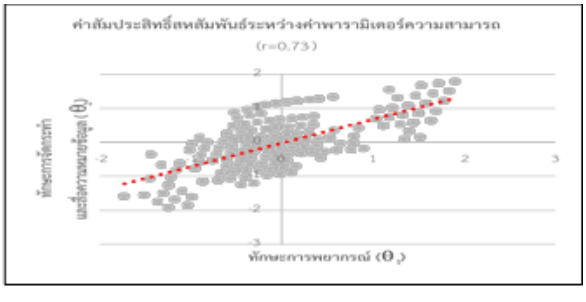
สำหรับหลักฐานความตรงเกี่ยวกับกระบวนการตอบนั้น ผู้วิจัยนำข้อคำถามไปทดลองใช้ และสัมภาษณ์นักเรียนภายหลังจากทำข้อคำถามแล้ว เพื่อตรวจสอบว่านักเรียนเข้าใจเนื้อหาของข้อคำถามตรงกับผู้วิจัยหรือไม่ พบว่า นักเรียนเข้าใจคำถามตรงตามความตั้งใจของผู้วิจัย นอกจากนี้ผู้วิจัยได้ข้อมูลเหล่านี้ไปใช้ปรับปรุงข้อคำถามก่อนนำไปใช้จริงด้วย

ผลการตรวจสอบความตรงโครงสร้างภายใน พบว่า โมเดลพหุมิติสอดคล้องกลมกลืนกับข้อมูลคำตอบดีกว่าโมเดลเอกมิติอย่างมีนัยสำคัญทางสถิติ (χ^2= 332.48, df = 2, p < .05) โดยมีค่า AIC (Akaike, 1974) และ BIC (Schwarz, 1978) ต่ำกว่า ดังแสดงในตาราง 1

ตาราง 1 ผลการตรวจสอบความสอดคล้องกลมกลืนของโมเดลพหุมิติ

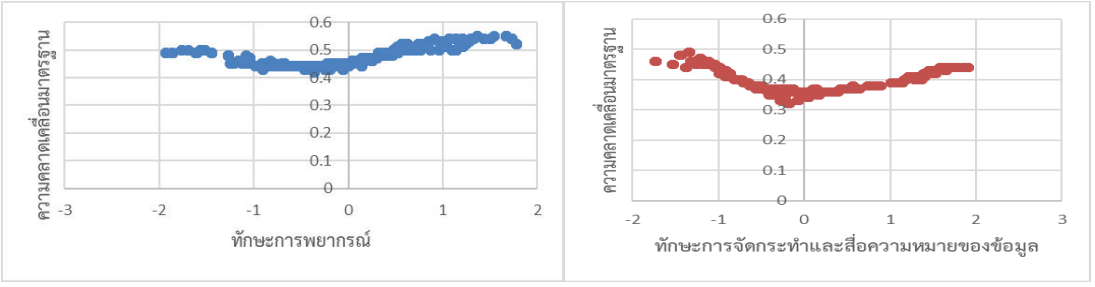
โมเดล	Deviance	จำนวนพารามิเตอร์	AIC	BIC
โมเดลเอกมิติ	38769.07	43	38855.07	39033.88
โมเดลพหุมิติ	38436.59	45	38526.59	38713.72

นอกจากนี้เมื่อพิจารณาความสัมพันธ์ของค่าประมาณทักษะกระบวนการทางวิทยาศาสตร์ระหว่างด้านการพยากรณ์และด้านการจัดกระทำและสื่อความหมายของข้อมูล พบว่า มีค่าเท่ากับ .73 แสดงให้เห็นว่าความสามารถของนักเรียนส่วนใหญ่เป็นไปในทิศทางเดียวกัน กล่าวคือ นักเรียนมีคะแนนด้านใดด้านหนึ่งสูงจะมีคะแนนอีกด้านสูงตามไปด้วย ดังแสดงในภาพประกอบ 8



ภาพประกอบ 8 ความสัมพันธ์ของค่าประมาณทักษะกระบวนการทางวิทยาศาสตร์แต่ละด้าน

ข้อคำถามที่พัฒนาขึ้นมีค่าความเที่ยงแบบ EAP/PV ด้านการพยากรณ์และด้านการจัดกระทำและสื่อความหมายเท่ากับ 0.78 และ 0.70 ตามลำดับ ซึ่งอยู่ในเกณฑ์ที่ยอมรับได้ (Adams, 2005) และค่าความเที่ยงแบบสอดคล้องภายใน เท่ากับ 0.80 ซึ่งอยู่ในเกณฑ์ที่ยอมรับได้เช่นกัน นอกจากนี้เมื่อพิจารณากราฟความคลาดเคลื่อนมาตรฐานของการวัด พบว่า ค่าประมาณทักษะกระบวนการทั้ง 2 ด้าน มีค่าความคลาดเคลื่อนต่ำ (Wilson, 2005) ดังภาพประกอบ 9 ผลวิเคราะห์ดังกล่าวเป็นหลักฐานสนับสนุนความคงเส้นคงวาของการประมาณค่าทักษะกระบวนการทางวิทยาศาสตร์ของทั้งสองด้าน



ภาพประกอบ 9 กราฟความคลาดเคลื่อนมาตรฐานของการวัด

อภิปรายผล

จากผลการวิจัยในครั้งนี้ พบว่า แผนที่ภาวะสันนิษฐานทักษะกระบวนทางวิทยาศาสตร์ทั้งสองด้านไม่เป็นไปตามที่กำหนดไว้ตั้งแต่ต้น โดยผู้วิจัยจำเป็นต้องยุบระดับสูงสุด 2 ระดับเข้าด้วยกัน เนื่องจากเกณฑ์การตรวจให้คะแนนไม่สามารถจำแนกความแตกต่างของนักเรียนที่อยู่ในระดับ 5 และ 6 ได้ และมีจำนวนนักเรียนน้อยมากในสองระดับนี้ แสดงให้เห็นว่า ระดับทักษะกระบวนการสูงสุดไม่สอดคล้องกับการออกแบบการประเมินในครั้งนี้ นอกจากนี้เมื่อพิจารณาในระดับ 1 และ 2 พบว่า เกณฑ์การให้คะแนนไม่สามารถจำแนกนักเรียนสองระดับนี้ได้เช่นกัน จึงมีความจำเป็นต้องยุบระดับ 1 และ 2 เข้าด้วยกัน เพื่อให้แผนที่ภาวะสันนิษฐานมีระดับทักษะสอดคล้องกับข้อมูลเชิงประจักษ์มากขึ้น ข้อค้นพบดังกล่าวสอดคล้องกับข้อเสนอแนะที่ว่า แผนที่ภาวะสันนิษฐานควรสร้างให้ครอบคลุมทุกระดับความสามารถ ไม่ควรมีจำนวนระดับมากจนเกินไปจนนำไปใช้จริงไม่ได้ (Cizek, 2012)

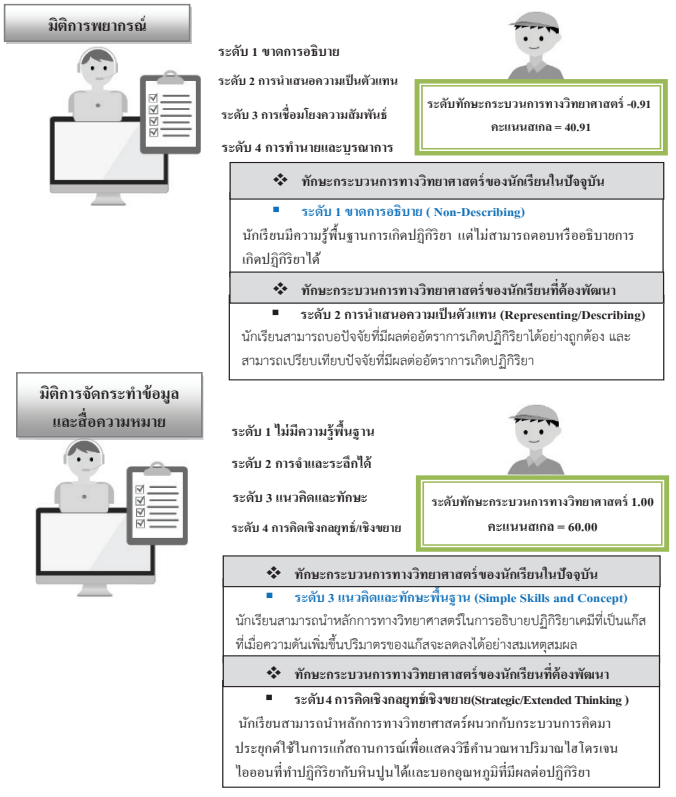
เมื่อพิจารณาแผนที่ภาวะสันนิษฐานด้านการจัดกระทำและสื่อความหมายข้อมูล พบว่า ระดับขั้นไม่ตอบหรือตอบไม่สอดคล้อง สะท้อนว่านักเรียนไม่มีความรู้พื้นฐานด้วย ดังนั้น ระดับ 1 และ 2 จึงมีสารสนเทศแบบเดียวกัน และไม่สามารถจำแนกความแตกต่างได้ ผลวิจัยดังกล่าวสอดคล้องกับผลการพัฒนาแผนที่ภาวะสันนิษฐานการจัดกระทำและสื่อความหมายข้อมูล ในรูปแบบการแสดงแทนทางคณิตศาสตร์ของ Junpeng et al. (2018) โดยงานวิจัยนี้เน้นกระบวนการแก้ปัญหาทางคณิตศาสตร์ โดยนำแนวคิดความลึกของความรู้ (depth of knowledge: Webb, 1997) มาเป็นฐานในการพัฒนาแผนที่ภาวะสันนิษฐาน และยุบรวมระดับไม่ตอบหรือตอบไม่เกี่ยวข้อง รวมกับระดับไม่มีความรู้พื้นฐาน เช่นเดียวกัน

เมื่อพิจารณาความเหมาะสมของจุดเปลี่ยนผ่านพบว่า สามารถนำไปใช้ประเมินทักษะกระบวนการทางวิทยาศาสตร์ได้ อย่างไรก็ตาม ด้านการจัดกระทำและสื่อความหมายข้อมูล มีจุดเปลี่ยนผ่านที่สุดโต่งอย่างชัดเจน โดยมีจุดเปลี่ยนผ่านจากระดับที่ 3 ไประดับที่ 4 เท่ากับ 3.55 โดยมีคะแนนสเกลเท่ากับ 85.5 จึงส่งผลทำให้ไม่มีนักเรียนคนใดเลยอยู่ในระดับดังกล่าว ซึ่งเมื่อพิจารณาจากแผนที่ของ Wright พบว่า ควรเพิ่มจำนวนข้อคำถามที่วัดระดับสูงมากขึ้น รวมถึงปรับคำบรรยายของระดับทักษะในแผนที่ภาวะสันนิษฐานสอดคล้องกับการเรียนรู้ในชั้นเรียนมากขึ้นด้วย

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

1.1 ครูและผู้เกี่ยวข้องสามารถนำข้อมูลผลการประเมินระดับทักษะกระบวนการทางวิทยาศาสตร์ไปวินิจฉัยนักเรียน เพื่อพัฒนานักเรียนให้มีระดับที่สูงขึ้น และเป็นการเพิ่มศักยภาพในการให้ข้อมูลย้อนกลับแก่นักเรียนในระบบสารสนเทศ และนำไปสู่การปรับปรุงการเรียนการสอนให้มีประสิทธิภาพสูงขึ้น ตัวอย่างการรายงานผลรายบุคคลของนักเรียน ดังภาพประกอบ 10



ภาพประกอบ 10 ตัวอย่างรายงานผลการประเมินระดับทักษะกระบวนการทางวิทยาศาสตร์รายบุคคล

1.2 แผนที่ภาวะสันนิษฐานแบบพหุมิติ ซึ่งมีคะแนนจุดเปลี่ยนผ่านทั้ง 4 ระดับ อยู่บนพื้นฐานของทฤษฎีการตอบสนองข้อสอบ และสามารถตีความทักษะกระบวนการทางวิทยาศาสตร์ของนักเรียนเป็นระดับได้ ครูจึงสามารถนำผลการประเมินการเรียนรู้ในชั้นเรียนมาพิจารณาร่วมกับแผนที่ภาวะสันนิษฐานนี้ได้ โดยพิจารณาว่า ภาระงานแต่ละชิ้นที่ใช้ในห้องเรียนตรงกับระดับทักษะกระบวนการใด รวมทั้งข้อคำถามที่ครูสร้างขึ้นด้วย เพื่อเป็นข้อมูลย้อนกลับแก่นักเรียนให้นักเรียนกำหนดทิศทางการเรียนรู้ได้ด้วยตนเอง

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 การวิจัยนี้ตรวจสอบความตรงเพียง 3 ด้าน ดังนั้น การวิจัยครั้งต่อไปควรตรวจสอบความตรงเกี่ยวกับความสัมพันธ์กับตัวแปรอื่น โดยพิจารณาจากค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนกับผลการประเมินอื่น ที่เกี่ยวข้อง เป็นต้น และความตรงเกี่ยวกับผลสืบเนื่องจากการนำข้อคำถามที่พัฒนาขึ้นไปใช้ เป็นการเก็บข้อมูลเกี่ยวกับผลกระทบทางบวกและทางลบจากการประเมินที่ผู้วิจัยพัฒนาขึ้น เช่น สัมภาษณ์ครูผู้สอนหลังจากใช้การประเมินนี้

2.2 ควรวิเคราะห์อิทธิพลของผู้ตรวจให้คะแนน โดยใช้เทคนิคโมเดลหลายฟาเซต (multi – facet model: Linacre, 1994)

2.3 การพัฒนาเครื่องมือและการรายงานผลการประเมินระดับทักษะกระบวนการทางวิทยาศาสตร์นั้น ในการวิเคราะห์ค่าคะแนนจุดเปลี่ยนผ่านและการประมาณค่าพารามิเตอร์ความสามารถยังเป็นสเกลที่เข้าใจยากในทางปฏิบัติ จึงจำเป็นอย่างยิ่งที่ต้องพัฒนาเทคโนโลยีในการประมวลผลที่ง่ายต่อการนำไปใช้ของครู รวมถึงผู้เกี่ยวข้อง มีการออกแบบวิธีการรายงานผลการประเมินนักเรียนให้สามารถวินิจฉัยระดับทักษะกระบวนการทางวิทยาศาสตร์ได้ทั้งรายบุคคลและภาพรวม เพื่อเป็นประโยชน์ทั้งครู ศึกษานิเทศก์ เขตพื้นที่การศึกษา โดยเฉพาะหน่วยงานในระดับชาติในการนำผลการประเมินไปปรับปรุงและพัฒนานักเรียนได้อย่างทันท่วงที

เอกสารอ้างอิง

กระทรวงศึกษาธิการ. (2551). *หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551*. กรุงเทพมหานคร: โรงพิมพ์คุรุสภาลาดพร้าว.

พนิดา กระทุ่มนอก และ พัชรี จันทร์เพ็ง. (2560). *ความก้าวหน้าในการเรียนรู้มโนมติทางวิทยาศาสตร์ของนักเรียน ระดับชั้นมัธยมศึกษาปีที่ 4 เรื่อง รูปร่างโมเลกุลโควาเลนต์*. ขอนแก่น: มหาวิทยาลัยขอนแก่น.

พัชรี จันทร์เพ็ง. (2561). *การประยุกต์ทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเพื่อการวิจัย.* ขอนแก่น: โรงพิมพ์มหาวิทยาลัยขอนแก่น.

วีรภัทร์ สุขศิริ และ ชนม์ชกรณ์ วรอินทร์. (2559). *การตรวจคะแนนจุดตัดขั้นต้นสำหรับกลุ่มสาระการเรียนรู้วิทยาศาสตร์จากคะแนนสอบระดับชาติขั้นพื้นฐานด้วยวิธีทำแผนที่ความสามารถแฝงเชิงประจักษ์: การวิเคราะห์เพื่อการตัดสิน*. กรุงเทพมหานคร: สถาบันทดสอบทางการศึกษาแห่งชาติ.

สำนักวิชาการและมาตรฐานการศึกษา. (2551). *ตัวชี้วัดและสาระการเรียนรู้แกนกลาง กลุ่มสาระการเรียนรู้วิทยาศาสตร์ ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551*. กรุงเทพมหานคร: โรงพิมพ์ชุมนุมสหกรณ์การเกษตรแห่งประเทศไทย

Adams, R. J. (2005). Reliability as a measurement design effect. *Studies in Educational Evaluation, 31*, 162-172.

Adams, R. J., & Khoo, S. T. (1996). *Quest.* Melbourne, Australia: Australian Council for Educational Research.

Adams, R. J., Wilson, M., & Wang, W. C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement, 21*(1), 1-23

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transaction on Automatic Control, 19*(6), 716-723.

American Educational Research Association, American Psychological Association, & Nation Council on Measurement in Education. (2014). *Standards for education and psychological testing*. Washington, DC: American Educational Research Association.

American Association for the Advancement of Science. (1993). *American Association for the Advancement of Science Project 2061: Science for all Americans*. Washington, DC: American Association for the Advancement of Science

Arantika, J., Saputro, S., & Mulyani, S. (2019). Effectiveness of guided inquiry-based module to improve science process skills. *Journal of Physics: Conference Series, 1157*(4).

Azizah, N., Ibrahim, M., & Widodo, W. (2017). Process skill assessment instrument: Innovation to measure student's learning result holistically. *Journal of Physics Conference, 947*(1).

Chiappetta, E. L., & Koballa, T. R. (2002). *Science instruction in the middle and secondary schools.* Upper Saddle River, NJ: Merrill/Prentice Hall.

Cizek, G. J. (2012). Defining and distinguishing validity: Interpretations of score meaning and justifications of test use. *Psychological Methods*, 17(1), 31.

DeMars, C. (2010). *Item response theory: Understanding statistics measurem*ent. Oxford: Oxford University Press.

Junpeng, P., Inprasitha, M., & Wilson, M. (2018). Modelling of the open-ended items for assessing multiple proficiencies in mathematical problem solving. *The Turkish Online Journal of Educational Technology (TOJET), 17*.

Linacre, J. M. (1994). Sample size and item calibration [or person measure] stability. *Rasch Measurement Transactions, 7*(4), 328.

_____. (1998). *Many-facet Rasch measurement*. Chicago: MESA Press.

Ongowo, R. O., & Indoshi, F. C. (2013). Science process skills in the Kenya certificate of secondary education biology practical examinations. *Creative Education, 11*(4), 713-717.

Ozgelen, S. (2012). Scientists' science process skills within a cognitive domain framework. *Eurasia Journal of Mathematics, Science & Technology Education, 8*, 283-292.

Padilla, M. J. (1990). *The science process skills: Research matters to the science teacher*. Reston, VA: National Association for Research in Science Teaching. http://www.narst.org/publications/research/skill.cfm

Rambuda, A. M., & Fraser, W. J. (2004). Perceptions of teachers of the application of science process skills in teaching geography in secondary schools in the free state province in South African. *Journal of Education, 24*, 10-17.

Rasch, G. (1980). *Probabilistic models for some intelligence and attainment tests*. Chicago: University of Chicago Press. (Original work published 1960).

Sass D. A., Schmitt T. A., & Walker C. M. (2008). Estimating non-normal latent trait distributions with item response theory using true and estimated item parameters. *Applied Measurement in Education, 21*, 65-88.

Schwarz, G. (1978). *Estimating the Dimension of a Model*. *The Annals of Statistics, 6*(2), 461-464.

Webb, N. (1997). *Criteria for alignment of expectations and assessments on mathematics and science education.* Washington, DC: CCSSO

Wilson, M. (2005). Constructing measures: An item response modeling approach. Mahwah, NJ: Routledge.

Wilson, M., & Draney, K. (2002). *A technique for setting standards and maintaining them over the time.* In S. Nishisato, Y. Baba, H. Bozdogan, & K. Kanefugi (Eds.), Measurement and multivariate analysis (pp. 325-332). New York, NY: SpringVerlag

Wu, M. L., Adams, R. J., Wilson, M. R., & Haldane, S. A. (2007). *ACER ConQuest: Generalised Item Response Modelling Software* [Computer software]. Version 2. Camberwell, Victoria: Australian Council for Educational Research.

Wright, B. D., & Stone, M. H. (1979). Best test design. Chicago, IL: Mesa Press.