

การออกแบบวิธีการตรวจให้คะแนนอัตโนมัติเพื่อวินิจฉัยมโนทัศน์ ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์ผ่านการเรียนรู้ของเครื่อง

Designing Automated Scoring Approach to Diagnose Scientific Competency Misconceptions through Machine Learning

ประไพพิมพ์ ทองเปราะ^{1*} พัทธี จันทรเพ็ง² ชัยวัฒน์ ทะวะรุ่งเรือง³ และ ธนพงศ์ อินทะ⁴

Prapaipim Thongprowh^{1*}, Putcharee Junpeng², Chaiwat Tawarungruang³ and Thanapong Intharah⁴

(Received: March 31, 2022; Revised: May 23, 2022; Accepted: May 27, 2022)

บทคัดย่อ

การออกแบบวิธีการตรวจให้คะแนนอัตโนมัติเพื่อวินิจฉัยมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์ผ่านการเรียนรู้ของเครื่อง (machine learning) เป็นการออกแบบวิธีการตรวจให้คะแนนแบบสอบวินิจฉัยรูปแบบผสมเพื่อประเมินระดับสมรรถนะทางวิทยาศาสตร์โดยใช้เครื่องตรวจคะแนนในขั้นตอนเดียวนวัตกรรมนี้สามารถรายงานผลและให้ข้อมูลป้อนกลับแก่ผู้เรียนแบบเรียลไทม์ ดังนั้นการวิจัยครั้งนี้ได้กำหนดความมุ่งหมายเพื่อนำไปสู่การออกแบบนวัตกรรมต้นแบบดังนี้ (1) เพื่อวิเคราะห์ผลการตอบของผู้เรียนสู่การออกแบบนวัตกรรมต้นแบบ ประกอบด้วยรูปแบบและระดับของมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์ การกำหนดจุดตัดจากผลการตอบแบบวินิจฉัยสมรรถนะวิทยาศาสตร์ และการกำหนดตัวแปรเพื่อสร้างสมการทำนายจากข้อมูลพื้นฐานของนักเรียน และ (2) เพื่อออกแบบวิธีการตรวจให้คะแนนอัตโนมัติในการวินิจฉัยมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์ผ่านการเรียนรู้ของเครื่อง การศึกษาครั้งนี้ใช้ข้อมูลพฤติกรรมจากกลุ่มผู้สอบคือนักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 จำนวน 847 คน วิเคราะห์ข้อมูลตามโมเดลพหุมิติสัมพันธ์การสุ่มแบบโลจิก (MRCMLM) และกำหนดจุดตัดบน wright map ผลการศึกษาพบว่า

¹ นักศึกษาปริญญาโท สาขาวิชาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

² รองศาสตราจารย์ สาขาวิชาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

³ นักวิจัย ศูนย์จัดการข้อมูลและวิเคราะห์ทางสถิติ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น

⁴ อาจารย์ ภาควิชาสถิติ คณะวิทยาศาสตร์ มหาวิทยาลัยขอนแก่น

งานวิจัยนี้ได้รับทุนอรรถกฤตภาพนักวิจัยรุ่นกลาง ประจำปี 2565 จากสำนักงานการวิจัยแห่งชาติ (วช.) รหัสโครงการวิจัย N42A650236

¹ Master's degree student, Educational Measurement and Evaluation Program, Faculty of Education, Khon Kaen University

² Associate Professor, Department of Measurement and Evaluation Program, Faculty of Education, Khon Kaen University

³ Researcher, Data Management and Statistical Analysis Center, Faculty of Public Health, Khon Kaen University

⁴ Lecturer, Department of Statistics, Faculty of Science, Khon Kaen University

The research received a research grant for Talented Mid-Career Researchers 2022 from the National Research Council of Thailand (NRCT) Project No N42A650236

* Corresponding Author E-mail: prapaipim_th@kkumail.com

(1) เกิดมโนทัศน์ที่คลาดเคลื่อนขึ้น 5 รูปแบบ และการหาจุดตัดทางสมรรถนะวิทยาศาสตร์ทั้ง 3 มิติ พบว่า มิติด้านการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ มีจุดตัดที่ -0.49 0.27 และ 0.83 มิติด้านการประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์ มีจุดตัดที่ -0.09 0.54 และ 1.72 ตามลำดับ และมิติด้านการแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์มีจุดตัดที่ -0.78 0.47 และ 2.98 ตามลำดับ นอกจากนี้พบตัวแปรต้นที่สำคัญ คือ (1) เกรดเฉลี่ยรวมเทอมที่ผ่านมา (2) เกรดเฉลี่ยวิชาวิทยาศาสตร์เทอมที่ผ่านมา และ (3) จำนวนชั่วโมงที่ศึกษาวิชาวิทยาศาสตร์ด้วยตนเอง/สัปดาห์

(2) การออกแบบวิธีการตรวจให้คะแนนอัตโนมัติเพื่อวินิจฉัยสมรรถนะวิทยาศาสตร์ในข้อสอบอัตนัย ใช้วิธีการจับคำ (keywords) และใช้วิธีการแปลงคำตอบจากการสังเคราะห์ผลการตอบในข้อมูลชุดเดียวให้เป็นตัวเลือกเพื่อช่วยให้การตรวจให้คะแนนและประมวลผลระดับสมรรถนะและการรายงานระดับมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์สามารถทำได้แบบอัตโนมัติและเรียลไทม์ สำหรับผลการประเมินคุณภาพการออกแบบวิธีการตรวจให้คะแนนจากผู้เชี่ยวชาญจำนวน 5 คน ครอบคลุมทั้ง 5 มาตรฐาน คือ ด้านอรรถประโยชน์ ด้านความเป็นไปได้ ด้านความเหมาะสมขอรับรอง ด้านความถูกต้อง และด้านความรับผิดชอบการประเมิน

คำสำคัญ: การเรียนรู้ของเครื่อง วิธีการตรวจให้คะแนนอัตโนมัติ มโนทัศน์ที่คลาดเคลื่อน สมรรถนะวิทยาศาสตร์ โมเดลพหุมิติสัมประสิทธิ์การสุ่มแบบโลจิท

Abstract

Designing automated scoring to diagnose misconceptions in scientific competencies through machine learning is a scoring method for a mixed-format diagnostic test to assess scientific competency levels through a scoring machine in a single step. This innovation can report results and provide real-time feedback to learners. Therefore, to design the prototypical innovation, the objectives of the present study were defined as follows: (1) to analyze the students' response patterns to the design of the prototypical innovation, including identifying types and levels of misconceptions in scientific competencies, defining cut scores based on response patterns in the scientific competency diagnostic test, and determining variables to create the predictive equation based on students' background data; and (2) to design automated scoring for diagnosis of misconceptions in scientific competencies through machine learning. This study drew on secondary data of a sample of 847 grade 7 students. The data were analyzed using a multidimensional random coefficients multinomial logit model (MRCMLM) while cut scores on the Wright map were defined. The results are reported below.

(1) There were 5 types of misconceptions in scientific competencies. Besides, on determining cut scores of 3 scientific competencies, it was found that the competency to explain

phenomena scientifically had the cut scores of -0.49, 0.27 and 0.83; the cut scores of the competency to evaluate and design scientific enquiry were -0.09, 0.54 and 1.72 respectively; and those of the competency to interpret data and use evidence scientifically were -0.78, 0.47 and 2.98 respectively. In addition, the important independent variables were discovered as follows: (1) the total GPA of the previous semester; (2) the GPA in the science subject of the previous semester; and (3) the number of hours of self-study on a science subject per week.

(2) The designed scoring to diagnose scientific competencies in subjective tests employed identified keywords and converted answers from the synthesis of response patterns in the secondary data into choices to allow for scoring, processing scientific competency levels, and reporting the levels of misconceptions automatically and in real time. The results also showed that, based on five experts' evaluation, the design of automated scoring was found to cover 5 standards, namely utility, feasibility, propriety, accuracy and accountability.

Keywords: machine learning, automated scoring, misconception, scientific competency, multidimensional random coefficients multinomial logit model

บทนำ

การทดสอบสมรรถนะวิทยาศาสตร์ระดับนานาชาติด้วยโปรแกรมประเมินสมรรถนะนักเรียนตามมาตรฐานสากล (programmed for international student assessment; PISA) เน้นความฉลาดรู้ด้านวิทยาศาสตร์ (scientific literacy) ภายใต้กรอบการประเมินสมรรถนะทางวิทยาศาสตร์ (scientific competencies) ประกอบด้วย (1) การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ (2) การประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์ และ (3) การแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์ (Organization for Economic Co-operation and Development, 2019) ซึ่งผลการประเมินนักเรียนไทยที่ผ่านพบว่ามีคะแนนระดับสมรรถนะต่ำกว่าค่าเฉลี่ยที่กำหนด เมื่อพิจารณาเป็นรายสาระพบว่าสาระที่ 2 มีคะแนนต่ำกว่าค่าเฉลี่ยรวมของวิชาวิทยาศาสตร์ โดยเฉพาะอย่างยิ่งในเนื้อหา เรื่อง สารและสมบัติของสาร สอดคล้องกับการเรียนวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 1 ที่ครูผู้สอนพบว่าในสาระดังกล่าวนักเรียนมักเกิดมโนทัศน์ที่คลาดเคลื่อนต่อการเรียนรู้ที่ส่งผลต่อระดับความสามารถทางวิทยาศาสตร์ของนักเรียนโดยตรง โดยการวัดระดับความรู้ในโรงเรียนและ PISA ต่างมุ่งเน้นการประเมินด้านวิทยาศาสตร์ของนักเรียน สำหรับแนวทางที่จะช่วยลดการเกิดมโนทัศน์ที่คลาดเคลื่อนคือ “การวินิจฉัย” เพื่อหาความคลาดเคลื่อนจากการเรียนรู้ของนักเรียน (Khawsuk & Srihaset, 2016)

การวินิจฉัย (diagnosis) เป็นเครื่องมือประเมินผลการเรียนรู้ที่ครูใช้วัดและประเมินผลนักเรียน เพราะการวินิจฉัยสามารถชี้ให้เห็นข้อบกพร่องของนักเรียนเป็นรายบุคคล (Insawat, 2016) จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่าแบบทดสอบเพื่อการวินิจฉัยมีหลายรูปแบบ เช่น ปรนัย อัตนัย และผสม

โดยเฉพาะแบบผสมนั้นสามารถลดข้อจำกัดของปรนัยและอัตนัย และมีจุดเด่นคือสามารถวัดทักษะของผู้เรียนทุกด้านในฉบับเดียว (Boonjun, 2019) ปัจจุบันทฤษฎีการตอบสนองข้อสอบแบบพหุมิติเป็นทฤษฎีที่ใช้กันอย่างกว้างขวางเพื่อสร้างแบบวินิจัย โดยมีโมเดลที่มุ่งไปที่ปฏิสัมพันธ์ระหว่างผู้สอบและข้อสอบและพิจารณาจากผลการตอบของผู้สอบเพื่อประมาณค่าพารามิเตอร์ของผู้สอบในหลายมิติ (Junpeng, 2018) ซึ่งคุณลักษณะของข้อสอบพหุมิติมีอยู่ 2 ลักษณะ คือ ความเป็นพหุมิติระหว่างข้อคำถามและความเป็นพหุมิติภายในข้อคำถาม (Chianchana, 2009; Rawanprakhon & Tayraukham, 2017) สำหรับโมเดลพหุมิติสัมพันธ์ การสุ่มแบบโลจิต (MRCMLM) เป็นหนึ่งในโมเดลของทฤษฎีนี้ที่นำมาประยุกต์ใช้ในการวิเคราะห์โครงสร้างข้อสอบเพื่อตรวจสอบคุณภาพของเครื่องมือทั้งคุณภาพรายข้อ และคุณภาพของเครื่องมือทั้งฉบับ ทั้งสามารถตรวจสอบความตรงเชิงโครงสร้างภายในโดยสะท้อนออกมาเป็น wright map ที่เป็นหัวใจสำคัญของการตรวจสอบคุณภาพซึ่งนำไปสู่การตีความของนักเรียนรายบุคคล โดยมีงานวิจัยที่โมเดลนี้สร้างเครื่องมือวัดเพื่อประเมินผู้เรียนในหลายด้าน เช่น การพัฒนาแบบวัดพหุมิติทักษะการแก้ปัญหาแบบร่วมมือสำหรับนักเรียนชั้นมัธยมศึกษาตอนต้น (Wongphummueang & Piyakun, 2021) และการพัฒนาข้อสอบพหุมิติเพื่อวัดสมรรถนะทางวิทยาศาสตร์ของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 (Intasoi, 2020)

งานวิจัยที่ผ่านมาข้อสอบรูปแบบผสมพบในเครื่องมือวัดแบบออนไลน์มากขึ้นแต่การตรวจให้คะแนนในเครื่องมือออนไลน์ยังมีจุดอ่อนคือในข้อสอบอัตนัยต้องรอให้คนตรวจจึงสามารถรายงานผลการประเมินได้ ดังนั้นปัญหาที่เกิดขึ้นของการตรวจให้คะแนนในส่วนข้ออัตนัยคือไม่สามารถป้อนข้อมูลกลับสู่ผู้สอบได้ทันที (Fiothong et al., 2022) สำหรับบริบทของสมรรถนะทางวิทยาศาสตร์พบว่ามุ่งเน้นไปที่การประเมินระดับสมรรถนะทางวิทยาศาสตร์แต่ขาดการวินิจัยมิติที่คลาดเคลื่อนที่เกิดขึ้น เช่น งานวิจัยของ Intasoi et al. (2020); Intasoi (2020) ที่พัฒนาข้อสอบแบบพหุมิติผ่านเครื่องมือดิจิทัลในบทเรียนเรื่อง สารบริสุทธิ์ ครอบคลุมการวัดระดับสมรรถนะทางวิทยาศาสตร์ใน 3 มิติ ภายใต้กรอบของ PISA โดยประยุกต์ใช้โมเดลเชิงโครงสร้างและใช้เทคนิคการวิเคราะห์ข้อมูลตามโมเดลการตอบสนองข้อสอบแบบพหุมิติในการวัดสมรรถนะ งานวิจัยดังกล่าวได้ให้ข้อมูลเพื่อประเมินระดับสมรรถนะของนักเรียนแต่มีข้อจำกัดคือ แยกขั้นตอนเป็น 3 ส่วน คือ การทำแบบทดสอบของผู้เรียนผ่าน google form (2) การตรวจให้คะแนน และ (3) การรายงานผลการประเมินกลับแต่ในข้อที่ตรวจให้คะแนนหลายค่ายังต้องรอให้ผู้ประเมินมาตรวจทำให้ไม่สามารถวินิจัยและให้ข้อมูลป้อนกลับได้ทันทีประกอบกับยังขาดการให้สารสนเทศเพื่อวิเคราะห์มิติที่คลาดเคลื่อนทางสมรรถนะทางวิทยาศาสตร์ ดังนั้นจึงต้องหาวิธีการที่สามารถดำเนินการวินิจัยระดับสมรรถนะและรายงานระดับมิติที่คลาดเคลื่อนในขั้นตอนเดียว

ปัจจุบันการนำปัญญาประดิษฐ์ (artificial intelligence, AI) มาช่วยพัฒนาการตรวจให้คะแนนนั้นเรียกว่า “ระบบการตรวจให้คะแนนอัตโนมัติ” หรือ “automatic scoring system” (Sudjai & Mungsing, 2021) และหนึ่งในวิธีการตรวจแบบอัตโนมัตินี้คือการนำการเรียนรู้ของเครื่อง (Machine learning, ML) ที่เป็นส่วนย่อยของปัญญาประดิษฐ์เข้ามาช่วยพัฒนาโดยใช้อัลกอริทึมมาคาดการณ์ซึ่งอาศัยหลักการเรียนรู้จากกลไกของตัวเครื่อง ในด้านการศึกษาได้นำการเรียนรู้ของเครื่องสร้างเครื่องมือวัดที่สามารถตรวจให้คะแนนอัตโนมัติและรายงานผลการเรียนรู้สู่ผู้เรียนทันที เช่น การให้คะแนนโดยใช้ระบบตอบกลับอัตโนมัติ (Lee et al., 2021) และการตรวจให้คะแนนข้อสอบพหุมิติในวิชาเคมีและฟิสิกส์ (Maestrales et al., 2021) ซึ่งหลักการ

ตรวจให้คะแนนอัตโนมัติผ่านการเรียนรู้ของเครื่องจะอาศัยหลักการจำแนก (classification) และหลักการทำนาย (prognosis) ที่ต้องอาศัยตัวแปรเพื่อนำมาสร้างสมการทำนายปรากฏการณ์ที่จะเกิดขึ้น (Lintean et al., 2012) ทำให้ผลการวิเคราะห์โดยการเรียนรู้ของเครื่องถูกต้องและไม่ต่างจากการใช้คนตรวจ มีความแม่นยำสามารถลดขั้นตอนระยะเวลา และให้ผลการวินิจฉัยได้แบบเรียลไทม์

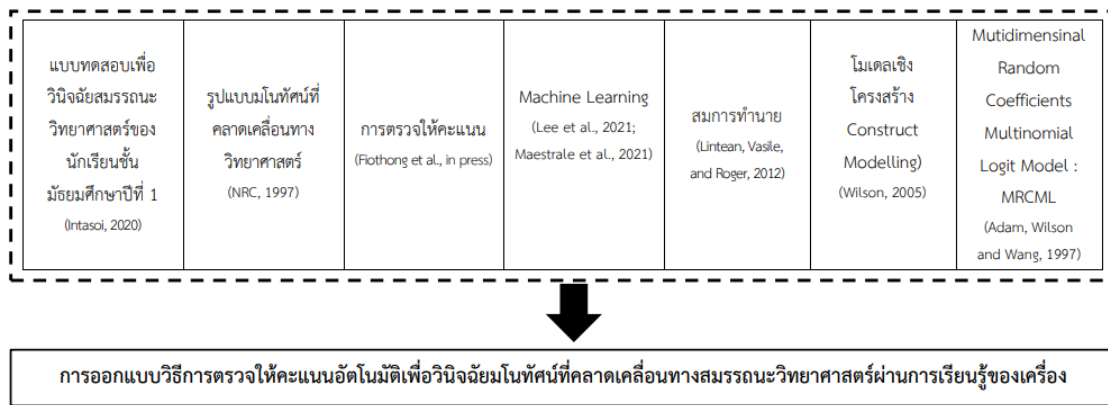
ดังนั้นเพื่อต่อยอดงานวิจัยของ Intasoi et al. (2020); Intasoi (2020) ที่มีจุดอ่อนในเรื่องการตรวจให้คะแนนเนื่องจากไม่สามารถตรวจอัตโนมัติแบบเรียลไทม์และเพื่อหาแนวทางสำคัญในการพัฒนาแพลตฟอร์มการตรวจให้คะแนนอัตโนมัติในขั้นตอนเดียวที่ถูกต้องและที่มีความคงเส้นคงวาในการตรวจให้คะแนนจากการศึกษาที่ผ่านมาพบว่าหากนำทฤษฎีของ MRCMLM และหลักการทำงานของการทำงานของเครื่องมาสร้างเครื่องมือวัด โดยใช้ทฤษฎี MRCMLM เป็นฐานในการวิเคราะห์ชุดข้อมูลขนาดใหญ่นำไปสู่การวินิจฉัยระดับความสามารถของนักเรียน โดยอาศัยวิธีการจำแนกกับหลักการทำนายของการเรียนรู้ของเครื่องในการประมวลผลเพื่อการตรวจให้คะแนนอัตโนมัติ จากที่กล่าวมาหากนำมาประยุกต์เพื่อสร้างระบบการตรวจให้คะแนนผ่านการเรียนรู้ของเครื่องจะทำให้การตรวจให้คะแนนสามารถทำได้ไม่ต่างจากการใช้คนตรวจและยังรายงานผลการวินิจฉัยได้แบบเรียลไทม์ ด้วยเหตุนี้คณะผู้วิจัยจึงสนใจออกแบบวิธีการตรวจให้คะแนนอัตโนมัติเพื่อการวินิจฉัยมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์ผ่านการเรียนรู้ของเครื่อง โดยนำข้อมูลทฤษฎีจากงานวิจัยของ Intasoi (2020) มาวิเคราะห์ซึ่งประกอบด้วย รูปแบบและระดับของมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์ การกำหนดจุดตัดจากผลการตอบของนักเรียนเพื่อใช้เป็นเกณฑ์ในการวินิจฉัยระดับความสามารถของนักเรียน การวิเคราะห์ข้อมูลพื้นฐานของนักเรียนเพื่อกำหนดตัวแปรต้นและคะแนนในแต่ละมิติเป็นตัวแปรตามในการสร้างสมการทำนาย รวมถึงการออกแบบวิธีการตรวจให้คะแนนอัตโนมัติผ่านการเรียนรู้ของเครื่อง ซึ่งงานวิจัยนี้จะเป็นประโยชน์อย่างยิ่งต่อการตรวจให้คะแนนเพื่อรายงานผลการวินิจฉัยระดับการเรียนรู้แก่นักเรียนเป็นรายบุคคล เพราะช่วยลดระยะเวลาในการตรวจและลดทรัพยากรบุคคลในการทำงานเนื่องจากมีระบบตรวจให้คะแนนที่ตรงและแม่นยำเทียบเท่าบุคคลตรวจ ทั้งยังเพิ่มความสะดวกในการใช้งานเพราะเป็นระบบออนไลน์ที่ครูและนักเรียนสามารถเข้าถึงได้ง่าย

ความมุ่งหมายของการวิจัย

1. เพื่อวิเคราะห์ผลการตอบของนักเรียนสู่การออกแบบนวัตกรรมต้นแบบ ประกอบด้วย รูปแบบและระดับของมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์ การกำหนดจุดตัดจากผลการตอบแบบวินิจฉัยสมรรถนะวิทยาศาสตร์ และการกำหนดตัวแปรเพื่อสร้างสมการทำนายจากข้อมูลพื้นฐานของนักเรียน
2. เพื่อออกแบบวิธีการตรวจให้คะแนนอัตโนมัติในการวินิจฉัยมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์ผ่านการเรียนรู้ของเครื่อง

กรอบแนวคิดในการวิจัย

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง ซึ่งสามารถกำหนดกรอบแนวคิดการวิจัยได้ดังภาพ 1



ภาพ 1 กรอบแนวคิดการวิจัย

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยการออกแบบ (Design Research: DR) (Reeves, 2006; Wongwanit, 2020) มีขั้นตอนในการดำเนินการ 4 ระยะ คือระยะที่ 1 วิเคราะห์ข้อมูลชุดข้อมูล ระยะที่ 2 ออกแบบวิธีการตรวจให้คะแนนผ่านการเรียนรู้ของเครื่อง ระยะที่ 3 นำนวัตกรรมใหม่ไปทดลองใช้ และระยะที่ 4 นำเสนอหลักการออกแบบระบบการตรวจให้คะแนนอัตโนมัติผ่านการเรียนรู้ของเครื่อง ซึ่งบทความวิจัยนี้เป็นการศึกษานำร่องของ 2 ระยะแรก โดยมีรายละเอียดทั้ง 2 ระยะ ดังนี้

ระยะที่ 1 วิเคราะห์ข้อมูลชุดข้อมูลจากการตอบแบบวินิจฉัยสมรรถนะวิทยาศาสตร์ของนักเรียนชั้นมัธยมศึกษาปีที่ 1

มุ่งเน้นการวิเคราะห์ข้อมูลชุดข้อมูลจากการตอบแบบวินิจฉัยของนักเรียน โดยมีรายละเอียดดังนี้

1. ประชากรและตัวอย่างวิจัย

ประชากรและตัวอย่างวิจัยของงานวิจัยครั้งนี้มาจากข้อมูลชุดข้อมูลจากงานวิจัยเรื่องการพัฒนาเครื่องมือดิจิทัลเพื่อประเมินสมรรถนะทางวิทยาศาสตร์แบบพหุมิติของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 ปีพุทธศักราช 2562 (Intasoi et al., 2020; Intasoi, 2020) ประชากรคือนักเรียนชั้นมัธยมศึกษาปีที่ 1 ปีการศึกษา 2562 จำนวน 5,685 คน ใช้วิธีกำหนดตัวอย่างวิจัยขั้นต่ำของ Wright & Stone (1979) เพื่อกระจายความสามารถของตัวอย่าง โดยการสุ่มแบบหลายขั้นตอนจำนวน 847 คน

2. เครื่องมือที่ใช้ในการวิจัย

1) แบบทดสอบวินิจฉัยพหุมิติแบบผสมเพื่อวัดสมรรถนะวิทยาศาสตร์ที่พัฒนาโดย Intasoi (2020) จำนวน 22 ข้อ โดยมีมิติการวัด 3 มิติ ภายใต้กรอบของ PISA ประกอบด้วย มิติการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ จำนวน 10 ข้อ มิติการประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์ จำนวน 4 ข้อ และมิติการแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์ จำนวน 8 ข้อ มีค่าดัชนีความตรงเชิงเนื้อหารายข้อ (I-CVI) อยู่ระหว่าง 0.50 – 1.00 และค่าดัชนีความตรงเชิงเนื้อหาทั้งฉบับ (S-CVI) เท่ากับ 0.90 ค่าสัมประสิทธิ์ความเที่ยงแบบ EAP/PV ในมิติที่ 1 2 และ 3 มีค่าเท่ากับ 0.71 0.61 และ 0.69 ตามลำดับ ค่าความเหมาะสมรายข้อ และค่า INFIT MNSQ อยู่ระหว่าง 0.82 - 1.28

2) แบบฟอร์มวิเคราะห์หมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์จาก National Research Council (1997) ร่วมกับหลักแนวคิดความเข้าใจทางวิทยาศาสตร์ (Desstya et al., 2019)

3.การเก็บรวบรวมข้อมูลและการวิเคราะห์ข้อมูล

ผู้วิจัยดำเนินการศึกษาข้อมูลที่ใช้ประกอบการสร้างจุดตัดระดับสมรรถนะทางวิทยาศาสตร์ มีรายละเอียดดังขั้นตอนต่อไปนี้

1) ศึกษาข้อมูลทุติยภูมิเพื่อกำหนดรูปแบบและระดับของหมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์ และวิเคราะห์แผนที่โครงสร้างระดับผลลัพธ์การเรียนรู้ ดังภาพ 2

ระดับ	มิติการอธิบายปรากฏการณ์ ในเชิงวิทยาศาสตร์	มิติการประเมินและออกแบบกระบวนการสืบเสาะหา ความรู้ทางวิทยาศาสตร์	มิติการแปลความหมายข้อมูลและการใช้ ประจักษ์พยานในเชิงวิทยาศาสตร์
4 ↑ สูง (High Level)	- สามารถเขียนหรืออธิบายระดับขอบเขตความสัมพันธ์ของแนวคิด และโมเดลทางวิทยาศาสตร์ เพื่อสร้างคำอธิบายของปรากฏการณ์ เหตุการณ์และกระบวนการ - ใช้ความรู้ที่เหนือกว่าหลักสูตรวิทยาศาสตร์และแสดงให้เห็นถึงวิธีการได้มาซึ่งความรู้ที่เหมาะสม	- สามารถประเมินการออกแบบการแข่งขันที่ซับซ้อน ศึกษาภาคสนาม หรือสถานการณ์จำลอง และการปรับตัวเลือก	- สามารถประเมินความแข็งแกร่งจากข้อมูลจากสมมติฐานและปรับข้อสรุปโดยใช้โมเดลทางวิทยาศาสตร์ที่เป็นนามธรรม - สามารถแยกความแตกต่างระหว่างข้อมูลที่เกี่ยวข้องและไม่เกี่ยวข้อง และสร้างคำอธิบายได้
3 ปานกลาง (Intermediate Level)	- จัดจำ หรือใช้แนวคิดทางวิทยาศาสตร์ที่ได้รับเพื่อสร้างคำอธิบายของเหตุการณ์และกระบวนการที่ค่อนข้างซับซ้อน - ใช้แนวคิดหรือโมเดลทางวิทยาศาสตร์ที่เป็นนามธรรมเพื่ออธิบาย	- ทำการทดลองเกี่ยวกับตัวแปรอิสระสองตัวหรือมากกว่า โดยแสดงขั้นตอน และสามารถตีความข้อมูล และตั้งข้อสรุปที่เหมาะสม - ใช้ข้อมูลจากบริบทที่มีความคุ้นเคยน้อย เพื่อระบุแนวโน้มและทำนาย - ประเมินการออกแบบหรือการตีความข้อมูล และการปรับตัวเลือก ระบุข้อจำกัดของการตีความของชุดข้อมูล	- สามารถตีความ และจัดการกับชุดข้อมูลที่มีความซับซ้อนในระดับปานกลางเพื่อเลือกหรือแสดงข้อสรุปที่เหมาะสม - สามารถแยกความแตกต่างระหว่างปัญหาทางวิทยาศาสตร์และสังคมหรือส่วนบุคคล - สามารถตีความชุดข้อมูลที่มีความซับซ้อนในระดับปานกลางโดยใช้โมเดลทางวิทยาศาสตร์
2 พื้นฐาน (Basic Level)	- จัดจำ และใช้ข้อเท็จจริงและแนวคิดทางวิทยาศาสตร์อย่างง่าย หรือเลือกคำอธิบายวิทยาศาสตร์อย่างง่าย - เขียนข้อเท็จจริงและแนวคิดทางวิทยาศาสตร์ ในการระบุหรือสร้างคำอธิบายง่าย ๆ ที่คุ้นเคยอย่างเหมาะสม	- เขียนขั้นตอน ในการออกแบบการทดลองอย่างง่าย - รวบรวมและตีความข้อมูล แยกความแตกต่างระหว่างคำถามที่ไม่ใช่วิทยาศาสตร์และวิทยาศาสตร์ - เขียนขั้นตอนในการออกแบบ และปรับมุมมองของการออกแบบการทดลองสามารถแยกแยะระหว่างประเด็นทางวิทยาศาสตร์ และไม่ใช่วิทยาศาสตร์	- สามารถระบุ ข้อมูลที่สนับสนุนคำตอบเพื่อเลือกคำอธิบายที่เกี่ยวข้อง - สามารถตีความและแปลงข้อมูลเพื่อสนับสนุนคำตอบหรือข้อสรุปทางวิทยาศาสตร์
1 ↓ ต่ำ (Low Level)	- รับรู้ศัพท์ทางวิทยาศาสตร์ และใช้ข้อเท็จจริงที่ใกล้เคียงกับประสบการณ์ส่วนตัว - เลือกตัวอย่างที่เหมาะสมของโมเดลทางวิทยาศาสตร์อย่างง่ายหรือระบุคำอธิบายทางวิทยาศาสตร์ที่เหมาะสมซึ่งสอดคล้องกับข้อมูลที่ได้รับ	- สามารถทำตามคำแนะนำขั้นตอนทางวิทยาศาสตร์ - ทำการทดลองง่าย ๆ ที่เกี่ยวข้องกับตัวแปรต้นหนึ่งตัว และตัวแปรตามหนึ่งตัว	- ระบุนข้อมูลในรูปแบบอย่างง่าย - สามารถทำการตีความชุดข้อมูลอย่างง่าย

ภาพ 2 แผนที่โครงสร้างเพื่อประเมินสมรรถนะทางวิทยาศาสตร์
ที่มา Intasoi (2020)

2) วิเคราะห์โครงสร้างของข้อสอบ รูปแบบแนวทางการให้คะแนนในแต่ละข้อตามระดับแผนที่โครงสร้าง ข้อสอบมีลักษณะเป็นพหุมิติระหว่างข้อคำถาม (between items multidimensionality) ตามโมเดลการตอบสนองข้อสอบแบบพหุมิติ (MRCMLM) แบ่งเป็น 3 รูปแบบ ได้แก่ (1) แบบเชิงซ้อน จำนวน 8 ข้อ (2) แบบเลือกตอบ จำนวน 6 ข้อ และ (3) แบบเขียนตอบ จำนวน 8 ข้อ

3) กำหนดจุดตัดด้วยการกำหนดเกณฑ์พื้นที่บน wright map โปรแกรม acer conquest version 2.0 (Wu et al., 2007) และประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบด้วยวิธี MLE พร้อมนำเสนอเป็น wright map ดังภาพ 3 การกำหนดเกณฑ์พื้นที่บน wright map พิจารณาจากค่าเฉลี่ยของความยากหรือระดับ threshold เดียวกันหารด้วยจำนวนข้อสอบและแยกตามมิติที่ศึกษา ผลการคำนวณจุดตัดปรากฏดังตาราง 1

ตาราง 1 ค่าความยากและค่า threshold ของข้อคำถาม ในแต่ละมิติเพื่อนำไปสู่การกำหนดจุดตัด

มิติด้านการอธิบายปรากฏการณ์ ในเชิงวิทยาศาสตร์					มิติด้านการประเมินและออกแบบกระบวนการ สืบเสาะหาความรู้ทางวิทยาศาสตร์					มิติด้านการแปลความหมายข้อมูลและการใช้ ประจักษ์พยานในเชิงวิทยาศาสตร์				
ข้อ	ค่าความยาก (b)	Threshold			ข้อ	ค่าความ ยาก (b)	Threshold			ข้อ	ค่าความยาก (b)	Threshold		
		1	2	3			1	2	3			1	2	3
1	0.932			0.93	5	-0.094	-0.09			8	1.530		0.08	2.98
2	0.053		0.05		7	0.099		0.10		10	-0.605	-0.60		
3	-0.722	-0.72			9	0.864	-0.08	0.98	1.72	11	-1.014	-1.02		
4	-0.914	-0.91			16	0.531		0.53		12	-1.252	-1.25		
6	0.096	-0.22	0.41							13	-0.333	-0.34		
15	-0.114	-0.12								14	-0.713	-0.71		
17	1.017			1.02						20	-0.114		0.77	
18	0.748			0.75						21	0.549		0.55	
19	0.623			0.63										
22	0.348		0.35											
ค่าเฉลี่ย		-0.49	0.27	0.83			-0.09	0.54	1.72			-0.78	0.47	2.98

5) นำคะแนนจุดตัดที่กำหนดในพื้นที่บน wright map และเปลี่ยนให้เป็นคะแนนสเกล (θ_d) เพื่อใช้ในการนำไปใช้ โดยใช้สูตร คะแนนสเกล = $50 + 10(\theta_d)$ เมื่อ θ_d แทนค่าความสามารถของผู้สอบซึ่งประมาณได้จากโมเดล MRCML ในมิติที่ทำการวินิจฉัย กำหนดให้มีค่าเฉลี่ยเท่ากับ 50 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 ซึ่งคะแนนสเกลจะใช้ในการตีความเพื่อนำไปสู่ผลการวินิจฉัยได้อย่างมีความหมาย (Junpeng, 2018)

6) วิเคราะห์ข้อคำถามและตีความระดับมโนทัศน์ในแต่ละมิติและกำหนดตัวแปรสำหรับสร้างสมการทำนายจากข้อมูลพื้นฐานของนักเรียน

ระยะที่ 2 พัฒนารูปแบบการตรวจให้คะแนนแบบอัตโนมัติผ่านการเรียนรู้ของเครื่อง

ในระยะนี้เป็นการออกแบบเพื่อพัฒนานวัตกรรมการตรวจให้คะแนนอัตโนมัติ ระหว่างผู้วิจัยและทีมเทคโนโลยี โดยใช้ผลการวิเคราะห์ในระยะที่ 1 มาเป็นข้อมูลพื้นฐานในการออกแบบวิธีการตรวจให้คะแนนอัตโนมัติดังนี้

1. ผู้ให้ข้อมูล

ผู้เชี่ยวชาญด้านการวัดและประเมินผลการศึกษา ผู้เชี่ยวชาญในด้านการใช้เทคโนโลยี วิศวกรคอมพิวเตอร์และครูวิทยาศาสตร์ชั้นมัธยมศึกษาปีที่ 1 จำนวน 5 คน เพื่อทำการสนทนากลุ่ม (focus group)

2. เครื่องมือ

แบบประเมินแบบสำรวจรายการตามมาตรฐานการประเมินคุณภาพของนวัตกรรมต้นแบบตาม Joint Committee on Standards for Educational Evaluation (Yarbrough et al., 2010; Varasunun, 2011)

3. การเก็บรวบรวมข้อมูลและการวิเคราะห์ข้อมูล

นำผลการกำหนดจุดตัด ระดับมโนทัศน์ที่คลาดเคลื่อน และตัวแปรที่ใช้ในการสร้างสมการทำนาย ใช้ออกแบบนวัตกรรมต้นแบบที่ใช้ในการตรวจให้คะแนนอัตโนมัติผ่านการเรียนรู้ของเครื่อง พิจารณาคุณภาพของนวัตกรรมต้นแบบที่พัฒนาขึ้นก่อนนำไปทดลองใช้โดยผู้เชี่ยวชาญที่เกี่ยวข้อง และร่วมกันสะท้อนความคิดและปรับปรุงก่อนนำไปทดลองใช้ในลำดับต่อไป

ผลการวิจัย

1. ผลการวิเคราะห์ผลการตอบของนักเรียนสู่การออกแบบนวัตกรรมต้นแบบ ประกอบด้วยรูปแบบและระดับของมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์ การกำหนดจุดตัดจากผลการตอบแบบวินิจฉัยสมรรถนะวิทยาศาสตร์ และการกำหนดตัวแปรเพื่อสร้างสมการทำนายจากข้อมูลพื้นฐานของนักเรียน

1.1 ผลการวิเคราะห์รูปแบบและระดับมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์

จากการวิเคราะห์ข้อมูลหัตถ์ภูมิสามารถจำแนกรูปแบบของมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์ ได้ 5 รูปแบบตามลักษณะข้อคำถามจากแบบทดสอบจำนวน 22 ข้อ โดยผลการวิเคราะห์รูปแบบมโนทัศน์ที่คลาดเคลื่อน แสดงในตาราง 2

ตาราง 2 มโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์

รูปแบบมโนทัศน์ที่คลาดเคลื่อน	ลักษณะของข้อคำถาม	เกณฑ์การตอบ	ข้อคำถาม
อุปาทาน (preconceived notion)	ข้อคำถามที่รูปแบบของคำตอบต้องอาศัยความคิดเห็นส่วนบุคคลควบคู่กับพื้นฐานความรู้วิทยาศาสตร์	คำตอบไม่ถูกต้องเพราะเกิดจากการนึกคิดและใช้ความเห็นส่วนตัวโดยไม่มีหลักฐานเพียงพอ	1,2,3,4,5,6,7,8,9,10,11,12,13,14, 15,16,17,18, 19,20 และ 21
ความเชื่อที่ไม่เป็นวิทยาศาสตร์ (nonscientific belief)	ข้อคำถามที่รูปแบบของคำตอบต้องสอดคล้องกับหลักการและทฤษฎีทางวิทยาศาสตร์ และอาศัยแหล่งข้อมูลมาพิจารณาก่อนตอบคำถาม โดยแหล่งข้อมูลนั้นคือ ตาราง แผนภูมิ รูปภาพ และผลการทดลอง	คำตอบไม่ถูกต้องตามกฎและหลักการที่เป็นบรรทัดฐานทางวิทยาศาสตร์	ทุกข้อคำถาม
ความเข้าใจผิดเกี่ยวกับมโนทัศน์ (conceptual misunderstanding)	ข้อคำถามที่รูปแบบของคำตอบต้องใช้ความรู้ที่ถูกต้องเท่านั้น ประกอบด้วยเนื้อหา ดังนี้ (1) สมบัติของธาตุ (2) โลหะ กึ่งโลหะ อโลหะ (3) ความดันบรรยากาศ (4) อุณหภูมิของสาร (5) ความหนาแน่น (6) สารเนื้อเดียว (7) โครงสร้างอะตอมของธาตุ (8) อุณหภูมิกับการเปลี่ยนสถานะของสาร และ (9) ความรู้ทั่วไป	1.คำตอบไม่ถูกต้องตามหลักและเนื้อหาที่ใช้ตั้งข้อคำถาม 2.อธิบายคำตอบได้แต่การอธิบายนั้นแสดงความคลาดเคลื่อนจากการเรียนรู้ที่ผ่านมา 3.ลักษณะของคำตอบกำกวมและนักเรียนไม่มั่นใจในคำตอบ	ทุกข้อคำถาม
มโนทัศน์ที่คลาดเคลื่อนเนื่องจากภาษา (vernacular misconception)	ข้อคำถามที่ต้องใช้การแปลความหมายจาก (1) สูตรเคมี (2) ตัวเลข (3) การคำนวณ (4) อุปกรณ์วิทยาศาสตร์ (5) สัญลักษณ์นิวเคลียร์ (6) แบบจำลองอนุภาคของสาร และ (7) บทความ	1.คำตอบแสดงความคลาดเคลื่อนจากการแปลความหมาย หรือตีความ จากลักษณะของข้อคำถาม	1,2,3,4,5,6,7,8,9,10,11,12,13,14, 15,16,17,18, 19,20 และ 21
มโนทัศน์ที่คลาดเคลื่อนเกี่ยวกับข้อเท็จจริง (factual misconception)	ข้อคำถามที่ต้องวิเคราะห์ข้อเท็จจริงจากบทเรียนเพื่อให้ได้คำตอบที่ถูกต้อง โดยพิจารณาจากข้อมูลตามรูปภาพ ตารางและสถานการณ์ที่กำหนด	1.อธิบายไร้เหตุผล 2.อธิบายคำตอบโดยแสดงความเข้าใจที่ผิดจากการเรียนรู้ที่ผ่านมา	1,2,3,4,6,8,9,11,12,13,14,15,16,17,18 และ 19

จากตาราง 2 พบรูปแบบความเชื่อที่ไม่เป็นวิทยาศาสตร์ (nonscientific belief) และความเข้าใจผิดเกี่ยวกับมโนทัศน์ (conceptual misunderstanding) เกิดในทุกข้อคำถาม โดยลักษณะของข้อคำถามนั้นมักเป็นข้อคำถามที่ต้องหาคำตอบโดยสามารถอธิบายจากหลักการและทฤษฎีทางวิทยาศาสตร์ได้ โดยการใช้ความรู้ที่ถูกต้องเท่านั้นในการตอบคำถาม และจากการวิเคราะห์รูปแบบของมโนทัศน์ที่คลาดเคลื่อนที่เกิดขึ้นสามารถจัดระดับมโนทัศน์ที่คลาดเคลื่อนจากการวิเคราะห์ผลการตอบของนักเรียนได้ โดยอิงเกณฑ์การจัดระดับตามคำอธิบายของแผนที่โครงสร้างทั้ง 4 ระดับ คลอบคลุมสมรรถนะทางวิทยาศาสตร์ใน 3 มิติ (Intasoi, 2020) ระดับมโนทัศน์ที่คลาดเคลื่อนแสดงในตาราง 3

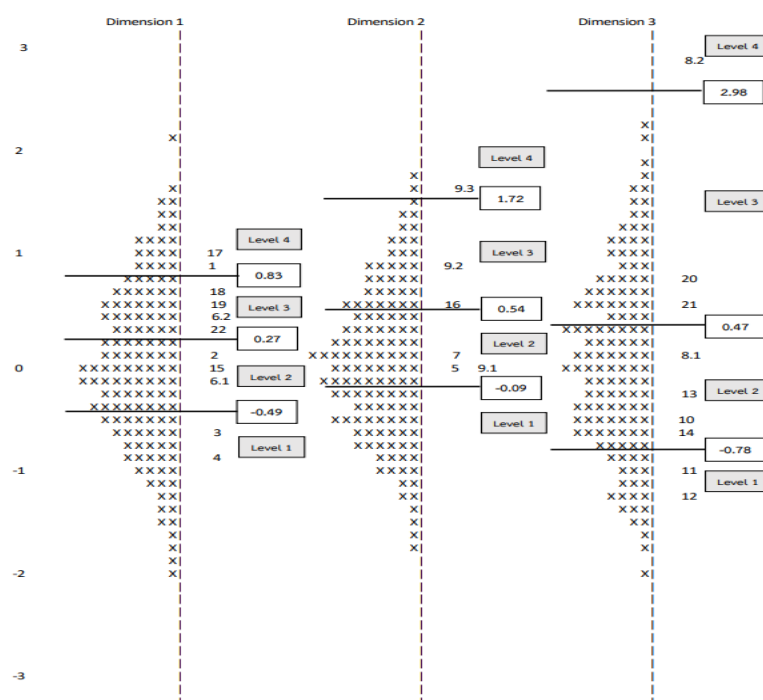
ตาราง 3 ระดับมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์

สมรรถนะทาง วิทยาศาสตร์	มโนทัศน์ที่คลาดเคลื่อน		
	มิติที่ 1	มิติที่ 2	มิติที่ 3
ระดับสูง (High Level)	-	-	-
ระดับปานกลาง (Intermediate Level)	- ความเชื่อที่ไม่เป็นวิทยาศาสตร์ - ความเข้าใจผิดเกี่ยวกับมโนทัศน์	- ความเข้าใจผิดเกี่ยวกับมโนทัศน์ - มโนทัศน์ที่คลาดเคลื่อนเนื่องจากภาษา	- อุปาทาน - ความเข้าใจผิดเกี่ยวกับมโนทัศน์ - มโนทัศน์ที่คลาดเคลื่อนเนื่องจากภาษา
ระดับพื้นฐาน (Basic Level)	- อุปาทาน - ความเชื่อที่ไม่เป็นวิทยาศาสตร์ - มโนทัศน์ที่คลาดเคลื่อนเกี่ยวกับข้อเท็จจริง	- อุปาทาน - ความเข้าใจผิดเกี่ยวกับมโนทัศน์ - มโนทัศน์ที่คลาดเคลื่อนเนื่องจากภาษา - มโนทัศน์ที่คลาดเคลื่อนเกี่ยวกับข้อเท็จจริง	- ความเชื่อที่ไม่เป็นวิทยาศาสตร์ - ความเข้าใจผิดเกี่ยวกับมโนทัศน์ - มโนทัศน์ที่คลาดเคลื่อนเนื่องจากภาษา - มโนทัศน์ที่คลาดเคลื่อนเกี่ยวกับข้อเท็จจริง
ระดับต่ำ (Low Level)	- อุปาทาน - ความเข้าใจผิดเกี่ยวกับมโนทัศน์ - ความเชื่อที่ไม่เป็นวิทยาศาสตร์ - มโนทัศน์ที่คลาดเคลื่อนเนื่องจากภาษา - มโนทัศน์ที่คลาดเคลื่อนเกี่ยวกับข้อเท็จจริง	- อุปาทาน - ความเข้าใจผิดเกี่ยวกับมโนทัศน์ - ความเชื่อที่ไม่เป็นวิทยาศาสตร์ - มโนทัศน์ที่คลาดเคลื่อนเนื่องจากภาษา	- อุปาทาน - ความเข้าใจผิดเกี่ยวกับมโนทัศน์ - ความเชื่อที่ไม่เป็นวิทยาศาสตร์ - มโนทัศน์ที่คลาดเคลื่อนเนื่องจากภาษา

จากตาราง 3 พบว่ารูปแบบมโนทัศน์ที่คลาดเคลื่อนที่พบมากในสมรรถนะระดับปานกลาง คือ ความเข้าใจผิดเกี่ยวกับมโนทัศน์ และระดับพื้นฐาน คือ มโนทัศน์ที่คลาดเคลื่อนเกี่ยวกับข้อเท็จจริง ส่วนรูปแบบมโนทัศน์ที่คลาดเคลื่อนที่ในระดับสมรรถนะต่ำมักพบ 4 รูปแบบ คือ อุปาทาน ความเข้าใจผิดเกี่ยวกับมโนทัศน์ ความเชื่อที่ไม่เป็นวิทยาศาสตร์ และมโนทัศน์ที่คลาดเคลื่อนเนื่องจากภาษา

1.2 ผลการวิเคราะห์ผลการตอบของผู้เรียนเพื่อพัฒนาและกำหนดจุดตัดในการวินิจฉัยมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์

ผลการกำหนดจุดตัดบน wright map ของสมรรถนะทางวิทยาศาสตร์ทั้ง 3 มิติ แบ่งจุดตัดได้ 4 ระดับสมรรถนะ ดังแสดงในภาพ 3



ภาพ 3 จุดตัดบน wright map ของสมรรถนะทางวิทยาศาสตร์ ในมิติการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์

(ซ้าย) มิติการประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์ (กลาง) และมิติการแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์ (ขวา)

จากภาพ 3 พบว่ามีมิติด้านการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์แบ่งได้ 4 ระดับ 3 จุดตัด คือ -0.49 0.27 และ 0.83 ตามลำดับ มิติด้านการประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์แบ่งได้ 4 ระดับ 3 จุดตัด คือ -0.09 0.54 และ 1.72 ตามลำดับ และมิติด้านการแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์แบ่งได้ 4 ระดับ 3 จุดตัด คือ -0.78 0.47 และ 2.98 ตามลำดับ จาก 3 จุดตัดในแต่ละมิติดังกล่าวข้างต้น นำไปสู่การกำหนดช่วงการประเมินระดับสมรรถนะทางวิทยาศาสตร์ในแต่ละมิติ เป็น 4 ช่วง ที่สามารถแปลงเป็นช่วงคะแนนสเกล และช่วงคะแนนดิบ เพื่อนำไปใช้ในการวินิจฉัยผู้เรียนเป็นรายบุคคล ได้ดังตาราง 4 เช่น ผู้สอบที่ประมาณค่าพารามิเตอร์ระดับความสามารถด้านการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ตั้งแต่ 0.83 ($\theta \geq 0.83$) แสดงว่า ผู้สอบจะได้รับการวินิจฉัยระดับสมรรถนะทางวิทยาศาสตร์อยู่ในระดับที่ 4 โดยมีคะแนนสเกลเท่ากับ 58.30 ขึ้นไป และเมื่อนำไปเทียบกับคะแนนดิบจะมีค่าเท่ากับ 8 คะแนน จากคะแนนเต็ม 11 คะแนน เป็นต้น ดังตาราง 4

ตาราง 4 เกณฑ์จุดเปลี่ยนผ่านระดับสมรรถนะทางวิทยาศาสตร์ของนักเรียนระดับชั้นมัธยมศึกษาปีที่ 1

มิติการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์					
ระดับ	ชื่อระดับ	จุดตัด	ช่วงระดับสมรรถนะทางวิทยาศาสตร์ (θ)	คะแนนสเกล	คะแนนดิบ (11 คะแนน)
4	ระดับสูง	$\theta = 0.83$	$\theta = 0.83$ ขึ้นไป	58.30 ขึ้นไป	8-11
3	ระดับปานกลาง	$\theta = 0.27$	$\theta = 0.27$ ถึง 0.82	52.70 ถึง 58.29	6-7
2	ระดับพื้นฐาน	$\theta = -0.49$	$\theta = -0.49$ ถึง 0.26	45.10 ถึง 52.69	4-5
1	ระดับต่ำ		ต่ำกว่า $\theta = -0.49$	ต่ำกว่า 45.10	0-3
มิติการประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์					
ระดับ	ชื่อระดับ	จุดตัด	ช่วงระดับสมรรถนะทางวิทยาศาสตร์ (θ)	คะแนนสเกล	คะแนนดิบ (6 คะแนน)
4	ระดับสูง	$\theta = 1.72$	$\theta = 1.72$ ขึ้นไป	67.20 ขึ้นไป	5-6
3	ระดับปานกลาง	$\theta = 0.54$	$\theta = 0.01$ ถึง 1.71	55.40 ถึง 67.19	3-4
2	ระดับพื้นฐาน	$\theta = -0.09$	$\theta = -0.09$ ถึง 0.00	49.10 ถึง 55.39	2
1	ระดับต่ำ		ต่ำกว่า $\theta = -0.09$	ต่ำกว่า 49.10	0-1
มิติการแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์					
ระดับ	ชื่อระดับ	จุดตัด	ช่วงระดับสมรรถนะทางวิทยาศาสตร์ (θ)	คะแนนสเกล	คะแนนดิบ (9 คะแนน)
4	ระดับสูง	$\theta = 2.98$	$\theta = 2.98$ ขึ้นไป	79.80 ขึ้นไป	+9
3	ระดับปานกลาง	$\theta = 0.47$	$\theta = 0.47$ ถึง 2.97	54.70 ถึง 79.79	6-8
2	ระดับพื้นฐาน	$\theta = -0.78$	$\theta = -0.78$ ถึง 0.46	42.20 ถึง 54.69	4-5
1	ระดับต่ำ		ต่ำกว่า $\theta = -0.78$	ต่ำกว่า 42.20	0-3

2.2 ผลการวิเคราะห์ตัวแปรที่จะใช้ในการสร้างสมการทำนายจากข้อมูลชุดเดียว

การวิเคราะห์ข้อมูลพื้นฐานของนักเรียนเพื่อทำนาย (prognosis) ระดับสมรรถนะทางวิทยาศาสตร์ต้องกำหนดตัวแปรเพื่อสร้างสมการทำนายโดยการพิจารณาถึงตัวแปรที่จะส่งผลต่อระดับสมรรถนะทางวิทยาศาสตร์ ซึ่งได้กำหนดตัวแปรที่เหมาะสมดังตาราง 5

ตาราง 5 ตัวแปรอิสระ ตัวแปรตาม คำนิยามและสัญลักษณ์ของตัวแปร

ชนิดของตัวแปร	ข้อมูลทฤษฎี	คำนิยาม	สัญลักษณ์ตัวแปร
ตัวแปรต้น	เกรดเฉลี่ยรวมเทอมที่ผ่านมา	เกรดเฉลี่ยรวมทุกวิชาของนักเรียนจากตลอดเทอมการศึกษาที่ผ่านมา	X_1
	เกรดเฉลี่ยวิชาวิทยาศาสตร์เทอมที่ผ่านมา	เกรดเฉลี่ยวิชาวิทยาศาสตร์ของนักเรียนจากตลอดเทอมการศึกษาที่ผ่านมา โดยมีระดับเกรดจากที่โรงเรียนกำหนด	X_2
	จำนวนชั่วโมงที่ศึกษาวิชาวิทยาศาสตร์ด้วยตนเอง/สัปดาห์	จำนวนชั่วโมงที่นักเรียนใช้ในการศึกษาวิชาวิทยาศาสตร์ด้วยตนเอง ซึ่งมีเกณฑ์ดังนี้ (1) น้อยกว่า 1 ชั่วโมง (2) 1-2 ชั่วโมง (3) 2-3 ชั่วโมง (4) 3-4 ชั่วโมง (5) 4-5 ชั่วโมง (6) ตั้งแต่ 5 ชั่วโมงขึ้นไป	X_3
ตัวแปรตาม	คะแนนมิติ 1	มิติการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์	Y_1
ตัวแปรตาม	คะแนนมิติ 2	มิติการประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์	Y_2
ตัวแปรตาม	คะแนนมิติ 3	มิติการแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์	Y_3

จากตาราง 5 พบว่าตัวแปรต้นที่จะส่งผลต่อการสร้างสมการทำนาย ได้แก่ เกรดเฉลี่ยรวมเทอมที่ผ่านมา เกรดเฉลี่ยวิชาวิทยาศาสตร์เทอมที่ผ่านมา และจำนวนชั่วโมงที่ศึกษาวิชาวิทยาศาสตร์ด้วยตนเอง/สัปดาห์ ซึ่งการกำหนดตัวแปรขึ้นเพื่อสร้างสมการทำนายนี้จะช่วยสนับสนุนและยืนยันแนวโน้มของผลการประเมินสมรรถนะทางวิทยาศาสตร์ผ่านแบบทดสอบวินิจฉัยได้เป็นอย่างดี

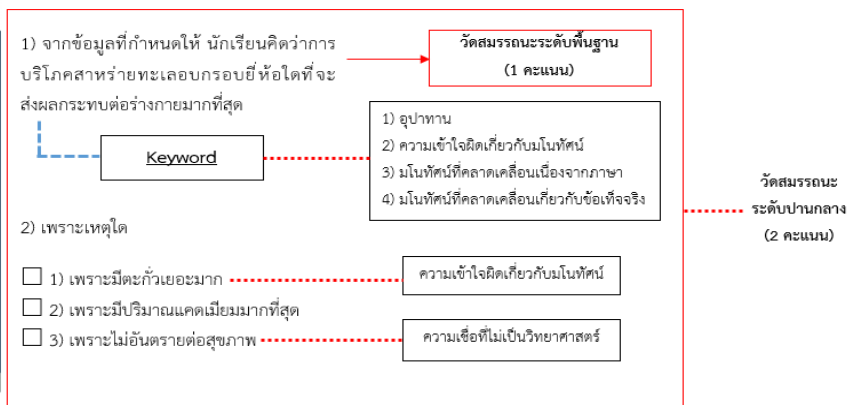
2. ผลการออกแบบวิธีการตรวจให้คะแนนอัตโนมัติในการวินิจฉัยมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์ผ่านการเรียนรู้ของเครื่อง

จากผลการวิเคราะห์รูปแบบ ระดับมโนทัศน์ที่คลาดเคลื่อน การกำหนดจุดตัดและการกำหนดตัวแปรข้อมูลพื้นฐานของนักเรียนเพื่อการสร้างสมการทำนายนั้น นำไปสู่การออกแบบการตรวจให้คะแนนอัตโนมัติผ่านการเรียนรู้ของเครื่องเพื่อตีความระดับสมรรถนะวิทยาศาสตร์ โดยนำเสนอผลการออกแบบดังนี้

2.1 ผลการออกแบบวิธีการตรวจให้คะแนนผ่านการเรียนรู้ของเครื่อง

ผลการออกแบบรูปแบบการตรวจให้คะแนนใช้โมเดลเชิงโครงสร้าง (construct modeling) เป็นแนวทางในการพัฒนาการตรวจให้คะแนน โดยมีส่วนข้อมูลป้อนกลับแก่ผู้เรียนคือระดับมโนทัศน์ที่คลาดเคลื่อน และส่วนรายงานผลการประเมินคือระดับสมรรถนะวิทยาศาสตร์ของนักเรียน ตัวอย่างการตรวจแสดงในภาพ 4

6) ศูนย์ทดสอบฉลาดซื้อ มูลนิธิเพื่อผู้บริโภค (มพบ.) เผยผลทดสอบ สหราชอาณาจักร รอบ ดังภาพ



ภาพ 4 รูปแบบการตรวจให้คะแนนผ่านระบบ

จากภาพ 4 ตัวอย่างข้อสอบเป็นรูปแบบอัตนัยที่วัดสมรรถนะระดับปานกลาง (2 คะแนน) ในการออกแบบการตรวจให้คะแนนได้ออกแบบให้มีการตรวจใน 2 ลักษณะ คือ (1) ใช้วิธีการตรวจจับคำ (keyword) และ (2) ใช้วิธีการแปลงคำตอบจากข้อมูลหัตถ์เขียนให้เป็นตัวเลือก และจากการออกแบบวิธีการตรวจให้คะแนนนี้ นำไปสู่การร่างระบบตรวจให้คะแนน ดังภาพ 5



ภาพ 5 ระบบการตรวจให้คะแนนอัตโนมัติผ่านการเรียนรู้ของเครื่อง

รูปแบบระบบการตรวจให้คะแนนนั้นประกอบด้วย 5 ส่วนสำคัญ คือ (1) ส่วนข้อมูลนำเข้า ประกอบด้วยช่วงคะแนนดิบและคะแนนสเกล คำอธิบายระดับความสามารถกำหนดเป็นเกณฑ์การให้คะแนน (2) ส่วนประมวลผล เน้นไปที่กระบวนการทำงานของอัลกอริทึมในการตรวจให้คะแนนในแบบสอบปรนัยและอัตนัย (3) ส่วนแสดงผล จะรายงานระดับสมรรถนะวิทยาศาสตร์รวมถึงระดับมโนทัศน์ที่คลาดเคลื่อนที่เกิดขึ้นจากการเรียนรู้ (4) ส่วนให้ข้อมูลป้อนกลับแก่ผู้เรียน จะให้ข้อเสนอแนะเพื่อให้พัฒนาตนเองและรายงานระดับมโนทัศน์ที่คลาดเคลื่อน และ (5) ส่วนรายงานผลการประเมิน เป็นนำผลการประเมินมารายงานให้ผู้ใช้งานระบบเข้าใจได้ง่ายมากยิ่งขึ้น

การประเมินคุณภาพเบื้องต้นของผลการออกแบบการตรวจให้คะแนนอัตโนมัติ ด้วยวิธีการสนทนากลุ่ม (focus group) ได้ผลการประเมินเชิงคุณภาพจากผู้เชี่ยวชาญจำนวน 5 คน ซึ่งผู้เชี่ยวชาญได้ให้ข้อเสนอแนะในการออกแบบระบบ โดยครอบคลุม 5 มาตรฐาน คือ ด้านอรรถประโยชน์ ด้านความเป็นไปได้ ด้านความเหมาะสมชอบธรรม ด้านความถูกต้อง และด้านความรับผิดชอบการประเมิน และจากการสนทน พบว่าการออกแบบเพื่อสร้างนวัตกรรมการตรวจให้คะแนนอัตโนมัติเป็นเครื่องมือวัดที่ดีเพราะเป็นการนำเทคโนโลยีสมัยใหม่เข้ามาช่วยในการออกแบบเครื่องมือวัดที่เหมาะสมกับสถานการณ์ในยุคปัจจุบัน เนื่องจากสามารถตรวจให้คะแนนแทนคนตรวจได้และรายงานผลการเรียนรู้ให้แก่ผู้เรียนได้แบบเรียลไทม์

อภิปรายผล

1. ผลวิเคราะห์รูปแบบมโนทัศน์ที่คลาดเคลื่อนทางสมรรถนะวิทยาศาสตร์จากข้อมูลหัตถ์เขียนในเนื้อหาเรื่อง สารบริสุทธิ์ พบรูปแบบมโนทัศน์ที่คลาดเคลื่อนขึ้น 5 รูปแบบ คือ อุปาทาน ความเชื่อที่ไม่เป็นวิทยาศาสตร์ ความเข้าใจผิดเกี่ยวกับมโนทัศน์ มโนทัศน์ที่คลาดเคลื่อนเนื่องจากภาษา และมโนทัศน์ที่คลาดเคลื่อนเกี่ยวกับข้อเท็จจริง ทั้งนี้อาจเนื่องมาจากการประเมินสมรรถนะวิทยาศาสตร์ภายใต้บทเรียนเรื่อง สารบริสุทธิ์ นั้นต้องพิจารณาข้อมูลต่างๆ เช่น รูปภาพ กราฟ และสถานการณ์ร่วมด้วย ประกอบกับการทำแบบทดสอบต้องใช้ความเข้าใจที่แท้จริงเพื่อหาคำตอบเป็นเหตุให้นักเรียนที่ไม่มีพื้นฐานหรือ

เข้าใจไม่ลึกซึ้งเกิดแนวความคิดคลาดเคลื่อน ไม่สามารถสรุปข้อเท็จจริงในเนื้อหาสาระได้ซึ่งสอดคล้องกับสถาบันส่งเสริมการสอนวิทยาศาสตร์และเทคโนโลยีที่ระบุว่านักเรียนมีแนวความคิดคลาดเคลื่อนเนื่องจากไม่มีความรู้เดิมหรือความรู้พื้นฐานทำให้เกิดมโนทัศน์ที่คลาดเคลื่อนในบทเรียนเรื่องสารบริสุทธิ์ (Institute for the promotion of technology science and technology, 2017) และจากผลการวิเคราะห์รูปแบบมโนทัศน์ที่เกิดสอดคล้องกับแนวคิดมโนทัศน์ที่คลาดเคลื่อนทางวิทยาศาสตร์ของสภาวิจัยแห่งชาติ National Research Council (1997) และการศึกษาของ Kamtet (2017)

จากการศึกษาพบว่าเกิดระดับมโนทัศน์ที่คลาดเคลื่อนตามระดับสมรรถนะทางวิทยาศาสตร์แบ่งออกเป็น 4 ระดับ ได้แก่ สูง ปานกลาง พื้นฐาน และต่ำ ซึ่งพบว่าสอดคล้องกับงานวิจัยของ Intasoi (2020) โดยในระดับสมรรถนะปานกลางและพื้นฐาน พบรูปแบบมโนทัศน์ที่คลาดเคลื่อนที่พบมากที่สุดคือความเข้าใจผิดเกี่ยวกับมโนทัศน์ และมโนทัศน์ที่คลาดเคลื่อนเกี่ยวกับข้อเท็จจริง ตามลำดับ ซึ่งสาเหตุดังกล่าวมาจากการที่นักเรียนไม่มีความรู้เบื้องต้นหรือขาดการคิดเชื่อมโยง และการให้เหตุผลที่ไม่สมบูรณ์ รวมถึงขาดความสนใจในการเรียนวิทยาศาสตร์ ซึ่งผลเหล่านี้แสดงให้เห็นว่ามโนทัศน์เป็นอุปสรรคต่อการเรียนรู้และเป็นปัจจัยสำคัญที่ขัดขวางไม่ให้เกิดการเรียนรู้วิทยาศาสตร์อย่างถาวร การรายงานรูปแบบและระดับมโนทัศน์แก่นักเรียนช่วยให้นักเรียนพัฒนาได้ตรงจุดมากขึ้นเพราะการเกิดมโนทัศน์นั้นแสดงว่านักเรียนมีความรู้บางส่วนแต่เกิดการคลาดเคลื่อนระหว่างการเรียนรู้ ดังนั้นนักเรียนต้องได้รับข้อมูลย้อนกลับที่แสดงถึงจุดบกพร่องที่ต้องพัฒนาหลังจากการเรียนรู้ สอดคล้องกับ Anuntasawat (2017) พบว่าการรายงานผลพร้อมให้ข้อมูลย้อนกลับถึงจุดเด่นจุดด้อยทำให้นักเรียนทราบความสามารถของตนเอง ช่วยเสริมแรงบันดาลใจในการเรียน และทราบจุดที่ควรพัฒนาไปสู่เป้าหมายที่ตั้งไว้

2. จากผลการวิจัยพบว่าจุดตัดที่ได้จากการใช้ข้อมูลชุดข้อมูลมีส่งผลต่อการประมาณค่าในการตรวจให้คะแนน เพราะแสดงค่าความสามารถของนักเรียนในแต่ละมิติและกระจายได้ครอบคลุมทุกระดับความสามารถสอดคล้องกับแผนที่โครงสร้างที่ได้กำหนดไว้ เพราะการศึกษาในครั้งนี้มีการวิเคราะห์ผลการตอบที่ได้มาจากฐานข้อมูลขนาดใหญ่ที่มีตัวแทนในทุกระดับความสามารถของนักเรียน ซึ่งการคัดเลือกกลุ่มผู้สอบนั้นใช้เทคนิคการกำหนดตัวอย่างวิจัยขั้นต่ำของ Wright & Stone (1979) เพื่อให้ได้การกระจายความสามารถของกลุ่มตัวอย่างทำให้สารสนเทศมีคุณภาพ สอดคล้องกับ Junpeng et al. (2019); Junpeng et al. (2020) ที่พบว่าถ้าการกระจายความสามารถของผู้สอบไม่เหมาะสมจะส่งผลต่อการกำหนดจุดตัดเพื่อใช้ในการวินิจฉัยระดับความสามารถของผู้เรียน ส่งผลให้การตีความเพื่อวินิจฉัยขาดความถูกต้องแม่นยำในการประมาณค่า และจะส่งผลกระทบต่อผลการพิจารณาระดับของมโนทัศน์ที่คลาดเคลื่อนที่เกิดขึ้นของผู้เรียน

3. กระบวนการสำคัญที่นำไปสู่การตรวจให้คะแนนอัตโนมัติผ่านการเรียนรู้ของเครื่อง คือการวิเคราะห์รูปแบบและระดับมโนทัศน์ การกำหนดจุดตัด และการกำหนดตัวแปรจากการสังเคราะห์ข้อมูลพื้นฐานเพื่อสร้างสมการทำนายที่เหมาะสมเพราะสมการทำนายจะช่วยทำนายผลการวินิจฉัยระดับความสามารถได้อย่างถูกต้องและแม่นยำ ซึ่งงานวิจัยของ Kladchuen & Sanrach (2018) พบว่า การเลือกใช้ตัวแปรเพื่อสร้างสมการที่เหมาะสมในการทำนายผลสัมฤทธิ์ทางการเรียนของนักศึกษาระดับอาชีวศึกษาช่วยให้การทำนายผลสัมฤทธิ์ทางการเรียนแม่นยำ นอกจากนี้ในข้อสอบอัตนัยได้ออกแบบให้ การตอบอยู่ในลักษณะ

ของการเติมคำ หรือวลีสั้นๆ เพื่อให้ระบบการตรวจที่จะพัฒนาขึ้นตรวจในลักษณะ การจับคำ (keywords) เพราะระบบจะตรวจจับคำได้ง่ายสอดคล้องกับงานวิจัยที่ผ่านมาได้นำการเรียนรู้ของเครื่องมาตรวจจับคำและวิธีการตอบของนักเรียนในวิชาเคมีและฟิสิกส์เพื่อประเมินระดับความสามารถของนักเรียน (Maestrales et al., 2021)

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

1. ครูผู้สอนสามารถนำรูปแบบโมเดลที่คลาดเคลื่อนเกิดขึ้นทั้ง 5 รูปแบบ ไปวางแผนการจัดการเรียนรู้ที่เน้นการส่งเสริมสมรรถนะทางวิทยาศาสตร์อันครอบคลุมทั้ง 3 ด้าน ได้แก่ ด้านการอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ การประเมินและออกแบบกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์ และการแปลความหมายข้อมูลและการใช้ประจักษ์พยานในเชิงวิทยาศาสตร์ เพื่อลดและแก้ไขปัญหาการเกิดโมเดลที่คลาดเคลื่อนในสมรรถนะวิทยาศาสตร์ได้ตรงจุด

2. การตรวจให้คะแนนผ่านการเรียนรู้ของเครื่องมีความจำเป็นอย่างยิ่งที่ควรให้ความสำคัญกับการตรวจให้คะแนนตามแผนที่โครงสร้างในแต่ละมิติ โดยเฉพาะอย่างยิ่งในส่วนนำเข้าข้อมูลที่จำเป็นต้องได้จากการวิเคราะห์ข้อมูลทฤษฎีซึ่งเป็นส่วนสำคัญในการนำไปสู่ส่วนการให้ข้อมูลป้อนกลับและส่วนรายงานผลในลำดับถัดไป

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

1. งานวิจัยนี้เป็นการออกแบบวิธีการตรวจให้คะแนนอัตโนมัติผ่านการเรียนรู้ของเครื่อง โดยในระยะเวลาการศึกษานี้ได้กำหนดตัวแปรไว้สร้างสมการทำนายเพื่อเป็นอัลกอริทึมในการเรียนรู้ของเครื่อง ดังนั้นเพื่อให้งานวิจัยมีความสมบูรณ์มากยิ่งขึ้นขั้นตอนต่อไปควรนำตัวแปรที่กำหนดมาสร้างสมการทำนายเพื่อป้อนสู่ระบบการตรวจให้คะแนนอัตโนมัติผ่านการเรียนรู้ของเครื่อง

2. ในการทำวิจัยครั้งต่อไปเมื่อพัฒนาระบบตรวจให้คะแนนผ่านการเรียนรู้ของเครื่องเรียบร้อยแล้วควรตรวจสอบคุณภาพของระบบก่อนโดยการพิจารณาความเที่ยงของผู้ประเมินระหว่างการตรวจของเครื่องและใช้คนตรวจ เพื่อประเมินประสิทธิภาพในการตรวจให้คะแนนอัตโนมัติและค้นหาข้อบกพร่องเพื่อปรับปรุงระบบต่อไป

References

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

Desstya, A., Prasetyo, Z.K., Suyanta., Susila, I., & Irwanto. (2019). Developing an Instrument to Detect Science Misconception of an Elementary School Teacher. *International Journal of Instruction*, 12(3), 201-218.

- Intasoi, S., Jungpeng P., Ngang Tang K., Ketchatturat J., Zhang Y., & Wilson M. (2020). Developing an assessment framework of multidimensional scientific competencies. *International Journal of Evaluation and Research in Education*, 9(4), 963-970.
- Junpeng, P., Krotha, J., Chanayota, K., Tang, K. N., & Wilson, M. (2019). Constructing Progress Maps of Digital Technology for Diagnosing Mathematical Proficiency. *Journal of Education and Learning*, 8(6), 90-102.
- Junpeng, P., Marwiang, M., Chiajunthuk, S., Suwannatrai, P., Chanayota, K., Pongboriboon, K., Tang, K. N., & Wilson, M. (2020). Validation of a digital tool for diagnosing mathematical proficiency. *International Journal of Evaluation and Research in Education*, 9(3), 665-674.
- Lee, H. S., Gweon, G. H., Lord, T., Paessel, N., Pallant, A., & Pryputniewicz, S. (2021). Machine Learning-Enabled Automated Feedback: Supporting Students' Revision of Scientific Arguments Based on Data Drawn from Simulation. *Journal of Science Education and Technology*, 30(2), 168-192.
- Lintean, M., Vasile, R., & Roger, A. (2012). Automatic Detection of Student Mental Models Based on Natural Language Student Input during Metacognitive Skill Training. *International Journal of Artificial Intelligence in Education*, 21(3), 169-190.
- Maestrales, S., Zhai, X., Touitou, I., Baker, Q., Schneider, B., & Krajcik, J. (2021). Using Machine Learning to Score Multi-Dimensional Assessments of Chemistry and Physics. *Journal of Science Education and Technology*, 30(2), 239-254.
- National Research Council. (1997). *Science teaching reconsidered: A handbook*. National Academies Press.
- Organisation for Economic Co-operation and Development. (2019). *PISA 2018 Assessment and Analytical Framework*. PISA, OECD Publishing.
- Reeves, T. (2006). *Design research from a technology perspective*. In J.V.D. Akker, K. Gravemeijer, S. McKenney & N. Nieveen (Eds.). *Educational design research* (pp. 52–66). Routledge.
- Varasunun, P. (2011). The Program Evaluation Standards (3rd Ed.) in 2010. *Journal of Research Methodology*, 24(2), 273-278.
- Wilson, M. (2005). *Constructing Measures: an item response modeling approach*. Lawrence Erlbaum Associates.
- Wright, B.D., & Stone, M. H. (1979). *Best Test Design: Rasch Measurement*. Mesa Press.

Wu, M. L., Adums, R., Wilson, M., & Haldane, S. A. (2007). *ConQuest Version 2: Generalised Item Response Modelling Software*. ACER Press.

Yarbrough, D.B., Shulha, L.M., Hopson, R.K., and Caruthers, F.A. (2010). *The program Evaluation standards: A guide for evaluators and evaluation users (3rd Ed.)*. Sage.

Translate Thai References

Anuntasawat, S. (2017). *Development of a Misconception Diagnostic system in Using a Three-Tier Diagnostic Test with Computer Based Reflective Feedback for Tenth Grade students* [Doctoral dissertation]. Chulalongkorn University. (in Thai)

Chianchana, C. (2009). Multidimensional Analysis. *Journal of Education Khon Kaen University*, 32(4), 13-22. (in Thai)

Insawat, M. (2016). *A comparison of the quality of three-tier diagnostic test of mathematical misconception using different levels of confidence* [Master's thesis]. Chulalongkorn University. (in Thai)

Institute for the promotion of technology science and technology. (2017). *Teacher Guide Basic Science Volume 1*. Bangkok: office of the welfare promotion commission for teachers and educational personnel press. (in Thai)

Intasoi, S. (2020). *Developing a digital tool for assessing multidimensional scientific competencies of the seventh-grade students* [Master's thesis]. Khon Kaen University. (in Thai)

Junpeng, P. (2018). *Applications of the multidimensional item response theory of research* (4th Ed.). Khon Kaen University press. (in Thai)

Kamtet, W. (2017). Scientific Misconception: Type and Assessment Tool. *STOU Education Journal*. 10(2), 54-64. (in Thai)

Khawsuk, J., & Srihasat, K. (2016). The Development of a Mathematics Diagnostic Test for Polynomials at the Matthayom Sueksa One Level. *Veridian E-Journal, Silpakorn University*, 9(3), 1206-1220. (in Thai)

Kladchuen, R., & Sanrach, J. (2018). An Efficiency Comparison of Algorithms and Feature Selection Methods to Predict the Learning Achievement of Vocational Students. *Research Journal Rajamangala University of Technology Thanyaburi*, 17(1), 1-10. (in Thai)

- Rawanprakhon, P., & Tayraukham, S. (2017). The Development of Multidimensional Essay Test. *Journal of Community Development Research (Humanities and Social Sciences)*, 10(4), 13-23. (in Thai)
- Sudjai, M., & Mungsing, S. (2021). Development of an Automated subjective answer scoring System with full-text search technique and PHP text comparison function *Journal of Information Science and Technology*, 11(1), 8-17. (in Thai)
- Wongphummueang, L., & Piyakun, A. (2021). Development of Multidimensional Test of Collaborative Problem-Solving Skills for Lower Secondary School Students. *Journal of Educational Measurment, Mahasarakham University*, 27(1), 232-243. (in Thai)
- Wongwanit, S. (2020). *Design Research in Education*. Chulalongkorn University press. (in Thai)
- Boonjun, S. (2019). *A Comparison of Differential Item Funtioning for Mixed Format Tests by Tree Dffferentail Methods* [Doctoral dissertation]. Burapha University. (in Thai)
- Fiothong, A., Junpeng, P., Suwannatrai, P., Chinjunthuk, S., & Tawarungruang, C. (2022). Designing Open-Ended Question Scoring for Assessment of student Mathematical Proficiency Levels through Digital Technology. *Journal of Educational Measurement Mahasarakham University*, 28(1), 346-362. (in Thai)