

การออกแบบวิธีการตรวจให้คะแนนสำหรับแบบทดสอบอัตนัยเพื่อประเมินระดับความสามารถทางคณิตศาสตร์ ผ่านเทคโนโลยีดิจิทัล

Designing Open-Ended Question Scoring for Assessment of Student Mathematical Proficiency Levels Through Digital Technology

อภิญา ฝอยทอง^{1*} พชรี จันท์เพ็ง² ประภาวดี สุวรรณไตรย์³

สำรวน ชินจันท์³ และ ชัยวัฒน์ ทะวะรุ่งเรือง⁴

Apinya Fiothong^{1*}, Putcharee Junpeng², Prapawadee Suwannatrai³

Samruan Chinjunthuk³ and Chaiwat Tawarungruang⁴

(Received: April 30, 2021 ; Revised: June 14, 2021 ; Accepted: June 18, 2021)

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ (1) วิเคราะห์ผลการตอบของผู้เรียนในการสร้างจุดเปลี่ยนผ่านเพื่อประเมินระดับความสามารถทางคณิตศาสตร์แบบพหุมิติสาระการวัดและเรขาคณิต และ (2) ออกแบบและตรวจสอบคุณภาพวิธีการตรวจให้คะแนนสำหรับแบบทดสอบอัตนัยเพื่อประเมินระดับความสามารถทางคณิตศาสตร์ผ่านเทคโนโลยีดิจิทัล โดยใช้ระเบียบวิธีวิจัยการออกแบบ กลุ่มผู้สอบ คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 1 จำนวน 528 คน เครื่องมือที่ใช้ คือ แบบทดสอบอัตนัย สาระการวัดและเรขาคณิต ผ่านชุดเครื่องมือวินิจฉัยในระบบการทดสอบออนไลน์ “eMAT-Testing” วิเคราะห์ข้อมูลโดยใช้โมเดลการตอบสนองข้อสอบแบบพหุมิติ MRCML

¹ นักศึกษาปริญญาโท สาขาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

² รองศาสตราจารย์ สาขาการวัดและประเมินผลการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยขอนแก่น

³ อาจารย์ โรงเรียนสาธิตมหาวิทยาลัยขอนแก่น ฝ่ายมัธยมศึกษา (มอดินแดง)

⁴ นักวิจัย ศูนย์จัดการข้อมูลและวิเคราะห์ทางสถิติ คณะสาธารณสุขศาสตร์ มหาวิทยาลัยขอนแก่น

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

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

³ Lecturer, Demonstration School of Khon Kaen University, Secondary Division (Modindang)

⁴ Researcher, Data Management and Statistical Analysis Center, Faculty of Public Health, Khon Kaen University

ได้รับทุนอุดหนุนการทำกิจกรรมส่งเสริมและสนับสนุนการวิจัยและนวัตกรรมจากสำนักงานวิจัยแห่งชาติ

This research and innovation activity was funded by National Research Council of Thailand (NRCT)

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

ผลการวิจัยพบว่า

1. จุดเปลี่ยนผ่านระดับความสามารถทางคณิตศาสตร์สำหรับแบบทดสอบอัตนัยเพื่อประเมินระดับความสามารถทางคณิตศาสตร์ผ่านเทคโนโลยีดิจิทัล โดยการกำหนดเกณฑ์พื้นที่บน Wright Map พบว่า มิติกระบวนการทางคณิตศาสตร์ แบ่งได้ 5 ระดับ 4 จุดเปลี่ยนผ่าน จากระดับต่ำสุดไปสูงสุด ที่ -2.30, -0.43, 0.78 และ 1.15 ตามลำดับ และมีมิติโครงสร้างความคิดรวบยอด แบ่งได้ 5 ระดับ 4 จุดเปลี่ยนผ่าน มีระดับต่ำสุดไปสูงสุด ที่ -2.76, 0.11, 0.46 และ 1.16 ตามลำดับ ซึ่งจุดเปลี่ยนผ่านดังกล่าวสามารถนำไปสู่การกำหนดช่วงระดับความสามารถ คะแนนสเกล และคะแนนดิบเพื่อนำไปใช้เป็นเกณฑ์การประเมินระดับความสามารถทางคณิตศาสตร์ในแต่ละมิติ

2. ผลการออกแบบการตรวจให้คะแนนผ่านเทคโนโลยีดิจิทัลสามารถแบ่งออกเป็น 5 ส่วน ประกอบด้วย (1) ส่วนข้อมูลเข้า (2) ส่วนกระบวนการ (3) ส่วนประมวลผล (4) ส่วนแสดงผล และ (5) ส่วนรายงานผลการประเมิน ซึ่งการตรวจสอบคุณภาพของผลการออกแบบ โดยพิจารณาจากการประเมินแบบอิงมาตรฐาน และการประเมินแบบฮิวริสติก จากผู้เชี่ยวชาญ พบว่า (1) การประเมินแบบอิงมาตรฐานทั้ง 3 ด้าน คือ ด้านความถูกต้อง ความมีประโยชน์ และความเป็นไปได้มีระดับการประเมินอยู่ในระดับมากที่สุด และ (2) การประเมินแบบฮิวริสติก โดยภาพรวมของระบบ มีความเหมาะสมในระดับมากที่สุด โดยด้านที่มีระดับการประเมินสูงสุด คือ ด้านการมองเห็นสถานะของระบบ และด้านที่มีระดับผลการประเมินน้อยที่สุด คือ รูปแบบสวยงาม และเรียบง่าย

คำสำคัญ: การตรวจให้คะแนนสำหรับแบบทดสอบอัตนัย วิจัยการออกแบบ โมเดลการตอบสนองข้อสอบพหุมิติ ระดับความสามารถทางคณิตศาสตร์

Abstract

The study aimed to (1) analyze students' multidimensional response patterns for determining cut scores for assessment of mathematical proficiency levels on the topic of Measurement and Geometry, and (2) to design and assess the quality of open-ended question scoring for assessment of mathematical proficiency levels through digital technology. Design research was applied. The sample consisted of 528 grade 7 students. The research instrument was an open-ended question test on the topic of Measurement and Geometry through diagnostic tools in an online testing system—"eMAT-Testing." The analysis of the collected data employed the MRCML model.

The results were as follows:

1. On determining cut scores of mathematical proficiency levels by defining criterion zones on Wright Map, it was found that mathematical processes featured five levels with four cut scores, ranging from the lowest to highest as follows: -2.30, -0.43, 0.78, and 1.15,

respectively. Similarly, conceptual structures consisted of five levels with four cut scores, including -2.76, 0.11, 0.46, and 1.16, respectively. Such cut scores can be employed to determine proficiency ranges, scale scores, and raw scores as criteria for assessment of mathematical proficiency in each dimension.

2. In terms of designing the open-ended question scoring through digital technology, it featured five parts, namely (1) input, (2) process, (3) processing, (4) output, and (5) assessment reporting. The assessment of its quality through standards-based assessment and heuristic assessment conducted by experts showed that: (1) the standards-based assessment on all 3 aspects— accuracy, utility, and feasibility—were rated with the highest level of assessment. (2) Based on the heuristic assessment, the overall system had the highest level of suitability; visibility of system status was rated with the highest level of assessment, while aesthetic and minimalist design obtained the lowest level of assessment.

Keywords: open-ended question scoring, design research, multidimensional item response model, mathematical proficiency level

บทนำ

การดำรงชีวิตในปัจจุบัน คณิตศาสตร์มีบทบาทสำคัญยิ่ง และจำเป็นสำหรับพลเมืองในศตวรรษที่ 21 ในการพัฒนาความสามารถด้านคณิตศาสตร์ จึงมีความจำเป็นอย่างยิ่งในการพัฒนาคุณภาพให้ได้ตามมาตรฐานระดับสากล ความสามารถทางคณิตศาสตร์ สามารถแบ่งได้หลายองค์ประกอบด้วยกัน เน้นไปยัง 2 องค์ประกอบที่สำคัญ ประกอบด้วย (1) ด้านกระบวนการทางคณิตศาสตร์ และ (2) ด้านโครงสร้างความคิดรวบยอด (Junpeng et al., 2019) ในปัจจุบันจะเห็นได้ว่าการประเมินความสามารถทางคณิตศาสตร์ถือเป็นสิ่งสำคัญในการขับเคลื่อนการศึกษา โดยเฉพาะอย่างยิ่งยังมุ่งเน้นไปที่กระบวนการคิดทางคณิตศาสตร์ด้วยเช่นกัน (Chinjunthuk and Junpeng, 2020)

จากผลการประเมินความสามารถทางคณิตศาสตร์ของผู้เรียนในระดับนานาชาติและระดับชาติ พบว่า ในโครงการ PISA 2018 พบว่า คณิตศาสตร์มีคะแนนเพิ่มขึ้นประมาณ 3 คะแนน และ 4 คะแนน ตามลำดับ ซึ่งในการทดสอบทางสถิติถือว่าด้านคณิตศาสตร์ไม่มีการเปลี่ยนแปลงเมื่อเทียบกับรอบการประเมินที่ผ่านมา ยังคงอยู่ในระดับที่ต่ำ (The institute for the Promotion of Teaching Science and Technology (IPST), 2020) ระดับความสามารถทางคณิตศาสตร์เป็นระดับศักยภาพของผู้เรียนในการนำความรู้ความเข้าใจ และทักษะมาใช้ในการแก้ปัญหา สำหรับการประเมินระดับความสามารถทางคณิตศาสตร์ในปัจจุบันส่วนใหญ่พบว่า ใช้ข้อสอบแบบอัตนัย ซึ่งข้อสอบอัตนัยเป็นข้อสอบที่มุ่งให้ผู้เรียนค้นหาคำตอบด้วยตนเอง (Aungkaseraneeekul, 2012; Koyama et al., 2020)

การประเมินจึงเป็นบทบาทสำคัญในการศึกษา และได้มีการพัฒนาเป็นวิธีการในการประเมินความรู้และความสามารถของผู้เรียนด้วยระบบคอมพิวเตอร์ ในการตรวจคำตอบของผู้เรียนในประเทศต่าง ๆ ซึ่งเป็นระบบการตรวจผ่านเทคโนโลยีดิจิทัล เห็นว่ามีระบบที่สมบูรณ์ดีระหว่างการตรวจของผู้สอน และการตรวจที่ดำเนินการโดยระบบ สำหรับการตรวจให้คะแนนแบบอัตโนมัติจะอธิบายถึงฟังก์ชันต่าง ๆ ในแอปพลิเคชันเพื่อทำการประเมินอัตโนมัติของคำตอบแบบสั้น อันดับแรกเราจำแนกคำถาม และคำตอบของครูตามประเภท จากนั้นระบบตรวจการสะกดคำจะใช้กับคำตอบของนักเรียนเพื่อแก้ไขข้อผิดพลาดในการสะกดคำ ลำดับถัดไปทั้งคำตอบของครูและนักเรียนจะได้รับการประมวลผลผ่านการประมวลผลข้อความล่วงหน้าจำนวนมาก เป็นข้อความที่เป็นที่ยอมรับได้ คำตอบเหล่านี้จะเปลี่ยนเป็นข้อความให้อยู่ในรูปแบบที่ยอมรับได้ และสำหรับการตรวจให้คะแนนแบบอัตโนมัติของประเทศไทย ในปี พ.ศ. 2557 ถึงปี พ.ศ. 2558 พบปัญหาในข้อจำกัดการกรอกข้อมูลเป็นชนิดภาษาไทยเท่านั้น การทดลองใช้งานพบปัญหาการเขียนภาษาไทยที่สะกดผิด ส่งผลให้การตรวจคะแนนเกิดความผิดพลาด (Berggren et al., 2019; Rodrigues and Araújo, 2012; Wang and Brown, 2007)

ในปัจจุบันการวัดระดับความสามารถทางคณิตศาสตร์ของผู้เรียนไม่ได้มุ่งวัดแค่เพียงความรู้ ความจำ และความเข้าใจ ในการทดสอบส่วนใหญ่จึงจำเป็นต้องใช้แบบทดสอบอัตนัยในการวัดทักษะความสามารถทางคณิตศาสตร์ ซึ่งแบบทดสอบอัตนัยสามารถตรวจได้ 2 วิธี คือ การตรวจให้คะแนน 2 ค่า หรือการตรวจให้คะแนนแบบหลายค่า (Suksiri and Worain, 2016) อย่างไรก็ตามในการประเมินทางด้านคณิตศาสตร์ที่จำเป็นต้องสะท้อนถึงการแสดงวิธีทำและคำตอบ ที่แสดงให้เห็นถึงกระบวนการแก้ปัญหาทางคณิตศาสตร์ และความคิดรวบยอดของผู้เรียน พบว่ายังเป็นข้อจำกัด ขาดความเป็นปรนัยของผู้ตรวจให้คะแนน และผลการประเมินที่ได้ ยังไม่เพียงพอในการนำไปใช้ตีความหมายเพื่อปรับปรุงพัฒนาผู้เรียน อันเป็นสารสนเทศได้ทันท่วงทีว่าผู้เรียนมีความสามารถระดับใด สิ่งใดที่ผู้เรียนทำได้ และจะพัฒนาให้ดีขึ้นได้อย่างไรบ้าง (Junpeng et al., 2020a) ดังนั้นจึงมีความจำเป็นอย่างยิ่งที่ต้องมีการพัฒนานวัตกรรมในการตรวจให้คะแนนสำหรับแบบทดสอบอัตนัยผ่านเทคโนโลยีดิจิทัลเพื่อนำไปสู่การตีความอย่างมีความหมาย โดยเฉพาะในกรณีที่มีการตรวจให้คะแนนหลายค่า (Rodrigues, 2012; European Language Resources Association (ELRA), 2020)

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องพบว่ามีความพยายามในการนำเทคโนโลยีดิจิทัลมาช่วยในการตรวจให้คะแนนหลายค่า แต่โดยส่วนใหญ่มีมุมมองเน้นเนื้อหาที่เป็นผลสัมฤทธิ์ทางการเรียนตามเนื้อหาในบทเรียนมากกว่าการประเมินระดับความสามารถที่เป็นการคิดขั้นสูง หากพิจารณาในบริบทของการพัฒนาความสามารถทางคณิตศาสตร์ พบว่า มีงานวิจัยที่พัฒนานวัตกรรมการประเมินดังกล่าวไม่มากนักทั้งในบริบทของต่างประเทศและในประเทศ เช่น การวิจัยของ Junpeng et al. (2020a) ได้มีความพยายามในการพัฒนาการตรวจให้คะแนนความสามารถทางคณิตศาสตร์ ในสาระการวัดและเรขาคณิต ของนักเรียนระดับมัธยมศึกษาปีที่ 1 แต่อย่างไรก็ตามนวัตกรรมการประเมินดังกล่าวยังมีข้อจำกัด คือยังไม่สามารถตรวจให้คะแนนสำหรับแบบทดสอบอัตนัยผ่านเทคโนโลยีดิจิทัล จึงยังไม่สามารถให้ข้อมูลป้อนกลับแบบโต้ตอบอัตโนมัติแบบเรียลไทม์เพื่อแนะนำแหล่งการเรียนรู้เพิ่มเติมให้กับผู้เรียน ได้ปรับปรุงแก้ไขในทันที จึงเป็นจุดอ่อนที่สำคัญของการวิจัยดังกล่าว นอกจากนี้ในการตรวจให้คะแนนแบบอัตนัย จำเป็นต้องอาศัยแนวคิด

โมเดลเชิงโครงสร้าง (construct modelling) (Wilson, 2005) มาเป็นฐานในการออกแบบการให้คะแนน และโมเดลการตอบสนองข้อสอบแบบพหุมิติ (multidimensional random coefficients multinomial logit: MRCML) (Adams et al., 1997) มาใช้เป็นแนวทางในการตรวจให้คะแนนสำหรับแบบทดสอบอัตนัยที่มีการตรวจคะแนนหลายค่า และการหาคุณภาพ เพื่อเป็นหลักฐานที่สะท้อนถึงค่าความตรงและความเที่ยงที่มีคุณภาพสูง

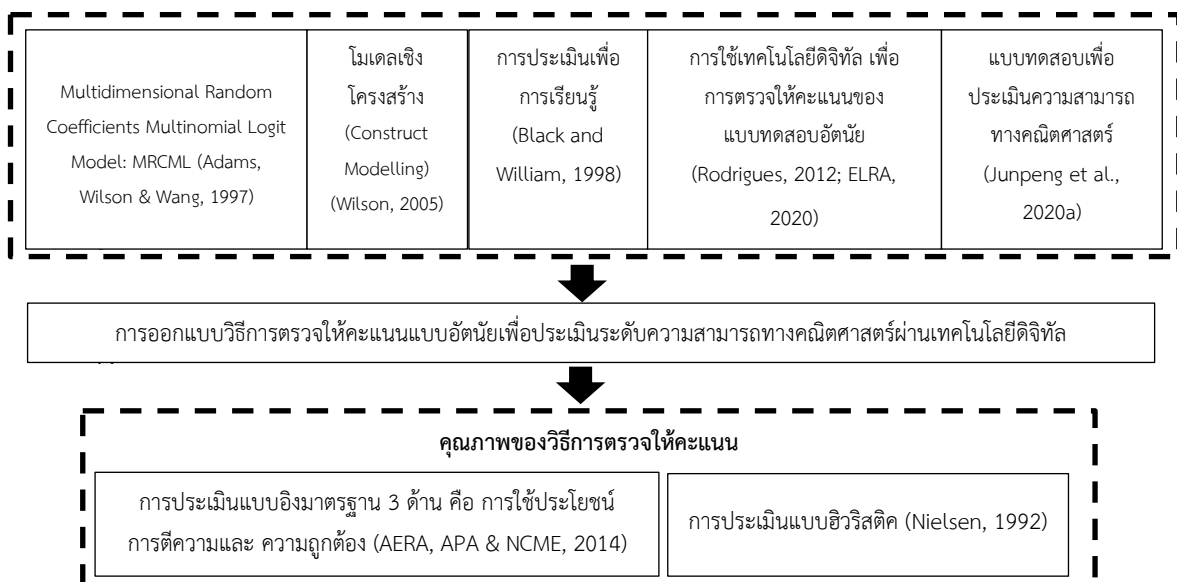
ดังนั้นผู้วิจัยจึงสนใจออกแบบการตรวจให้คะแนนสำหรับแบบทดสอบอัตนัย ผ่านเทคโนโลยีดิจิทัล โดยใช้ข้อมูลทฤษฎีจากโครงการวิจัยของ Junpeng et al. (2020a) มาเป็นฐานในการพัฒนานวัตกรรมต้นแบบ ร่วมกับการนำแนวคิดโมเดลเชิงโครงสร้างและแนวคิดทฤษฎีการตอบสนองข้อสอบแบบพหุมิติ มาเป็นฐานในการพัฒนาการตรวจให้คะแนน เพื่อช่วยในการตรวจคำตอบแบบอัตนัย ที่มีจำนวนมากให้มีความถูกต้องและรวดเร็วมากยิ่งขึ้น โดยมีการพิจารณาคุณภาพของการออกแบบนวัตกรรมต้นแบบในครั้งนี้จากการประเมินแบบอิงมาตรฐาน 3 ด้าน คือ การใช้ประโยชน์ การตีความและ ความถูกต้อง (AERA, APA & NCME, 2014) และการประเมินแบบฮิวริสติก (Nielsen, 1992)

ความมุ่งหมายของการวิจัย

1. เพื่อวิเคราะห์ผลการตอบของผู้เรียนในการสร้างจุดเปลี่ยนผ่านเพื่อประเมินระดับความสามารถทางคณิตศาสตร์แบบพหุมิติ สาระการวัดและเรขาคณิต
2. เพื่อออกแบบและตรวจสอบคุณภาพวิธีการตรวจให้คะแนนสำหรับแบบทดสอบอัตนัยเพื่อประเมินระดับความสามารถทางคณิตศาสตร์ผ่านเทคโนโลยีดิจิทัล

กรอบแนวคิดการวิจัย

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้อง สามารถกำหนดกรอบแนวคิดการวิจัยได้ดังภาพ 1



ภาพ 1 กรอบแนวคิดการวิจัย

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้เป็นการวิจัยการออกแบบ (design research: DR) (Wongwanit, 2020) ซึ่งมีขั้นตอนในการดำเนินการ 4 ระยะ คือ ระยะที่ 1 วิเคราะห์ผลการตอบของผู้เรียนในการสร้างจุดเปลี่ยนผ่าน ระยะที่ 2 การออกแบบและตรวจสอบคุณภาพของวิธีการตรวจให้คะแนน ระยะที่ 3 นำเครื่องมือต้นแบบ (prototype) ไปทดสอบ และระยะที่ 4 นวัตกรรมใหม่และหลักการในการพัฒนาวิธีการตรวจให้คะแนน แต่ในการวิจัยครั้งนี้เป็นการศึกษานำร่องซึ่งจะนำเสนอ 2 ระยะ คือ ระยะที่ 1 และระยะที่ 2 มีรายละเอียด ดังนี้

ระยะที่ 1 วิเคราะห์ผลการตอบของผู้เรียนในการสร้างจุดเปลี่ยนผ่าน

ระยะนี้มุ่งเน้นการวิเคราะห์ผลการตอบของผู้เรียน ที่ได้จากข้อมูลทฤษฎีภูมิโนโครงการการสร้างเครื่องมือวินิจฉัยระดับความสามารถทางคณิตศาสตร์ของผู้เรียนโดยใช้เทคโนโลยีสารสนเทศในการรายงานผลการประเมินเพื่อการเรียนรู้ (Junpeng et al., 2020a) มีรายละเอียดดังนี้

1. กลุ่มผู้สอบ

การศึกษาครั้งนี้ได้กำหนดผู้สอบใช้ในการวิเคราะห์ผลการตอบซึ่งเป็นข้อมูลทฤษฎีภูมิโนโครงการการสร้างเครื่องมือวินิจฉัยระดับความสามารถทางคณิตศาสตร์ของผู้เรียนโดยใช้เทคโนโลยีสารสนเทศในการรายงานผลการประเมินเพื่อการเรียนรู้ (Junpeng et al., 2020a) คือ ผู้เรียนระดับชั้นมัธยมศึกษาปีที่ 1 ภาคเรียนที่ 2 ปีการศึกษา 2562 ของโรงเรียน 4 ภูมิภาค คือ ภาคเหนือ ภาคตะวันออกเฉียงเหนือ ภาคกลาง และภาคใต้ จำนวน 528 คน กำหนดขนาดกลุ่มตัวอย่างให้เพียงพอต่อการประมาณค่าพารามิเตอร์ด้วยโมเดล MRCML (Adams et al., 1997) จึงให้ความสำคัญกับการกำหนดขนาดตัวอย่างและการได้มาของตัวอย่างที่ครอบคลุมเพียงพอในทุกระดับความสามารถ (θ) ในแต่ละมิติที่ทำการศึกษา มากกว่ามุ่งเน้นการสุ่มตัวอย่างให้มีความเป็นตัวแทนของประชากรเหมือนการออกแบบการวิจัยที่มุ่งเน้นการสรุปอ้างอิงนัยทั่วไป (generalizability) จากกลุ่มผู้สอบไปยังประชากร (Demars, 2010) ในการศึกษาครั้งนี้เป็นการศึกษาแบบพหุมิติ ดังนั้นจำนวนพารามิเตอร์ที่ใช้ในการประมาณค่าเพิ่มมากขึ้น มีการตรวจให้คะแนน 2 ค่า และมีการตรวจให้คะแนนมากกว่าสองค่า จำนวนกลุ่มผู้สอบจึงควรมีค่าเพิ่มมากขึ้น กลุ่มผู้สอบควรใช้ขั้นต่ำอย่างน้อย 250-500 คน จึงจะส่งผลให้การประมาณค่ามีความแม่นยำ (Wright and Stone, 1979)

2. เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้มาจากโครงการการสร้างเครื่องมือวินิจฉัยระดับความสามารถทางคณิตศาสตร์ของผู้เรียนโดยใช้เทคโนโลยีสารสนเทศในการรายงานผลการประเมินเพื่อการเรียนรู้ ผ่านชุดเครื่องมือวินิจฉัยเรียกว่า “eMAT-Testing” (Junpeng et al., 2020a) ซึ่ง “eMAT-Testing” เป็นชุดเครื่องมือในระบบออนไลน์แบบเรียลไทม์ผ่าน web application แบบตอบสนองได้โดยตรง (interactive) รองรับระบบ iOS Android และ Window ซึ่งเป็นแบบทดสอบอัตโนมัติ กลุ่มสาระการเรียนรู้คณิตศาสตร์ชั้นมัธยมศึกษาปีที่ 1 ในสาระที่ 2 การวัดและเรขาคณิต ที่ประกอบด้วย 6 สถานการณ์ 13 ข้อย่อย แบ่งเป็นมิติกระบวนการทางคณิตศาสตร์ จำนวน 6 ข้อ และมิติโครงสร้างความคิดรวบยอด จำนวน 7 ข้อ คุณภาพของเครื่องมือในครั้งนี้เป็นไปตามเกณฑ์ที่ยอมรับได้ตามมาตรฐานการทดสอบทางการศึกษาและจิตวิทยา (AERA, APA, and NCME, 2014) ปรากฏรายละเอียดดังหัวข้อ 3.4 การวิเคราะห์โมเดลเชิงโครงสร้างและการแปลผล

3. การเก็บรวบรวมข้อมูล

ผู้วิจัยดำเนินการศึกษาข้อมูลที่ใช้ประกอบการสร้างจุดเปลี่ยนผ่านระดับความสามารถทางคณิตศาสตร์ โดยมีรายละเอียดดังนี้

3.1 ศึกษาแผนที่โครงสร้างระดับความสามารถทางคณิตศาสตร์ของโครงการการสร้างเครื่องมือวินิจฉัยระดับความสามารถทางคณิตศาสตร์ของผู้เรียนโดยใช้เทคโนโลยีสารสนเทศในการรายงานผลการประเมินเพื่อการเรียนรู้ ซึ่งทำการสร้างแผนที่โครงสร้างของระดับผลลัพธ์การเรียนรู้เพื่อใช้ในการประเมินระดับความสามารถทางคณิตศาสตร์ 2 มิติ (Junpeng et al., 2020b) คือ (1) ด้านกระบวนการทางคณิตศาสตร์ และ (2) ด้านโครงสร้างความคิดรวบยอด โดยแต่ละมิติจะแบ่งระดับคุณลักษณะภายในที่มีลักษณะเป็นระดับและมีความเข้มข้นของการคิดเชิงคณิตศาสตร์ลดหลั่นกันจากน้อยไปมาก รายละเอียดดังภาพ 2



ภาพ 2 แผนที่โครงสร้างและคำอธิบายในแต่ละด้านเพื่อประเมินความสามารถทางคณิตศาสตร์
ที่มา : Junpeng et al. (2020b)

จากภาพ 2 แผนที่โครงสร้างของระดับผลลัพธ์การเรียนรู้เพื่อใช้ในการวินิจฉัยระดับความสามารถทางคณิตศาสตร์ 2 มิติ คือ (1) มิติกระบวนการทางคณิตศาสตร์ แบ่งเป็น 5 ระดับ และ (2) มิติโครงสร้างความคิดรวบยอด แบ่งเป็น 5 ระดับ

3.2 การพัฒนาข้อคำถามตามแผนที่โครงสร้าง เป็นการออกแบบข้อคำถามเพื่อให้ได้ข้อมูลตามที่บรรยายไว้ในแผนที่โครงสร้าง โดยข้อสอบแต่ละข้อตรวจให้คะแนนแบบหลายค่า (polytomously scoring)

3.3 การให้คะแนนของผลลัพธ์การเรียนรู้ เป็นการให้คะแนนคำตอบของผู้เรียนที่มีต่อข้อคำถามที่สร้างขึ้น โดยเกณฑ์การให้คะแนนคำตอบจะสัมพันธ์กับระดับของทักษะ ตามแผนที่โครงสร้างที่พัฒนาขึ้น คะแนนที่ผู้เรียน จะได้รับมีตั้งแต่ 0 ถึง 4 คะแนน ในแต่ละมิติ และสอดคล้องกับผลการตอบของผู้เรียนในบริบทจริง โดยมีการพิจารณาความเที่ยงของการตรวจให้คะแนน จากผู้ตรวจให้คะแนน จำนวน 2 ท่าน

ซึ่งเป็นครุคณิตศาสตร์และมีความเชี่ยวชาญในการตรวจให้คะแนนตามแผนที่เชิงโครงสร้าง ให้มีความเห็นสอดคล้องกันตรงกันทั้งสองคน

3.4 การวิเคราะห์โมเดลเชิงโครงสร้างและการแปลผล (measurement model/Wright map) เป็นขั้นตอนที่ผู้วิจัยนำคะแนนคำตอบที่ผู้เรียน แต่ละคนที่ได้รับการประมาณค่าพารามิเตอร์ระดับความสามารถทางคณิตศาสตร์ และประมาณค่าพารามิเตอร์คุณลักษณะข้อสอบ โดยผู้วิจัยประมาณค่าพารามิเตอร์ตามโมเดลการวัดแบบพหุมิติ ด้วยโปรแกรม ConQuest 2.0 (Wu et al., 2007)

จากภาพ 4 (a) โมเดลการวัดระดับความสามารถทางคณิตศาสตร์ประกอบด้วย มิติด้านกระบวนการทางคณิตศาสตร์ (MAP) มีข้อคำถามจำนวน 6 ข้อ ประกอบด้วยข้อคำถามที่ 5,6,7,9,10 และ 11 และมิติด้านโครงสร้างความคิดรวบยอด (SLO) ข้อคำถามจำนวน 7 ข้อ ประกอบด้วยข้อคำถามที่ 1, 2, 3, 4, 8,12 และ 13 โดยคุณภาพของเครื่องมือเป็นไปตามเกณฑ์ที่ยอมรับได้ตามมาตรฐานการทดสอบทางการศึกษาและจิตวิทยา (AERA, APA, and NCME, 2014) คือ หลักฐานความตรงเชิงโครงสร้างภายใน มีความสอดคล้องกับข้อมูลเชิงประจักษ์ ($\chi^2 = 5297.80$, $df=11$, $p=.01$) เมื่อทดสอบด้วยสถิติทดสอบ Likelihood-Ratio พบว่า การตรวจให้คะแนนแบบหลายค่าสอดคล้องกลมกลืนกับข้อมูล ($G^2 = 11516.86$, $AIC = 11596.86$, $BIC = 11625.77$) สะท้อนให้เห็นว่าการประเมินระดับความสามารถทางคณิตศาสตร์เหมาะสมสำหรับการวัดแบบพหุมิติ และเหมาะสมสำหรับการตรวจให้คะแนนหลายค่าตามแผนที่โครงสร้างที่กำหนดไว้ (2) หลักฐานความเที่ยงแบบ Expected A Posteriori (EAP/PV) มีค่าความเที่ยงในมิติที่ 1 และ 2 เท่ากับ 0.88 และ 0.87 สะท้อนให้เห็นว่า เครื่องมือดังกล่าวมีความคงเส้นคงวาในการประมาณค่าความสามารถในแต่ละมิติ และ (3) ความเหมาะสมรายข้อ พบว่า มีค่า INFIT MNSQ อยู่ระหว่าง 0.79-1.27 ซึ่งมีค่าอยู่ในช่วง 0.75-1.33 ตามเกณฑ์ที่กำหนดไว้ จากหลักฐานที่กล่าวมาสะท้อนให้เห็นว่าเครื่องมือมีคุณภาพ สามารถนำไปประเมินระดับความสามารถทางคณิตศาสตร์ของผู้เรียนได้อย่างถูกต้องและมีความคงเส้นคงวา

4. ขั้นตอนการเก็บรวบรวมข้อมูลและการวิเคราะห์ข้อมูล

4.1 การให้คะแนนของผลลัพธ์การเรียนรู้ มีการกำหนดระดับคะแนนให้สอดคล้องกับแผนที่โครงสร้างของระดับผลลัพธ์การเรียนรู้ ซึ่งข้อสอบที่ผู้วิจัยนำมาวิเคราะห์ผลการตอบเป็นข้อสอบอัตนัย มีการตรวจให้คะแนนแบบหลายค่า (0,1,2,3,4)

4.2 การสร้างจุดเปลี่ยนผ่านเพื่อประเมินระดับความสามารถทางคณิตศาสตร์

1) วิเคราะห์โมเดลการตอบสนองข้อสอบแบบพหุมิติสัมพันธ์การสุ่มแบบโลจิท (MRCML) โดยใช้โปรแกรม ACER ConQuest Version 2.0 (Wu et al., 2007) ในการประมาณค่าพารามิเตอร์ของข้อสอบและผู้สอบแบบ MLE พร้อมนำเสนอเป็น Wright Map โดยด้านซ้ายมีนำเสนอค่าความสามารถของผู้เรียนแต่ละข้อ ส่วนด้านขวามีนำเสนอค่าระดับความสามารถทางคณิตศาสตร์แต่ละขั้นของการตอบ โดยนำเสนอแยกตามมิติ

2) ทำการกำหนดจุดเปลี่ยนผ่าน โดยนำค่าขั้นระดับความสามารถทางคณิตศาสตร์ (threshold) ในแต่ละขั้นของการตอบแบบเดียวกันรวมกันแล้วหาค่าเฉลี่ยในแต่ละมิติ ตัวอย่างเช่น นำค่าขั้นระดับความสามารถทางคณิตศาสตร์จากระดับที่ 1 ไปยังระดับที่ 2 มารวมกันแล้วหาค่าเฉลี่ย เพื่อกำหนดเป็นจุดเปลี่ยนผ่านในการแบ่งระดับความสามารถทางคณิตศาสตร์ระดับที่ 1

3) แปลงคะแนนสเกลให้สามารถเปรียบเทียบกันเนื่องจากเมื่อพิจารณาการนำคะแนนระดับความสามารถทางคณิตศาสตร์ที่แท้จริงนำเสนออาจยากต่อการนำไปใช้ในทางปฏิบัติ เนื่องจากไม่เป็นที่แพร่หลาย ใช้โดยทั่วไปในสถานการณ์ เพื่อให้ง่ายต่อการนำไปตีความและไปใช้ในทางปฏิบัติ (Junpeng et al., 2019) ได้มีการคำนวณคะแนนความสามารถรายบุคคล โดยให้มีค่าเฉลี่ยเท่ากับ 50 และส่วนเบี่ยงเบนมาตรฐานเท่ากับ 10 สามารถเขียนเป็นสมการ คะแนนสเกล = $50 + 10(\theta)$ เมื่อ θ แทนระดับความสามารถในแต่ละมิติ

ระยะที่ 2 ผลการออกแบบและตรวจสอบคุณภาพวิธีการตรวจให้คะแนนผ่านเทคโนโลยีดิจิทัล

ผู้วิจัยร่วมกับครูและวิศวกรคอมพิวเตอร์ ผู้เชี่ยวชาญในด้านการวัดและประเมินผล ผู้เชี่ยวชาญในด้านการใช้เทคโนโลยีเพื่อร่วมกันออกแบบการตรวจให้คะแนนสำหรับแบบทดสอบอัตโนมัติผ่านเทคโนโลยีดิจิทัล โดยมีขั้นตอน ดังนี้

1. ผู้ให้ข้อมูล

ผู้เชี่ยวชาญด้านการวัดและประเมินผลการศึกษา ผู้เชี่ยวชาญในด้านการใช้เทคโนโลยีและครุคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 1 เป็นผู้เชี่ยวชาญด้านละ 3 คน ที่มีประสบการณ์อย่างน้อย 5 ปี

2. เครื่องมือ

เครื่องมือคือแบบประเมินแบบอิงมาตรฐาน (AERA, APA & NCME, 2014) และแบบประเมินแบบฮิวริสติก (Nielsen, 1992) โดยมีการดัดแปลงมาจากเครื่องมือที่พัฒนาแล้วของ Nielsen (1992) เพื่อสะท้อนความถูกต้อง ความเหมาะสม และความเป็นไปได้ ในการนำไปใช้ตีความ

3. ขั้นตอนการเก็บรวบรวมข้อมูลและวิเคราะห์ข้อมูล

3.1 นำผลการสร้างจุดเปลี่ยนผ่านระดับความสามารถทางคณิตศาสตร์มาร่วมกันออกแบบนวัตกรรมต้นแบบวิธีการตรวจให้คะแนนสำหรับแบบทดสอบอัตโนมัติผ่านเทคโนโลยีดิจิทัล โดยอาศัยโมเดลเชิงโครงสร้าง (construct modelling) มาเป็นแนวทางในการพัฒนา เพื่อประเมินระดับความสามารถทางคณิตศาสตร์ของผู้เรียนรายบุคคล

3.2 พิจารณาคุณภาพของนวัตกรรมต้นแบบที่พัฒนาขึ้นโดยผู้เชี่ยวชาญด้านการวัดและประเมินผลการศึกษา ด้านการใช้เทคโนโลยีนักเรียนและครุคณิตศาสตร์ระดับชั้นมัธยมศึกษาปีที่ 1 เพื่อพิจารณาความถูกต้อง ความเหมาะสม และความเป็นไปได้ในการนำไปใช้ตีความ

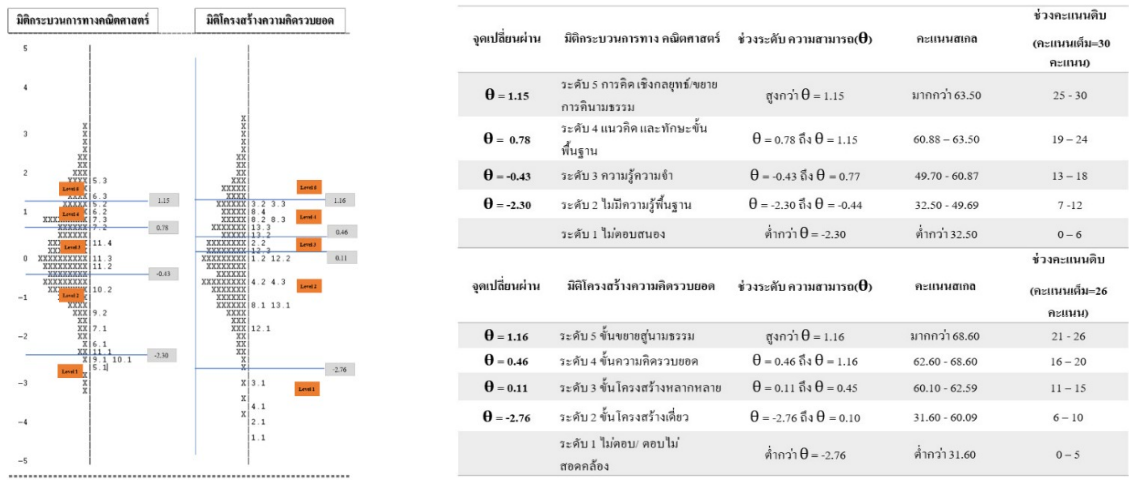
3.3 ร่วมกันสะท้อนคิดและนำไปทดลองใช้ในลำดับต่อไป

ผลการวิจัย

1. การสร้างจุดเปลี่ยนผ่านเพื่อประเมินระดับความสามารถทางคณิตศาสตร์จากการวิเคราะห์ผลการตอบของผู้เรียน

จุดเปลี่ยนผ่านระดับความสามารถทางคณิตศาสตร์สำหรับแบบทดสอบอัตโนมัติเพื่อประเมินระดับความสามารถทางคณิตศาสตร์ผ่านเทคโนโลยีดิจิทัล โดยการกำหนดเกณฑ์พื้นที่บน Wright Map พบว่า มิติกระบวนการทางคณิตศาสตร์ แบ่งได้ 5 ระดับ 4 จุดเปลี่ยนผ่าน จากระดับต่ำสุดไปสูงสุด ที่ -2.30, -0.43,

0.78 และ 1.15 ตามลำดับ และมิติโครงสร้างความคิดรวบยอด แบ่งได้ 5 ระดับ 4 จุดเปลี่ยนผ่าน มีระดับต่ำสุดไปสูงสุด ที่ -2.76, 0.11, 0.46 และ 1.16 ตามลำดับ ซึ่งจุดเปลี่ยนผ่านดังกล่าวสามารถนำไปสู่การกำหนดช่วงระดับความสามารถ คะแนนสเกล และคะแนนดิบ ซึ่งเป็นเกณฑ์การประเมินระดับความสามารถทางคณิตศาสตร์ในแต่ละมิติ รายละเอียดปรากฏผลดังภาพ 3



(a) Wright Map การตรวจให้คะแนน (b) เกณฑ์จุดเปลี่ยนผ่านระดับความสามารถทางคณิตศาสตร์ในแต่ละมิติ

ภาพ 3 Wright Map การตรวจให้คะแนนในแต่ละมิติ และเกณฑ์จุดเปลี่ยนผ่านระดับความสามารถทางคณิตศาสตร์ในแต่ละมิติ

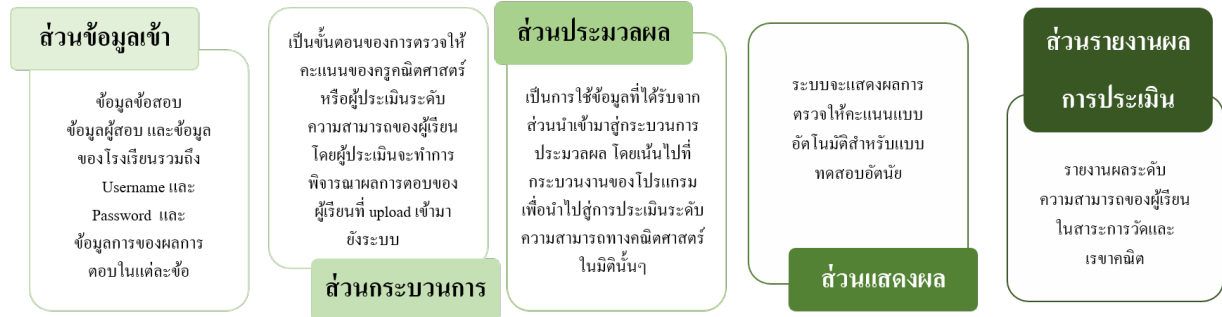
จากภาพ 3 เป็น Wright Map แสดงพื้นที่กำหนดขอบเขต ของระดับความสามารถทางคณิตศาสตร์แบบพหุมิติ ในมิติโครงสร้างทางความคิดรวบยอด ข้อคำถามแต่ละข้อประกอบด้วยการวินิจฉัย 5 ระดับ ที่สอดคล้องกับแผนที่โครงสร้าง ประกอบด้วย 4 Threshold โดย Threshold จะเป็นจุดแบ่งระดับความสามารถทางคณิตศาสตร์ ในที่นี้จะแทนตำแหน่ง Threshold ในแต่ละลำดับขั้นของการตอบแต่ละข้อด้วยจุดทึบ และ Threshold ของลำดับขั้นการตอบในแต่ละข้อจะลดหลั่นกันตามแนวคอลัมน์ โดยผู้วิจัยกำหนดขอบเขตด้วยเส้นแบ่ง จำนวน 4 เส้น จะสังเกตเห็นได้ว่า Threshold ในแต่ละกลุ่มถูกแบ่งออกจากกันได้อย่างชัดเจน และไม่พบว่ามี Threshold ใดที่มีส่วนซ้อนทับกันระหว่างกลุ่มของ Threshold เช่น ถ้าเป็นกลุ่ม Threshold ที่ 1 (พิจารณาในแนวนอน) ทั้ง 2 ข้อ จะต้องมีการกระจายของจุดแสดง Threshold ที่ 1 ในระดับที่ ใกล้เคียงกันทั้ง 2 ข้อ และจะต้องไม่ห่างจากเส้นแบ่งมากจนเกินไป (เส้นที่ลากผ่านค่าเฉลี่ยของค่า Threshold ที่ 1) หรือ ต้องไม่ห่างออกไปจนไปใกล้อยู่กับอีกกลุ่ม Threshold อื่น เมื่อพิจารณาทั้ง 4 Threshold ของแผนภาพนี้ถือว่า มีความเหมาะสมในการกำหนดจุดเปลี่ยนผ่านได้

2. ผลการออกแบบและตรวจสอบคุณภาพการตรวจให้คะแนนสำหรับแบบทดสอบอัตนัยเพื่อประเมินระดับความสามารถทางคณิตศาสตร์ผ่านเทคโนโลยีดิจิทัล

2.1 ผลการออกแบบวิธีการตรวจให้คะแนนสำหรับแบบทดสอบอัตนัยผ่านเทคโนโลยีดิจิทัล

ผลการออกแบบการตรวจให้คะแนนสำหรับแบบทดสอบอัตนัยผ่านเทคโนโลยีดิจิทัล พบว่าควรมีการตรวจให้คะแนนแบบหลายค่าตามแผนที่โครงสร้างในแต่ละมิติ โดยมีระบบการตรวจให้คะแนน

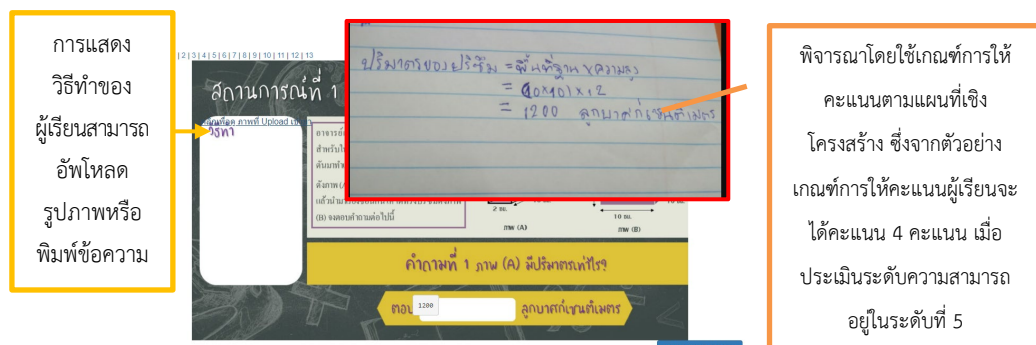
สำหรับแบบทดสอบอัตโนมัติผ่านเทคโนโลยีดิจิทัล ซึ่งจะแสดงข้อมูลส่วนตัว คะแนนที่ผู้เรียนทำได้ ระดับความสามารถปัจจุบันของผู้เรียน สิ่งที่คุณเรียนต้องพัฒนาไปสู่ระดับที่สูงขึ้น โดยสามารถแบ่งออกเป็น 5 ส่วน ประกอบด้วย (1) ส่วนข้อมูลเข้า (2) ส่วนกระบวนการ (3) ส่วนประมวลผล (4) ส่วนแสดงผล และ (5) ส่วนรายงานผลการประเมิน ดังภาพ 5



ภาพ 5 แสดงการออกแบบการตรวจให้คะแนนสำหรับแบบทดสอบอัตโนมัติผ่านเทคโนโลยีดิจิทัล

รายละเอียดขั้นตอนของการออกแบบการตรวจให้คะแนนสำหรับแบบทดสอบอัตโนมัติผ่านเทคโนโลยีดิจิทัล ดังนี้

(1) ส่วนข้อมูลเข้า เป็นการนำจุดเปลี่ยนผ่านเพื่อประเมินระดับความสามารถทางคณิตศาสตร์ที่ได้จากระยะที่ 1 ซึ่งประกอบด้วย ช่วงคะแนนดิบและคะแนนสเกล พร้อมคำอธิบายระดับความสามารถในแต่ละมิติ จากนั้นมีการกำหนดคำตอบที่ถูกต้องของข้อสอบอัตโนมัติแต่ละข้อ เพื่อเป็นประโยชน์สำหรับผู้ใช้ที่ต้องการตรวจให้คะแนนของข้อสอบอัตโนมัติแบบอัตโนมัติ และสำหรับครูที่จะใช้เป็นแนวทางในการตรวจให้คะแนนผ่านระบบดิจิทัล ในส่วนนี้ระบบจำเป็นต้องอาศัยข้อมูล รายชื่อโรงเรียน รหัสนักเรียน เพื่อเตรียมพร้อมในการสอบแต่ละคน หลังจากนั้นมีการกำหนดห้องสอบเพื่อให้นักเรียนสามารถเข้าสู่ระบบการสอบได้ โดยการสร้าง QR Code ที่สามารถสแกนไปไปยังระบบได้ทันที โดยผู้เรียนต้องมีการ login ก่อนดำเนินการสอบ โดยขณะที่ทำการทดสอบข้อสอบอัตโนมัติแต่ละข้อ ผู้เรียนสามารถ upload ผลการตอบได้เป็น 2 ลักษณะ คือ (1) การพิมพ์เพื่อแสดงวิธีทำและเติมคำตอบในระบบ และ (2) สามารถ upload ไฟล์ รูปภาพนามสกุล .jpg เพื่อแสดงวิธีทำ โดยสามารถนำภาพที่ได้จากการแสดงวิธีทำบนกระดาน หรือ โปรแกรมที่สามารถเขียนแสดงวิธีทำได้ ดังภาพ 6



ภาพ 6 แสดงตัวอย่างผลการตอบของผู้เรียนโดยการอัปโหลดรูปภาพของการแสดงวิธีทำ

(2) ส่วนกระบวนการ เป็นขั้นตอนของการตรวจให้คะแนนของครูคณิตศาสตร์ หรือ ผู้ประเมินระดับความสามารถของผู้เรียน โดยใช้เกณฑ์การให้คะแนนตามแผนที่เชิงโครงสร้างที่กำหนดไว้ในแต่ละมิติ โดยผู้ประเมินจะทำการพิจารณาผลการตอบของผู้เรียนที่ upload เข้ามายังระบบ โดยพิจารณาการให้คะแนนทีละข้อ จากนั้นผู้ประเมินจะทำการกรอกคะแนนในช่องของการคะแนนแต่ละข้อ ตั้งแต่ 0-4 คะแนน ในแต่ละมิติ ตัวอย่างการให้คะแนน เช่น จากภาพ 8 ในกระบวนการตรวจให้คะแนนจะพิจารณาโดยใช้เกณฑ์การให้คะแนนตามแผนที่เชิงโครงสร้าง นักเรียนสามารถนำความคิดรวบยอดมาเชื่อมโยงและสามารถสรุปอ้างอิงเป็นองค์ความรู้เชิงนามธรรมและสร้างองค์ความรู้ จากแนวความคิดของตนเองได้

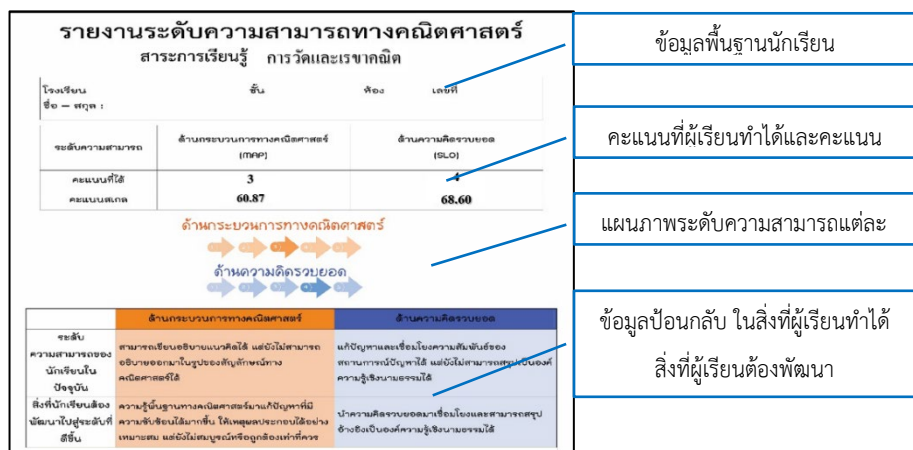
(3) ส่วนประมวลผล เป็นการใช้ข้อมูลที่ได้รับจากส่วนนำเข้ามาสู่กระบวนการประมวลผล โดยเน้นไปที่กระบวนการของโปรแกรม (algorithm) เช่น สูตรการคำนวณ การแปลผลและตีความ เป็นต้น ซึ่งระบบดิจิทัลจะทำการรวมคะแนนดิบของข้อสอบทุกข้อในแต่ละมิติ ไปเทียบกับจุดเปลี่ยนผ่านที่กำหนดไว้ จากนั้นระบบจะนำไปประมวลผลเทียบกับคะแนนสากล และนำไปสู่การประเมินระดับความสามารถทางคณิตศาสตร์ ในมิตินั้นๆ โดยเมื่อระบบทำการประเมินผู้เรียนแต่ละคนแล้วว่าอยู่ในระดับใด

(4) ส่วนแสดงผล เป็นการแสดงผลการตรวจให้คะแนนสำหรับแบบทดสอบอัตโนมัติ ผ่านเทคโนโลยีดิจิทัลของการตรวจให้คะแนนแบบหลายค่า ในขั้นตอนนี้ผู้เรียน สามารถรู้ผลคะแนนได้ 2 ลักษณะ คือ คะแนนที่ได้จากการตรวจผ่านระบบดิจิทัลแบบอัตโนมัติ โดยรายงานคะแนนโดยพิจารณาจากคำตอบที่ถูกต้อง แต่อย่างไรก็ตามการรายงานระดับความสามารถทางคณิตศาสตร์ในแต่ละมิติ จำเป็นต้องรอผลการตรวจคะแนนที่ยืนยันจากผู้ประเมิน ดังนั้น การกำหนดว่าผู้เรียนสามารถดูคะแนนได้ช่วงใดขึ้นอยู่กับการออกแบบของครูหรือผู้ประเมิน ตัวอย่างการแสดงผลคะแนนแบบอัตโนมัติ

(4) ส่วนรายงานผลการประเมิน

ในส่วนนี้เป็นการรายงานผลการทดสอบเพื่อประเมินระดับความสามารถทางคณิตศาสตร์ เป็นการนำผลจากการประมวลผลมารายงานให้ผู้ใช้งานเข้าใจได้ง่ายมากยิ่งขึ้น การรายงานในส่วนนี้จะเป็นการรายงานผลจากการประเมินและจากการวิเคราะห์ของระบบโดยจะแสดงผลเป็นรายบุคคลสำหรับผู้เรียน โดยจะสามารถดูผลประเมินได้เพียงของตัวเองเท่านั้น ในการรายงานจะประกอบไปด้วย

- 1) ข้อมูลส่วนบุคคล
- 2) ข้อมูลสารสนเทศและสถิติของการประเมิน
- 3) ข้อมูลจากการประเมินผลระดับความสามารถทางคณิตศาสตร์ ดังภาพ 7



ภาพ 7 ตัวอย่างหน้าสำหรับรายงานผลการประเมินระดับความสามารถของผู้เรียนรายบุคคล

2.2 ผลการตรวจสอบคุณภาพวิธีการตรวจให้คะแนนสำหรับแบบทดสอบอัตนัย ผ่านเทคโนโลยีดิจิทัล

สำหรับการวิจัยครั้งนี้การตรวจสอบคุณภาพของวิธีการตรวจให้คะแนนสำหรับแบบทดสอบอัตนัย โดยประเมิน 2 แบบ คือ แบบประเมินอิงมาตรฐาน และแบบประเมินอิวิริสติก โดยผ่านการประเมินจากผู้เชี่ยวชาญ ดังรายละเอียดดังนี้

2.2.1 ผลประเมินวิธีการตรวจให้คะแนนแบบสำหรับแบบทดสอบอัตนัยผ่านเทคโนโลยีดิจิทัลด้วยแบบประเมินอิงมาตรฐาน โดยมีผู้เชี่ยวชาญด้านวัดและประเมินผลการศึกษา ทั้งหมด 3 ท่าน พบว่าทั้ง 3 ด้าน คือ ด้านความถูกต้อง ความมีประโยชน์ และความเป็นไปได้มีระดับการประเมินอยู่ในระดับมากที่สุด

2.2.2 ผลการประเมินวิธีการตรวจให้คะแนนแบบสำหรับแบบทดสอบอัตนัยผ่านเทคโนโลยีดิจิทัลด้วยแบบประเมินอิวิริสติก โดยมีผู้เชี่ยวชาญในด้านเทคโนโลยีทั้งหมด 3 ท่าน พบว่าโดยภาพรวมของระบบ มีความเหมาะสมในระดับมากที่สุด ($\bar{X} = 4.83$, $S.D. = 0.23$) ด้านที่มีระดับการประเมินสูงสุด คือ ด้านการมองเห็นสถานะของระบบ มีความเหมาะสมในระดับมากที่สุด ($\bar{X} = 5$, $S.D. = 0$) และน้อยที่สุด คือ รูปแบบสวยงาม และเรียบง่าย มีความเหมาะสมในระดับมาก ($\bar{X} = 4.33$, $S.D. = 0.57$)

อภิปรายผล

จากผลการวิจัยในครั้งนี้พบว่าสามารถนำประเด็นมาอภิปรายได้หลายประเด็น ดังนี้

1. ผลการวิจัยในครั้งนี้พบว่า จุดเปลี่ยนผ่านระดับความสามารถทางคณิตศาสตร์สำหรับแบบทดสอบอัตนัย มีความเหมาะสมสามารถประเมินความสามารถทางคณิตศาสตร์ได้โดยพิจารณาจากเกณฑ์การกำหนดพื้นที่บนแผนที่โครงสร้าง มีการกระจายระดับของแผนภาพที่แตกต่างกันอย่างชัดเจนในแต่ละจุดเปลี่ยนผ่าน แสดงให้เห็นว่าผู้เรียนใช้ความสามารถในการทำข้อสอบได้แตกต่างกัน ข้อสอบในส่วนของระดับความสามารถทางคณิตศาสตร์ทั้ง 5 ชั้นนี้ จึงสามารถแยกผู้เรียนออกจากกันตามระดับความสามารถทางคณิตศาสตร์ได้อย่างชัดเจนหรือคำบรรยายความสามารถและคำบรรยายคำตอบสามารถแบ่ง เป็นการสะท้อนว่าสามารถนำไปใช้เป็นมาตรฐานในการให้คะแนนเพื่อประเมินระดับความสามารถทางคณิตศาสตร์ได้ (Junpeng et al, 2019; Junpeng et al, 2020b) ซึ่งสอดคล้องกับมาตรฐานของการวัดทางการศึกษาและจิตวิทยา (AERA, APA, & NCME, 2014) ที่กล่าวว่า จุดเปลี่ยนผ่านควรครอบคลุมกับระดับความสามารถของผู้สอบในแต่ละมิติที่ทำการศึกษา โดยในที่นี้ พิจารณาจาก Wright Map ทางด้านซ้ายมือที่กระจายให้เห็นถึง สะท้อนให้เห็นถึงความถูกต้องและแสดงให้เห็นถึงหลักฐานความตรงเชิงโครงสร้างภายใน (internal structure)

2. จากผลการออกแบบการตรวจให้คะแนน พบว่า ส่วนข้อมูลเข้าถือเป็นหัวใจสำคัญที่นำไปสู่การตรวจให้คะแนนในส่วนกระบวนการ ส่วนประมวลผล ส่วนแสดงผล และส่วนรายงานผลการประเมินมีความถูกต้องและคงเส้นคงวาในการตรวจให้คะแนนให้ตรงกับระดับความสามารถในแต่ละมิติของผู้เรียนแต่ละคน ดังนั้น ก่อนที่จะนำแบบทดสอบไปใช้และตรวจให้คะแนนของผู้ประเมิน ควรมีการตรวจสอบข้อมูลดังกล่าว และออกแบบให้ง่ายในการนำไปใช้จริงมากขึ้น โดยเฉพาะการนำเสนอแสดงวิธีทำของผู้เรียนแต่ละคนที่มีได้หลายวิธี ซึ่งผู้ดำเนินการสอบควรเตรียมกระดาษไว้ให้ผู้เรียนแสดงวิธีทำสำหรับกรณีผู้เรียน upload

คำตอบเป็น JPG. เข้าระบบ รวมถึง ความพร้อมของอุปกรณ์ต่างๆ เช่น อุปกรณ์ทางคอมพิวเตอร์ มือถือ ระบบ Internet เป็นต้น ซึ่งสอดคล้องกับ Jaihuek and Mungsing (2020) ที่ให้ความสำคัญกับการนำเข้าสู่ข้อมูล เช่น ปัญหาการสะกดคำผิด การแสดงวิธีทำ การแสดงคำตอบ ระบบการทดสอบลุ่ม เนื่องจากขาดการเชื่อมต่อ อินเทอร์เน็ตกระทันหัน เป็นต้น

3. ในขั้นตอนของการตรวจให้คะแนนจำเป็นที่ผู้ประเมินต้องพิจารณาเกณฑ์และแนวทางให้คะแนน ตามแผนที่เชิงโครงสร้างที่กำหนดไว้ในแต่ละมิติเป็นหลัก โดยมีข้อคำนึงโดยเฉพาอย่างยิ่งในมิติกระบวนการทางคณิตศาสตร์ ผู้ประเมินควรพิจารณาการแสดงวิธีทำควบคู่ไปกับการแสดงคำตอบ โดยคำตอบที่ไม่ถูกต้อง อาจไม่ได้สะท้อนว่าผู้เรียนมีระดับความสามารถในระดับต่ำ ผู้ประเมินจำเป็นต้องให้ความสำคัญกับ กระบวนการแสดงวิธีทำเพื่อสะท้อนถึงระดับของกระบวนการแก้ปัญหาประกอบการประเมิน และเพื่อให้ผล การตรวจให้คะแนนมีความถูกต้องและน่าเชื่อถือมากยิ่งขึ้นจำเป็นต้องมีการใช้ผู้ประเมินมากกว่า 1 คน แล้วนำ ผลการประเมินของแต่ละคนมาพิจารณาถึงความสอดคล้องของการให้คะแนน (Berggren et al., 2019; Rodrigues and Araújo, 2012; Wang and Brown, 2007)

4. ผลการตรวจสอบคุณภาพการตรวจให้คะแนนผ่านเทคโนโลยีดิจิทัล พบว่า ด้านที่เป็นจุดเด่น สำหรับการพัฒนานวัตกรรมต้นแบบในครั้งนี้ คือ ด้านความมีประโยชน์ ทั้งนี้อาจจะเป็นเนื่องมาจาก การให้ คะแนนในครั้งนี้มุ่งเน้นไปที่การประเมินระหว่างเรียน (formative feedback) ดังนั้นจึงมุ่งเน้นในการออกแบบ ครั้งนี้ จึงมุ่งเน้นให้ผู้ประเมินหรือครูสามารถนำผลการประเมินที่ได้ไปสู่การวินิจฉัย หรือการให้ข้อมูลป้อนกลับ แก่ผู้เรียนเพื่อนำไปสู่การปรับปรุงแก้ไขผู้เรียน และพัฒนาความสามารถทางคณิตศาสตร์ให้สูงขึ้นเป็นสำคัญ มากกว่าการนำผลที่ได้ไปตัดสินผู้เรียนหรือตัดสินเลื่อนชั้นเรียน (Black and William, 1998) จุดเด่น ที่สำคัญอีกด้านคือ ด้านความถูกต้อง มีระดับการประเมินสูงเช่นเดียวกัน อาจจะเป็นเนื่องมาจากการตรวจสอบ ที่มีลักษณะอัตโนมัติถ้าผู้ประเมินหรือครูไม่เข้าใจเกณฑ์การให้คะแนนตามแผนที่โครงสร้างที่กำหนดไว้ อาจส่งผล อย่างยิ่งต่อคะแนนดิบเพื่อนำไปใช้ในการประมวลผลและประเมินผู้เรียนต่อไป ดังนั้นในขั้นตอนดังกล่าว จึงจำเป็นต้องอาศัยความระมัดระวังเป็นอย่างยิ่ง และทำความเข้าใจการให้คะแนน (Rodrigues and Araújo, 2012; ELRA, 2020) โดยแนวทางการตรวจสอบว่าการตรวจให้คะแนนมีความถูกต้องและน่าเชื่อถือหรือไม่ ผู้ประเมินหรือครูอาจนำผลการประเมินที่ได้ภายหลังจากการตรวจให้คะแนนไปหาค่าสัมประสิทธิ์สหสัมพันธ์ ร่วมกับเกณฑ์อื่นที่เกี่ยวข้องกับระดับความสามารถทางคณิตศาสตร์ เช่น คะแนนผลสัมฤทธิ์ทางการเรียน เป็นต้น ซึ่งเป็นหลักฐานความตรงที่มีความสัมพันธ์กับเกณฑ์อื่นตามมาตรฐานของการวัดทางการศึกษาและ จิตวิทยา (AERA, APA, & NCME, 2014)

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 การแสดงผลการตรวจให้คะแนนสำหรับแบบทดสอบอัตโนมัติผ่านเทคโนโลยีดิจิทัล เพื่อกำหนด ว่าผู้เรียนสามารถทราบผลการประเมินได้ช่วงใดขึ้นอยู่กับกระบวนการออกแบบของผู้ใช้ระบบและจุดมุ่งหมายของ การประเมินเป็นสำคัญ โดยการวิจัยครั้งนี้ผู้เรียนสามารถรู้ผลคะแนนได้ 2 ลักษณะ คือ (1) คะแนนที่ได้จาก การตรวจผ่านระบบดิจิทัลแบบอัตโนมัติแบบเรียลไทม์ โดยรายงานคะแนนที่พิจารณาจากคำตอบของผู้เรียน

แล้วเทียบว่าคำตอบนั้นตรงกับแผนที่โครงสร้างระดับใดตามที่กำหนดไว้ และ (2) การรายงานผลการตรวจให้คะแนนอย่างเป็นทางการภายหลังการดำเนินการทดสอบเสร็จสิ้น ตามการรายงานระดับความสามารถทางคณิตศาสตร์ในแต่ละมิติที่สะท้อนถึงกระบวนการแก้ปัญหาของผู้เรียน จำเป็นต้องรอผลการตรวจคะแนนที่ยืนยันจากผู้ประเมิน เนื่องจากจำเป็นต้องมีครูหรือผู้ประเมิน เข้ามาร่วมตรวจสอบความถูกต้องของการแสดงวิธีทำ พร้อมสูตรหรือสัญลักษณ์ต่างๆ ที่ผู้เรียนใช้

1.2 ครูและผู้เกี่ยวข้องสามารถนำข้อมูลผลการประเมินระดับความสามารถทางคณิตศาสตร์ไปวินิจฉัยผู้เรียน เพื่อพัฒนาผู้เรียนให้มีระดับที่สูงขึ้น และเป็นการเพิ่มศักยภาพในการให้ข้อมูลย้อนกลับแก่ผู้เรียนในระบบสารสนเทศ และนำไปสู่การปรับปรุงการเรียนการสอนให้มีประสิทธิภาพสูงขึ้น

1.3 แนวทางการใช้งานระบบการตรวจให้คะแนนผ่านระบบ information technology assessment report (ITAR) โดยครูทำการเข้าระบบ ITAR ผ่านเว็บไซต์ www.itassess.com/site/login 201 และลงทะเบียน โดยการ login เข้าสู่ระบบ ตามด้วยการกรอกข้อมูล username และ password ซึ่งครูสามารถศึกษาเอกสารตามคู่มือการใช้งานได้

2. ข้อเสนอแนะในการทำวิจัยครั้งต่อไป

2.1 ในการวิจัยครั้งนี้เป็นการตรวจให้คะแนนแบบอัตโนมัติสำหรับแบบทดสอบอัตนัยเท่านั้น ดังนั้นการวิจัยครั้งต่อไปควรมีการพัฒนาเพิ่มการให้ข้อมูลย้อนกลับเพื่อการปรับปรุงและพัฒนาอย่างทันที่ ควรพิจารณาให้ครอบคลุมการให้ข้อมูลย้อนกลับทั้ง 3 ลักษณะ คือ การให้ข้อมูลกระตุ้นการเรียนรู้ (feed up), การให้ข้อมูลย้อนกลับ (feedback) และการให้ข้อมูลเพื่อการเรียนรู้ต่อยอด (feed forward)

2.2 ในการวิจัยครั้งต่อไปควรพัฒนาในส่วนการรายงานผลการประเมิน ที่เน้นการจัดทำรายงานผลที่สามารถนำไปใช้การพัฒนาผู้เรียนรายบุคคล เพื่อให้ข้อมูลสารสนเทศสำคัญว่าผู้เรียนมีความรู้และทักษะใดที่ผู้เรียนจำเป็นต้องได้รับการปรับปรุงพัฒนา โดยข้อมูลที่ได้เป็นประโยชน์ทั้งกับผู้เรียน ครู สถานศึกษา และเขตพื้นที่การศึกษา

References

- AERA, APA, & NCME. (2014). *Standards for Educational and Psychological Testing* (6th ed.). American Educational Research Association.
- Adams, R. J., Wilson, M., and Wang, W.C. (1997). The multidimensional random coefficients multinomial logit model. *Applied Psychological Measurement*, 21(1), 1-23.
- Berggren, S. J., Rama, T., and Ovreid, L. (2019). *Regression or classification? Automated Essay Scoring for Norwegian*. <https://www.aclweb.org/anthology/W19-4409.pdf>
- Black, P., and William, D. (1998). Inside the black box: raising standards through classroom assessment. *Phi Delta Kappan*, 8(2), 139-148.
- Demars, C. (2010). *Item Response Theory: Understanding Statistics Measurement*. Oxford University Press.

- European Language Resources Association (ELRA). (2020). *Proceedings of the 12th Conference on Language Resources and Evaluation (LREC 2020)*. Tokyo Metropolitan University.
- 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., Chaijunthuk, S., Suwannatrai, P., Chanayota, K., Pongboriboon, K., Tang, K. N., Wilson, M. (2020b). Validation of a digital tool for diagnosing mathematical proficiency. *International Journal of Evaluation and Research in Education (IJERE)*, 9(3), 665-674.
- Koyama, Kiyuna, Kobayashi, Arai, and Komachi. (2020). *Proceedings of the 12th conference on language resources and evaluation (LREC 2020)*. France.
- Nielsen, J. (1992). *Finding Usability Problems through Heuristic Evaluation*. Paper presented at the ACM CHI'92, Monterey, CA.
- Rodrigues, and Araújo. (2012, April). *Automatic assessment of short free text answers*. https://www.researchgate.net/profile/Fatima_Rodrigues3/publication/234023013_Automatic_Assessment_of_Short_Free_Text_Answers/links/552d8aa90cf2e089a3ad78af/Automatic-Assessment-of-Short-Free-Text-Answers.pdf.
- Wang, J., and Brown, M.S. (2007). *Automated Essay Scoring Versus Human Scoring: A Comparative Study*. *Journal of Technology, Learning, and Assessment* (2). <http://www.jtla.org>
- Wilson, M. (2005). *Constructing measures: An item response modeling approach*. Routledge.
- Wright, B. D., and Stone, M. H. (1979). *Best test design: Rasch measurement*. Mesa Press.
- Wu, Adams, Wilson, and Haldane. (2007). *ACER ConQuest version 2.0*. ACER Press.

Translate Thai References

- Aungkaseraneekul, S. (2012). Automated thai-language essay scoring. [Unpublished master's thesis]. Kasetsart University. (in Thai)
- Chinjunthuk, S., Junpeng, P. (2020). Assessment Guidelines for Student's Personalized Mathematical Proficiency Development. *Journal of Educational Measurement, Mahasarakram University*, 26(1), 47- 64. (in Thai)
- Jaihuek, S., and Mungsing, S. (2020). Scoring Thai Language Subjective Answer Automatic System by Sematic. *Information Technology Journal*, 16(1), 15-23. (in Thai)

- Junpeng, P., Marwiang, M., Chinjunthuk, S., Suwannatrai, P., Krotha, J., Chanayota, K., Tawarungruang, C., Thuanman, J., Tang K. N., and Wilson M. (2020a). *Developing Students' Mathematical Proficiency Level Diagnostic Tools through Information Technology in Assessment for Learning Report*. The Thailand Research Fund and Khon Kaen University. (in Thai)
- Suksiri, W. and Worain, C. (2016). *Investigating Tentative Cut scores for Science Learning Area on the Ordinary National Educational Test Scores using the Construct Mapping Method: An Analysis for Further Judgments*. National Institute of Educational Testing Service (Public Organization). (in Thai)
- The institute for the Promotion of Teaching Science and Technology (IPST). (2020). *PISA 2021 with assessment mathematical literacy*. <https://pisathailand.ipst.ac.th/issue-2020-53> (in Thai)
- Wongwanit, S. (2020). *Design Research in Education* (1st ed.). Chulalongkorn University Press. (in Thai)