



JEM-MSU

Journal of Educational Measurement
Mahasarakham University

ISSN 2985-0614 (Online)

วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม

Journal of Educational Measurement

Mahasarakham University

ปีที่ 32 ฉบับที่ 1 (มกราคม - มิถุนายน 2569)

Vol. 32 No. 1 (January-June 2026)



วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม

Journal of Educational Measurement Mahasarakham University

ปีที่ 32 ฉบับที่ 1: มกราคม - มิถุนายน 2569

Vol. 32 No. 1: January-June 2026

ISSN 2985-0614 (Online)

วารสารตีพิมพ์ปีละ 2 ฉบับ โดยมีกำหนดการพิมพ์เผยแพร่
ฉบับที่ 1: มกราคม - มิถุนายน
ฉบับที่ 2: กรกฎาคม - ธันวาคม

ภาควิชาวิจัยและพัฒนาการศึกษา

คณะศึกษาศาสตร์

มหาวิทยาลัยมหาสารคาม

วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม
ปีที่ 32 ฉบับที่ 1: มกราคม - มิถุนายน 2569
Journal of Educational Measurement Mahasarakham University
Vol. 32 No. 1: January-June 2026

Aim & Scope

กองบรรณาธิการจะรับพิจารณาบทความที่เกี่ยวข้องกับศาสตร์การวัดและประเมินผลการศึกษา การวิจัยทางการศึกษา สถิติและจิตวิทยาการศึกษาหรือที่เกี่ยวข้อง ทั้งนี้ ขึ้นอยู่กับดุลยพินิจของกองบรรณาธิการในการพิจารณา

วัตถุประสงค์

1. เพื่อส่งเสริมและเผยแพร่การศึกษาค้นคว้าการวิจัยและความรู้ใหม่ๆ ในสาขาวิชาการวัดผลและการวิจัยการศึกษา ตลอดจนวิชาอื่น ๆ ที่เกี่ยวข้อง
2. เพื่อเป็นสื่อแลกเปลี่ยนความคิดเห็นทางวิชาการ ทั้งทางทฤษฎีและการปฏิบัติระหว่างอาจารย์นักวิชาการและนิสิต
3. เพื่อส่งเสริมการเสนอผลงานทางวิชาการของอาจารย์และนิสิต
4. เพื่อเป็นเอกสารค้นคว้าอ้างอิงของอาจารย์ นิสิตและบุคลากร ผู้สนใจในความก้าวหน้าทางวิชาการ

ออกแบบปก

นางสาวอรุณภา คำบัวโคตร

ติดต่อวารสาร

ภาควิชาวิจัยและพัฒนาการศึกษา อาคารวิทยพัฒนา ชั้น 8 คณะศึกษาศาสตร์
มหาวิทยาลัยมหาสารคาม ถ.นครสวรรค์ ต.ตลาด อ.เมืองมหาสารคาม จ.มหาสารคาม 44000
โทรศัพท์: 043-754321-40 ต่อ 6272
E-mail: Jem2msu@gmail.com

บทความทุกบทความที่ตีพิมพ์ในวารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม
ผ่านการตรวจประเมินคุณภาพจากผู้ทรงคุณวุฒิจำนวน 3 ท่าน
ในรูปแบบ Double – Blinded Review

บทบรรณาธิการ

วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม ISSN 2985-0614 (Online) ผ่านการประเมินคุณภาพวารสารวิชาการในฐานข้อมูล TCI รอบที่ 5 พ.ศ. 2568 – 2572 โดยถูกจัดให้เป็นวารสารกลุ่มที่ 2 ทั้งนี้วารสารฯ ยังคงรักษาคุณภาพและมาตรฐานเพื่อให้ผู้อ่านได้รับประโยชน์สูงสุด

สำหรับวารสารปีที่ ปีที่ 32 ฉบับที่ 1: มกราคม - มิถุนายน 2569 มีบทความวิจัยเผยแพร่ทั้งสิ้น 14 เรื่อง โดยแบ่งเป็นบทความวิชาการ 1 เรื่อง และบทความวิจัย 13 เรื่อง ซึ่งประกอบด้วยบทความภาษาอังกฤษ 5 เรื่อง และบทความภาษาไทย 9 เรื่อง โดยบทความทุกเรื่องผ่านการพิจารณากลั่นกรองจากผู้ทรงคุณวุฒิจำนวน 3 ท่าน จากหลากหลายสถาบัน ในรูปแบบ Double – Blinded Review และทุกกระบวนการอยู่ภายใต้การควบคุมจรรยาบรรณอย่างเคร่งครัด

กองบรรณาธิการขอขอบพระคุณผู้เขียนทุกท่านที่ให้ความไว้วางใจส่งบทความมาพิจารณาเผยแพร่ในวารสาร ขอขอบพระคุณผู้ทรงคุณวุฒิที่ได้กรุณาให้ข้อเสนอแนะอย่างละเอียดและสร้างสรรค์ อันมีส่วนสำคัญต่อการยกระดับคุณภาพของบทความ

กองบรรณาธิการหวังเป็นอย่างยิ่งว่าวารสารฉบับนี้จะเป็นประโยชน์ทางวิชาการสำหรับผู้อ่านทุกท่าน หากท่านมีข้อเสนอแนะประการใด กองบรรณาธิการยินดีน้อมรับคำแนะนำเพื่อจะได้นำไปปรับปรุงและพัฒนาคุณภาพวารสารให้มีมาตรฐาน ยิ่งขึ้นต่อไป

ผู้ช่วยศาสตราจารย์ ดร.ทัศนศิริพันธ์ สว่างบุญ

บรรณาธิการ



ทีมบรรณาธิการ

ที่ปรึกษาทั่วไป

อธิการบดี มหาวิทยาลัยมหาสารคาม
คณบดีคณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม
หัวหน้าภาควิชาวิจัยและพัฒนการศึกษา

ที่ปรึกษากองบรรณาธิการ

ศ.ดร.ศิริชัย กาญจนวาสี
จุฬาลงกรณ์มหาวิทยาลัย
รศ.ดร.ประวิต เอราวรรณ
สำนักงานเลขาธิการสภาการศึกษา

รศ.สมนึก ภัททิยธนี
ข้าราชการเกษียณ

บรรณาธิการ

ผศ.ดร.ทัศน์ศิริรินทร์ สว่างบุญ
มหาวิทยาลัยมหาสารคาม

ผู้ช่วยบรรณาธิการ

อ.ดร.ศุภานดา ด้านกุลประเสริฐ
มหาวิทยาลัยมหาสารคาม

ฝ่ายประสานงาน

นางสาวเอมอร สิงห์พันธ์

กองบรรณาธิการ

รศ.ดร.ญาณภัทร สีหะมงคล	มหาวิทยาลัยมหาสารคาม
รศ.ดร.ทรงศักดิ์ ภูสีอ่อน	มหาวิทยาลัยมหาสารคาม
รศ.ดร.ประเสริฐ เรือนนระการ	มหาวิทยาลัยมหาสารคาม
รศ.ดร.วราพร เอราวรรณ	มหาวิทยาลัยมหาสารคาม
รศ.ดร.อัมพันธ์ พูลพุดา	มหาวิทยาลัยมหาสารคาม
อ.ดร.สาวิตรี ราญมิชัย	มหาวิทยาลัยมหาสารคาม
รศ.ดร.ณัฐภรณ์ หลาวทอง	จุฬาลงกรณ์มหาวิทยาลัย
รศ.ดร.ชนะศึก นิชาพันธ์	มหาวิทยาลัยสวนดุสิต
รศ.ดร.สมบัติ ห้ายเรือคำ	มหาวิทยาลัยเชียงใหม่
รศ.ดร.สุนทรพจน์ ดำรงพานิช	มหาวิทยาลัยเชียงใหม่
ผศ.ดร.ศักดิ์สิทธิ์ ฤทธิสัน	มหาวิทยาลัยกาฬสินธุ์
ผศ.ดร.สาธิตา สกลรัตน์กุลชัย	มหาวิทยาลัยธรรมศาสตร์
ผศ.ดร.พินดา วราสุนันท์	มหาวิทยาลัยเกษตรศาสตร์
รศ.ดร.จตุภูมิ เขตจัตุรัส	มหาวิทยาลัยขอนแก่น
ผศ.ดร.สุกัญญา รุจิเมธธาส	มหาวิทยาลัยราชภัฏอุดรดิตถ์
ผศ.ดร.สรัญญา จันทร์ชูสกุล	มหาวิทยาลัยศิลปากร
รศ.ดร.กมลทิพย์ ศรีหาเศษ	มหาวิทยาลัยรามคำแหง
อ.ดร.ชลี ภัทรพิชญธรรม Southern Methodist University, USA	
ผศ.ดร.อรนุช วราอศวปติ ศรีสะอาด ข้าราชการเกษียณ	

สารบัญ

บทความ	หน้า
บทความวิชาการ	
<i>การประยุกต์ใช้ปัญญาประดิษฐ์เชิงสร้างสรรค์ในการพัฒนาเครื่องมือวิจัยทางการศึกษา: กรอบแนวคิดเชิงวิธีวิทยา</i>	
<i>Applying Generative Artificial Intelligence in Educational Research Instrument Development: A Methodological Framework</i>	
ปวีศ สารมะโน.....	1
บทความวิจัยภาษาอังกฤษ	
<i>Factors Correlated with Digital Competency Among Youth in the 21st Century: A Meta-Analysis</i>	
<i>Somkiat Kaewkohsaba and Sirirut Jumnaksarn.....</i>	25
<i>Establishing CEFR-Referenced Cut Scores for the Royal Police Cadet Academy Test of English Proficiency (RPCA-TEP) Using the Modified Angoff Method</i>	
<i>Komjit Chaiyo and Pitupoom Nakprada.....</i>	40
<i>Psychometric Evaluation of the Emotion Regulation Instrument for Teacher Education Institution Students Utilizing AI in Scientific Writing</i>	
<i>Rizki Mustikasari, Ratri Harida and Cutiana Windri Astuti.....</i>	51
<i>Development and Validation of an Academic Engagement Scale for Undergraduate Students in Distance Learning</i>	
<i>Supunsa Langprasert and Pisutthipa Metheekul.....</i>	71
<i>Development Guidelines for Lifelong Learning Among Retirees in Guangdong Province, People's Republic of China: A Practitioner-oriented Inquiry</i>	
<i>Zhou Hong, Teerawat Monthaisong and Wacharapatr Techawattanasiridumrong.....</i>	86

สารบัญ (ต่อ)

บทความ

หน้า

บทความวิจัยภาษาไทย

บทบาทการส่งผ่านของความไว้วางใจระหว่างภาวะผู้นำของผู้บริหารกับความพึงพอใจในการทำงานและการสอนของครู : หลักฐานเชิงประจักษ์จากข้อมูล PISA 2022 ของฮ่องกง <i>The Mediating Role of Trust in the Relationship between Principal Leadership and Teachers' Job and Teaching Satisfaction: Evidence from Hong Kong's 2022 PISA Data</i> ธัญชนก ทองนพแก้ว, สุภชิต ผดุงผล, อีรณันท์ ศรีวิทัศน์ และณรัตน์ ศิริภาพ.....	103
การพัฒนดัชนีสถานการณ์ความเสี่ยงทางการศึกษาที่เชื่อมโยงกับสภาพภูมิอากาศของประเทศไทย <i>Development of a Climate-Mapped Educational Risk Index for Thailand</i> สิวะโชติ ศรีสุทธยากร, กนิษฐ์ ศรีเคลือบ, ประภาศิริ รัชชประภาพรกุล และ วาทีณี อมรไพศาลเลิศ.....	119
การประเมินอภิมานรายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย <i>Meta-evaluation of Final Evaluation Reports by Save the Children Australia</i> วรวรรณ จิรธนาพิวัฒน์, ศิวะพร ภูพันธ์ และ อุทัยทิพย์ เจียวิวรรณกุล.....	138
การศึกษาสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 โรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษาพิจิตร <i>The Study of Competency-Based Learning Assessment Problems of Grade 6 Mathematics Teachers in Schools under the Kalasin Primary Educational Service Area Office</i> ศักดิ์สิทธิ์ ฤทธิลัน, อมร มะลาศรี, คมสันที ขจรปัญญาไพศาล, สุวรรณวัฒน์ เทียนนุทธกุล และ สมใจ ภูครองทุ่ง.....	155
การพัฒนาครูเชิงระบบเพื่อยกระดับสมรรถนะความฉลาดรู้ด้านวิทยาศาสตร์ของนักเรียน ชั้นประถมศึกษาปีที่ 6 ตามกรอบการประเมิน PISA 2025 ด้วยรูปแบบ PCACK-4D: การวิจัย เชิงวินิจฉัยและวิเคราะห์ความต้องการจำเป็น <i>Systemic Teacher Development for Enhancing Scientific Literacy Competency of Grade 6 Students Based on the PISA 2025 Framework Using the PCACK-4D Model: Diagnostic Research and Needs Assessment</i> จตุพล แสนสุข และ สติติ คลาดโรค.....	172

สารบัญ (ต่อ)

บทความ	หน้า
การพัฒนาแบบวัดสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ เรื่อง คลื่นเสียง สำหรับนักเรียน ระดับมัธยมศึกษาตอนปลาย <i>Development of a Scientific Explanation Competency Test on Sound Waves for Upper Secondary School Students</i> กฤษฎา ทองประไฟ.....	187
การพัฒนาเครื่องมือวัดความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK ของครู มัธยมศึกษาในจังหวัดพิษณุโลก <i>Development of an Instrument to Measure Artificial Intelligence Technological Pedagogical Content Knowledge (AI-TPACK) Competencies of Secondary School Teachers in Phitsanulok Province</i> ธีรลักษณ์ จันทร์ดี และ ณัฐกานต์ ประจันบาน.....	203
การประเมินโครงการคณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชน กรณีศึกษาโรงเรียนบ้านหนองเล <i>Evaluation of Mathematics for Life and Community through the Sufficiency Economy Project Case Study of Ban Nongle School</i> วริศรา วงศ์หล่มแก้ว และ สายฝน วิบูลรังสรรค์.....	218
ภาคผนวก.....	233
รายชื่อผู้ทรงคุณวุฒิพิจารณาบทความ วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม ปีที่ 32 ฉบับที่ 1: มกราคม - มิถุนายน 2569.....	234
หลักเกณฑ์และคำแนะนำสำหรับผู้เขียน.....	235



บทความวิชาการ

Academic Article

การประยุกต์ใช้ปัญญาประดิษฐ์เชิงสร้างสรรค์ในการพัฒนาเครื่องมือวิจัยทางการศึกษา: กรอบแนวคิดเชิงวิธีวิทยา

Applying Generative Artificial Intelligence in Educational Research Instrument Development: A Methodological Framework

ปาวริศ สารมะโน^{1*}

Pawaris Saramano^{1*}

(Received: February 3, 2026; Revised: April 25, 2026; Accepted: April 27, 2026)

บทคัดย่อ

การพัฒนาเครื่องมือวิจัยทางการศึกษาเป็นกระบวนการสำคัญที่ส่งผลโดยตรงต่อคุณภาพความน่าเชื่อถือ และความหมายของผลการวิจัย อย่างไรก็ตาม นักวิจัยจำนวนมากยังประสบปัญหาในการพัฒนาเครื่องมือวิจัยที่มีความสอดคล้องเชิงทฤษฎีและมีคุณภาพด้านการวัดและประเมินผล ขณะเดียวกัน ความก้าวหน้าของปัญญาประดิษฐ์เชิงสร้างสรรค์ (Generative Artificial Intelligence: GenAI) โดยเฉพาะเครื่องมือที่อาศัยแบบจำลองภาษาขนาดใหญ่ (Large Language Models: LLMs) และการประมวลผลภาษาธรรมชาติ (Natural Language Processing: NLP) ได้เปิดโอกาสใหม่ในการสนับสนุนการนิยามตัวแปร การกำหนดตัวชี้วัด การร่างข้อคำถาม และการตรวจสอบความชัดเจนของภาษาเบื้องต้น แต่การประยุกต์ใช้เทคโนโลยีดังกล่าวในการพัฒนาเครื่องมือวิจัยยังขาดกรอบเชิงวิธีวิทยา (methodological framework) ที่เชื่อมโยงกับหลักการวัดและประเมินผลทางการศึกษาอย่างเป็นระบบ

บทความเชิงแนวคิด (conceptual paper) นี้มีวัตถุประสงค์เพื่อเสนอกรอบแนวคิดเชิงวิธีวิทยาสำหรับการประยุกต์ใช้ปัญญาประดิษฐ์เชิงสร้างสรรค์ในการพัฒนาเครื่องมือวิจัยทางการศึกษา โดยบูรณาการ 4 องค์ประกอบสำคัญ ได้แก่ ฐานแนวคิดการวัดและประเมินผลทางการศึกษา กระบวนการพัฒนาเครื่องมือวิจัย บทบาทของปัญญาประดิษฐ์ในฐานะกลไกสนับสนุนเชิงวิธีวิทยา และการกำกับการตัดสินใจโดยมนุษย์ (human-in-the-loop) กรอบที่นำเสนอวางตำแหน่งของปัญญาประดิษฐ์ให้เป็นผู้ช่วยเชิงวิธีวิทยา มิใช่ผู้แทนการตัดสินใจทางวิชาการ โดยสนับสนุนการกำหนดโครงสร้างเชิงแนวคิด การพัฒนาตัวชี้วัด การสร้างข้อคำถามเบื้องต้น และการตรวจสอบคุณภาพเบื้องต้น ขณะที่การตัดสินใจเกี่ยวกับความตรง ความเที่ยง ความแม่นยำของการวัด ความเป็นธรรมของการวัด และความเหมาะสมเชิงจริยธรรมยังคงเป็นหน้าที่ของผู้วิจัยและผู้เชี่ยวชาญ

¹ อาจารย์ สาขาวิชาวิจัยและพัฒนาหลักสูตร คณะครุศาสตร์ มหาวิทยาลัยราชภัฏมหาสารคาม

¹ Lecturer, Research and Curriculum Development Program, Faculty of Education, Mahasarakham Rajabhat University

* Corresponding Author E-mail: Pawaris.sa@rmu.ac.th

กรอบแนวคิดดังกล่าวช่วยทำให้การประยุกต์ใช้ปัญญาประดิษฐ์ในการพัฒนาเครื่องมือวิจัยมีความเป็นระบบ โปร่งใส และตรวจสอบได้มากขึ้น อีกทั้งช่วยเน้นย้ำว่า การใช้ปัญญาประดิษฐ์ในงานพัฒนาเครื่องมือวิจัยควรตั้งอยู่บนหลักการวัดและประเมินผลที่มั่นคง ควบคู่กับการกำกับโดยมนุษย์อย่างรับผิดชอบ บทความนี้จึงมีสถานะเป็นข้อเสนอเชิงวิวิธวิทยาเบื้องต้นสำหรับนักวิจัยทางการศึกษา ผู้สอนวิจัย และนักศึกษาระดับบัณฑิตศึกษา และเป็นฐานสำหรับการตรวจสอบเชิงประจักษ์ของกรอบดังกล่าวในงานวิจัยต่อไป

คำสำคัญ: ปัญญาประดิษฐ์เชิงสร้างสรรค์ แบบจำลองภาษาขนาดใหญ่ การพัฒนาเครื่องมือวิจัยทางการศึกษา การวัดและประเมินผลทางการศึกษา กรอบแนวคิดเชิงวิวิธวิทยา

Abstract

The development of educational research instruments is critical to the quality, credibility, and interpretability of research findings. However, many researchers continue to face challenges in constructing instruments that are both theoretically grounded and methodologically sound. At the same time, recent advances in Generative Artificial Intelligence (GenAI), particularly tools based on Large Language Models (LLMs) and Natural Language Processing (NLP), have created new opportunities to support construct definition, indicator specification, preliminary item drafting, and early-stage language review. Nevertheless, the application of these technologies to educational research instrument development still lacks a clearly articulated methodological framework grounded in educational measurement and evaluation principles.

This conceptual paper proposes a methodological framework for applying Generative Artificial Intelligence to educational research instrument development. The framework integrates four core components: educational measurement and evaluation foundations, instrument development processes, the role of Artificial Intelligence as a methodological support mechanism, and Human-in-the-Loop decision-making. Within this framework, AI is positioned as a methodological assistant rather than a substitute for scholarly judgment. It may support construct clarification, indicator development, preliminary item generation, and early-stage quality review, while responsibility for decisions concerning validity, reliability, measurement precision, measurement fairness, and ethical appropriateness remains with researchers and domain experts.

The proposed framework offers a structured, transparent, and theoretically grounded approach to integrating AI into instrument development. It further emphasizes that the academic value of AI-assisted instrument development depends not merely on procedural efficiency, but on its alignment with sound measurement and evaluation principles and

responsible human oversight. This paper thus provides a preliminary methodological contribution for educational researchers, research instructors, and graduate students, while also offering a foundation for future empirical validation of AI-assisted instrument development practices in education.

Keywords: Generative Artificial Intelligence, large language models, educational research instrument development; educational measurement and evaluation; methodological framework

1. บทนำ

การพัฒนาเครื่องมือวิจัยทางการศึกษาเป็นกระบวนการสำคัญที่ส่งผลโดยตรงต่อคุณภาพความน่าเชื่อถือ และความสมเหตุสมผลของผลการวิจัย เครื่องมือที่มีคุณภาพจำเป็นต้องสะท้อนโครงสร้างเชิงทฤษฎีของตัวแปรที่ศึกษา มีความตรง (validity) และความเที่ยง (reliability) ที่เหมาะสม ตลอดจนสามารถนำไปใช้วัดพฤติกรรมหรือคุณลักษณะของกลุ่มเป้าหมายได้อย่างถูกต้องและเป็นธรรม (Cohen et al., 2018; Ebel & Frisbie, 1991) อย่างไรก็ตาม ในทางปฏิบัติ นักวิจัยจำนวนไม่น้อย โดยเฉพาะบุคลากรทางการศึกษา นักศึกษาระดับบัณฑิตศึกษา และนักวิจัยมือใหม่ มักประสบปัญหาในการออกแบบและพัฒนาเครื่องมือวิจัยที่มีคุณภาพ ทั้งในด้านการนิยามตัวแปรเชิงปฏิบัติการ การสร้างข้อคำถามให้สอดคล้องกับโครงสร้างเชิงทฤษฎี และการตรวจสอบคุณภาพของเครื่องมือก่อนนำไปใช้จริง

ในช่วงไม่กี่ปีที่ผ่านมา ความก้าวหน้าของปัญญาประดิษฐ์เชิงสร้างสรรค์ โดยเฉพาะเครื่องมือที่อาศัยแบบจำลองภาษาขนาดใหญ่และการประมวลผลภาษาธรรมชาติ ได้ทำให้บทบาทของปัญญาประดิษฐ์ในกระบวนการวิจัยขยับจากการสนับสนุนการวิเคราะห์ข้อมูลในระยะท้ายของกระบวนการวิจัย ไปสู่การสนับสนุนงานในระยะต้นของกระบวนการวิจัยมากขึ้น เช่น การสังเคราะห์แนวคิด การนิยามตัวแปร การกำหนดตัวชี้วัด การร่างข้อคำถามเบื้องต้น และการตรวจสอบความชัดเจนของภาษา (Kasneci et al., 2023; UNESCO, 2023) อย่างไรก็ตาม ศักยภาพของเครื่องมือดังกล่าวมิได้คงที่ตายตัว หากแปรผันตามแพลตฟอร์ม โมเดล รุ่น และช่วงเวลาที่ใช้งาน จึงทำให้การอภิปรายบทบาทของปัญญาประดิษฐ์ในงานวิจัยจำเป็นต้องมีความเฉพาะเจาะจง โปร่งใส และระมัดระวังทางวิชาการมากกว่าที่ผ่านมา (Flanagin et al., 2024)

ในบริบทของประเทศไทย การประยุกต์ใช้ปัญญาประดิษฐ์มิได้เป็นเพียงแนวโน้มทางเทคโนโลยีระดับสากล หากแต่เชื่อมโยงกับทิศทางการพัฒนากำลังคน ระบบนิเวศการวิจัย และการยกระดับศักยภาพทางวิชาการของประเทศตามกรอบยุทธศาสตร์ปัญญาประดิษฐ์ระดับชาติด้วย (AI Thailand, 2022) ในภาคการศึกษาและการวิจัย เครื่องมือปัญญาประดิษฐ์เชิงสร้างสรรค์เริ่มเข้ามามีบทบาทมากขึ้นในการสืบค้นสังเคราะห์ เรียบเรียง และตรวจสอบสาระทางวิชาการเบื้องต้น แต่การใช้งานดังกล่าวยังเผชิญคำถามสำคัญเกี่ยวกับความน่าเชื่อถือของผลลัพธ์ ความโปร่งใสในการรายงานการใช้เครื่องมือ ความเสี่ยงด้านอคติ และขอบเขตความรับผิดชอบของผู้วิจัย (Flanagin et al., 2024; UNESCO, 2023) ประเด็นเหล่านี้ยิ่งมีความสำคัญ

มากขึ้นเมื่อปัญญาประดิษฐ์ถูกนำมาใช้ในขั้นตอนการพัฒนาเครื่องมือซึ่งเป็นขั้นตอนที่เกี่ยวข้องโดยตรงกับความหมายของการวัดและคุณภาพของข้อสรุปจากงานวิจัย

อย่างไรก็ดี บทบาทของปัญญาประดิษฐ์ในกระบวนการวิจัยทางการศึกษาส่วนใหญ่ยังคงมุ่งเน้นอยู่ที่ขั้นตอนหลังการเก็บข้อมูล เช่น การวิเคราะห์ข้อมูลเชิงปริมาณและเชิงคุณภาพ การจำแนกกลุ่มข้อมูล หรือการคาดการณ์ผลลัพธ์ ขณะที่การประยุกต์ใช้ปัญญาประดิษฐ์ในระยะต้นของกระบวนการวิจัย โดยเฉพาะการสร้างและพัฒนาเครื่องมือวิจัย ยังได้รับการศึกษาและนำเสนอในเชิงวิชาการค่อนข้างจำกัด ทั้งที่ขั้นตอนดังกล่าวถือเป็นรากฐานสำคัญของคุณภาพงานวิจัยโดยรวม (Downing, 2006; Haladyna & Rodriguez, 2013) ยิ่งไปกว่านั้น งานเขียนจำนวนหนึ่งยังกล่าวถึง “ปัญญาประดิษฐ์” ในลักษณะกว้าง โดยมีได้แยกให้ชัดว่าหมายถึงระบบประเภทใด ใช้เพื่อวัตถุประสงค์ใด และมีข้อจำกัดอย่างไร ส่งผลให้การอภิปรายศักยภาพของปัญญาประดิษฐ์ในการพัฒนาเครื่องมือวิจัยยังขาดความแม่นยำตรงเชิงวิธีวิทยาและยากต่อการตรวจสอบซ้ำในเชิงวิชาการ

ในบทความนี้ ผู้เขียนจึงใช้คำว่า “ปัญญาประดิษฐ์” ในความหมายที่เฉพาะเจาะจงยิ่งขึ้น กล่าวคือ มุ่งอภิปรายปัญญาประดิษฐ์เชิงสร้างสรรค์ที่อาศัยแบบจำลองภาษาขนาดใหญ่และเครื่องมือประมวลผลภาษาธรรมชาติ ซึ่งสามารถสนับสนุนการทำงานกับภาษา ข้อความ และโครงสร้างเชิงแนวคิดในกระบวนการพัฒนาเครื่องมือวิจัยได้โดยตรง บทความฉบับนี้มีได้มุ่งเปรียบเทียบประสิทธิภาพของระบบปัญญาประดิษฐ์รายแพลตฟอร์มหรือรายเวอร์ชัน หากแต่มุ่งเสนอกรอบแนวคิดเชิงวิธีวิทยาสำหรับการใช้เทคโนโลยีกลุ่มดังกล่าวอย่างเป็นระบบ ภายใต้หลักการด้านการวัดและประเมินผลทางการศึกษา (Flanagin et al., 2024)

ในบริบทนี้ การประยุกต์ใช้ปัญญาประดิษฐ์เพื่อสนับสนุนกระบวนการพัฒนาเครื่องมือวิจัยทางการศึกษา ไม่ควรถูกมองเพียงในฐานะเครื่องมืออัตโนมัติที่เข้ามาแทนที่การตัดสินใจของนักวิจัย หากแต่ควรถูกพิจารณาในฐานะกลไกสนับสนุนเชิงวิธีวิทยา (methodological support mechanism) ที่ช่วยเสริมสร้างความเป็นระบบ ความสอดคล้องทางทฤษฎี และความโปร่งใสของกระบวนการพัฒนาเครื่องมือวิจัย ภายใต้หลักการสำคัญที่ยังคงยึดผู้วิจัยเป็นผู้รับผิดชอบต่อการตัดสินใจเชิงวิชาการและจริยธรรมของงานวิจัย กล่าวอีกนัยหนึ่ง ปัญญาประดิษฐ์อาจมีบทบาทในการช่วยสังเคราะห์นิยามของโครงสร้างเชิงแนวคิด ตรวจสอบความครอบคลุมของตัวชี้วัด เสนอข้อคำถามเบื้องต้น หรือสะท้อนความกำกวมของภาษา แต่ไม่อาจทำหน้าที่แทนการตัดสินใจคุณภาพของการวัด การยืนยันความตรงของเครื่องมือ หรือการรับผิดชอบต่อการใช้คะแนนที่เกิดขึ้นจากเครื่องมือดังกล่าวได้ (UNESCO, 2023)

ด้วยเหตุนี้ บทความนี้จึงตั้งอยู่บนข้อพิจารณาสำคัญ 2 ประการ คือ ประการแรก ศักยภาพของปัญญาประดิษฐ์เชิงสร้างสรรค์และแบบจำลองภาษาขนาดใหญ่ทำให้การใช้ปัญญาประดิษฐ์ในการพัฒนาเครื่องมือวิจัยมีความเป็นไปได้มากขึ้นอย่างมีนัยสำคัญ และประการที่สอง การใช้ศักยภาพดังกล่าวอย่างมีความหมายทางวิชาการจำเป็นต้องอาศัยกรอบเชิงวิธีวิทยาที่เชื่อมเทคโนโลยีเข้ากับหลักการวัดและประเมินผลอย่างเป็นระบบ ดังนั้น บทความนี้จึงมีวัตถุประสงค์เพื่อเสนอกรอบแนวคิดเชิงวิธีวิทยาสำหรับการประยุกต์ใช้ปัญญาประดิษฐ์ในการพัฒนาเครื่องมือวิจัยทางการศึกษา โดยมุ่งเน้นปัญญาประดิษฐ์เชิงสร้างสรรค์ที่อาศัยแบบจำลองภาษาขนาดใหญ่และการประมวลผลภาษาธรรมชาติเป็นสำคัญ ภายใต้กรอบแนวคิดการวัดและประเมินผลทางการศึกษาที่ได้รับการยอมรับในระดับสากล กรอบดังกล่าวมุ่งอธิบายบทบาทของ

ปัญญาประดิษฐ์ในฐานะผู้ช่วยเชิงวิธีวิทยาในแต่ละขั้นตอนของการพัฒนาเครื่องมือ ตั้งแต่การนิยามตัวแปร การกำหนดตัวชี้วัด การออกแบบข้อคำถาม ไปจนถึงการตรวจสอบคุณภาพเบื้องต้นของเครื่องมือ โดยยังคงให้การตัดสินใจเชิงวิชาการและจริยธรรมเป็นความรับผิดชอบของผู้วิจัยและผู้เชี่ยวชาญ

2. ช่องว่างของการวิจัย

จากการทบทวนวรรณกรรมพบว่า งานวิจัยเกี่ยวกับการประยุกต์ใช้ปัญญาประดิษฐ์ในงานวิจัยทางการศึกษาส่วนใหญ่ให้ความสำคัญกับการใช้ปัญญาประดิษฐ์ในการวิเคราะห์ข้อมูล การประเมินผลการเรียนรู้ และการสนับสนุนการเรียนการสอน (Holmes et al., 2019; Zawacki-Richter et al., 2019) แม้ว่าจะมีงานดังกล่าวจะช่วยอธิบายบทบาทของปัญญาประดิษฐ์ในแวดวงการศึกษได้ในระดับหนึ่ง แต่เมื่อพิจารณาภายใต้บริบทของการเปลี่ยนแปลงอย่างรวดเร็วในช่วงหลังการเข้ามาของปัญญาประดิษฐ์เชิงสร้างสรรค์ โดยเฉพาะเครื่องมือที่อาศัยแบบจำลองภาษาขนาดใหญ่และการประมวลผลภาษาธรรมชาติ จะเห็นได้ว่าช่องว่างที่เคยปรากฏในงานศึกษาก่อนหน้านี้จำเป็นต้องได้รับการทบทวนใหม่ ทั้งนี้ เพราะบทบาทของปัญญาประดิษฐ์มิได้จำกัดอยู่เพียงการช่วยวิเคราะห์หรือจัดการข้อมูลในระยะท้ายของการวิจัยอีกต่อไป หากแต่ขยายเข้าสู่ระยะต้นของการพัฒนาองค์ความรู้และการออกแบบงานวิจัย เช่น การสังเคราะห์แนวคิด การนิยามตัวแปร การกำหนดตัวชี้วัด การร่างข้อคำถามเบื้องต้น และการตรวจสอบความชัดเจนของภาษา (Kasneci et al., 2023; UNESCO, 2023)

อย่างไรก็ตาม แม้เครื่องมือปัญญาประดิษฐ์เชิงสร้างสรรค์จะเปิดโอกาสใหม่ให้นักวิจัยทางการศึกษาอย่างมีนัยสำคัญ แต่งานวิชาการที่มุ่งศึกษาการใช้ปัญญาประดิษฐ์ในกระบวนการพัฒนาเครื่องมือวิจัยยังมีจำนวนจำกัด และมักนำเสนอในลักษณะการใช้เครื่องมือเฉพาะกิจ เช่น การช่วยร่างข้อคำถาม การสรุปนิยามของตัวแปร หรือการตรวจสอบภาษาเบื้องต้น โดยยังขาดกรอบเชิงวิธีวิทยาที่ชัดเจนสำหรับอธิบายว่า ปัญญาประดิษฐ์ควรเข้ามามีบทบาทในขั้นตอนใด ภายใต้เงื่อนไขใด และด้วยขอบเขตความรับผิดชอบระดับใดของผู้วิจัย กล่าวอีกนัยหนึ่ง งานที่มีอยู่ยังให้ภาพของการใช้ปัญญาประดิษฐ์เพื่อช่วยทำงานมากกว่าการอธิบายการใช้ปัญญาประดิษฐ์อย่างมีระบบภายใต้ตรรกะของการพัฒนาเครื่องมือวิจัย

ประการแรก งานวิจัยส่วนใหญ่ยังไม่เชื่อมโยงบทบาทของปัญญาประดิษฐ์เข้ากับกรอบแนวคิดด้านการวัดและประเมินผลทางการศึกษาอย่างเป็นระบบ เช่น ความตรงเชิงเนื้อหา ความตรงเชิงโครงสร้าง หรือหลักฐานความตรงตามแนวคิดความตรงแบบบูรณาการ ทำให้การใช้ปัญญาประดิษฐ์ในการพัฒนาเครื่องมือยังขาดฐานทฤษฎีที่มั่นคง (Kane, 2013; Messick, 1995) แม้ปัญญาประดิษฐ์จะสามารถช่วยเสนอคำอธิบายของโครงสร้างเชิงแนวคิด จัดกลุ่มตัวชี้วัด หรือร่างข้อคำถามเบื้องต้นได้ แต่การตัดสินใจว่าเนื้อหาดังกล่าวสะท้อนสิ่งที่มุ่งวัดอย่างครอบคลุมหรือไม่ ยังคงต้องอาศัยหลักการทางการวัดและประเมินผล รวมทั้งการพิจารณาของผู้วิจัยและผู้เชี่ยวชาญเป็นสำคัญ งานเขียนจำนวนมากยังไม่ได้อธิบายอย่างชัดเจนว่าผลลัพธ์จากปัญญาประดิษฐ์จะถูกนำมาตรวจสอบหรือกลั่นกรองให้สอดคล้องกับคุณภาพของการวัดอย่างไร จึงทำให้การประยุกต์ใช้ปัญญาประดิษฐ์ในขั้นตอนนี้ยังมีลักษณะเป็นการใช้งานเชิงเทคนิคมากกว่าการใช้งานเชิงวิธีวิทยา

ประการที่สอง งานศึกษาหลายชิ้นยังกล่าวถึง “ปัญญาประดิษฐ์” ในลักษณะกว้าง โดยมิได้แยกให้ชัดว่าหมายถึงระบบประเภทใด โมเดลใด หรือใช้เพื่อจุดประสงค์ใดในกระบวนการพัฒนาเครื่องมือวิจัย ภายใต้

บริบทของปัญญาประดิษฐ์เชิงสร้างสรรค์ ประเด็นนี้มีความสำคัญอย่างยิ่ง เพราะความสามารถ ข้อจำกัด และความเสี่ยงของระบบแต่ละแพลตฟอร์มอาจแตกต่างกันมากตามรุ่น โมเดล และช่วงเวลาที่ใช้งาน เมื่อไม่มีการกำหนดขอบเขตของปัญญาประดิษฐ์อย่างเฉพาะเจาะจง การอภิปรายศักยภาพของปัญญาประดิษฐ์ในการพัฒนาเครื่องมือวิจัยจึงเสี่ยงต่อการสรุปกว้างเกินฐานหลักฐาน และทำให้ยากต่อการตรวจสอบซ้ำในเชิงวิชาการ (Flanagin et al., 2024)

ประการที่สาม แม้งานบางส่วนจะเสนอให้ปัญญาประดิษฐ์มีบทบาทในฐานะผู้ช่วยด้านการวิจัย แต่ยังมีข้อจำกัดในการอธิบายบทบาทของนักวิจัยในฐานะผู้กำกับดูแลและตัดสินใจเชิงวิชาการ ทำให้การใช้ปัญญาประดิษฐ์เสี่ยงต่อการพึ่งพาเครื่องมือโดยขาดการตรวจสอบอย่างเพียงพอ (Floridi et al., 2018; UNESCO, 2023) ในบริบทของการพัฒนาเครื่องมือวิจัย ประเด็นนี้มีความสำคัญเป็นพิเศษ เพราะการตัดสินใจเกี่ยวกับการนิยามสิ่งที่ต้องการวัด การคัดเลือกตัวชี้วัด ความเหมาะสมของข้อคำถาม ความเป็นธรรมของการวัด และการตีความผลจากเครื่องมือ ล้วนเป็นการตัดสินใจที่มีนัยเชิงวิชาการและจริยธรรมสูง ดังนั้น แม้ปัญญาประดิษฐ์จะช่วยเสนอทางเลือกหรือสะท้อนข้อสังเกตเบื้องต้นได้ แต่ยังไม่อาจทำหน้าที่แทนการตัดสินใจเชิงวิชาการของมนุษย์ได้โดยสมบูรณ์

ประการที่สี่ เมื่อพิจารณาในบริบทของนักวิจัยทางการศึกษาในประเทศไทย ยังไม่พบบทความเชิงแนวคิดที่เสนอกรอบการประยุกต์ใช้ปัญญาประดิษฐ์เชิงสร้างสรรค์หรือเครื่องมือที่อาศัยแบบจำลองภาษาขนาดใหญ่ในการพัฒนาเครื่องมือวิจัยอย่างเป็นระบบ โดยเชื่อมโยงทั้งมิติของการวัดและประเมินผล กระบวนการพัฒนาเครื่องมือ ความโปร่งใสในการรายงานการใช้ปัญญาประดิษฐ์ และบทบาทการกำกับของผู้วิจัยเข้าด้วยกัน ช่องว่างดังกล่าวมีนัยสำคัญ เนื่องจากนักวิจัยไทยจำนวนมากเริ่มเข้าถึงและใช้งานเครื่องมือปัญญาประดิษฐ์ในงานวิชาการมากขึ้น แต่ยังขาดกรอบอ้างอิงเชิงวิธีวิทยาที่ช่วยให้การใช้เทคโนโลยีเหล่านี้สอดคล้องกับหลักการทางวิชาการและจริยธรรมของการวิจัยทางการศึกษา

จากช่องว่างดังกล่าว จะเห็นได้ว่า ประเด็นสำคัญมิได้อยู่ที่การมีหรือไม่มีปัญญาประดิษฐ์ในกระบวนการพัฒนาเครื่องมือวิจัย หากแต่อยู่ที่การขาดกรอบเชิงวิธีวิทยาที่สามารถเชื่อมศักยภาพของปัญญาประดิษฐ์เชิงสร้างสรรค์เข้ากับหลักการวัดและประเมินผลทางการศึกษา กระบวนการพัฒนาเครื่องมือวิจัย และการกำกับการตัดสินใจโดยมนุษย์อย่างเป็นระบบ ด้วยเหตุนี้ บทความนี้จึงมุ่งเติมเต็มช่องว่างดังกล่าวโดยเสนอกรอบแนวคิดเชิงวิธีวิทยาสำหรับการประยุกต์ใช้ปัญญาประดิษฐ์ในการพัฒนาเครื่องมือวิจัยทางการศึกษา ซึ่งวางตำแหน่งปัญญาประดิษฐ์ในฐานะกลไกสนับสนุนเชิงวิธีวิทยา ภายใต้ฐานแนวคิดด้านการวัดและประเมินผลที่ชัดเจน และภายใต้หลักการที่ผู้วิจัยยังคงเป็นผู้รับผิดชอบต่อการตัดสินใจเชิงวิชาการและจริยธรรมของกระบวนการทั้งหมด

3. กรอบแนวคิดเชิงวิธีวิทยา

กรอบแนวคิดเชิงวิธีวิทยาสำหรับการประยุกต์ใช้ปัญญาประดิษฐ์ในการพัฒนาเครื่องมือวิจัยทางการศึกษาที่นำเสนอในบทความนี้ ตั้งอยู่บนหลักการสำคัญว่า การพัฒนาเครื่องมือวิจัยมิใช่เพียงการจัดทำชุดข้อคำถามหรืองานประเมิน หากแต่เป็นกระบวนการเชิงระบบที่ต้องอาศัยความสอดคล้องระหว่างฐานทฤษฎี

ของสิ่งที่มีงูวัด หลักการวัดและประเมินผล กระบวนการพัฒนาเครื่องมือ และการตัดสินใจเชิงวิชาการของผู้วิจัยอย่างต่อเนื่อง ภายใต้กรอบนี้ ปัญญาประดิษฐ์เชิงสร้างสรรค์ โดยเฉพาะเครื่องมือที่อาศัยแบบจำลองภาษาขนาดใหญ่และการประมวลผลภาษาธรรมชาติ ถูกวางตำแหน่งให้เป็นกลไกสนับสนุนเชิงวิธีวิทยา มากกว่าจะเป็นผู้แทนการตัดสินใจคุณภาพของการวัดโดยอัตโนมัติ

การพัฒนากรอบนี้ตั้งอยู่บนความพยายามที่จะเชื่อม 4 มิติที่มักถูกอภิปรายแยกส่วนในวรรณกรรมให้ทำงานร่วมกันอย่างเป็นระบบ ได้แก่ (1) ฐานแนวคิดการวัดและประเมินผลทางการศึกษา (2) กระบวนการพัฒนาเครื่องมือวิจัย (3) บทบาทของปัญญาประดิษฐ์ในฐานะเครื่องมือสนับสนุนเชิงวิธีวิทยา และ (4) การกำกับการตัดสินใจโดยมนุษย์ ทั้ง 4 มิตินี้มีได้ทำงานแบบเรียงลำดับเชิงกลไก หากแต่เชื่อมโยงกันในลักษณะระบบเชิงปฏิสัมพันธ์ที่ต้องอาศัยการสะท้อนกลับ การทบทวน และการตัดสินใจเชิงวิชาการในทุกช่วงของการพัฒนาเครื่องมือ

3.1 ฐานคิดเชิงวิธีวิทยา

กรอบแนวคิดนี้ตั้งอยู่บนฐานคิดเชิงวิธีวิทยา 3 ประการที่เป็นแกนในการออกแบบและตีความบทบาทของปัญญาประดิษฐ์ในกระบวนการพัฒนาเครื่องมือวิจัย

ประการแรก เครื่องมือวิจัยมิใช่เพียงชุดของข้อความหรืองานประเมิน แต่เป็นตัวแทนของโครงสร้างเชิงแนวคิด (construct) ที่ต้องสะท้อนสิ่งที่มีงูวัดอย่างมีความหมาย ดังนั้น การพัฒนาเครื่องมือจึงต้องเริ่มจากการนิยามโครงสร้างเชิงแนวคิด (construct definition) อย่างชัดเจน กำหนดขอบเขตของการวัด และแปลงสู่ตัวชี้วัดหรือนิยามเชิงปฏิบัติการ (operational definition) อย่างสมเหตุสมผล (Ebel & Frisbie, 1991; Haladyna & Rodriguez, 2013)

ประการที่สอง คุณภาพของเครื่องมือวิจัยสะท้อนคุณภาพของกระบวนการพัฒนา มิใช่พิจารณาเฉพาะค่าทางสถิติในระยะท้ายเท่านั้น แนวคิดนี้สอดคล้องกับมุมมองความตรงแบบบูรณาการ ซึ่งมองความตรงเป็นระดับความเพียงพอของหลักฐานที่สนับสนุนการตีความและการใช้คะแนน มากกว่าการเป็นคุณสมบัติคงที่ของเครื่องมือ (Kane, 2013; Messick, 1995) ด้วยเหตุนี้ การใช้ปัญญาประดิษฐ์ในการพัฒนาเครื่องมือจึงควรถูกประเมินจากความสามารถในการช่วยให้กระบวนการสร้างหลักฐานคุณภาพการวัดมีความเป็นระบบ โปร่งใส และตรวจสอบได้ มากกว่าการพิจารณาเพียงว่าระบบสามารถสร้างข้อความหรือข้อคำถามได้รวดเร็วเพียงใด

ประการที่สาม ปัญญาประดิษฐ์เชิงสร้างสรรค์ควรถูกวางบทบาทในฐานะการเสริมพลังเชิงวิธีวิทยา กล่าวคือ เป็นกลไกที่ช่วยเพิ่มความเป็นระบบ ความสม่ำเสมอ ความครอบคลุม และความสามารถในการสะท้อนทางเลือกเชิงวิธีวิทยาแก่ผู้วิจัย โดยไม่ลดทอนความรับผิดชอบเชิงวิชาการและจริยธรรมของมนุษย์ ฐานคิดข้อนี้ทำให้กรอบแนวคิดฉบับนี้แตกต่างจากการมองปัญญาประดิษฐ์เป็นเพียงเครื่องมืออัตโนมัติหรือผู้ช่วยทางเทคนิคอย่างผิวเผิน

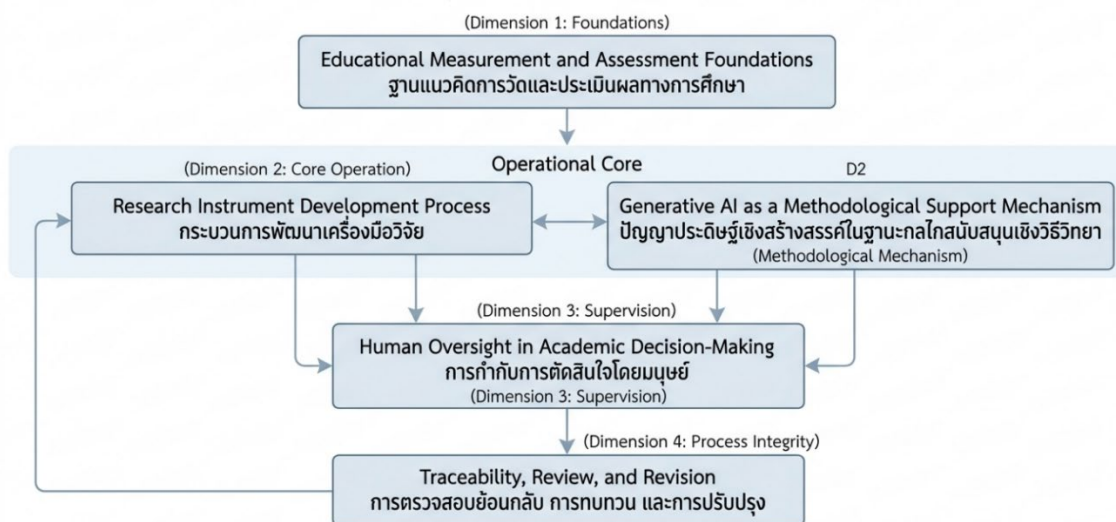
3.2 โครงสร้างของกรอบแนวคิดเชิงวิธีวิทยา

กรอบแนวคิดเชิงวิธีวิทยาที่น่าเสนอในบทความนี้ประกอบด้วยองค์ประกอบหลัก 4 ส่วนที่เชื่อมโยงกันอย่างเป็นระบบ ได้แก่ ฐานแนวคิดการวัดและประเมินผลทางการศึกษา กระบวนการพัฒนาเครื่องมือวิจัย กลไกการสนับสนุนของปัญญาประดิษฐ์ และการกำกับการตัดสินใจโดยมนุษย์

องค์ประกอบทั้งสี่ส่วนนี้ไม่ควรถูกเข้าใจว่าเป็นเพียงรายการแนวคิดที่จัดวางเรียงกัน แต่ควรถูกมองเป็นระบบเชิงปฏิสัมพันธ์ที่แต่ละส่วนทำหน้าที่กำกับและตรวจสอบกันเอง กล่าวคือ ฐานแนวคิดการวัดและประเมินผลทำหน้าที่กำหนดเกณฑ์ของคุณภาพการวัด กระบวนการพัฒนาเครื่องมือทำหน้าที่จัดลำดับขั้นตอนของการแปลงทฤษฎีสู่การวัด ปัญญาประดิษฐ์ทำหน้าที่เป็นเครื่องมือสนับสนุนในบางภารกิจที่เกี่ยวข้องกับการสังเคราะห์ข้อความ ภาษา และรูปแบบข้อความ ขณะที่การกำกับการตัดสินใจโดยมนุษย์ทำหน้าที่คงไว้ซึ่งความรับผิดชอบทางวิชาการและจริยธรรมของกระบวนการทั้งหมด

ในมิติของบทความฉบับนี้ “ปัญญาประดิษฐ์” หมายถึงเครื่องมือปัญญาประดิษฐ์เชิงสร้างสรรค์ที่อาศัยแบบจำลองภาษาขนาดใหญ่และการประมวลผลภาษาธรรมชาติเป็นสำคัญ มิได้ครอบคลุมปัญญาประดิษฐ์ทุกประเภท และได้มุ่งเปรียบเทียบประสิทธิภาพของแพลตฟอร์มหรือโมเดลรายย่อยที่ห่อหุ้มที่นำเสนอจึงเป็นข้อเสนอเชิงวิวิธยาระดับหลักการสำหรับการใช้เครื่องมือกลุ่มดังกล่าวในการพัฒนาเครื่องมือวิจัยภายใต้หลักการวัดและประเมินผลทางการศึกษา

4-Dimensional Conceptual Framework for the Application of Generative AI in Educational Research กรอบแนวคิด 4 มิติ สำหรับการประยุกต์ใช้ปัญญาประดิษฐ์เชิงสร้างสรรค์ในการวิจัยทางการศึกษา



ภาพ 1 กรอบแนวคิดเชิงวิวิธยาระดับหลักการสำหรับการประยุกต์ใช้ปัญญาประดิษฐ์ในการพัฒนาเครื่องมือวิจัยทางการศึกษา

3.3 ฐานทฤษฎีของกรอบแนวคิดเชิงวิวิธยาระดับหลักการ

3.3.1 ฐานแนวคิดการวัดและประเมินผลทางการศึกษา

ฐานของกรอบแนวคิดนี้เริ่มจากหลักการสำคัญในศาสตร์การวัดและประเมินผล ซึ่งถือว่าเครื่องมือวิจัยที่ดีต้องสามารถเชื่อมโยงระหว่างโครงสร้างเชิงแนวคิดที่มุ่งวัด ตัวบ่งชี้หรือนิยามเชิงปฏิบัติการ หลักฐานความตรงของการวัด ความเที่ยงของการวัด ความแม่นยำของการวัด (measurement precision) และความเป็นธรรมของการวัด (fairness) ได้อย่างมีเหตุผล (AERA, APA, & NCME, 2014; Kane, 2013; Messick, 1995) กล่าวอีกนัยหนึ่ง เครื่องมือมิใช่เพียงผลิตผลของการเขียนข้อคำถาม แต่เป็นผลลัพธ์ของกระบวนการสร้างและตรวจสอบความหมายของการวัดอย่างต่อเนื่อง

ภายใต้ฐานคิดนี้ ปัญญาประดิษฐ์ไม่ได้ถูกใช้เพื่อตัดสินว่าเครื่องมือมีความตรงหรือไม่ หากแต่ใช้เพื่อช่วยสนับสนุนภารกิจบางประการที่เกี่ยวข้องกับการจัดระเบียบข้อมูลเชิงทฤษฎี การตรวจสอบความครอบคลุมของมิติ การสะท้อนความกำกวมของภาษา หรือการเสนอทางเลือกเบื้องต้นในการออกแบบข้อคำถาม ทั้งหมดนี้ยังต้องถูกตรวจสอบภายใต้หลักการของความตรง ความเที่ยง ความแม่นยำของการวัด และความเป็นธรรมชาติของการวัด โดยผู้วิจัยและผู้เชี่ยวชาญเสมอ

3.3.2 ฐานคิดด้านกระบวนการพัฒนาเครื่องมือวิจัย

กระบวนการพัฒนาเครื่องมือวิจัยในกรอบนี้ตั้งอยู่บนลำดับตรรกะจากการกำหนดสิ่งที่ต้องการวัดไปสู่การพัฒนาตัวชี้วัด การออกแบบข้อคำถามหรืองานประเมิน การตรวจสอบคุณภาพเบื้องต้น และการเตรียมสู่การตรวจสอบเชิงประจักษ์ (DeVellis & Thorpe, 2021; Downing, 2006; Haladyna & Rodriguez, 2013) กระบวนการดังกล่าวมีลักษณะเป็นกระบวนการวนกลับมาเกินกว่าการทำงานแบบขั้นบันไดตายตัว กล่าวคือ ผลที่ได้จากการตรวจสอบในแต่ละช่วงสามารถย้อนกลับไปปรับนิยามโครงสร้างเชิงแนวคิด ตัวชี้วัด หรือรูปแบบข้อคำถามได้เสมอ

ปัญญาประดิษฐ์จึงมีประโยชน์เมื่อถูกใช้ในฐานเครื่องมือช่วยสังเคราะห์ ตรวจสอบ และสะท้อนข้อเสนอเบื้องต้นในแต่ละช่วงของกระบวนการ แต่จะมีความหมายเชิงวิธีวิทยาได้ก็ต่อเมื่อผู้วิจัยใช้ผลลัพธ์ดังกล่าวภายใต้ตรรกะของการพัฒนาเครื่องมือที่มีฐานรองรับ ไม่ใช่รับข้อเสนอจากระบบมาใช้โดยตรง

3.3.3 ฐานคิดด้านบทบาทของปัญญาประดิษฐ์ในฐานะผู้ช่วยเชิงวิธีวิทยา

ในกรอบนี้ ปัญญาประดิษฐ์เชิงสร้างสรรค์ถูกวางบทบาทเป็นผู้ช่วยเชิงวิธีวิทยา ไม่ใช่ผู้ตัดสินใจแทนนักวิจัย โดยทำหน้าที่สนับสนุนการสังเคราะห์วรรณกรรม การจัดกลุ่มตัวชี้วัด การเสนอร่างข้อคำถาม และการตรวจจุดกำกวมของภาษาในเบื้องต้น การจัดวางเช่นนี้สอดคล้องกับแนวคิดที่มอง AI เป็นเครื่องมือสนับสนุนภายใต้กรอบที่ยึดมนุษย์เป็นศูนย์กลาง มีความโปร่งใส และมีความรับผิดชอบต่อผลลัพธ์ที่นำไปใช้ (Holmes et al., 2019; UNESCO, 2023; Flanagan et al., 2024; Floridi et al., 2018)

3.3.4 ฐานคิดด้านการกำกับการตัดสินใจโดยมนุษย์

การกำกับการตัดสินใจโดยมนุษย์เป็นหลักการกลางของกรอบนี้ เพราะทุกขั้นตอนของการพัฒนาเครื่องมือมีนัยเชิงทฤษฎี วิธีวิทยา และจริยธรรมที่ไม่อาจถ่ายโอนไปยังระบบอัตโนมัติได้โดยสมบูรณ์ ผู้วิจัยจึงต้องเป็นผู้กำหนดความหมายของสิ่งที่มุ่งวัด คัดเลือกข้อเสนอจากระบบ และตัดสินความเหมาะสมของการตีความและการใช้ผลการวัด ในความหมายนี้ การกำกับโดยมนุษย์จึงเป็นทั้งหลักจริยธรรมและกลไกเชิงวิธีวิทยาที่ทำให้การวัดยังคงมีความชอบธรรม (Floridi et al., 2018; UNESCO, 2023; Messick, 1995; Kane, 2013; AERA, APA, & NCME, 2014)

3.4 กระบวนการพัฒนาเครื่องมือวิจัยภายใต้กรอบแนวคิด

กรอบนี้มองการพัฒนาเครื่องมือวิจัยเป็นกระบวนการต่อเนื่องจากโครงสร้างเชิงแนวคิดไปสู่นิยามเชิงปฏิบัติการ ตัวชี้วัด ข้อคำถาม และการตรวจสอบคุณภาพเบื้องต้นอย่างเป็นระบบ ปัญญาประดิษฐ์อาจช่วยสังเคราะห์ข้อมูล เปรียบเทียบทางเลือก และจัดระบบข้อเสนอในระดับต้นของกระบวนการวิจัยได้ แต่การตัดสินความเหมาะสมเชิงวิธีวิทยาในแต่ละขั้นยังต้องอยู่ภายใต้การกำกับของผู้วิจัยและผู้เชี่ยวชาญ เพื่อให้เครื่องมือ

มีฐานรองรับทั้งในเชิงทฤษฎีและเชิงการวัด (Downing, 2006; Haladyna & Rodriguez, 2013; DeVellis & Thorpe, 2021; AERA, APA, & NCME, 2014)

3.4.1 การกำหนดโครงสร้างเชิงแนวคิดและขอบเขตการวัด

ขั้นตอนแรกคือการกำหนดโครงสร้างเชิงแนวคิดที่มุ่งวัดอย่างชัดเจนจากทฤษฎีและวรรณกรรมที่เกี่ยวข้อง ปัญหาประดิษฐ์สามารถช่วยสังเคราะห์นิยาม ช่วยจัดกลุ่มมิติ และสะท้อนช่องว่างหรือความทับซ้อนของแนวคิดได้ อย่างไรก็ตาม การตัดสินใจว่ามิติใดควรถูกรวมหรือตัดออกยังต้องอาศัยการพิจารณาของผู้วิจัยภายใต้ฐานทฤษฎีและบริบทการใช้เครื่องมือ (Ebel & Frisbie, 1991; DeVellis & Thorpe, 2021)

3.4.2 การนิยามเชิงปฏิบัติการและการออกแบบตัวชี้วัด

เมื่อได้โครงสร้างเชิงแนวคิดแล้ว ขั้นตอนต่อมาคือการแปลงมิติที่กำหนดไว้ให้เป็นนิยามเชิงปฏิบัติการและตัวชี้วัดที่สามารถวัดได้ ปัญหาประดิษฐ์อาจช่วยตรวจสอบความสอดคล้องของภาษา ช่วยจัดกลุ่มตัวชี้วัด และชี้จุดที่นิยามยังไม่ชัดหรือซ้ำซ้อน อย่างไรก็ตาม การตัดสินใจว่าตัวชี้วัดใดสะท้อนสิ่งที่มุ่งวัดได้เพียงพอและเหมาะสมกับกลุ่มเป้าหมาย ยังคงเป็นความรับผิดชอบของผู้วิจัยและผู้เชี่ยวชาญ (Downing, 2006; DeVellis & Thorpe, 2021)

3.4.3 การสร้างข้อคำถามและโครงสร้างเครื่องมือ

ในขั้นนี้ ปัญหาประดิษฐ์สามารถช่วยเสนอรูปแบบข้อคำถามหลายทางเลือก ตรวจสอบความชัดเจนของภาษา และลดความกำกวมของถ้อยคำได้ อย่างไรก็ตาม ข้อเสนอที่ระบบสร้างขึ้นยังไม่อาจถือเป็นข้อคำถามที่พร้อมใช้โดยอัตโนมัติ เพราะยังต้องผ่านการกลั่นกรองตามฐานทฤษฎี ความเหมาะสมเชิงบริบท และเจตนาของการวัดโดยผู้วิจัยและผู้เชี่ยวชาญเสมอ (Haladyna & Rodriguez, 2013; Downing, 2006)

3.4.4 การตรวจสอบคุณภาพเบื้องต้นของเครื่องมือ

กรอบนี้ให้ความสำคัญกับการตรวจสอบคุณภาพเบื้องต้นก่อนการทดลองใช้จริง เพื่อคัดกรองจุดอ่อนของเครื่องมือในระยะเบื้องต้น ปัญหาประดิษฐ์สามารถช่วยชี้จุดเสี่ยงของความกำกวมทางภาษา ความไม่สม่ำเสมอของรูปแบบข้อคำถาม การทับซ้อนของเนื้อหา หรืออคติเบื้องต้นได้ อย่างไรก็ตาม หลักฐานที่ได้ในขั้นนี้ยังเป็นเพียงข้อมูลเบื้องต้น มิอาจใช้แทนการตรวจสอบโดยผู้เชี่ยวชาญหรือการวิเคราะห์เชิงประจักษ์ในขั้นถัดไปได้ (AERA, APA, & NCME, 2014; Downing, 2006)

3.5 การกำกับการตัดสินใจโดยมนุษย์ในแต่ละขั้นตอน

หัวใจของกรอบนี้คือการคงบทบาทของผู้วิจัยในฐานะผู้ตัดสินใจเชิงวิชาการและจริยธรรมตลอดกระบวนการพัฒนาเครื่องมือวิจัย ผู้วิจัยเป็นผู้กำหนดขอบเขตของสิ่งที่มุ่งวัด ตัดสินความสอดคล้องของตัวชี้วัด ประเมินความเหมาะสมของข้อคำถาม และเลือกหลักฐานที่ใช้ยืนยันคุณภาพของเครื่องมือ ดังนั้น การกำกับโดยมนุษย์จึงมิใช่เพียงการตรวจทานผลลัพธ์ภายหลัง แต่เป็นการรับผิดชอบต่อความหมายของการวัดในทุกๆ ระยะ (Messick, 1995; Kane, 2013; UNESCO, 2023)

3.6 การสนับสนุนหลักฐานความตรงของการวัด

ภายใต้แนวคิดความตรงแบบบูรณาการ ความตรงของการวัดเป็นระดับความเพียงพอของหลักฐานที่สนับสนุนการตีความและการใช้ผลการวัดอย่างเหมาะสม กรอบนี้จึงเสนอว่า ปัญหาประดิษฐ์อาจช่วยสนับสนุน

การสร้างหลักฐานความตรงได้ในบางมิติ เช่น การสังเคราะห์วรรณกรรมเพื่อตรวจสอบความครอบคลุมของเนื้อหา การทบทวนความชัดเจนของภาษา และการช่วยจัดระบบข้อมูลเบื้องต้นสำหรับพิจารณาโครงสร้างภายในของเครื่องมือ อย่างไรก็ตาม การตัดสินความเพียงพอของหลักฐานและการตีความผลยังต้องอาศัยผู้วิจัยและผู้เชี่ยวชาญเป็นหลัก (Messick, 1995; Kane, 2013; AERA, APA, & NCME, 2014)

3.7 ความเที่ยงและความแม่นยำของการวัด

คุณภาพของเครื่องมือยังต้องพิจารณาความเที่ยงและความแม่นยำของการวัดควบคู่กับหลักฐานความตรง ความเที่ยงสะท้อนความสม่ำเสมอของผลการวัด ส่วนความแม่นยำสะท้อนระดับความคลาดเคลื่อนของคะแนนหรือข้อมูลที่ได้ ปัญญาประดิษฐ์อาจช่วยตรวจสอบความสม่ำเสมอของข้อคำถามและจัดเตรียมข้อมูลเบื้องต้นสำหรับการวิเคราะห์ได้ แต่การเลือกดัชนี การแปลความหมายของผลการวิเคราะห์ และการตัดสินว่าเครื่องมือมีคุณภาพเพียงพอหรือไม่ ยังคงต้องอาศัยการพิจารณาเชิงวิชาการของผู้วิจัย (AERA, APA, & NCME, 2014; DeVellis & Thorpe, 2021)

3.8 ความเป็นธรรมและความรับผิดชอบในการวัด

ความเป็นธรรมของการวัดเชื่อมโยงโดยตรงกับจริยธรรมของการวิจัยและการใช้คะแนน กรอบนี้จึงกำหนดให้ปัญญาประดิษฐ์อยู่ภายใต้การกำกับของมนุษย์ เพื่อป้องกันความเสี่ยงจากอคติทางภาษา อคติทางวัฒนธรรม และการตีความคะแนนที่ไม่เหมาะสมต่อกลุ่มผู้ตอบที่มีความหลากหลาย แม้ระบบอาจช่วยตรวจสอบจุดเสี่ยงในเบื้องต้นได้ แต่การตัดสินว่าความแตกต่างที่พบเป็นผลจากอคติของเครื่องมือหรือไม่ ยังต้องอาศัยการพิจารณาเชิงวิชาการและเชิงจริยธรรมของผู้วิจัยเป็นหลัก (AERA, APA, & NCME, 2014; Floridi et al., 2018; UNESCO, 2023)

3.9 ผลเชิงวิวิธวิทยาของกรอบแนวคิด

โดยสรุป กรอบแนวคิดนี้ช่วยให้การพัฒนาเครื่องมือวิจัยมีความเป็นระบบ ตรวจสอบย้อนกลับได้ และเชื่อมโยงกับฐานทฤษฎีและหลักการวัดอย่างชัดเจนยิ่งขึ้น อีกทั้งยังช่วยลดความคลาดเคลื่อนเชิงเทคนิค และกำหนดตำแหน่งของปัญญาประดิษฐ์ให้เป็นผู้ช่วยเชิงวิวิธวิทยาภายใต้การกำกับโดยมนุษย์ อย่างไรก็ตาม กรอบนี้ควรถูกเข้าใจในฐานะข้อเสนอเชิงวิวิธวิทยาสำหรับการออกแบบกระบวนการพัฒนาเครื่องมือ มิใช่ข้อยืนยันเชิงประจักษ์ว่าเครื่องมือที่ใช้ปัญญาประดิษฐ์สนับสนุนจะมีคุณภาพสูงกว่าเสมอไป (AERA, APA, & NCME, 2014; Kane, 2013; DeVellis & Thorpe, 2021; UNESCO, 2023; Flanagan et al., 2024)

4. ตารางสังเคราะห์กรอบแนวคิดเชิงวิวิธวิทยา

เพื่อให้โครงสร้างของกรอบแนวคิดเชิงวิวิธวิทยาที่เสนอในบทความนี้มีความชัดเจนและตรวจสอบได้มากขึ้น ผู้เขียนได้นำเสนอความสัมพันธ์ระหว่าง 4 มิติสำคัญของกรอบ ได้แก่ ฐานแนวคิดการวัดและประเมินผลทางการศึกษา กระบวนการพัฒนาเครื่องมือวิจัย บทบาทของปัญญาประดิษฐ์ในฐานะกลไกสนับสนุนเชิงวิวิธวิทยา และการกำกับการตัดสินใจโดยมนุษย์ ในรูปของตารางสังเคราะห์ดังตาราง 1 ทั้งนี้ ตารางดังกล่าวมิได้มีวัตถุประสงค์เพื่อกำหนดขั้นตอนตายตัวสำหรับการใช้ปัญญาประดิษฐ์ในการพัฒนาเครื่องมือวิจัย หากแต่ใช้เพื่อแสดงตรรกะเชิงวิวิธวิทยาว่า ปัญญาประดิษฐ์ควรถูกวางตำแหน่งและกำกับอย่างไรในแต่ละช่วงของการพัฒนาเครื่องมือ ภายใต้หลักการวัดและประเมินผลที่เหมาะสม

ตาราง 1 กรอบแนวคิดเชิงวิธีวิทยาสำหรับการประยุกต์ใช้ปัญญาประดิษฐ์เชิงสร้างสรรค์ในการพัฒนาเครื่องมือวิจัยทางการศึกษา

ขั้นตอนการพัฒนาเครื่องมือ	ฐานแนวคิด/หลักฐานคุณภาพการวัดที่เกี่ยวข้อง	บทบาทของปัญญาประดิษฐ์เชิงสนับสนุน	การกำกับการตัดสินใจโดยมนุษย์	ผลลัพธ์เชิงวิธีวิทยา
1. การกำหนดโครงสร้างเชิงแนวคิด	หลักฐานด้านเนื้อหาและความสอดคล้องกับฐานทฤษฎี	ช่วยสังเคราะห์วรรณกรรม จัดกลุ่มมิติของโครงสร้างเชิงแนวคิด ตรวจสอบความครอบคลุมขององค์ประกอบ และสะท้อนความเข้าใจหรือความคลุมเครือของนิยามเบื้องต้น	กำหนดขอบเขตของโครงสร้างเชิงแนวคิด ตัดสินความเหมาะสมของมิติ และยืนยันความสอดคล้องกับฐานทฤษฎี และบริบทการวิจัย	โครงสร้างเชิงแนวคิดที่ชัดเจน และสอดคล้องกับฐานทฤษฎี
2. การกำหนดตัวชี้วัดและนิยามเชิงปฏิบัติการ	หลักฐานด้านเนื้อหาและความสอดคล้องเชิงความหมายของการวัด	ช่วยจัดกลุ่มตัวชี้วัด เสนอถ้อยคำสำหรับนิยามเชิงปฏิบัติการ ตรวจสอบความเข้าใจของตัวชี้วัด และสะท้อนช่องว่างของมิติที่อาจยังไม่ครอบคลุม	ยืนยันว่าตัวชี้วัดสะท้อนโครงสร้างเชิงแนวคิดอย่างเหมาะสม ตัดสินความชัดเจนของนิยาม และพิจารณาความเหมาะสมต่อกลุ่มเป้าหมาย	ตัวชี้วัดที่สอดคล้องกับโครงสร้างเชิงแนวคิดและพร้อมต่อการพัฒนาเครื่องมือ
3. การออกแบบข้อคำถามหรืองานประเมิน	หลักฐานด้านเนื้อหาและกระบวนการตอบสนอง	เสนอร่างข้อคำถามหลายรูปแบบ ช่วยปรับระดับภาษา เสนอถ้อยคำทางเลือก ตรวจสอบความกำกวม และช่วยให้ข้อคำถามสอดคล้องกับตัวชี้วัดในระดับต้น	คัดเลือก ปรับแก้ หรือตัดข้อคำถามตามหลักการวัด เจตนาของการวัด และบริบทของผู้ตอบ	ร่างเครื่องมือที่มีความสอดคล้องระหว่างข้อคำถาม ตัวชี้วัด และจุดมุ่งหมายการวัด
4. การตรวจสอบความตรงเชิงเนื้อหาเบื้องต้น	หลักฐานด้านเนื้อหา	ช่วยจัดทำตารางความเชื่อมโยงระหว่างโครงสร้างเชิงแนวคิด-ตัวชี้วัด-ข้อคำถาม สะท้อนข้อคำถามที่อาจไม่ครอบคลุมหรือซ้ำซ้อน และช่วยเตรียมข้อมูลเพื่อเสนอผู้เชี่ยวชาญพิจารณา	ประเมินความสอดคล้องเชิงเนื้อหา รวบรวมความเห็นผู้เชี่ยวชาญ และตัดสินใจคง/ปรับ/ตัดข้อคำถาม	เครื่องมือที่มีความชัดเจนด้านความตรงเชิงเนื้อหาในระดับเบื้องต้น
5. การตรวจสอบกระบวนการตอบสนองและความชัดเจนของภาษา	หลักฐานด้านกระบวนการตอบสนอง	ตรวจสอบระดับภาษา ความกำกวมของถ้อยคำ ความซับซ้อนของประโยค และเสนอข้อสังเกตเกี่ยวกับความเป็นไปได้ของความเข้าใจคลาดเคลื่อนในระดับต้น	พิจารณาว่าผู้ตอบตีความข้อคำถามสอดคล้องกับสิ่งที่มุ่งวัดหรือไม่ โดยอาศัยการทดสอบเบื้องต้น การสัมภาษณ์ หรือการพิจารณาเชิงคุณภาพ	ข้อคำถามที่มีความชัดเจนและสอดคล้องกับกระบวนการคิดของผู้ตอบมากขึ้น

ตาราง 1 (ต่อ)

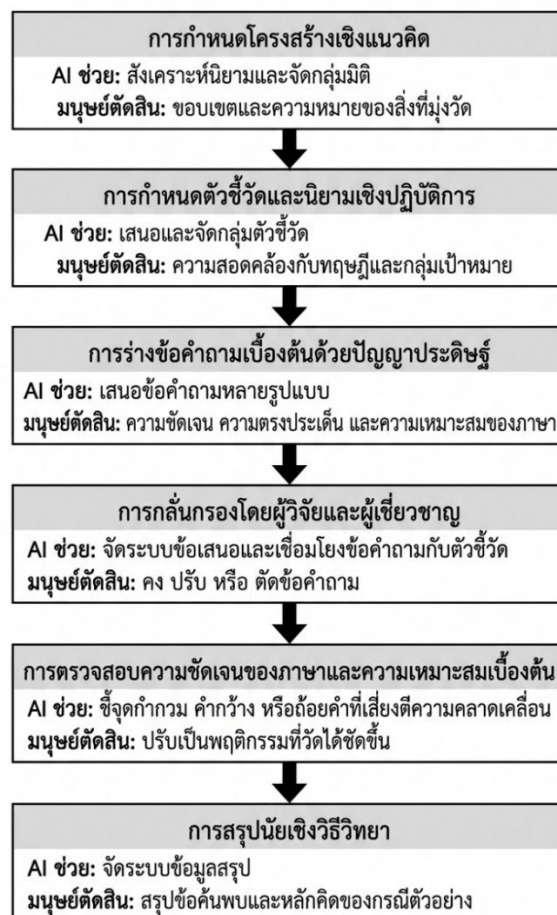
ขั้นตอนการ พัฒนาเครื่องมือ	ฐานแนวคิด/ หลักฐานคุณภาพ การวัดที่เกี่ยวข้อง	บทบาทของปัญญาประดิษฐ์ เชิงสนับสนุน	การกำกับการตัดสินใจ โดยมนุษย์	ผลลัพธ์เชิงวิธี วิทยา
6. การตรวจสอบ โครงสร้างภายใน เบื้องต้น	หลักฐานด้าน โครงสร้างภายใน	ช่วยเตรียมข้อมูล ตรวจสอบ รูปแบบความสัมพันธ์เบื้องต้น ของข้อความ หรือจำลอง ทางเลือกในการวิเคราะห์เชิง สำรวจในระดับเบื้องต้น	เลือกเทคนิคการวิเคราะห์ ที่เหมาะสม ตีความผล ตามฐานทฤษฎี และ ตัดสินใจว่าผลสอดคล้องกับ โครงสร้างที่ตั้งไว้หรือไม่	หลักฐานเบื้องต้น ว่าข้อความ สะท้อนโครงสร้าง เชิงแนวคิดอย่างมี เหตุผล
7. การตรวจสอบ ความเที่ยง และ ความแม่นยำของ การวัด	ความเที่ยง และ ความแม่นยำของ การวัด	ช่วยตรวจสอบความสม่ำเสมอ ของรูปแบบข้อความ ระบุข้อ คำถามที่อาจมีความแปรปรวน ผิดปกติ และช่วยจัดเตรียม ข้อมูลสำหรับการวิเคราะห์ ความเที่ยง ในระดับต้น	เลือกดัชนีความเที่ยง ที่เหมาะสม ตีความผล และตัดสินใจความเพียงพอ ของระดับความเที่ยง และความแม่นยำตาม ลักษณะของเครื่องมือ	เครื่องมือที่มีความ สม่ำเสมอและมี ความแม่นยำของ การวัดในระดับที่ เหมาะสม
8. การพิจารณา ความเป็นธรรมชาติ ความเหมาะสมเชิง จริยธรรม	ความเป็นธรรมชาติของ การวัดและความ รับผิดชอบในการใช้ คะแนน	ช่วยตรวจสอบถ้อยคำที่อาจมี อคติทางภาษา วัฒนธรรม หรือ บริบท ชี้จุดที่อาจสร้างความ เสียเปรียบแก่ผู้ตอบบางกลุ่ม และสะท้อนความไม่สมดุลของ รูปแบบข้อความในเชิงสำรวจ	ตัดสินใจว่าความแตกต่างที่ พบเป็นความแตกต่างของ โครงสร้างเชิงแนวคิดหรือ เป็นอคติของเครื่องมือ พิจารณาความเหมาะสม เชิงจริยธรรมและผลกระทบ ของการใช้คะแนน	เครื่องมือที่มีความ เหมาะสมต่อการ ใช้ในบริบทจริง มากขึ้น และลด ความเสี่ยงด้าน อคติ
9. การตัดสินใจ ความพร้อมของ เครื่องมือก่อน การใช้จริง	การบูรณาการ หลักฐานคุณภาพ การวัดเพื่อ สนับสนุนการใช้ เครื่องมือ	ช่วยสรุปประเด็นที่ควรปรับแก้ รวบรวมข้อค้นพบจากหลาย ขั้นตอน และจัดระเบียบข้อมูล เพื่อสนับสนุนการทบทวนขั้น สุดท้ายของผู้วิจัย	ตัดสินใจความพร้อมของ เครื่องมือ กำหนดว่าควร ปรับปรุงส่วนใดเพิ่มเติม และรับผิดชอบต่อผู้นำ เครื่องมือไปใช้ใน ภาคสนาม	เครื่องมือที่ผ่าน การกลั่นกรองเชิง วิธีวิทยาอย่างเป็น ระบบก่อนการใช้ จริง

จากตาราง 1 จะเห็นได้ว่า บทบาทของปัญญาประดิษฐ์ภายใต้กรอบแนวคิดเชิงวิธีวิทยานี้ถูกจัดวางให้ทำหน้าที่สนับสนุนการสังเคราะห์ การสะท้อน และการจัดระบบข้อมูลเชิงแนวคิดและเชิงภาษาในแต่ละช่วงของการพัฒนาเครื่องมือ ขณะที่การตัดสินใจคุณภาพของการวัดยังคงตั้งอยู่บนหลักการทางการวัดและประเมินผล และอยู่ภายใต้ความรับผิดชอบของผู้วิจัยและผู้เชี่ยวชาญ ในความหมายนี้ ตารางสังเคราะห์นี้ได้เสนอให้ปัญญาประดิษฐ์เป็นผู้รับรองคุณภาพของเครื่องมือโดยอัตโนมัติ หากแต่เสนอให้ปัญญาประดิษฐ์เป็นกลไกที่ช่วยเสริมความเป็นระบบ ความโปร่งใส และความสามารถในการตรวจสอบย้อนกลับของกระบวนการพัฒนาเครื่องมือวิจัยทางการศึกษา

5. ตัวอย่างการประยุกต์กรอบแนวคิดเชิงวิธีวิทยาในการพัฒนาเครื่องมือวิจัย

เพื่อให้กรอบแนวคิดเชิงวิธีวิทยาที่เสนอในบทความนี้มีความเป็นรูปธรรมมากขึ้น หัวข้อนี้จึงนำเสนอกรณีตัวอย่างเชิงสาคิตของการพัฒนาแบบสอบถามทางการศึกษา โดยมุ่งแสดงให้เห็นลำดับการทำงานตั้งแต่การกำหนดโครงสร้างเชิงแนวคิด การกำหนดตัวชี้วัด การร่างข้อคำถามด้วยปัญญาประดิษฐ์ การกลั่นกรองโดยผู้วิจัยและผู้เชี่ยวชาญ ไปจนถึงการตรวจสอบความชัดเจนของภาษาในเบื้องต้น ทั้งนี้จุดมุ่งหมายของหัวข้อนี้มีใช้การยืนยันผลเชิงประจักษ์ของกรอบ หากแต่เพื่อแสดงให้เห็นอย่างเป็นระบบว่าปัญญาประดิษฐ์สามารถทำหน้าที่เป็นผู้ช่วยเชิงวิธีวิทยาได้อย่างไร ภายใต้การกำกับกับการตัดสินใจโดยมนุษย์

กรณีตัวอย่างที่ใช้คือการพัฒนาแบบสอบถามความสามารถในการกำกับตนเองในการเรียนรู้ของนักศึกษาระดับอุดมศึกษา ซึ่งเป็นโครงสร้างเชิงแนวคิดที่มีฐานทฤษฎีรองรับและเหมาะต่อการสาคิตบทบาทของปัญญาประดิษฐ์ในงานสังเคราะห์ การร่าง และการตรวจสอบเบื้องต้น โดยไม่ลดทอนบทบาทของผู้วิจัยในฐานะผู้ตัดสินใจเชิงวิชาการ



ภาพ 2 ลำดับการประยุกต์กรอบแนวคิดเชิงวิธีวิทยาในกรณีตัวอย่างการพัฒนาแบบสอบถาม

ภาพ 2 แสดงให้เห็นว่า ปัญญาประดิษฐ์มีบทบาทหลักในฐานะเครื่องมือช่วยสังเคราะห์ ร่าง และตรวจสอบเบื้องต้น ขณะที่การตัดสินใจเชิงวิชาการและเชิงจริยธรรมยังคงเป็นหน้าที่ของผู้วิจัยและผู้เชี่ยวชาญในทุกขั้นตอน

5.1 การกำหนดโครงสร้างเชิงแนวคิดและขอบเขตการวัด

ขั้นตอนแรกของกรณีตัวอย่างคือการทำให้งานวัดมีความชัดเจนในระดับโครงสร้างเชิงแนวคิดก่อนจะพัฒนาไปสู่ตัวชี้วัดและข้อคำถาม

5.2 การกำหนดตัวชี้วัดและนิยามเชิงปฏิบัติการ

เมื่อได้ขอบเขตของโครงสร้างเชิงแนวคิดแล้ว ขั้นตอนถัดมาคือการแปลงมติดังกล่าวให้เป็นตัวชี้วัดเชิงปฏิบัติการที่สามารถนำไปพัฒนาเป็นข้อคำถามได้อย่างเป็นระบบ

5.3 การร่างข้อคำถามเบื้องต้นด้วยปัญญาประดิษฐ์

เมื่อได้ตัวชี้วัดเบื้องต้นแล้ว ปัญญาประดิษฐ์สามารถเข้ามามีบทบาทในการเสนอข้อคำถามหลายทางเลือก เพื่อเปิดพื้นที่ให้ผู้วิจัยเปรียบเทียบความชัดเจนของภาษาและความเหมาะสมของการวัด

กล่องตัวอย่างที่ 1 ตัวอย่างการร่างและกลั่นกรองข้อคำถามจากตัวชี้วัดเดียวกัน

ตัวชี้วัด: ผู้เรียนมีการกำหนดเป้าหมายการเรียนรู้ก่อนเริ่มกิจกรรม

ข้อเสนอจากปัญญาประดิษฐ์

1. ฉันกำหนดเป้าหมายการเรียนรู้ของตนเองก่อนเริ่มทำกิจกรรมการเรียนรู้ทุกครั้ง
2. ก่อนเริ่มเรียน ฉันมักคิดว่าล่วงหน้าว่าต้องการบรรลุอะไรจากการเรียนครั้งนั้น
3. ฉันมีเป้าหมายชัดเจนว่าต้องการเรียนรู้อะไรในแต่ละครั้ง

เกณฑ์ที่ผู้วิจัยใช้พิจารณา

- ข้อคำถามใดสื่อพฤติกรรมได้ชัด
- ข้อคำถามใดใช้ภาษาตรงและตีความได้ใกล้เคียงกัน
- ข้อคำถามใดเสี่ยงวัดทัศนคติหรือการประเมินตนเองกว้างเกินไป

ข้อสรุปเชิงวิธีวิทยา

ข้อเสนอจากปัญญาประดิษฐ์มีประโยชน์ต่อการเปิดทางเลือกเชิงภาษา แต่การคัดเลือกข้อคำถามสุดท้ายยังต้องอาศัยการกลั่นกรองโดยผู้วิจัยตามหลักความชัดเจนของการวัดและความเหมาะสมต่อกลุ่มเป้าหมาย

ตัวอย่างการปรับจากข้อคำถามที่ยังไม่ชัด

ข้อเสนอเดิมจากปัญญาประดิษฐ์: ฉันวางแผนและติดตามตนเองอย่างเป็นระบบเมื่อเรียน

ปัญหาที่พบ: วัดมากกว่า 1 พฤติกรรมในข้อเดียว

แนวทางปรับ: แยกเป็น 2 ข้อ ได้แก่

- ฉันวางแผนการเรียนรู้ของตนเองก่อนเริ่มกิจกรรม
- ฉันตรวจสอบความก้าวหน้าของตนเองระหว่างการเรียนอย่างสม่ำเสมอ

5.4 การกลั่นกรองข้อคำถามโดยผู้วิจัยและผู้เชี่ยวชาญ

อย่างไรก็ดี ข้อเสนอจากปัญญาประดิษฐ์ยังไม่ใช่ว่าข้อคำถามที่พร้อมใช้โดยตรง หากต้องผ่านการกลั่นกรองตามหลักการวัดและการพิจารณาเชิงเนื้อหาโดยมนุษย์ก่อนเสมอ

5.5 การตรวจสอบความชัดเจนของภาษาและความเหมาะสมเบื้องต้น

ภายหลังการกลั่นกรองเชิงเนื้อหาแล้ว ยังจำเป็นต้องตรวจสอบระดับภาษาและความกำกวมของข้อความอีกครั้ง เพื่อให้ข้อความมีความเหมาะสมต่อกลุ่มเป้าหมายจริง

ตาราง 2 การประยุกต์ใช้กรอบแนวคิดเชิงวิธีวิทยาในกรณีตัวอย่าง

ขั้นตอน	บทบาทของปัญญาประดิษฐ์	การกำกับการตัดสินใจโดยมนุษย์	ผลลัพธ์ที่ได้
5.1 กำหนดโครงสร้างเชิงแนวคิด	ช่วยสังเคราะห์นิยามและจัดกลุ่มมิติจากวรรณกรรม	ตัดสินใจขอบเขตและความหมายของสิ่งที่มุ่งวัด	โครงสร้างเชิงแนวคิดที่ชัดเจน
5.2 กำหนดตัวชี้วัดเบื้องต้น	ช่วยเสนอและจัดกลุ่มตัวชี้วัดเบื้องต้น	ตัดสินใจสอดคล้องกับทฤษฎีและบริบทกลุ่มเป้าหมาย	ตัวชี้วัดที่พร้อมพัฒนาเป็นข้อคำถาม
5.3 ร่างข้อคำถามเบื้องต้น	ช่วยเสนอข้อคำถามหลายรูปแบบ	คัดเลือกข้อคำถามที่ชัดและตรงประเด็น	ร่างข้อคำถามเบื้องต้น
5.4 กลั่นกรองโดยผู้วิจัย/ผู้เชี่ยวชาญ	ช่วยจัดระบบข้อเสนอแนะเชื่อมโยงข้อคำถามกับตัวชี้วัด	คง/ปรับ/ตัดข้อคำถามตามหลักการวัด	ข้อคำถามที่เหมาะสมเชิงเนื้อหา
5.5 ตรวจสอบภาษาและความเหมาะสมเบื้องต้น	ช่วยชี้จุดกำกวม คำกว้าง หรือถ้อยคำที่เสี่ยงตีความคลาดเคลื่อน	ปรับข้อความให้เป็นพฤติกรรมที่วัดได้ชัดเจน	ข้อคำถามที่พร้อมเข้าสู่การตรวจสอบขั้นถัดไป

5.6 นัยเชิงวิธีวิทยาจากกรณีตัวอย่าง

กรณีตัวอย่างนี้สะท้อนนัยเชิงวิธีวิทยาสำคัญอย่างน้อย 3 ประการ ได้แก่ **ประการแรก** ปัญญาประดิษฐ์มีคุณค่าในฐานะเครื่องมือช่วยเปิดทางเลือกเชิงแนวคิดและเชิงภาษา มากกว่าการทำหน้าที่สร้างข้อคำถามแทนผู้วิจัย **ประการที่สอง** คุณภาพของเครื่องมือยังคงขึ้นอยู่กับกรกลั่นกรองโดยมนุษย์ในทุกขั้นตอนที่เกี่ยวข้องกับโครงสร้างเชิงแนวคิด ตัวชี้วัด และข้อคำถาม และ**ประการที่สาม** การใช้ปัญญาประดิษฐ์ภายใต้กรอบที่ชัดเจนช่วยให้กระบวนการพัฒนาเครื่องมือมีความเป็นระบบ โปร่งใส และตรวจสอบย้อนกลับได้มากขึ้น ในความหมายนี้ ปัญญาประดิษฐ์จึงมีสถานะเป็นผู้ช่วยเชิงวิธีวิทยา มิใช่ผู้แทนการตัดสินใจเชิงวิชาการของผู้วิจัย

จากกรณีตัวอย่างข้างต้น จะเห็นได้ว่า แม้ปัญญาประดิษฐ์จะช่วยลดภาระในงานสังเคราะห์และงานร่างเบื้องต้นได้ แต่ทุกขั้นตอนยังต้องอาศัยการกำกับการตัดสินใจโดยมนุษย์อย่างเข้มแข็ง ประเด็นดังกล่าวจึงนำไปสู่ข้อพิจารณาสำคัญในหัวข้อถัดไปเกี่ยวกับจริยธรรม ความโปร่งใส และการใช้ปัญญาประดิษฐ์อย่างมีความรับผิดชอบในการพัฒนาเครื่องมือวิจัย

6. ประเด็นจริยธรรมและการใช้ปัญญาประดิษฐ์อย่างมีความรับผิดชอบ

การประยุกต์ใช้ปัญญาประดิษฐ์ในการพัฒนาเครื่องมือวิจัยทางการศึกษา มิได้เป็นเพียงประเด็นด้านประสิทธิภาพของการทำงานหรือความสะดวกในการร่างข้อความเท่านั้น หากยังเกี่ยวข้องโดยตรงกับความน่าเชื่อถือทางวิชาการ ความชอบธรรมของการวัด และความรับผิดชอบต่อการใช้ผลการวัดอย่างหลีกเลี่ยงไม่ได้ วรรณกรรมร่วมสมัยด้านปัญญาประดิษฐ์ในการศึกษาและการวิจัยชี้ตรงกันว่า การใช้ปัญญาประดิษฐ์เชิงสร้างสรรค์

ควรตั้งอยู่บนหลักการสำคัญอย่างน้อย ได้แก่ ความโปร่งใส ความรับผิดชอบของมนุษย์ ความเป็นธรรม การคุ้มครองข้อมูล และความซื่อสัตย์ทางวิชาการ มิใช่พิจารณาเฉพาะความสามารถของระบบในการสร้างข้อความหรือสังเคราะห์เนื้อหาได้อย่างลื่นไหลเท่านั้น (Floridi et al., 2018; UNESCO, 2023; Flanagan et al., 2024) ในบริบทของการพัฒนาเครื่องมือวิจัย ประเด็นดังกล่าวยิ่งมีความสำคัญมากขึ้น เพราะข้อเสนอที่ปัญญาประดิษฐ์สร้างขึ้นอาจดูสมเหตุสมผลในระดับภาษา แต่ยังไม่เพียงพอที่จะยืนยันความถูกต้องเชิงทฤษฎี ความเพียงพอของการวัด หรือความเหมาะสมเชิงจริยธรรมได้โดยอัตโนมัติ

หลักการประการแรก คือ ความโปร่งใสในการใช้ปัญญาประดิษฐ์ ผู้วิจัยควรเปิดเผยอย่างชัดเจนว่า มีการใช้ปัญญาประดิษฐ์ในขั้นตอนใดของการพัฒนาเครื่องมือ ใช้เพื่อวัตถุประสงค์ใด ใช้แพลตฟอร์มหรือเครื่องมือใด และผลลัพธ์ที่ได้ถูกนำไปใช้ในลักษณะใด หลักการนี้มีความสำคัญอย่างยิ่งในบริบทของปัญญาประดิษฐ์เชิงสร้างสรรค์ เพราะศักยภาพและข้อจำกัดของระบบอาจแตกต่างกันตามรุ่นของโมเดล การออกแบบของแพลตฟอร์ม และช่วงเวลาที่ใช้งาน หากผู้วิจัยไม่เปิดเผยข้อมูลดังกล่าวอย่างเพียงพอ ผู้อ่านย่อมไม่สามารถประเมินความเหมาะสมของกระบวนการ หรือพิจารณาความสามารถในการตรวจสอบย้อนกลับของงานวิจัยได้อย่างครบถ้วน (UNESCO, 2023; Flanagan et al., 2024) ดังนั้น ความโปร่งใสจึงมิใช่เพียงข้อพึงปฏิบัติด้านจริยธรรมทั่วไป หากเป็นเงื่อนไขพื้นฐานของความน่าเชื่อถือทางวิชาการในงานที่มีการใช้ปัญญาประดิษฐ์เข้ามาเกี่ยวข้อง

หลักการประการที่สอง คือ ความรับผิดชอบทางวิชาการของผู้วิจัยและการกำกับการตัดสินใจโดยมนุษย์ แม้ปัญญาประดิษฐ์จะช่วยสังเคราะห์นิยามของตัวแปร ช่วยจัดกลุ่มตัวชี้วัด หรือช่วยร่างข้อคำถามเบื้องต้นได้ แต่การตัดสินใจเกี่ยวกับขอบเขตของโครงสร้างเชิงแนวคิด ความสอดคล้องของข้อคำถามกับตัวชี้วัด การประเมินหลักฐานความตรงของการวัด และการตัดสินใจพร้อมของเครื่องมือก่อนนำไปใช้จริง ยังคงต้องเป็นความรับผิดชอบของผู้วิจัยและผู้เชี่ยวชาญเสมอ (Floridi et al., 2018; UNESCO, 2023) หลักการนี้สอดคล้องกับแนวคิดเรื่อง Human Oversight และยังสอดคล้องกับมุมมองด้านการวัดที่เห็นว่า ความชอบธรรมของการตีความและการใช้ผลการวัดต้องตั้งอยู่บนข้อโต้แย้งและหลักฐานที่มนุษย์สามารถอธิบายและรับผิดชอบได้ ไม่อาจถ่ายโอนไปยังระบบอัตโนมัติได้โดยสมบูรณ์ (Messick, 1995; Kane, 2013) ในความหมายนี้ ปัญญาประดิษฐ์ไม่มีสถานะเป็นผู้ร่วมตัดสินคุณภาพของเครื่องมือ หากเป็นเพียงเครื่องมือช่วยคิดที่ผู้วิจัยต้องกลั่นกรองด้วยวิจารณญาณเชิงวิชาการ

หลักการประการที่สาม คือ ความสอดคล้องกับหลักการวัดและประเมินผลทางการศึกษา การใช้ปัญญาประดิษฐ์ในการพัฒนาเครื่องมือวิจัยไม่ควรถูกประเมินจากความรวดเร็วในการร่างข้อความหรือจำนวนข้อคำถามที่ระบบสามารถสร้างได้เท่านั้น แต่ควรถูกประเมินด้วยคำถามสำคัญว่า ผลลัพธ์ที่ระบบเสนอช่วยเพิ่มความชัดเจน ความครอบคลุม และความสมเหตุสมผลของการวัดจริงหรือไม่ มาตรฐานการทดสอบทางการศึกษาและจิตวิทยาเน้นว่าคุณภาพของเครื่องมือวัดต้องเชื่อมโยงกับหลักฐานความตรง ความเที่ยง ความแม่นยำของการวัด และความเป็นธรรมของการใช้ผลการวัดอย่างเป็นระบบ (AERA, APA, & NCME, 2014) ดังนั้น แม้ข้อคำถามที่ปัญญาประดิษฐ์สร้างขึ้นอาจดูเป็นวิชาการและสละสลวยในระดับภาษา แต่หากไม่

สอดคล้องกับโครงสร้างเชิงแนวคิด ไม่ครอบคลุมตัวชี้วัดที่สำคัญ หรือมีความกำกวมทางความหมาย ข้อคำถามดังกล่าวก็อาจก่อให้เกิดปัญหาต่อการตีความผลการวัดได้ (Messick, 1995; Kane, 2013)

หลักการประการที่สี่ คือ ความเป็นธรรมและการลดอคติ วรรณกรรมด้านจริยธรรมของปัญญาประดิษฐ์ชี้ให้เห็นอย่างชัดเจนว่า ระบบอัจฉริยะอาจสะท้อนอคติที่แฝงอยู่ในข้อมูลฝึก แบบแผนทางภาษา หรือกรอบสมมติฐานที่ผู้ใช้ไม่สามารถมองเห็นได้โดยตรง (Floridi et al., 2018; UNESCO, 2023) ในกรณีของการพัฒนาเครื่องมือวิจัยทางการศึกษา ความเสี่ยงดังกล่าวมีความสำคัญมาก เพราะข้อคำถามที่เอนเอียงทางภาษา วัฒนธรรม เพศ หรือบริบททางสังคม อาจทำให้ผู้ตอบบางกลุ่มเสียเปรียบในการตีความและการตอบคำถามโดยไม่เกี่ยวกับสิ่งที่มุ่งวัดจริง มาตรฐานด้านการวัดและประเมินผลยังย้ำว่าความเป็นธรรมเป็นองค์ประกอบพื้นฐานของการพัฒนาและการใช้เครื่องมือ มิใช่ประเด็นที่นำมาพิจารณาภายหลังเมื่อพบปัญหาแล้วเท่านั้น (AERA, APA, & NCME, 2014) ด้วยเหตุนี้ แม้ปัญญาประดิษฐ์จะช่วยตรวจสอบถ้อยคำที่อาจเอนเอียงหรือชี้จุดเสี่ยงในเชิงสำรวจได้ แต่การตัดสินว่าความแตกต่างที่พบเป็นผลจากอคติของเครื่องมือหรือไม่ ยังคงต้องอาศัยการพิจารณาของผู้วิจัยภายใต้กรอบความเป็นธรรมของการวัดอย่างเคร่งครัด

หลักการประการที่ห้า คือ การคุ้มครองข้อมูลและความเหมาะสมของข้อมูลที่ป้อนเข้าสู่ระบบ ในกระบวนการพัฒนาเครื่องมือวิจัย ผู้วิจัยบางกรณีอาจใช้ปัญญาประดิษฐ์เพื่อวิเคราะห์ข้อความจากเอกสารความเห็นของผู้เชี่ยวชาญ หรือข้อมูลตัวอย่างจากกลุ่มเป้าหมายเบื้องต้น จึงจำเป็นต้องพิจารณาอย่างรอบคอบว่า ข้อมูลใดเหมาะสมต่อการป้อนเข้าสู่ระบบ และข้อมูลใดไม่ควรถูกนำไปใช้ โดยเฉพาะข้อมูลที่มีลักษณะระบุตัวบุคคล ข้อมูลอ่อนไหว หรือข้อมูลที่อยู่ภายใต้ข้อตกลงด้านจริยธรรมการวิจัย UNESCO (2023) เน้นว่า การใช้ Generative AI ในการศึกษาและการวิจัยต้องคำนึงถึงความปลอดภัยของข้อมูล สิทธิของผู้เกี่ยวข้อง และความรับผิดชอบต่อผลกระทบที่อาจเกิดขึ้นจากการใช้ระบบ ดังนั้น การคุ้มครองข้อมูลจึงไม่ใช่ประเด็นทางเทคนิคที่แยกขาดจากการพัฒนาเครื่องมือ หากเป็นส่วนหนึ่งของความรับผิดชอบต่อเชิงวิธีวิทยาและเชิงจริยธรรมของผู้วิจัยโดยตรง

หลักการประการที่หก คือ ความซื่อสัตย์ทางวิชาการและความถูกต้องของการอ้างอิง หนึ่งในความเสี่ยงสำคัญของปัญญาประดิษฐ์เชิงสร้างสรรค์คือการสร้างข้อความที่ดูน่าเชื่อถือแต่มีการอ้างอิงคลาดเคลื่อน สร้างแหล่งอ้างอิงที่ไม่มีอยู่จริง หรือสรุปสาระจากทฤษฎีและงานวิจัยอย่างผิดเพี้ยน ซึ่งอาจทำให้กระบวนการนิยามโครงสร้างเชิงแนวคิด การกำหนดตัวชี้วัด และการสร้างเหตุผลรองรับข้อคำถามตั้งอยู่บนฐานข้อมูลที่ไม่ถูกต้องได้ (Flanagin et al., 2024; UNESCO, 2023) ในบริบทของการพัฒนาเครื่องมือวิจัย ความคลาดเคลื่อนเช่นนี้มีนัยรุนแรงกว่าการผิดพลาดในเชิงถ้อยคำทั่วไป เพราะส่งผลต่อฐานทฤษฎีของการวัดโดยตรง ผู้วิจัยจึงจำเป็นต้องตรวจสอบแหล่งอ้างอิงทุกครั้ง ทวนสอบความถูกต้องของสาระที่ระบบสรุป และไม่ใช้ผลลัพธ์จากปัญญาประดิษฐ์เป็นฐานอ้างอิงโดยปราศจากการตรวจสอบจากแหล่งวิชาการต้นทาง

เมื่อพิจารณาโดยรวม การใช้ปัญญาประดิษฐ์อย่างมีความรับผิดชอบในงานสร้างเครื่องมือวิจัยทางการศึกษา จึงมิได้หมายถึงการใช้เทคโนโลยีอย่างระมัดระวังในความหมายกว้างเท่านั้น แต่หมายถึงการทำให้ทุกขั้นตอนของการใช้ปัญญาประดิษฐ์สามารถอธิบายได้ ตรวจสอบได้ และเชื่อมโยงกับหลักการทางวิชาการอย่างเป็นระบบ กล่าวอีกนัยหนึ่ง ผู้วิจัยควรสามารถตอบได้อย่างชัดเจนว่า ใช้ปัญญาประดิษฐ์เพื่ออะไร เหตุใด

จึงใช้ ผลลัพธ์ที่ได้ถูกกลั่นกรองอย่างไร หลักการใดถูกใช้ในการตัดสินใจ และใครเป็นผู้รับผิดชอบต่อการสรุปที่เกิดขึ้นจากเครื่องมือดังกล่าว ในกรอบแนวคิดของบทความนี้ ความรับผิดชอบดังกล่าวมิได้เป็นประเด็นภายนอกของกระบวนการวิจัย แต่เป็นส่วนหนึ่งของความเข้มแข็งเชิงวิธีวิทยาของการพัฒนาเครื่องมือเอง (AERA, APA, & NCME, 2014; Floridi et al., 2018; UNESCO, 2023; Flanagan et al., 2024)

7. ข้อพิจารณาเบื้องต้นในการประยุกต์ใช้กรอบ

แม้กรอบแนวคิดเชิงวิธีวิทยาที่น่าเสนอในบทความนี้จะช่วยจัดระเบียบตรรกะของการประยุกต์ใช้ปัญญาประดิษฐ์ในการพัฒนาเครื่องมือวิจัยทางการศึกษาให้มีความชัดเจนยิ่งขึ้น แต่การนำกรอบดังกล่าวไปใช้ในทางปฏิบัติยังควรดำเนินการด้วยความระมัดระวังและอยู่ภายใต้เงื่อนไขทางวิชาการที่เหมาะสม ทั้งนี้ เพราะกรอบที่เสนอมีสถานะเป็นข้อเสนอเชิงแนวคิดที่มุ่งอธิบายความสัมพันธ์ระหว่างหลักการวัดและประเมินผลกระบวนการพัฒนาเครื่องมือ บทบาทของปัญญาประดิษฐ์ และการกำกับการตัดสินใจโดยมนุษย์ มากกว่าจะเป็นกรอบที่ผ่านการตรวจสอบเชิงประจักษ์ในหลายบริบทแล้วอย่างครบถ้วน ดังนั้น การประยุกต์ใช้กรอบนี้จึงควรถูกมองเป็นแนวทางเบื้องต้นสำหรับการออกแบบกระบวนการทำงานอย่างมีหลักการ มิใช่สูตรสำเร็จที่สามารถนำไปใช้ได้โดยไม่ต้องปรับตามลักษณะของเครื่องมือ ประชากร และบริบทของการวิจัย (AERA, APA, & NCME, 2014; DeVellis & Thorpe, 2021; Kane, 2013)

ข้อพิจารณาประการแรก คือ การนำกรอบนี้ไปใช้ควรเริ่มต้นจากความชัดเจนของโครงสร้างเชิงแนวคิดและจุดมุ่งหมายของการวัด มิใช่เริ่มต้นจากความสามารถของระบบปัญญาประดิษฐ์ที่ผู้วิจัยเข้าถึงได้ ในขณะนั้น กล่าวคือ ผู้วิจัยควรกำหนดให้ชัดเจนว่า ต้องการวัดสิ่งใด ขอบเขตของสิ่งที่มุ่งวัดคืออะไร ตัวชี้วัดใดสะท้อนโครงสร้างเชิงแนวคิดดังกล่าว และผลการวัดจะถูกนำไปใช้เพื่อวัตถุประสงค์ใด จากนั้นจึงพิจารณาว่าปัญญาประดิษฐ์สามารถเข้ามาสนับสนุนกระบวนการใดได้อย่างเหมาะสม หลักคิดเช่นนี้สอดคล้องกับแนวทางการพัฒนาเครื่องมือที่ให้ความสำคัญกับการนิยามสิ่งที่มุ่งวัดให้ชัดเจนก่อนการสร้างตัวชี้วัดและข้อคำถาม และย้ำว่าคุณภาพของการวัดขึ้นอยู่กับความสมเหตุสมผลของความสัมพันธ์ระหว่างโครงสร้างเชิงแนวคิดกับหลักฐานที่ต้องการมากกว่าความสามารถเชิงเทคนิคของเครื่องมือที่ใช้ (DeVellis & Thorpe, 2021; Messick, 1995)

ข้อพิจารณาประการที่สอง คือ ผู้วิจัยควรกำหนดขอบเขตของบทบาทปัญญาประดิษฐ์อย่างเฉพาะเจาะจงในแต่ละขั้นตอนของการพัฒนาเครื่องมือ ภายใต้กรอบนี้ ปัญญาประดิษฐ์เหมาะสำหรับการทำหน้าที่ช่วยสังเคราะห์ ช่วยจัดระบบข้อมูลเชิงแนวคิด ช่วยเสนอทางเลือกของถ้อยคำ หรือช่วยสะท้อนความกำกวมทางภาษาในระดับต้น แต่ไม่ควรถูกใช้ในลักษณะที่แทนที่การตัดสินใจเกี่ยวกับความตรงของการวัด ความเหมาะสมของตัวชี้วัด หรือการตีความคะแนนที่ได้จากเครื่องมือ เหตุผลสำคัญคือ ข้อสรุปในระดับดังกล่าวเกี่ยวข้องกับตรงกับความชอบธรรมของการตีความและการใช้ผลการวัด ซึ่งต้องอาศัยการพิจารณาเชิงวิธีวิทยาและเชิงจริยธรรมของผู้วิจัยและผู้เชี่ยวชาญเป็นสำคัญ การกำหนดขอบเขตบทบาทของปัญญาประดิษฐ์ให้ชัดเจนจึงเป็นเงื่อนไขที่ช่วยรักษาหลักการกำกับการตัดสินใจโดยมนุษย์ไว้ได้อย่างเป็นรูปธรรม (AERA, APA, & NCME, 2014; Kane, 2013; UNESCO, 2023)

ข้อพิจารณาประการที่สาม คือ การประยุกต์ใช้กรอบนี้ควรดำเนินไปพร้อมกับการบันทึกและรายงานการใช้ปัญหาประดิษฐ์อย่างโปร่งใส ในทางปฏิบัติ ผู้วิจัยควรสามารถอธิบายได้ว่า ใช้ปัญหาประดิษฐ์ในขั้นตอนใด ใช้เพื่อวัตถุประสงค์ใด ใช้แพลตฟอร์มหรือโมเดลใด และผลลัพธ์ที่ได้ถูกกลั่นกรอง ปรับแก้ หรือคัดเลือกด้วยเกณฑ์ใดโดยผู้วิจัย ประเด็นนี้มีความสำคัญอย่างยิ่ง เพราะการเปิดเผยกระบวนการใช้ปัญหาประดิษฐ์มิได้เป็นเพียงข้อพึงปฏิบัติด้านจริยธรรมทั่วไป แต่ยังสัมพันธ์กับความสามารถในการตรวจสอบย้อนกลับ ความน่าเชื่อถือของรายงานวิจัย และความรับผิดชอบของผู้วิจัยต่อผลลัพธ์ที่นำเสนอ หากไม่มีการบันทึกและรายงานอย่างชัดเจน ผู้อ่านย่อมประเมินความเหมาะสมของกระบวนการและคุณภาพของการใช้ปัญหาประดิษฐ์ได้ยาก (Flanagin et al., 2024; UNESCO, 2023)

ข้อพิจารณาประการที่สี่ คือ การนำกรอบนี้ไปใช้ควรดำเนินควบคู่กับการตรวจสอบคุณภาพของเครื่องมือในเชิงประจักษ์ ไม่ควรหยุดอยู่เพียงการร่างหรือการกลั่นกรองเชิงข้อความที่ได้รับความช่วยเหลือจากปัญหาประดิษฐ์ แม้ปัญหาประดิษฐ์จะช่วยให้การกำหนดตัวชี้วัด การร่างข้อคำถาม และการตรวจสอบภาษาเบื้องต้นมีความเป็นระบบมากขึ้น แต่คุณภาพของเครื่องมือยังต้องได้รับการพิจารณาผ่านหลักฐานด้านความตรง ความเที่ยง ความแม่นยำของการวัด และความเป็นธรรมในการใช้เครื่องมือกับกลุ่มเป้าหมายจริง กล่าวอีกนัยหนึ่ง กรอบที่นำเสนอในบทความมีหน้าที่ช่วยเสริมความเป็นระบบของกระบวนการพัฒนาเครื่องมือ มิได้ทำหน้าที่แทนการตรวจสอบคุณภาพตามหลักการวัดและประเมินผล ซึ่งยังต้องอาศัยข้อมูลเชิงประจักษ์และการพิจารณาเชิงวิชาการในลำดับต่อไป (AERA, APA, & NCME, 2014; DeVellis & Thorpe, 2021; Messick, 1995)

ข้อพิจารณาประการที่ห้า คือ ผู้วิจัยควรตระหนักว่า ความเหมาะสมของกรอบนี้อาจแตกต่างกันตามลักษณะของเครื่องมือวิจัย ตัวอย่างเช่น เครื่องมือประเภทแบบสอบถามอาจเปิดโอกาสให้ปัญหาประดิษฐ์ช่วยในงานร่างข้อคำถามและตรวจสอบความชัดเจนของภาษาได้มากกว่าเครื่องมือประเภทภาระงานปฏิบัติ ระบุริคหรือแบบประเมินที่ต้องอาศัยการตัดสินใจเชิงคุณภาพหลายมิติอย่างละเอียด ในกรณีหลัง บทบาทของปัญหาประดิษฐ์อาจเหมาะกับการช่วยจัดระบบเกณฑ์ ช่วยตรวจสอบความสม่าเสมอของถ้อยคำ หรือช่วยเสนอทางเลือกของคำอธิบายระดับคุณภาพมากกว่าการเสนอข้อสรุปเชิงคุณภาพโดยตรง ดังนั้น การใช้กรอบนี้อย่างมีประสิทธิภาพจึงต้องคำนึงถึงธรรมชาติของเครื่องมือ ระดับของความซับซ้อนในการตัดสินใจ และลักษณะของหลักฐานที่ต้องการสร้างในแต่ละกรณี (DeVellis & Thorpe, 2021; UNESCO, 2023)

สำหรับนักวิจัยทางการศึกษาในประเทศไทย ข้อพิจารณาเบื้องต้นที่สำคัญอีกประการหนึ่ง คือ การใช้กรอบนี้ควรสอดคล้องกับบริบทของภาษา วัฒนธรรม และรูปแบบการเรียนรู้ของผู้ตอบในแต่ละกลุ่ม แม้ระบบปัญหาประดิษฐ์บางประเภทจะสามารถสร้างข้อความภาษาไทยได้ดีขึ้นอย่างต่อเนื่อง แต่ความสั่นไหวของภาษาไม่ได้รับประกันว่าข้อคำถามที่สร้างขึ้นจะมีความเหมาะสมต่อระดับภาษา ความเข้าใจ และบริบททางสังคมของผู้เรียนไทยทุกกลุ่มเสมอไป ประเด็นนี้มีความสัมพันธ์โดยตรงกับความเป็นธรรมของการวัดและความเสี่ยงด้านอคติทางภาษาและวัฒนธรรม ซึ่งผู้วิจัยต้องพิจารณาอย่างระมัดระวัง โดยเฉพาะเมื่อเครื่องมือถูกนำไปใช้กับผู้ตอบที่มีพื้นฐานทางภาษา ประสบการณ์การเรียนรู้ หรือบริบททางสังคมแตกต่างกันอย่างมีนัยสำคัญ (Floridi et al., 2018; UNESCO, 2023; AERA, APA, & NCME, 2014)

เมื่อสังเคราะห์โดยรวม ข้อพิจารณาเบื้องต้นในการประยุกต์ใช้กรอบนี้มีได้มุ่งเสนอแนวทางใช้ปัญญาประดิษฐ์ให้มากขึ้น หากแต่มุ่งเสนอเงื่อนไขของการใช้ปัญญาประดิษฐ์อย่างมีหลักการในการพัฒนาเครื่องมือวิจัยทางการศึกษา กล่าวคือ การใช้ปัญญาประดิษฐ์จะมีคุณค่าทางวิชาการได้ก็ต่อเมื่อผู้วิจัยสามารถรักษาความสมดุลระหว่างการใช้ศักยภาพของเทคโนโลยีกับการคงไว้ซึ่งฐานทฤษฎีของโครงสร้างเชิงแนวคิด คุณภาพของการวัด ความโปร่งใสของกระบวนการ และความรับผิดชอบของมนุษย์ในทุกขั้นตอนของการตัดสินใจ ทั้งนี้ จุดยืนดังกล่าวสอดคล้องกับแนวคิดการใช้ปัญญาประดิษฐ์อย่างมีความรับผิดชอบ ซึ่งเน้นการกำกับโดยมนุษย์ (human oversight) ความรับผิดชอบ (accountability) ความโปร่งใส (transparency) และความเป็นธรรม (fairness) เป็นแกนกลางของการประยุกต์ใช้เทคโนโลยีในงานวิชาการ (Floridi et al., 2018; Flanagin et al., 2024; UNESCO, 2023)

8. บทสรุป

การประยุกต์ใช้ปัญญาประดิษฐ์ (Artificial Intelligence: AI) ในการพัฒนาเครื่องมือวิจัยทางการศึกษาเป็นประเด็นที่มีความสำคัญเพิ่มขึ้นอย่างต่อเนื่องในบริบทของการวิจัยร่วมสมัย โดยเฉพาะเมื่อปัญญาประดิษฐ์เชิงสร้างสรรค์ (generative AI) และเครื่องมือที่อาศัยแบบจำลองภาษาขนาดใหญ่ (Large Language Models: LLMs) เริ่มเข้ามามีบทบาทในระยะเริ่มต้นของกระบวนการวิจัยมากกว่าที่ผ่านมา อย่างไรก็ตาม บทความนี้ชี้ให้เห็นว่า ศักยภาพของปัญญาประดิษฐ์ในการช่วยสังเคราะห์แนวคิด กำหนดตัวชี้วัด รวบรวมคำถาม และตรวจสอบความชัดเจนของภาษาในระดับเบื้องต้น จะมีความหมายทางวิชาการได้ก็ต่อเมื่อถูกจัดวางภายใต้กรอบเชิงวิธีวิทยา (methodological framework) ที่เชื่อมโยงเทคโนโลยีเข้ากับหลักการวัดและประเมินผลทางการศึกษาอย่างเป็นระบบ มิฉะนั้น การใช้ปัญญาประดิษฐ์อาจกลายเป็นเพียงการเพิ่มความเร็วของการทำงาน โดยไม่สามารถรับประกันคุณภาพของเครื่องมือหรือความชอบธรรมของการตีความผลการวัดได้อย่างเพียงพอ (AERA, APA, & NCME, 2014; Kane, 2013; UNESCO, 2023)

องค์ความรู้สำคัญที่บทความนี้เสนอ คือ กรอบแนวคิดเชิงวิธีวิทยา (methodological framework) สำหรับการประยุกต์ใช้ปัญญาประดิษฐ์เชิงสร้างสรรค์ในการพัฒนาเครื่องมือวิจัยทางการศึกษา ซึ่งบูรณาการ 4 องค์ประกอบหลัก ได้แก่ ฐานแนวคิดการวัดและประเมินผลทางการศึกษา กระบวนการพัฒนาเครื่องมือวิจัย บทบาทของปัญญาประดิษฐ์ในฐานะกลไกสนับสนุนเชิงวิธีวิทยา และการกำกับการตัดสินใจโดยมนุษย์ (human-in-the-loop) กรอบดังกล่าวช่วยให้บทบาทของปัญญาประดิษฐ์ถูกกำหนดขอบเขตอย่างชัดเจนขึ้น จากเดิมที่มักถูกกล่าวถึงในลักษณะกว้างและคลุมเครือ ให้กลายเป็นเครื่องมือช่วยสังเคราะห์ จัดระบบ และสะท้อนทางเลือกเชิงวิธีวิทยาในบางภารกิจ ขณะที่การตัดสินใจเกี่ยวกับโครงสร้างเชิงแนวคิด (construct) ตัวชี้วัด ความตรง (validity) ความเที่ยง (reliability) ความแม่นยำของการวัด (measurement precision) และความเป็นธรรมของการวัด (fairness) ยังคงเป็นความรับผิดชอบของผู้วิจัยและผู้เชี่ยวชาญ หลักคิดดังกล่าวสอดคล้องกับวรรณกรรมด้านการพัฒนาเครื่องมือและการตรวจสอบความตรงร่วมสมัย ซึ่งย้ำว่าคุณภาพของการวัดขึ้นอยู่กับ การเชื่อมโยงระหว่างโครงสร้างเชิงแนวคิด หลักฐาน และการตีความอย่างมีเหตุผล มากกว่าการพึ่งพาความสามารถของเครื่องมือทางเทคโนโลยีเพียงอย่างเดียว (DeVellis & Thorpe, 2021; Messick, 1995; Kane, 2013)

ในแง่นี้ บทความนี้ได้เสนอว่า ปัญญาประดิษฐ์เป็นคำตอบสำเร็จรูปของปัญหาการพัฒนาเครื่องมือวิจัย หากแต่เสนอว่า ปัญญาประดิษฐ์สามารถเป็นผู้ช่วยเชิงวิธีวิทยา (methodological assistant) ที่มีคุณค่าได้เมื่อการใช้งานถูกออกแบบอย่างมีหลักการ โปร่งใส (transparency) และตรวจสอบได้ กรอบที่เสนอมิใช่กรอบสำหรับแทนที่ผู้วิจัย แต่เป็นกรอบสำหรับกำกับการใช้ปัญญาประดิษฐ์ให้สอดคล้องกับมาตรฐานของการวิจัยทางการศึกษา จุดเน้นดังกล่าวมีความสำคัญอย่างยิ่งในบริบทที่เทคโนโลยีปัญญาประดิษฐ์เชิงสร้างสรรค์เปลี่ยนแปลงรวดเร็ว และอาจทำให้ผู้ใช้เกิดความเข้าใจคลาดเคลื่อนว่า แม้ระบบจะสามารถสร้างข้อความหรือสังเคราะห์ภาษาได้ดีในเชิงรูปแบบ แต่ก็ยังไม่เพียงพอต่อการพัฒนาเครื่องมือวิจัยที่มีคุณภาพ ทั้งที่ในความเป็นจริงคุณภาพของเครื่องมือยังต้องได้รับการพิจารณาภายใต้หลักฐานด้านการวัดอย่างรอบด้าน และยังคงอาศัยการกำกับการตัดสินใจโดยมนุษย์ในทุกจุดที่เกี่ยวข้องกับข้อสรุปทางวิชาการ (Holmes et al., 2019; Floridi et al., 2018; UNESCO, 2023)

ขณะเดียวกัน บทความนี้ยังย้ำให้เห็นว่า การใช้ปัญญาประดิษฐ์อย่างมีความรับผิดชอบ (responsible AI use) เป็นส่วนหนึ่งของความเข้มแข็งเชิงวิธีวิทยา มิใช่เพียงประเด็นเสริมด้านจริยธรรม ความโปร่งใสในการเปิดเผยการใช้ปัญญาประดิษฐ์ การกำหนดขอบเขตของบทบาทระบบอย่างชัดเจน การตรวจสอบความเหมาะสมของผลลัพธ์เชิงภาษา การพิจารณาความเป็นธรรมของข้อความ และการคงไว้ซึ่งความรับผิดชอบต่อวิชาการของมนุษย์ ล้วนเป็นเงื่อนไขที่ทำให้การประยุกต์ใช้ปัญญาประดิษฐ์ในการพัฒนาเครื่องมือวิจัยมีความชอบธรรมและน่าเชื่อถือ ประเด็นนี้สอดคล้องกับข้อเสนอด้านจริยธรรมของปัญญาประดิษฐ์และแนวทางการรายงานการใช้ปัญญาประดิษฐ์ในงานวิจัย ซึ่งเน้นการกำกับโดยมนุษย์ (human oversight) ความรับผิดชอบ (accountability) ความโปร่งใส (transparency) และความเป็นธรรม (fairness) เป็นหลักการสำคัญของการใช้เทคโนโลยีในงานวิชาการ (Floridi et al., 2018; Flanagin et al., 2024; UNESCO, 2023)

อย่างไรก็ดี ผู้เขียนตระหนักว่ากรอบแนวคิดที่นำเสนอในบทความนี้ยังมีข้อจำกัดตามธรรมชาติของบทความเชิงแนวคิด (conceptual article) กล่าวคือ แม้กรอบดังกล่าวจะช่วยจัดระเบียบตรรกะของการประยุกต์ใช้ปัญญาประดิษฐ์ในการพัฒนาเครื่องมือวิจัยได้อย่างเป็นระบบ แต่ยังมีข้อยืนยันเชิงประจักษ์ (empirical confirmation) ว่าการใช้กรอบนี้จะให้ผลดีกว่าวิธีอื่นในทุกบริบท นอกจากนี้ ศักยภาพของปัญญาประดิษฐ์ โดยเฉพาะเครื่องมือที่อาศัยแบบจำลองภาษาขนาดใหญ่ ยังแปรผันตามแพลตฟอร์ม โมเดล รุ่น และช่วงเวลาที่ใช้งาน จึงทำให้การนำกรอบนี้ไปประยุกต์ใช้จริงจำเป็นต้องกระทำอย่างระมัดระวัง พร้อมรายงานรายละเอียดของระบบที่ใช้และกระบวนการกลั่นกรองผลลัพธ์อย่างชัดเจน ทั้งนี้ เพื่อเปิดโอกาสให้เกิดการตรวจสอบย้อนกลับ (traceability) และการประเมินความเหมาะสมของกระบวนการในเชิงวิธีวิทยาได้อย่างเพียงพอ (Flanagin et al., 2024; UNESCO, 2023)

โดยสรุป บทความนี้เสนอว่า คุณค่าของการใช้ปัญญาประดิษฐ์ในการพัฒนาเครื่องมือวิจัยทางการศึกษา ไม่ได้อยู่ที่การทำให้กระบวนการเร็วขึ้นเพียงอย่างเดียว แต่อยู่ที่การทำให้กระบวนการดังกล่าวมีความเป็นระบบ โปร่งใส และเชื่อมโยงกับหลักการวัดและประเมินผลได้อย่างชัดเจนยิ่งขึ้น กรอบแนวคิดเชิงวิธีวิทยาที่นำเสนอจึงมีคุณูปการในฐานะจุดตั้งต้นทางวิชาการสำหรับนักวิจัย ผู้สอนวิจัย และนักศึกษาระดับบัณฑิตศึกษา

ที่ต้องการประยุกต์ใช้ปัญญาประดิษฐ์ในการพัฒนาเครื่องมือวิจัยอย่างรอบคอบ มีหลักการ และอยู่ภายใต้การกำกับ การตัดสินใจโดยมนุษย์อย่างรับผิดชอบ พร้อมทั้งเปิดพื้นที่สำหรับการศึกษาต่อยอดในอนาคต เพื่อทดสอบกรอบดังกล่าวในบริบทของเครื่องมือ ประชากร และสถานการณ์วิจัยที่หลากหลายยิ่งขึ้น (AERA, APA, & NCME, 2014; DeVellis & Thorpe, 2021; Floridi et al., 2018; UNESCO, 2023)

References

- AERA, APA, & NCME. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- AI Thailand. (2022). *Thailand National AI Strategy and Action Plan (2022–2027)*.
<https://www.ai.in.th/en/about-ai-thailand/>
- Cohen, L., Manion, L., & Morrison, K. (2018). *Research methods in education* (8th ed.). Routledge. <https://doi.org/10.4324/9781315456539>
- DeVellis, R. F., & Thorpe, C. T. (2021). *Scale development: Theory and applications* (5th ed.). SAGE.
- Downing, S. M. (2006). Twelve steps for effective test development. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 3–25). Lawrence Erlbaum Associates.
- Ebel, R. L., & Frisbie, D. A. (1991). *Essentials of educational measurement* (5th ed.). Prentice Hall.
- Flanagin, A., Pirracchio, R., Khera, R., Berkwits, M., Hswen, Y., & Bibbins-Domingo, K. (2024). Reporting use of AI in research and scholarly publication—JAMA Network guidance. *JAMA*, 331(13), 1096–1098. <https://doi.org/10.1001/jama.2024.3471>
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707. <https://doi.org/10.1007/s11023-018-9482-5>
- Haladyna, T. M., & Rodriguez, M. C. (2013). *Developing and validating test items*. Routledge. <https://doi.org/10.4324/9780203850381>
- Holmes, W., Bialik, M., & Fadel, C. (2019). *Artificial intelligence in education: Promises and implications for teaching and learning*. Center for Curriculum Redesign.
- Kane, M. T. (2013). Validating the interpretations and uses of test scores. *Journal of Educational Measurement*, 50(1), 1–73. <https://doi.org/10.1111/jedm.12000>

- Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, Article 102274. <https://doi.org/10.1016/j.lindif.2023.102274>
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749. <https://doi.org/10.1037/0003-066X.50.9.741>
- UNESCO. (2023). *Guidance for generative AI in education and research*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000386693>
- Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education. *International Journal of Educational Technology in Higher Education*, 16(1), Article 39. <https://doi.org/10.1186/s41239-019-0171-0>



บทความวิจัย

Research Article

Factors Correlated with Digital Competency Among Youth in the 21st Century: A Meta-Analysis

Somkiat Kaewkobsaba¹ and Sirirut Jumnaksarn^{1*}

(Received: September 29, 2025; Revised: January 3, 2026; Accepted: January 6, 2026)

Abstract

Digital competency (DC) is widely recognized as a foundational skill for 21st-century citizens, encompassing technical proficiency, critical thinking, and responsible use of digital tools. This study conducted a meta-analysis to quantitatively synthesize the relationship between various factors and youth DC, following the PRISMA guidelines. The meta-analysis included 42 eligible studies comprising 49 effect sizes focusing on youth (ages 12–24). A random-effects model revealed a large, significant positive correlation between the examined factors and youth digital competency ($r=.525$, Fisher's $z = 0.583$, $p<.001$). However, extremely high heterogeneity was detected ($I^2=99.49\%$). Key subgroup analysis was performed comparing factor types: Internal/Individual Variables (e.g., self-efficacy, cognitive skills) showed an estimated average correlation of $z=0.475$, which was comparable in magnitude to environmental, social norms, and other variables (e.g., parental mediation, SES), which showed $z=0.481$. The findings confirm the overall importance of both intrinsic and extrinsic factors in shaping DC, while the high variance suggests a strong need to identify contextual and methodological moderators.

Keywords: digital competency, digital literacy, youth, meta-analysis, correlates

Introduction

Digital Competency has become one of the most important foundational skills for citizens in the twenty-first century, as digital technology permeates every aspect of life. Beyond merely being technically proficient with digital tools, digital competency includes using digital tools for learning, work, and social interaction in a confident, critical, and responsible manner (Vuorikari et al., 2022). This competency is a combination of knowledge, abilities, and attitudes, per the Digital Competence Framework for Citizens (Dig Comp) developed by the European Commission (Vuorikari et al., 2022). Information and data literacy,

¹ Lecturer, Office of Registration, Records and Evaluation, Sukhothai Thammathirat Open University

* Corresponding Author E-mail: Sirirut.jum@stou.ac.th

communication and teamwork, digital content creation, safety, and problem solving are its five main categories (Vuorikari et al., 2022). A knowledge-based society is built on people's ability to access, assess, manage, and produce information (Rizza, 2014).

Digital competency is essential to the development of youth, a group that has grown up with digital technology, in a variety of areas, such as the intellectual, emotional, and social domains (Gutiérrez Ángel et al., 2022). Strong digital competency is essential for fostering online safety, advancing education (Learning.com Team, 2023), and preparing students for the workforce of the future. However, it has been widely refuted that teenagers are "Digital Natives" and should naturally be proficient in digital technology. According to a number of studies, many teenagers lack the profound comprehension and critical thinking abilities required to successfully negotiate the dangers and complexities of the digital world, including cyberbullying, addiction, and data security (Rizza, 2014). Therefore, a crucial educational and societal agenda is to comprehend the elements that either support or impede youth digital competency.

A significant amount of research has been conducted in the last ten years in an effort to determine the variables associated with teenage digital competency. These elements fall into three general categories: environmental, demographic/contextual, and individual. Cognitive (like critical thinking), affective (like attitudes and self-efficacy), and behavioral (like the frequency and kind of online activities) are examples of individual factors. Interestingly, one important predictor has been found to be internet self-efficacy (Lu & Gu, 2024).

Concurrently, opportunities for access, support, and use of digital technologies are still influenced by demographic and contextual factors like age, gender, and socioeconomic status (SES) (Shi et al., 2024). The "Digital Divide" of today encompasses differences in skills and the advantages of usage, not just access to devices or the internet (Fang & Nie, 2025). Additionally, the educational system, curriculum design, and the role of parental mediation are all important environmental factors that shape youth digital competency (Erstad et al., 2021). But even with the increasing amount of research on this subject, the results are still varied, disjointed, and occasionally incongruous (Kreuder et al., 2024). It is challenging to create a clear and complete picture because of differences in how digital competency is defined and measured, as well as the variety of samples and research settings. Thus, this research was conducted to perform a meta-analysis of correlation coefficients to determine the summary effect size of the relationship between various factors and the digital competency of youth and analyze and present the characteristics of the primary research studies in this field to map the current research landscape.

Rationale for the Present Study

Given this fragmentation of knowledge, this study identifies a significant gap: the lack of a comprehensive quantitative synthesis to evaluate the overall magnitude of the relationship between various factors and youth digital competency. A meta-analysis, a statistical tool for combining and analyzing results from multiple studies, allows for the estimation of a reliable average effect size and a systematic examination of the variance across study findings. Therefore, this research has two primary objectives:

1. To conduct a meta-analysis of correlation coefficients to determine the summary effect size of the relationship between various factors and the digital competency of youth.
2. To analyze and present the characteristics of the primary research studies in this field to map the current research landscape.

Methods

This study adheres to established guidelines for systematic reviews and meta-analyses, such as PRISMA (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) and the Cochrane Handbook, to ensure transparency and replicability (Shaw, 2018).

1. Search Strategy and Study Selection

A systematic literature search was conducted across Scopus, Web of Science, PsycINFO, and ERIC, covering publications up to December 2023. The search combined terms related to digital competency ("digital competence," "digital literacy," "ICT skills"), the target population ("youth," "youth," "student"), and research type ("correlation," "factors," "predictors"). The selection process involved two independent stages: an initial screening of titles and abstracts, followed by a full-text review of potentially relevant articles. Disagreements were resolved via discussion between the two researchers. Inclusion criteria required studies to: 1) use an youth sample (ages 12-24); 2) include a measure of digital competency or a related construct; 3) report the Pearson correlation coefficient (r) or provide sufficient data for its calculation; and 4) be published in a peer-reviewed journal. Exclusion criteria targeted qualitative, experimental (without baseline correlations), non-youth samples, and review/meta-analysis articles.

The initial search yielded 59 studies, of which 52 passed the initial title/abstract screening. Seven studies were excluded at this stage due to: irrelevant common themes ($n=2$), using a working-age sample ($n=3$), being abstract-only ($n=1$), and utilizing inappropriate

statistics ($n=1$). The remaining 52 studies proceeded to full-text review, which led to the exclusion of 10 more articles for the following reasons: the dependent variable's operational definition did not match the scope ($n=3$), insufficient statistical data to calculate the correlation coefficient ($n=4$), and using a non-target sample population ($n=3$). Consequently, a final total of 42 studies were included in the meta-analysis, comprising 38 studies that directly reported the correlation coefficient and 4 studies where the correlation coefficient was calculated from other reported statistics.

2. Statistical Analysis

2.1 Data extraction and effect size calculation

A standardized coding form was developed to extract key information from each eligible study, including bibliographic details, sample characteristics (sample size (N), mean age, educational level), research context (country), instruments used, and the necessary statistical data (correlation coefficient: r). For descriptive purposes, additional characteristics such as research field, sampling method, and data collection method were also coded (Shaw, 2018). The Pearson correlation coefficient (r) served as the primary effect size metric. Prior to the meta-analysis, each (r) value was transformed into Fisher's (z) using the standard Fisher's r -to- z transformation. This process normalizes the sampling distribution and stabilizes the variance, making the data suitable for pooling. The final pooled results were back-transformed to r for interpretation using the formula $r = \tanh(z)$ (Revelle, & Revelle, 2015).

2.2 Meta-analytic model and heterogeneity assessment

Given the observed diversity in study designs, samples, and measurement tools, a **Random-Effects Model** was employed, assuming that the true effect sizes vary across the included studies (Hedges & Olkin, 1985). The model estimates the mean of the distribution of these true effects. The between-studies variance (τ^2) was estimated using the Restricted Maximum-Likelihood (REML) method.

Heterogeneity was formally assessed using Cochran's Q -test and the I^2 statistic (Higgins et al., 2003). I^2 values of approximately 25%, 50%, and 75% were interpreted as low, moderate, and high levels of heterogeneity, respectively. A 95% prediction interval (PI) was also calculated to estimate the plausible range of true effect sizes in future studies.

2.3 Sensitivity and subgroup analyses

To investigate the potential for publication bias, funnel plot asymmetry was assessed visually, complemented by formal statistical tests: Begg and Mazumdar's rank correlation test and Egger's regression test (Shaw, 2018). To explore potential sources of

heterogeneity, a subgroup analysis was conducted to compare effect sizes across samples from internal/individual variables (e.g., self-efficacy, cognitive skills) and environmental, social Norms, and other Variables (e.g., parental mediation, SES), based on consistently extracted moderator data. All statistical analyses were performed using Jamovi Version 2.6, which utilizes the metafor package in R.

Results

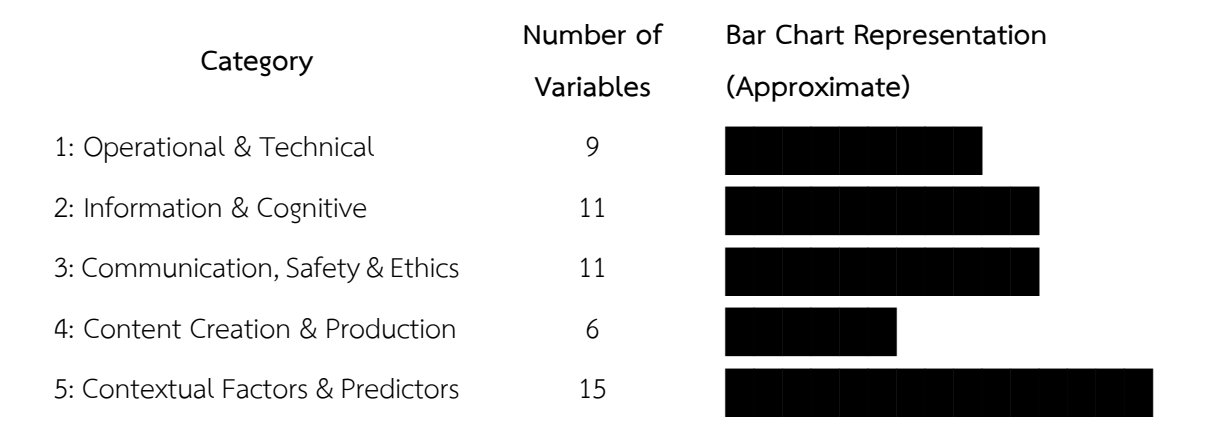
This section presents the findings according to the two research objectives, beginning with a descriptive overview of the selected studies, followed by the results of the meta-analysis.

1. Characteristics of Included Studies

Following the systematic review, 42 studies met the criteria for descriptive analysis (Objective 2), and 49 effect sizes (which could originate from the same article reporting multiple correlations) provided sufficient data for the meta-analysis. Summarizing the key characteristics of the 42 studies reveals several prominent features of the current research landscape. First, research is predominantly concentrated in Asia (54.8%) and Europe (35.7%), with very few studies from other continents. Second, the primary population studied is university students (57.1%), significantly outnumbering secondary school students (31.0%). Third, the majority of studies are situated within the behavioral and social sciences (71.4%) and favor online data collection methods (59.5%).

The summary consolidates and categorizes the variables related to Digital Literacy, Digital Competence, ICT Skills, and ICT Competency identified across the reviewed literature, organizing them into thematic dimensions commonly utilized in research frameworks. The variables are broadly classified into Digital Competence Dimensions and Associated Predictors and Outcomes (contextual factors, psychological states, and learning results). Categorization of digital and ICT-related variables based on the sources, variables can be systematically grouped into five major categories: category1 operational and technical skills (Instrumental Dimension) that encompasses fundamental knowledge and practical ability required to handle digital devices and tools, category2 information and cognitive literacy (Content and Evaluation) that focuses on the higher-order mental processes required to navigate, evaluate, and extract meaning from digital information, category3 communication, safety, and ethical/social dimension that includes interaction, collaboration, security awareness, and adherence to social/legal norms in digital environments, category4 digital

content creation and production that involves generating, adapting, and distributing new digital content, category 5 associated contextual factors, attitudes, and outcomes (Predictors) that variables influence, moderate, or are consequences of Digital/ICT Competence. The following bar chart illustrates the count of unique variables within each major research category, providing a clear overview of the conceptual emphasis in the literature provided.



2. Meta-Analytic Results of Correlation Coefficients for Overall Group

The primary meta-analysis included 49 effect sizes from the eligible studies. The results are summarized in table 1 and discussed in detail below.

Table 1 Random-effects meta-analysis of factors correlated with youth digital competency

Group	k	Fisher's z (95% CI)	r	τ^2	I ² (%)	Q (df)
Overall	49	0.583 [0.492, 0.675]	.525	0.098	99.49	7000.80 (48)***
Subgroups						
Secondary Education	25	0.534 [0.403, 0.666]	.488	0.099	99.56	5471.53 (24)***
Higher Education	24	0.635 [0.502, 0.769]	.562	0.096	98.62	1189.80 (23)***

*Note: k = number of effect sizes, CI = confidence interval, r = back-transformed correlation coefficient, τ^2 = between-studies variance, I² = heterogeneity index, Q = Cochran's Q statistic, df = degrees of freedom. **p<.001.

2.1 Overall Summary Effect Size

The random-effects model analysis of all 49 effect sizes yielded a mean Fisher-transformed correlation of z=0.583 (95% CI [0.492, 0.675], *p*<.001). When back-transformed to a Pearson correlation coefficient, this corresponds to r=.525. According to Cohen's (1988) criteria, this represents a large positive correlation. This result indicates that, on average, the factors examined in these studies have a significant and strong positive relationship with youth digital competency.

2.2 Heterogeneity

The analysis of heterogeneity is one of the most critical findings of this study. Cochran's Q statistic was large and statistically significant ($Q(48) = 7000.800$, $p < .001$), and the I^2 index was extremely high at 99.49%. This indicates that nearly all of the observed variance in effect sizes is due to true differences between studies, rather than sampling error alone. This finding is further corroborated by the 95% prediction interval, which ranged from -0.052 to 1.218. The wide span of this interval, which includes zero and slightly negative values, suggests that although the overall average effect is positive and large, the results of future studies could be highly variable, ranging from no correlation to a very strong positive correlation. (Table1 and Figure 2)

2.3 Risk of Bias

The methodological quality and risk of bias (RoB) for the included cross-sectional studies were assessed using the adapted Newcastle-Ottawa Scale for Cross-Sectional Studies (NOS-CS), with a maximum total score of 10 points. The overall RoB assessment revealed that the majority of studies (66.67%) were classified as having a low risk of bias (scoring 7–10 points), while the remaining third (33.33%) were categorized as having an unclear risk of bias (scoring 4–6 points), and importantly, no studies were rated as high risk (0.00%). Domain-level analysis indicated the highest methodological quality in the Outcome domain, with 71.43% of studies achieving a low risk score (3 points), whereas the Selection and Comparability domains showed low risk in 54.76% of the studies each. Furthermore, only a negligible number of studies demonstrated a high risk of bias across all domains: Selection (4.76%), Comparability (2.38%), and Outcome (2.38%), suggesting a generally robust and acceptable quality among the studies included in this meta-analysis (Table2 and Figure1).

Table 2 Risk of Bias (RoB) Assessment

RoB list	RoB level (points)	Percentage (%)
Selection	low risk (4-5)	54.76
	unclear risk (3)	40.48
	high risk (0-2)	4.76
comparability	low risk (2)	54.76
	unclear risk (1)	42.86
	high risk (0)	2.38
Outcome	low risk (3)	71.43
	unclear risk (2)	26.19
	high risk (0-1)	2.38
Overall RoB	low risk of bias (7-10)	66.67
	unclear risk of bias (4-6)	33.33
	high risk of bias (0-3)	0.00

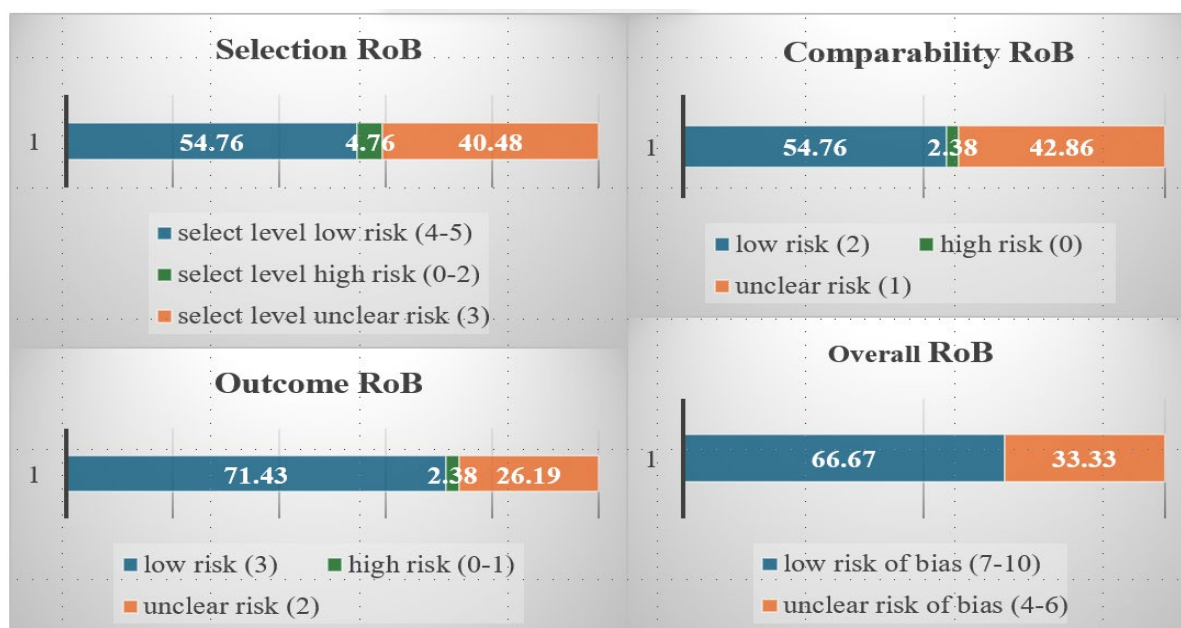


Figure 1 Risk of Bias (RoB) Assessment

2.4 Publication Bias

Publication bias was assessed via visual inspection of a funnel plot, which plots effect sizes against their standard errors. Visually, the funnel plot appeared reasonably symmetric. This observation was supported by statistical tests: both Begg's rank correlation test ($p=.856$) and Egger's regression test ($p=.431$) found no evidence of significant asymmetry. Therefore, it is unlikely that publication bias has substantially affected the validity of the summary effect size in this study and shown in table 3. Summary, the funnel plot indicates that while the results of this meta-analysis are likely not distorted by publication bias, they are characterized by a very high degree of heterogeneity, meaning the relationship between the factors and digital competency varies significantly from one study to another (Figure 2)

Table 3 Publication Bias Assessment

Test Name	value	p
Fail-Safe N	284113.000	<.001
Begg and Mazumdar Rank Correlation	-0.018	0.856
Egger's Regression	0.794	0.431
Trim and Fill Number of Studies	0.000	.

Note. Fail-safe N Calculation Using the Rosenthal Approach

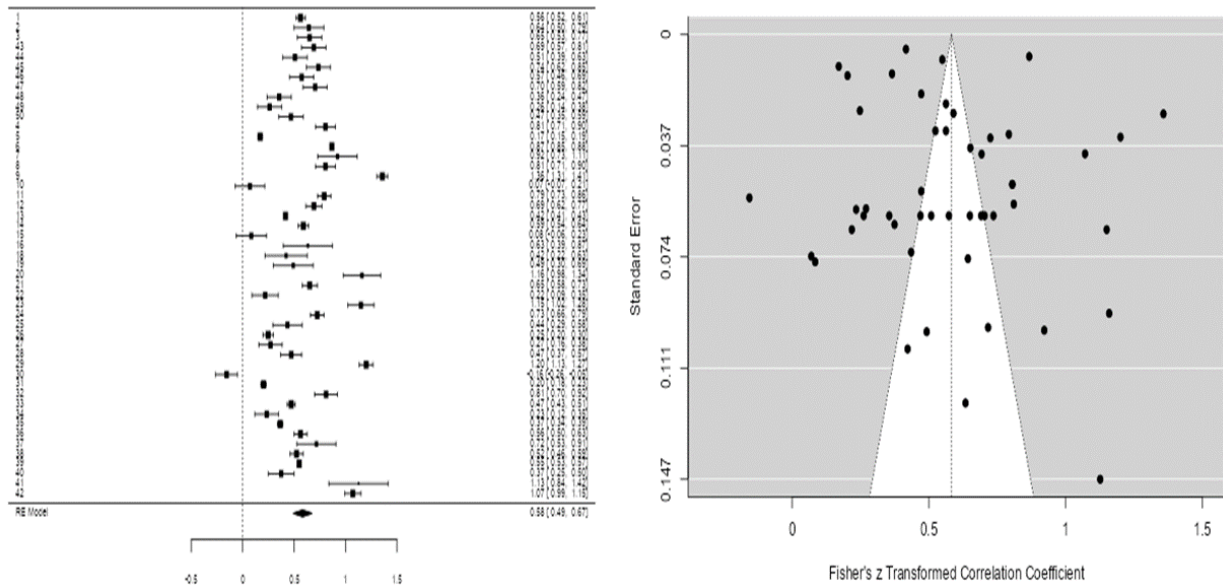


Figure 2 Funnel plot and Forest plot for Overall Group

3. Meta-Analytic Results of Correlation Coefficients for Subgroup

3.1 Subgroup Analysis 1: Internal/Individual Variables

Meta-Analytic Effect Size. This subgroup, representing internal variables (cognitive, affective, and skill domains) included $k_1 = 20$ effect sizes. The estimated average Fisher's r -to- z transformed correlation coefficient was 0.475 (95% CI: 0.436 to 0.514). The average outcome was highly significant ($z = 24.023$, $p < 0.000$).

Heterogeneity. Significant heterogeneity was again detected ($Q(19) = 7284.713$, $p < 0.000$), with a high I^2 of 98.19% and τ^2 of 0.041. The 95% prediction interval for the true outcomes was 0.078 to 0.872. Since the entire interval is positive, the true outcomes of the studies in this subgroup are generally in the same positive direction as the estimated average outcome.

Outliers and Publication Bias. No outliers ($+/-3.506$) or overly influential studies were identified. However, the analysis for publication bias yielded mixed results. The Begg and Mazumdar Rank Correlation Test indicated significant funnel plot asymmetry ($p = 0.046$), while the Egger's Regression Test did not ($p = 0.458$). The Fail-Safe N was very large at 838,818. The visual inspection of the funnel plot below suggests a potential for bias.

3.2 Subgroup Analysis 2: Environmental, Social Norms, and Other Variables

Meta-Analytic Effect Size. This subgroup included $k_2 = 29$ effect sizes focused on environmental variables, social norms, and other external factors. The estimated average

Fisher's r-to-z transformed correlation coefficient was 0.481 (95% CI: 0.411 to 0.550). This effect size is significant ($z=13.576$, $p<0.001$) and comparable in magnitude to subgroup 1.

Heterogeneity. Similar to the other analyses, there was a high degree of heterogeneity detected ($Q(28) = 5109.390$, $p < 0.001$), with an I^2 of 99.20% and τ^2 of 0.047. The 95% prediction interval was entirely positive, ranging from 0.052 to 0.909. This confirms that the true outcomes for studies in this subgroup are generally in the same positive direction as the estimated average outcome.

Outliers and Publication Bias. No outliers (± 3.2200) or influential studies were identified in this subgroup. For publication bias, neither the Begg and Mazumdar Rank Correlation Test ($p=0.827$) nor the Egger's Regression Test ($p=0.972$) indicated any statistically significant funnel plot asymmetry. The Fail-Safe N was 212,890.

Discussion

This study aimed to synthesize a quantitative overview of the factors correlated with youth digital competency. The findings reveal a complex and compelling picture, which will be discussed in the following sections.

1. Synthesizing the Findings: A Strong but Inconsistent Relationship

The primary finding of this meta-analysis is twofold and must be considered in tandem. On one hand, the summary effect size ($r = .525$) indicates a strong positive correlation on average. This provides quantitative confirmation that the factors researchers have focused on—be they individual, social, or environmental—have a significant connection to the development of youth digital competency. This result aligns with a substantial body of recent research that consistently links higher digital competency to factors such as internet self-efficacy, active parental mediation, and engagement in informal digital learning. On the other hand, an equally important finding is the extremely high level of heterogeneity ($I^2 = 99.5\%$), which serves as a crucial caveat that this average value masks immense variability among individual studies. The practical implication of this high heterogeneity is reflected in the wide 95% prediction interval that includes zero. This means we cannot confidently apply the average $r = .525$ to predict the relationship in any specific context, as the true effect in a particular situation could differ dramatically from this mean. Therefore, the main conclusion is not simply that "there is a positive relationship," but rather that "this positive relationship is highly variable and context-dependent."

2. Analyzing Heterogeneity: A Symptom of a Developing Field

This extreme heterogeneity is not merely a statistical limitation but a significant finding in itself, reflecting the current state of knowledge in this field. The sources of this heterogeneity can be attributed to several interconnected causes:

2.1 Conceptual Ambiguity. A primary cause is the lack of a standardized conceptual framework and operational definition for "digital competency." While frameworks like Dig Comp are widely acknowledged, their practical application varies. For instance, whereas one study might define competence through the lens of digital content creation, another might prioritize critical information evaluation and spotting **misinformation**, inevitably leading to different correlational patterns.

2.2 Methodological Diversity. The analysis revealed a wide variety of measurement instruments, often with inadequate reporting of their psychometric properties. This is a critical issue, as the choice of instrument—whether self-report questionnaires versus performance-based tasks—can significantly alter outcomes. Recent research by Hatlevik (2022) demonstrates that self-perceived competence often does not align with actual demonstrated skills, contributing significantly to measurement variance.

2.3 Diversity of Factors Studied. One of the key findings of this meta-analysis is the diversity of factors associated with youth digital competency. Internal factors (e.g., cognitive abilities, self-efficacy, and digital skills) and external factors (e.g., socioeconomic status, family support, educational environment, social norms, and technology access) produced comparable effect sizes ($z = 0.475$ and $z = 0.481$, respectively). This suggests that digital competency is a multidimensional construct shaped by the combined influence of personal and contextual factors rather than by a single dominant predictor. This finding is consistent with Scheerder et al. (2017), who concluded that digital skills are influenced by multiple interacting determinants, including demographic characteristics, educational opportunities, socioeconomic resources, and technology access. Likewise, Livingstone et al. (2023) emphasized that the benefits of digital skills depend on both learners' capabilities and the contexts in which those skills are developed and applied. Together, these studies support the view that the diversity of predictors identified in this meta-analysis reflects the multifaceted nature of digital competency. The high heterogeneity observed in both subgroups ($I^2 = 98.19\%$ and 99.20%) further suggests that the effects of these predictors vary across populations, educational settings, and research contexts. Future studies should

therefore adopt integrative models that consider both individual and contextual factors to better explain variations in youth digital competency.

2.4 Contextual Differences. Another important finding of this meta-analysis is the substantial variation in effect sizes across different educational, cultural, and socioeconomic contexts. Although both internal and external factors showed significant positive relationships with youth digital competency, the exceptionally high heterogeneity ($I^2 = 98.19\%$ and 99.20%) suggests that contextual conditions moderate these relationships. This finding is consistent with Helsper (2012), who argued that digital inclusion and exclusion are embedded within broader social and educational structures rather than determined solely by individual characteristics. Similarly, Scheerder et al. (2017) concluded that digital skills are influenced by multiple interacting determinants, including socioeconomic resources, educational opportunities, demographic characteristics, and technology access, with their effects varying across contexts. In addition, Gottschalk and Weise (2023) emphasized that digital competency develops within an educational ecosystem that includes infrastructure, teacher support, curriculum implementation, and equitable access to technology. Together, these findings suggest that contextual diversity is a plausible explanation for the substantial between-study heterogeneity observed in this meta-analysis. Therefore, interventions to enhance youth digital competency should be adapted to local educational and social conditions. Future research should further examine contextual moderators, such as educational systems, socioeconomic environments, and cultural characteristics, to better explain variations in digital competency across diverse settings.

Recommendation

1. Theoretical Implications

The findings highlight the need for greater theoretical and methodological consistency in the field. A clearer conceptual framework for digital competency is required, together with measurement instruments whose psychometric properties have been rigorously validated. Such instruments should reflect the multidimensional nature of digital competency, including cognitive, skill-based, and attitudinal components.

2. Practical and Policy Implications

For policymakers and educators, the findings highlight the need for caution when generalizing results from a single study to develop one-size-fits-all policies. The substantial heterogeneity indicates that educational interventions should be tailored to specific contexts

and informed by needs assessments before large-scale digital literacy programs are implemented. The findings also point to the need for further research among understudied groups, particularly rural and out-of-school youth, to promote greater equity.

Limitations and Future Research Directions

This study has several limitations that should be considered. The most significant limitation is the inability to conduct a meta-regression to statistically investigate the sources of heterogeneity, due to the diversity and inconsistent reporting of independent variables in the primary studies. Other limitations include restricting the analysis to English-language publications, which may introduce a language bias, and the aggregation of diverse factors, which necessitates careful interpretation of the mean effect size. Based on these limitations, several directions for future research can be proposed: 1) Primary studies must adhere to higher standards for reporting methodology and the psychometric properties of their instruments. 2) Future meta-analyses should be more narrowly focused on specific groups of predictors to draw clearer conclusions. 3) There is a critical need for more research in underrepresented regions and populations to achieve a more complete and global understanding.

Conclusion

This meta-analysis has synthesized the quantitative evidence on the relationship between various factors and the digital competency of youth. The results show that, on average, this relationship is positive and relatively large in magnitude. However, the most significant contribution of this study may not be the presentation of this average, but the revelation of extremely high heterogeneity, which reflects the diversity and complexity of the current research landscape. This finding indicates that the strength of the correlation is highly dependent on the factors studied, the population, and the surrounding context. Therefore, future research must be directed toward establishing clearer conceptual and methodological standards, which will serve as a foundation for developing policies and practices that can genuinely foster the digital competency essential for all youth to thrive in the 21st century.

References

- Cohen, J. (1988). Set correlation and contingency tables. *Applied Psychological Measurement, 12*(4), 425–434. <https://doi.org/10.1177/014662168801200410>
- Erstad, O., Kjällander, S., & Järvelä, S. (2021). Facing the challenges of ‘digital competence’: A Nordic agenda for curriculum development for the 21st century. *Nordic Journal of Digital Literacy, 16*(2), 77–87. <https://doi.org/10.18261/issn.1891-943x-2021-02-04>
- Fang, W., & Nie, C. (2025). Inequalities in digital literacy: Exploring the disparity in tangible outcomes of internet use among college students in China. *Frontiers in Communication, 10*, Article 1601240. <https://doi.org/10.3389/fcomm.2025.1601240>
- Gottschalk, F., & Weise, C. (2023). *Digital equity and inclusion in education: An overview of practice and policy in OECD countries* (OECD Education Working Papers No. 299). OECD Publishing. <https://doi.org/10.1787/7cb15030-en>
- Gutiérrez Ángel, N., Mercader Rubio, I., Trigueros Ramos, R., Oropesa Ruiz, N. F., García-Sánchez, J. N., & García Martín, J. (2022). Digital competence, use, actions and time dedicated to digital devices: Repercussions on the interpersonal relationships of Spanish youth. *International Journal of Environmental Research and Public Health, 19*(16), Article 10358. <https://doi.org/10.3390/ijerph191610358>
- Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis*. Academic Press.
- Helsper, E. J. (2012). A corresponding fields model for the links between social and digital exclusion. *Communication Theory, 22*(4), 403–426. <https://doi.org/10.1111/j.1468-2885.2012.01416.x>
- Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ, 327*(7414), 557–560. <https://doi.org/10.1136/bmj.327.7414.557>
- Kreuder, A., Frick, U., Rakoczy, K., & Schlittmeier, S. J. (2024). Digital competence in adolescents and young adults: A critical analysis of concomitant variables, methodologies and intervention strategies. *Humanities and Social Sciences Communications, 11*, Article 48. <https://doi.org/10.1057/s41599-023-02501-4>
- Learning.com Team. (2023, October 5). *Why digital literacy skills are important for students*. Learning.com. <https://www.learning.com/blog/reasons-digital-literacy-is-important-for-students/>
- Livingstone, S., Mascheroni, G., & Stoilova, M. (2023). The outcomes of gaining digital skills for young people’s lives and wellbeing: A systematic evidence review. *New Media & Society, 25*(5), 1176–1202. <https://doi.org/10.1177/14614448211043189>

- Lu, C., & Gu, M. M. (2024). A systematic review and meta-analysis of factors and outcomes of digital citizenship among youth. *Asia Pacific Journal of Education*, 45(4), 1–16.
<https://doi.org/10.1080/02188791.2023.2296352>
- Revelle, W., & Revelle, M. W. (2015). psych [Computer software]. Comprehensive R Archive Network. <https://cran.r-project.org/package=psych>
- Rizza, C. (2014). Digital competences. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 1614–1619). Springer. https://doi.org/10.1007/978-94-007-0753-5_731
- Scheerder, A., van Deursen, A., & van Dijk, J. (2017). Determinants of internet skills, uses and outcomes: A systematic review of the second- and third-level digital divide. *Telematics and Informatics*, 34(8), 1607–1624. <https://doi.org/10.1016/j.tele.2017.07.007>
- Shaw, A. (2018). *Meta analysis: An introduction* [PowerPoint slides]. The University of Sydney. <https://sydneyuni.atlassian.net/wiki/spaces/SIH/pages/390168761/Meta+analysis+an+introduction>
- Shi, X., He, J., & Niu, G. (2024). The association between family socioeconomic status and children's digital literacy: The explanatory role of parental mediation. *Adolescents*, 4(3), 386–395. <https://doi.org/10.3390/adolescents4030027>
- Vuorikari, R., Kluzer, S., & Punie, Y. (2022). *DigComp 2.2: The digital competence framework for citizens: With new examples of knowledge, skills and attitudes* (EUR 31006 EN). Publications Office of the European Union. <https://doi.org/10.2760/115376>

Establishing CEFR-Referenced Cut Scores for the Royal Police Cadet Academy Test of English Proficiency (RPCA-TEP) Using the Modified Angoff Method

Komjit Chaiyo¹ and Pitupoom Nakprada^{2*}

(Received: November 16, 2025; Revised: January 31, 2026; Accepted: February 11, 2026)

Abstract

The Royal Police Cadet Academy Test of English Proficiency (RPCA-TEP) was designed to evaluate the English language proficiency of police cadets in accordance with the mandate of the Higher Education Commission, which requires English proficiency testing aligned with the Common European Framework of Reference (CEFR). This study aimed to align RPCA-TEP scores with CEFR proficiency levels. Using the Modified Angoff method, 13 subject-matter experts provided judgments on the likelihood of borderline candidates at CEFR levels A2, B1, and B2 answering test items correctly. The analysis established cut-off scores out of 120, with candidates scoring 26–55 classified as A2, 56–82 as B1, and 83–120 as B2. The findings highlight the potential of the RPCA-TEP as a CEFR-referenced assessment tool and recommend further refinement using CEFR ability descriptors to guide future test development.

Keywords: English proficiency test, CEFR-referenced, cut-off scores, Modified Angoff method

Introduction

English proficiency is a core professional competence for modern police officers in Thailand. Officers routinely interact with foreign tourists, expatriates, international organizations, and transnational law-enforcement partners, making the ability to communicate accurately and appropriately in English operationally critical. Miscommunication in police work can have serious consequences, including failure to obtain accurate witness statements, mishandling of victims, and breakdowns in international cooperation (Suwannaphim, 2011; Sae-Jeng, 2018).

¹ Assistant Professor, Department of English, Faculty of Social Science, Royal Police Cadet Academy

² Lecturer, Department of English, Faculty of Social Science, Royal Police Cadet Academy

Research grant from Royal Police Cadet Academy

* Corresponding Author E-mail: pitupoom@gmail.com

To address these challenges, Thailand's higher-education policy requires universities and professional academies to define English proficiency standards using international reference frameworks, particularly the Common European Framework of Reference for Languages (CEFR) (Higher Education Standards Commission, 2024). The CEFR provides a six-level scale (A1–C2) that describes what learners “can do” in real-world communication, allowing institutions to make language requirements transparent, interpretable, and comparable across programs (Council of Europe, 2001).

However, adopting CEFR levels in policy does not automatically ensure that locally developed tests actually measure those levels. CEFR alignment requires that cut scores—the score points used to classify examinees into levels such as A2, B1, or B2—be empirically and theoretically justified. A cut score is defined as a threshold on a test score scale that separates performance categories corresponding to qualitatively different levels of ability (Cizek & Bunch, 2007). Without defensible cut scores, CEFR labels attached to test results become arbitrary and may misrepresent candidates' true communicative competence.

The Need for CEFR-referenced Cut Scores for RPCA-TEP

The Royal Police Cadet Academy (RPCA) uses the Royal Police Cadet Academy - Test of English Proficiency (RPCA-TEP) as one of the instruments for certifying cadets' English readiness for graduation and professional deployment. While cadets may also submit scores from external tests (e.g., TOEIC, IELTS), RPCA-TEP remains the only assessment specifically designed to reflect the linguistic demands of Thai police work. Despite its institutional importance, RPCA-TEP has not yet been formally aligned to CEFR levels through a standard-setting process.

This creates a critical knowledge and policy gap: RPCA currently uses RPCA-TEP scores for high-stakes decisions (graduation and certification), yet there is no defensible basis for interpreting those scores as representing CEFR levels such as A2, B1, or B2. Without CEFR-linked cut scores, it is unclear whether cadets who “pass” RPCA-TEP actually meet the communicative standards required by national higher-education policy or by international policing contexts. This study addresses that gap by establishing CEFR-referenced cut scores for RPCA-TEP.

Alignment of RPCA-TEP Construct with CEFR Descriptors in Police Communication Contexts

To clarify how the RPCA-TEP construct aligns with CEFR proficiency descriptors, Table 1 maps the test sections to CEFR levels A2, B1, and B2 and illustrates how these abilities correspond to communicative tasks encountered in police work. This mapping demonstrates that the RPCA-TEP construct covers receptive language abilities that are directly relevant to operational policing contexts, thereby supporting the rationale for linking RPCA-TEP performance to CEFR levels.

Table 1 Alignment of RPCA-TEP Sections with CEFR Descriptors and Police Communication Tasks

CEFR Level	CEFR Descriptor	RPCA-TEP Sections	Examples of Police Communication Tasks
A2	Can understand and use everyday expressions and simple sentences related to personal information, places, work, and daily life. Can exchange basic information and describe simple background details, needs, and surroundings.	Listening, Vocabulary, Reading	Understanding simple tourist inquiries, reading short notices, identifying personal information from documents, following simple spoken instructions from foreigners
B1	Can understand the main points of clear standard input on familiar matters such as work, school, and leisure. Can handle situations likely to arise while traveling. Can describe experiences, events, and give brief explanations.	Listening, Reading, Language Structure, Vocabulary	Understanding incident reports, reading statements, handling routine interactions with tourists, understanding complaints, explaining procedures in simple English
B2	Can understand extended speech and more complex texts. Can read and comprehend detailed written information. Can communicate fluently and accurately in most situations.	Listening, Reading (longer texts), Language Structure	Understanding detailed reports, reading legal or procedural documents, following complex spoken briefings, interpreting detailed written information from foreign sources

Standard Setting and the Choice of the Modified Angoff Method

This study focuses explicitly on standard setting, not score equating or test linking. Standard setting refers to the process of establishing cut scores that define performance levels on a test (Cizek & Bunch, 2007). Among available standard-setting methods, the Modified Angoff method was selected because it is theory-driven, judgment-based, and widely used for criterion-referenced and certification testing.

In the Modified Angoff procedure, trained subject-matter experts estimate the probability that a minimally qualified candidate at each CEFR level (e.g., A2, B1, B2) would answer each test item correctly. These item-level judgments are aggregated to produce cut scores that are directly anchored in performance expectations rather than statistical distributions (Hambleton & Pitoniak, 2006).

This approach is particularly appropriate for CEFR alignment because CEFR itself is a performance-descriptor framework, not a statistical scale. By asking experts to conceptualize minimally competent CEFR-level candidates and evaluate how they would perform on RPCA-TEP items, the Modified Angoff method operationalizes CEFR descriptors in a defensible psychometric way. Prior research has shown that Angoff-based methods provide reliable and interpretable cut scores when experts are well trained and the construct is clearly defined (Cizek & Bunch, 2007).

Objective

The objective of this study is to establish CEFR-referenced cut scores for RPCA-TEP at the A2, B1, and B2 levels using the Modified Angoff standard-setting method, thereby providing a defensible framework for interpreting RPCA-TEP scores.

Method

1. Population and sample

Thirteen panelists were recruited using purposive, criterion-based sampling, consistent with recommendations that Angoff panels include approximately 10–15 experts to ensure stability of cut-score estimates (Cizek, 2012; Cizek & Bunch, 2007; Tannenbaum & Cho, 2014).

To qualify, panelists were required to: (a) have at least five years of experience teaching English at the tertiary level; (b) be involved in English testing, curriculum design, or test development; (c) demonstrate familiarity with CEFR descriptors; and (d) have experience

with receptive-skills instruction. Preference was given to experts with experience in ESP or policing-related English. Collectively, the panel represented a balance of instructional, assessment, and domain expertise, which is considered essential for defensible standard-setting outcomes.

2. Research instruments

2.1 RPCA-TEP

The RPCA-TEP is a 120-item multiple-choice test comprising Listening (40), Grammar (20), Vocabulary (20), and Reading (40). Approximately 40% of items are in policing-related contexts, with the remainder drawn from general domains.

The test was piloted with 280 fourth-year RPCA cadets, and classical item statistics (item difficulty and discrimination) were computed. These statistics were used to support expert judgment in Round 2 of the Angoff procedure. In line with standard-setting best practice, the pilot data were used to inform but not to override expert judgment (Hambleton & Pitoniak, 2006; Cizek & Bunch, 2007).

Content validity was supported through alignment of test specifications and items with CEFR descriptors and RPCA curriculum objectives, and through expert review during test development.

2.2 Angoff rating materials

A structured rating booklet was developed containing test items, probability-rating fields, CEFR borderline profiles, and rating instructions. The materials were trialed using 10 practice items during training to ensure consistent interpretation of the task.

3. Data collection

Prior to the standard-setting procedure, panelists received training on CEFR descriptors, RPCA-TEP test specifications, and the Modified Angoff method. This was followed by a calibration activity using practice items to ensure a shared understanding of borderline candidate performance at CEFR levels A2, B1, and B2.

The Modified Angoff process was conducted across three rounds. In Round 1, panelists independently estimated the probability that borderline candidates at each CEFR level would answer each test item correctly. In Round 2, panelists were provided with summaries of their initial ratings together with pilot item statistics, including item difficulty and discrimination indices, and were invited to revise their estimates accordingly. In Round 3,

panelists reviewed the aggregated results, engaged in structured discussion to resolve discrepancies, and finalized their judgments.

The inclusion of empirical item statistics in Round 2 follows best practice in data-informed Modified Angoff procedures and has been shown to improve the consistency and realism of expert judgments (Hambleton & Pitoniak, 2006; Cizek & Bunch, 2007; Tannenbaum & Cho, 2014).

4. Data analysis

Final CEFR cut scores were derived from the panelists' Round 3 judgments through a structured aggregation process. For each test item and each target CEFR level (A2, B1, and B2), the probability estimates provided by the 13 panelists were compiled and averaged to obtain an item-level mean probability. This value represented the expected likelihood that a borderline candidate at the specified CEFR level would answer that item correctly.

Because the Angoff judgments addressed the question of what percentage of items borderline candidates would answer correctly, the Round 3 mean cut scores were first expressed as percentages (out of 100). These percentage values were then converted proportionally to the 120-point RPCA-TEP reporting scale to establish the operational cut scores. The CEFR score bands were subsequently defined using these cut scores as boundaries.

Finally, the stability of the standard-setting process was evaluated by examining changes in the dispersion of panelists' estimates across the three rating rounds. Reductions in variability from Round 1 to Round 3 were interpreted as evidence of increasing convergence of expert judgment.

Results

This section presents the outcomes of the Modified Angoff standard-setting procedure conducted across three rounds of expert judgment. The results first demonstrate the convergence of expert estimates across rounds and then present the operational CEFR-linked cut scores derived from Round 3.

Convergence of expert judgments across rounds

Across CEFR levels A2, B1, and B2, a consistent pattern of decreasing standard deviations was observed from Round 1 to Round 3, indicating increasing agreement among panelists as they became familiar with CEFR descriptors, reviewed item statistics, and engaged in structured discussion. Tables 2–4 present the mean estimates and standard deviations for each test section across rounds.

Table 2 Convergence of Expert Judgments Across Rounds for CEFR Level A2

Section	Round 1		Round 2		Round 3	
	Mean	SD	Mean	SD	Mean	SD
Listening	20.39	16.71	18.16	14.03	17.84	12.97
Language Structure	29.21	18.33	26.25	16.18	25.03	15.15
Vocabulary	28.08	15.29	24.78	14.20	24.32	12.10
Reading	22.73	19.36	21.92	18.74	20.36	16.95
Total	25.10	17.42	22.77	15.78	21.88	14.29

Table 3 Convergence of Expert Judgments Across Rounds for CEFR Level B1

Section	Round 1		Round 2		Round 3	
	Mean	SD	Mean	SD	Mean	SD
Listening	47.92	15.42	45.23	14.11	43.59	12.06
Language Structure	54.17	16.33	52.42	14.26	50.26	13.12
Vocabulary	53.10	14.87	51.57	13.74	50.93	11.37
Reading	45.79	17.11	43.63	15.98	43.14	14.53
Total	50.24	15.93	48.21	14.52	46.98	12.77

Table 4 Convergence of Expert Judgments Across Rounds for CEFR Level B2

Section	Round 1		Round 2		Round 3	
	Mean	SD	Mean	SD	Mean	SD
Listening	70.19	16.84	69.24	15.54	67.17	14.15
Language Structure	76.00	15.16	75.53	14.82	73.41	14.33
Vocabulary	73.64	16.58	72.32	14.14	70.68	12.92
Reading	67.38	17.45	66.94	16.01	64.43	15.81
Total	71.80	16.50	71.00	15.12	68.92	14.30

The decreasing standard deviations confirm convergence of expert judgments during the Modified Angoff process. Following this convergence, the Round 3 results were adopted as the operational cut scores.

Operational cut scores from Round 3

Table 5 presents the Round 3 mean percentage cut scores representing the expected proportion of test items that borderline candidates at each CEFR level would answer correctly.

Table 5 Post-discussion Operational Cut Scores from Round 3

CEFR Level	Round 3 Mean Cut Score (%)
A2	21.88
B1	46.98
B2	68.92

Because these values represent percentages (out of 100), they were proportionally converted to the 120-point RPCA-TEP reporting scale to establish operational cut scores, as shown in Table 6.

Table 6 Cut Scores After Proportional Conversion

Cut Score	CEFR Level
26	A2
56	B1
83	B2

These cut scores were then used to define CEFR score bands on the RPCA-TEP scale.

CEFR score bands for RPCA-TEP

Table 7 RPCA-TEP Score Ranges Corresponding to CEFR Levels

RPCA-TEP Score Range	CEFR Level
26 - 55	A2
56 - 82	B1
83 - 120	B2

RPCA-TEP scores within these ranges correspond to CEFR levels A2, B1, and B2 respectively.

Discussion

Interpretation of the CEFR-linked cut scores

The primary objective of this study was to establish CEFR-linked cut scores for the RPCA-TEP using the Modified Angoff method. This objective was successfully achieved. The operational score bands derived from Round 3 were: A2 (26–55), B1 (56–82), and B2 (83–120) on the 120-point RPCA-TEP scale. These score ranges represent the expected performance of

borderline candidates at each CEFR level on the RPCA-TEP. The cut scores therefore provide a meaningful basis for interpreting RPCA-TEP results in relation to CEFR proficiency descriptors, particularly in listening and reading, which constitute the core construct of the test.

Quality and defensibility of the cut scores

The quality of these cut scores is supported by the convergence of expert judgments across the three Angoff rounds. Standard deviations decreased consistently from Round 1 to Round 3 at all CEFR levels, indicating increasing agreement among panelists regarding borderline performance. This convergence reflects effective expert familiarization with CEFR descriptors, the use of empirical item statistics, and structured panel discussion.

Ricker (2006) highlights that iterative rounds, structured panel discussion, and presentation of empirical item statistics are key features that reduce variability among experts and improve agreement in cut-score estimation. The pattern observed in this study therefore aligns closely with these recommended practices in standard-setting research. Such convergence is widely recognized as evidence of stable and defensible cut scores rather than arbitrary estimates.

Uncertainty, fairness, and careful interpretation

Despite their defensibility, these cut scores should be interpreted as decision thresholds rather than precise indicators of overall language ability. Because RPCA-TEP currently assesses only receptive skills, it does not represent the full range of CEFR descriptors, particularly those related to speaking and writing. Users of the test must therefore avoid overgeneralizing that candidates classified at a CEFR level can perform all functions described at that level. Classification uncertainty is greatest for candidates whose scores fall near the cut points, and careful interpretation is required to ensure fairness in decision-making.

Practical use of the score bands at RPCA

The CEFR-linked score bands provide RPCA with a clear framework for interpreting test results. They can support graduation decisions, diagnostic placement, curriculum alignment, and internal policy development within the academy. However, these scores are intended for institutional use and should not yet be treated as directly equivalent to international tests such as IELTS or TOEFL without further external validation.

Limitations

Several limitations should be acknowledged when interpreting the findings of this study. The RPCA-TEP currently assesses only receptive language skills (listening and reading). As a result, the established cut scores reflect CEFR alignment within this limited construct and do not encompass productive skills such as speaking and writing, which are integral to full CEFR proficiency.

The standard-setting panel consisted of 13 experts with relevant experience in English language teaching, assessment, and police education. This panel size is fully consistent with recommended practice in Modified Angoff procedures, where panel quality and contextual expertise are considered more critical than sheer numbers. Importantly, the experts represented diverse professional backgrounds. Future studies may broaden stakeholder representation further, for example by including CEFR specialists or international assessors to enrich perspectives on descriptor interpretation.

The study did not include external validation against international CEFR-benchmarked English proficiency tests (e.g., IELTS, TOEFL) or independent performance indicators. Consequently, while the cut scores are defensible within the RPCA context, additional post-alignment validation is necessary to examine classification accuracy and broader test-use validity.

As with all judgment-based standard-setting procedures, classification uncertainty is greatest for candidates whose scores fall near the cut points. Although the convergence of expert judgments supports the reliability of the established thresholds, careful interpretation remains necessary for borderline cases.

Addressing these limitations through future research will further strengthen confidence in the RPCA-TEP as a CEFR-referenced assessment and support its responsible use in police education.

Recommendation

To strengthen the RPCA-TEP and guide future research, several actions are recommended. First, the present study was conducted within a single institution and among cadets with relatively similar academic backgrounds; therefore, future comparative and longitudinal studies across different faculties, cohorts, and police-education contexts are recommended to provide a broader understanding of how English proficiency and learner perceptions develop over time.

Second, for assessment improvement, post-alignment validation procedures should be systematically implemented. These include conducting content audits to ensure alignment with CEFR descriptors specific to policing tasks, examining correlations with external criteria such as international English test scores or supervisor performance ratings, and undertaking studies of decision accuracy and rater reliability to verify the technical soundness of the assessment.

Finally, ongoing attention to fairness, impact, and test-use validity is crucial. It is recommended that the RPCA routinely monitor these components to maintain transparency, support defensible decision-making, and uphold the long-term credibility of the RPCA-TEP.

References

- Cizek, G. J. (2012). *Setting performance standards: Foundations, methods, and innovations* (2nd ed.). Routledge.
- Cizek, G. J., & Bunch, M. B. (2007). *Standard setting: A guide to establishing and evaluating performance standards on tests*. Sage.
- Council of Europe. (2001). *Common European Framework of Reference for Languages: Learning, teaching, assessment*. Cambridge University Press.
- Hambleton, R. K., & Pitoniak, M. J. (2006). Setting performance standards. In R. L. Brennan (Ed.), *Educational measurement* (4th ed., pp. 433–470). Praeger.
- Ricker, K. L. (2006). *Setting cut-scores: A critical review of the Angoff and modified Angoff methods*. *The Alberta Journal of Educational Research*, 52(1), 53–64.
- Sae-Jeng, S. (2018). *Problems and needs of the occupational English language skills for Thai tourist police officers working at one of the international airports in Thailand* [Independent study]. Thammasat University.
https://ethesisarchive.library.tu.ac.th/thesis/2018/TU_2018_6021040271_10740_10283.pdf
- Suwannaphim, S. (2011). *English oral communication problems of police officers in high tourism areas* [Master's thesis]. Khon Kaen University.
- Tannenbaum, R. J., & Cho, Y. (2014). Critical thinking and the Angoff method: A study of standard-setting panelists' judgments. *Educational Measurement: Issues and Practice*, 33(2), 14–24. <https://doi.org/10.1111/emip.12033>

Psychometric Evaluation of the Emotion Regulation Instrument for Teacher Education Institution Students Utilizing AI in Scientific Writing

Rizki Mustikasari^{1*}, Ratri Harida², and Cutiana Windri Astuti³

(Received: January 7, 2026; Revised: April 6, 2026; Accepted: April 16, 2026)

Abstract

This study examined the validity and reliability of an emotion regulation measurement instrument for university students using artificial intelligence (AI) in scientific writing, involving a sample of 100 respondents. Data were analysed using Confirmatory Factor Analysis (CFA) within the Covariance-Based Structural Equation Modelling (CB-SEM) framework, implemented using AMOS with Maximum Likelihood estimation. The adapted instrument consists of 18 items across two dimensions: Cognitive Reappraisal (Items 1–13) and Expressive Suppression (Items 14–18), expanded contextually from the original 10-item Gross & John's ERQ. The measurement model demonstrated good overall fit ($\chi^2/df = 1.31$; RMSEA = 0.056; CFI = 0.955; TLI = 0.949; IFI = 0.956). All 18 indicators loaded significantly ($p < 0.001$) with standardized factor loadings ranging from 0.57 to 0.76, confirming convergent validity. Construct Reliability (CR) was also computed and found to meet the acceptable threshold ($CR \geq 0.70$). Overall, the instrument demonstrates adequate psychometric properties and provides a sound basis for assessing emotion regulation in the context of AI-assisted scientific writing.

Keywords: AI-assisted, emotion regulation, instrument, psychometric evaluation, scientific writing, TEI Student

Introduction

Universities hold a fundamental responsibility as centers for research and the advancement of knowledge. The Deputy Minister of Higher Education and Science and Technology of Indonesia, Prof. Stella Christie, Ph.D., stated that research is an important aspect in the development of science and technology in Indonesia. Her statement in a talk show on

¹ Lecturer, Early Childhood Teacher Education Program, STKIP PGRI Ponorogo Indonesia

² Lecturer, English Language Education, STKIP PGRI Ponorogo Indonesia

³ Lecturer, Indonesian Language and Literature Education, STKIP PGRI Ponorogo Indonesia

* Corresponding Author E-mail: mustikasari rizki@gmail.com

the IDN Times YouTube channel on October 23, 2024, emphasized that research is one of the pillars of the Tridharma (three main duties) of higher education, alongside education and community service.

The development of technology, particularly in the field of artificial intelligence (AI), has significantly revolutionized the landscape of higher education. The integration of AI in various academic processes—such as student thesis writing, automatic paraphrasing, reference compilation, and sentence structure improvement—is becoming increasingly widespread (Chubb et al., 2022; Costa et al., 2024; Črček et al., 2023). Although this technology offers convenience and efficiency, its use does not fully address the academic challenges faced by students, especially those related to psychological aspects. Many students experience confusion when using AI, feel less confident about the originality of their work, or even show excessive dependence on the technology. These conditions highlight the importance of understanding students' emotional processes, particularly how they regulate their emotions during academic writing activities. The ineffective emotional management may reduce the quality of scientific writing and negatively impact overall well-being.

The integration of AI tools into academic writing introduces a distinct set of emotional demands that extend beyond traditional academic stressors. When using AI to assist with scientific writing, having good emotion regulation is a crucial factor in coping with academic pressure (Gross, 2015). Students must navigate not only the cognitive complexity of scholarly writing itself, but also the uncertainty, self-doubt, and ethical ambiguity that frequently accompany AI-assisted work — such as concerns about academic authenticity, over-reliance on generated content, and the perceived diminishment of personal intellectual agency. These emotionally laden experiences require active and adaptive regulation strategies if students are to engage productively with AI tools without compromising their academic development or psychological well-being.

Emotion regulation is defined as an individual's process of monitoring, evaluating, and modifying emotional responses to specific situations (Gross, 1998; Sheppes et al., 2015; Gratz & Roemer, 2004; Gross, 2015). Within the process model proposed by Gross (1998), emotion regulation operates across multiple stages — from the selection and modification of emotionally relevant situations, to the cognitive reappraisal of their meaning, and ultimately to the modulation of expressive and physiological responses. This framework distinguishes between antecedent-focused strategies, such as cognitive reappraisal, which involve changing how a situation is interpreted before an emotional response is fully generated, and response-

focused strategies, such as expressive suppression, which involve modifying emotional expression after an emotional response has already been initiated (Gross & John, 2003). Cognitive reappraisal has consistently been associated with more adaptive outcomes — including greater positive affect, lower negative affect, and higher well-being — whereas expressive suppression has been linked to increased cognitive load and reduced social functioning (John & Gross, 2004). Understanding how students deploy these two regulation strategies in the context of AI-assisted writing is therefore theoretically and practically significant.

Students with strong emotional regulation skills are better equipped to deal with academic stress (Weiss et al., 2022), whereas students who are less emotionally regulated may struggle more as they experience decreased academic performance and an increased risk of sustained stress (Andrés et al., 2017). In the specific context of AI-assisted scientific writing, the emotional stakes are heightened by the iterative, multi-stage nature of the writing process — from formulating research questions and conducting literature reviews to analyzing data and constructing conclusions. Each stage presents distinct emotional challenges: topic selection may evoke uncertainty about academic identity; literature review may generate anxiety about the sufficiency of one's knowledge base; and data analysis may trigger frustration or self-doubt when AI-generated outputs do not meet expectations. Students who are unable to regulate these stage-specific emotional responses may disengage prematurely, avoid challenging tasks, or develop maladaptive coping behaviors such as uncritical acceptance of AI outputs or avoidance of independent intellectual engagement.

The ability to regulate emotions can therefore be utilized to manage academic demands and enhance the effectiveness of AI technology in supporting the scientific writing process (Gross, 2014). Rather than functioning merely as a passive buffer against stress, effective emotion regulation enables students to sustain motivation, maintain epistemic curiosity, and engage reflectively with AI-generated content — thereby preserving the integrity of their scholarly work. This position is consistent with the regulatory flexibility perspective advanced by (Bonanno & Burton, 2013), which holds that adaptive emotion regulation is not simply a matter of deploying a particular strategy, but of flexibly selecting and adjusting strategies in response to the specific demands of the situation. In the dynamic and often ambiguous environment of AI-assisted academic writing, such flexibility is essential. Consequently, the accurate measurement of emotion regulation in this context is not only theoretically warranted but also practically indispensable for informing evidence-based interventions in higher education.

The interaction between emotional regulation abilities and AI technology can be a key factor in supporting effective scientific writing. The interaction between technology engagement and emotional capacity in the academic environment needs to be examined to enhance students' focus and productivity in completing scientific writing (Zeidner et al., 2009). The use of AI in scientific writing has the potential to save time, energy, and cognitive effort; however, it may also trigger emotional responses such as uncertainty, overreliance, and self-doubt. In this context, the regulation of both positive and negative emotions becomes crucial to maintain optimal performance which falls within the scope of emotion regulation (Risman et al., 2020). Strengthening positive emotions such as self-confidence can enhance motivation, while managing negative emotions such as anxiety and frustration in AI-assisted scientific writing involves not only affective responses but also ethical considerations, self-confidence, and perceptions of academic authenticity.

Accurately examining emotion regulation in the context of AI-assisted scientific writing requires an instrument that is both valid and reliable. As AI becomes increasingly integrated into academic practices — offering efficiency in generating, organizing, and refining ideas — students are exposed to a new set of emotional dynamics that did not exist in traditional writing environments. These include uncertainty about the authenticity of AI-generated content, tension between efficiency and intellectual ownership, and anxiety arising at each stage of the writing process. Capturing these context-specific emotional experiences demands a measurement tool that is not only psychometrically sound but also sensitive to the situational demands of AI-assisted academic work.

Among the instruments most widely used to assess emotion regulation in the literature, two have received particular attention. The Difficulties in Emotion Regulation Scale (DERS), developed by Gratz & Roemer (2004) and subsequently validated across various cultural contexts (Dan-Glauser et al., 2013), measures multiple dimensions of emotion regulation dysfunction — including limited access to regulation strategies, lack of emotional clarity, and difficulties engaging in goal-directed behavior under distress. The Emotion Regulation Questionnaire (ERQ), developed by John & Gross (2004), and applied extensively in academic and cross-cultural research (Aldao et al., 2010; Balabdaoui et al., 2024), takes a different approach by assessing two adaptive regulation strategies (cognitive reappraisal and expressive suppression) through self-report measures that capture stable individual differences in how people manage their emotional responses (John & Gross, 2004). While the DERS is particularly useful for identifying regulation deficits and clinical vulnerabilities, the ERQ is better suited to

examining the strategic, context-sensitive regulation processes that characterize students' emotional experiences during complex academic tasks such as scientific writing.

This study adopts the ERQ due to its established validity and reliability across cultures (Radde et al., 2021; Messina et al., 2022), making it a suitable foundation for contextual adaptation. However, since the Indonesian version of the ERQ was originally designed for general contexts, it was considered insufficient to capture the domain-specific emotional regulation processes experienced by students during AI-assisted academic writing. To address this gap, the instrument was expanded based on focus group discussions (FGDs) with education experts, incorporating key stages of the scientific writing process: topic selection, title formulation, background development, literature review, research design, data analysis, and conclusion writing.

The development from 10 to 18 items was undertaken for two primary reasons. First, the original ERQ does not specifically address the anxiety-related emotional processes that emerge at each stage of academic writing, particularly in the context of AI assistance. Second, focus group discussions with education experts identified that domain-specific emotional experiences (particularly contextual anxiety during writing phases) were not adequately captured by the existing items. Theoretically, the adapted instrument maintains alignment with the two-factor structure of the ERQ (cognitive reappraisal and expressive suppression), while extending coverage to situational and contextual emotional processes relevant to AI-assisted scientific writing. Content validity was established through expert review during the FGD process, ensuring that newly developed items accurately reflected the target construct. This approach is also consistent with the notion that effective emotion regulation depends on individuals' ability to flexibly apply different strategies across varying situational demands. Prior to the main study, a pilot test was conducted with 100 participants outside the main research sample to evaluate the suitability and clarity of the adapted instrument (Retnawati, 2016; Kaewta et al., 2025). This pilot testing provided objective results for the adapted instrument and informed necessary refinements prior to main data collection. This psychometric validation is critical in the adaptation process to ensure that the instrument accurately measures the intended construct within a specific sociocultural context (Larantukan & Yudiarso, 2025). Validity testing is important to ensure that the instrument truly measures the intended construct (Meliala & Ahman, 2024). It is meant to ensure validity and reliability of the instrument for Indonesian students in teacher training institution.

Ample studies have explored the concept of emotion regulation among university students. Aryansah & Sari (2021) examined the role of students' emotional regulation in response to the school-from-home policy during the pandemic, revealing that the ability to manage emotions significantly influenced readiness for online learning. Furthermore, Kumala & Darmawanti (2012) investigated emotion regulation strategies among students with multiple roles and found that emotional regulation is crucial to maintain academic and social balance. Radde et al. (2021) developed a transdiagnostic mobile intervention aimed at enhancing adaptive emotional regulation skills among university students. Meanwhile, López González et al. (2025) explored the relationship between emotion regulation and major policy decisions using an Aristotelian-Thomistic framework across cross-continental student samples, broadening theoretical insights within ethical and global educational contexts.

Despite the abundance of prior research, there remains a gap in the accurate and contextual measurement of emotion regulation, particularly among students in teacher education institutions (TEIs) who are increasingly exposed to the integration of artificial intelligence (AI) in scientific writing processes. Existing instruments are often insufficient to capture context-specific emotional regulation processes and are rarely validated within local Indonesian contexts and the current AI-driven academic environment.

Therefore, this study aims to conduct a psychometric evaluation of an emotion regulation measurement instrument tailored for TEI students in the context of AI-assisted scientific writing. The instrument is adapted from the Emotion Regulation Questionnaire (ERQ), which has demonstrated strong validity and reliability across cultures, and is expanded to reflect domain-specific emotional regulation processes identified through focus group discussions with education experts. The expansion incorporates key stages of the scientific writing process, including topic selection, title formulation, background development, literature review, research design, data analysis, and conclusion writing. The findings are expected to contribute to the development of a valid, reliable, and contextually grounded instrument, thereby supporting data-driven interventions in higher education.

Objectives

1. To examine the validity of an instrument measuring emotion regulation among students in Teacher Education Institutions who use AI tools in academic writing.
2. To examine the reliability of an instrument measuring emotion regulation among students in Teacher Education Institutions who use AI tools in academic writing.

Method

This research employed a quantitative survey approach and is part of a larger study investigating emotion regulation among teacher education students (TEI) who use Artificial Intelligence (AI) in academic writing. This stage focused on the validation and reliability testing of an emotion regulation instrument adapted for AI-based writing contexts. The study was conducted in the even semester of the 2024/2025 academic year at four teacher education institutions (LPTK) in Ponorogo Regency, East Java, Indonesia.

1. Population and sample

The research targets were undergraduate students in teacher education institutions (LPTK) who utilize AI tools (e.g., ChatGPT, Grammarly, Quillbot) to assist in writing their academic papers, particularly in relation to their emotion regulation processes during academic writing. The participants were active undergraduate students in the 2024/2025 academic year, drawn from various levels and institutions, particularly those in the fourth to final semesters.

This study involved 100 respondents ($N = 100$). This sample size meets the minimum requirement for confirmatory factor analysis (CFA), as recommended by Hair et al., (2010), which suggests at least 100 participants or a ratio of 5–10 subjects per item. With 18 items in the adapted instrument, a sample of 100 respondents yields a ratio of approximately 5.6 subjects per item, which is within the acceptable range for stable parameter estimation in CFA (Hair et al., 2010). While larger samples are generally preferable (particularly for more complex models) the current sample size is considered adequate for the scope of this instrument validation study.

2. Research instrument

The instrument psychometrically examined in this study is the emotion regulation measurement developed by Gross & John (2003), which consists of 10 items, shown in Figure 1.

1. *I control my emotions by changing the way I think about the situation I'm in*
2. *When I want to feel less negative emotion, I change the way I'm thinking about the situation.*
3. *When I want to feel more positive emotion, I change the way I'm thinking about the situation.*
4. *When I want to feel more positive emotion (such as joy or amusement), I change what I'm thinking about.*
5. *When I want to feel less negative emotion (such as sadness or anger), I change what I'm thinking about.*
6. *When I'm faced with a stressful situation, I make myself think about it in a way that helps me stay calm.*
7. *I control my emotions by not expressing them.*
8. *When I am feeling negative emotions, I make sure not to express them.*
9. *I keep my emotions to myself.*
10. *When I am feeling positive emotions, I am careful not to express them.*

Figure 1 The Emotion Regulation Measure Developed by Gross and John (2003)

The instrument examined in this study is adapted from the Emotion Regulation Questionnaire (ERQ) developed by Gross & John (Radde et al., 2021), which originally consists of 10 items measuring two emotion regulation strategies: Cognitive Reappraisal (6 items) and Expressive Suppression (4 items). The ERQ has demonstrated strong psychometric properties in the original English version (Gross & John, 2003) and has been validated across various cultural contexts, including the Indonesian version validated by Radde et al. (2021), which reported satisfactory factorial validity and internal consistency (Cronbach's $\alpha \geq 0.70$ for both subscales).

In the present study, the instrument was adapted to measure emotion regulation among university students who utilize AI in writing academic works. The adaptation was guided by focus group discussions (FGDs) with education experts and expanded the instrument from 10 to 18 items to better reflect the domain-specific emotional processes encountered across stages of AI-assisted scientific writing. The adapted instrument retains the two-factor structure: (1) Cognitive Reappraisal (Items 1–13) and (2) Expressive Suppression (Items 14–18).

The expansion was conducted to improve content validity by capturing situational emotion regulation processes (particularly those associated with academic anxiety) across distinct phases of the scientific writing process: topic selection, title formulation, background development, literature review, research design, data analysis, and conclusion writing. These additions were informed by expert review and maintain theoretical alignment with the underlying dimensions of the ERQ.

The instrument was administered in the Indonesian language, as altering the language of administration may affect item meaning and measurement accuracy. English translations are provided in this manuscript solely for explanatory purposes to facilitate comprehension by international readers. Table 1 presents the adaptation process. The specific contextual adaptations applied are presented in the following table.

Table 1 Adaptation of the ERQ Instrument for AI-Assisted Scientific Writing Context

No	Strategies	Original's Instrument (Radde et al., 2021)	Adapted Instrument
1	Cognitive reappraisal	1. Saya mengendalikan emosi dengan mengubah pola pikir saya sesuai dengan situasi di lingkungan sekitar. 2. Ketika saya ingin merasakan lebih sedikit emosi negatif, saya mengubah pola pikir berdasarkan	1. Saya mengendalikan emosi dengan mengubah pola pikir saya ketika harus mencari judul untuk karya tulis ilmiah 2. Saya mengendalikan emosi dengan mengubah pola pikir saya ketika harus menyusun pendahuluan

No	Strategies	Original's Instrument (Radde et al., 2021)	Adapted Instrument
		situasi yang ada. 3. Ketika saya ingin merasakan lebih banyak emosi positif, saya mengubah pola pikir berdasarkan situasi yang ada. 4. Ketika saya ingin merasakan lebih banyak emosi positif, maka saya mengubah pola pikir saya. 5. Ketika saya ingin merasakan lebih sedikit emosi negatif, maka saya mengubah pola pikir saya. 6. Ketika saya dihadapkan pada situasi yang penuh tekanan, saya berusaha untuk tetap tenang.	untuk karya tulis ilmiah 3. Saya mengendalikan emosi dengan mengubah pola pikir saya ketika harus mencari referensi terkait karya tulis ilmiah 4. Ketika saya ingin mengurangi emosi negatif, saya mengubah pola pikir ketika harus mencari judul untuk karya tulis ilmiah 5. Ketika saya ingin mengurangi emosi negatif, saya mengubah pola pikir ketika harus menyusun pendahuluan untuk karya tulis ilmiah 6. Ketika saya ingin mengurangi emosi negatif, saya mengubah pola pikir ketika harus mencari referensi terkait karya tulis ilmiah 7. Ketika saya ingin merasakan lebih banyak emosi positif, saya mengubah pola pikir ketika harus mencari judul untuk karya tulis ilmiah 8. Ketika saya ingin merasakan lebih banyak emosi positif, saya mengubah pola pikir ketika harus menyusun pendahuluan untuk karya tulis ilmiah 9. Ketika saya ingin merasakan lebih banyak emosi positif, saya mengubah pola pikir ketika harus mencari referensi terkait karya tulis ilmiah 10. Ketika saya dihadapkan pada situasi yang penuh tekanan karena harus mencari judul karya tulis ilmiah, saya berusaha untuk tetap tenang. 11. Ketika saya dihadapkan pada situasi yang penuh tekanan karena harus menyusun pendahuluan karya tulis ilmiah, saya berusaha untuk tetap tenang.

No	Strategies	Original's Instrument (Radde et al., 2021)	Adapted Instrument
			12. Ketika saya dihadapkan pada situasi yang penuh tekanan karena harus mencari referensi untuk karya tulis ilmiah, saya berusaha untuk tetap tenang. 13. Ketika merasakan kesulitan dalam mencari judul untuk karya tulis ilmiah, saya mengendalikan emosi dengan tidak mengungkapkannya
2.	Expressive supression	7. Saya mengendalikan emosi dengan tidak mengungkapkannya. 8. Ketika saya merasakan emosi negatif, saya tidak mengungkapkannya. 9. Saya memendam perasaan untuk diri sendiri. 10. Ketika saya merasakan emosi positif, saya berhati-hati untuk tidak mengungkapkannya	14. Ketika merasakan kesulitan dalam menyusun pendahuluan untuk karya tulis ilmiah, saya mengendalikan emosi dengan tidak mengungkapkannya 15. Ketika merasakan kesulitan dalam mencari referensi untuk karya tulis ilmiah, saya mengendalikan emosi dengan tidak mengungkapkannya 16. Ketika dapat menemukan judul yang tepat untuk karya tulis ilmiah, saya berhati-hati dalam mengekspresikannya 17. Ketika dapat menyusun pendahuluan untuk karya tulis ilmiah, saya berhati-hati dalam mengekspresikannya. 18. Ketika dapat menemukan referensi terkait karya tulis ilmiah, saya berhati-hati dalam mengekspresikannya.

No linguistic translation was applied during data collection, as altering the language of the items may affect their meaning and measurement accuracy. English translations of the questionnaire items are provided in the manuscript solely for explanatory purposes to facilitate understanding for international readers and were not used as an adapted version of the instrument. The English translation of the instrument is presented below for reference.

1. I regulate my emotions by changing the way I think when I need to determine a title for a scientific manuscript.

2. I regulate my emotions by changing the way I think when I need to write the introduction of a scientific manuscript.

3. I regulate my emotions by changing the way I think when I need to search for references related to a scientific manuscript.

4. When I want to reduce negative emotions, I change the way I think when I need to determine a title for a scientific manuscript.

5. When I want to reduce negative emotions, I change the way I think when I need to write the introduction of a scientific manuscript.

6. When I want to reduce negative emotions, I change the way I think when I need to search for references related to a scientific manuscript.

7. When I want to experience more positive emotions, I change the way I think when I need to determine a title for a scientific manuscript.

8. When I want to experience more positive emotions, I change the way I think when I need to write the introduction of a scientific manuscript.

9. When I want to experience more positive emotions, I change the way I think when I need to search for references related to a scientific manuscript.

10. When I am faced with a stressful situation because I need to determine a title for a scientific manuscript, I try to remain calm.

11. When I am faced with a stressful situation because I need to write the introduction of a scientific manuscript, I try to remain calm.

12. When I am faced with a stressful situation because I need to search for references for a scientific manuscript, I try to remain calm.

13. When I experience difficulty determining a title for a scientific manuscript, I regulate my emotions by not expressing them.

14. When I experience difficulty writing the introduction of a scientific manuscript, I regulate my emotions by not expressing them.

15. When I experience difficulty searching for references for a scientific manuscript, I regulate my emotions by not expressing them.

16. When I am able to determine an appropriate title for a scientific manuscript, I am careful about how I express it.

17. When I am able to write the introduction of a scientific manuscript, I am careful about how I express it.

18. When I am able to find references related to a scientific manuscript, I am careful about how I express it.

3. Data Collection

Data were collected through a self-report questionnaire, distributed in printed and online formats (Google forms). Participants responded to the adapted instrument voluntarily and anonymously.

4. Data analysis

Data analysis was conducted using Confirmatory Factor Analysis (CFA) within the Covariance-Based Structural Equation Modeling (CB-SEM) framework, implemented with AMOS (Analysis of Moment Structures) using Maximum Likelihood (ML) estimation. Prior to CFA, the assumption of normality was assessed using Kolmogorov–Smirnov and Shapiro–Wilk tests, which indicated that the data significantly deviated from normal distribution ($p < 0.001$). Given that violations of univariate normality may indicate potential multivariate non-normality, Bollen-Stine bootstrap with 2,000 resamples was applied to obtain robust parameter estimates and to ensure the appropriateness of ML estimation. As this study focused on instrument adaptation rather than instrument development, CFA was employed to test the factorial structure previously established in the original ERQ, and the resulting measurement model was subsequently evaluated using established goodness-of-fit criteria.

The measurement model was evaluated using multiple goodness-of-fit indices, including the Chi-square statistic, the ratio of Chi-square to degrees of freedom ($\chi^2/df \leq 3.00$), Root Mean Square Error of Approximation (RMSEA ≤ 0.08), and incremental fit indices such as the Comparative Fit Index (CFI ≥ 0.90), Tucker-Lewis Index (TLI ≥ 0.90), and Incremental Fit Index (IFI ≥ 0.90). Convergent validity was assessed based on standardized factor loadings (≥ 0.50) and Average Variance Extracted (AVE ≥ 0.50).

Instrument reliability was evaluated using Construct Reliability (CR ≥ 0.70), computed from the standardized factor loadings and error variances (Hair et al., 2010). The stability of parameter estimates and the absence of estimation problems (e.g., negative error variances) were also examined as supplementary indicators of model adequacy (Goodwin, 1995).

Results

The measurement model for emotion regulation was evaluated using Confirmatory Factor Analysis (CFA) with Maximum Likelihood estimation in AMOS. The model consisted of one latent construct measured by 18 observed indicators and was estimated using data from 100 respondents. The estimation process converged successfully, with 36 free parameters and

135 degrees of freedom, indicating that the model was over-identified and statistically testable. Although the Chi-square statistic was significant ($\chi^2 = 176.82$, $df = 135$, $p = .009$), the ratio of Chi-square to degrees of freedom was within an acceptable range ($\chi^2/df = 1.31$), suggesting an adequate model fit given the sample size.

All indicators loaded significantly on the latent construct ($p < 0.001$), with standardized factor loadings ranging from 0.57 to 0.76, exceeding the recommended minimum threshold of 0.50. These results demonstrate that the indicators adequately represent the construct of emotion regulation and provide evidence of convergent validity. The variances of the latent construct and measurement errors were positive and significant, indicating stable parameter estimation without estimation problems such as negative variances.

Overall model fit indices supported the adequacy of the measurement model. The RMSEA value was 0.056 with a 90% confidence interval of 0.029 to 0.078 and a PCLOSE value of 0.326, indicating a close fit to the data. Incremental fit indices also showed strong model fit (CFI = 0.955, TLI = 0.949, IFI = 0.956). Taken together, these findings indicate that the proposed measurement model demonstrates satisfactory fit and provides a valid and reliable representation of the emotion regulation construct.

Table 2 Model Fit Indices of the Measurement Model

No	Fit Index	Threshold	Obtained Value	Conclusion
1	χ^2/df	≤ 3.00	1.31	Good fit
2	RMSEA	≤ 0.08	0.056	Good fit
3	PCLOSE	> 0.05	0.326	Good fit
4	CFI	≥ 0.90	0.955	Good fit
5	TLI	≥ 0.90	0.949	Good fit
6	IFI	≥ 0.90	0.956	Good fit

Construct Reliability (CR) was computed based on the standardized factor loadings and error variances of all 18 indicators. The CR value met the recommended threshold of ≥ 0.70 , confirming the internal consistency and reliability of the instrument. To further assess convergent validity, the Average Variance Extracted (AVE) was calculated. The AVE value exceeded the recommended threshold of ≥ 0.50 , providing additional evidence that the indicators adequately represent the latent construct of emotion regulation. Table 3 presents the psychometric quality summary.

Table 3 Psychometric Quality Summary of the Measurement Model

No	Parameter	Threshold	Obtained Value	Conclusion
1	Standardized Factor Loadings	≥ 0.50	0.57 – 0.76	Adequate
2	Average Variance Extracted (AVE)	≥ 0.50	0.50	Adequate
3	Construct Reliability (CR)	≥ 0.70	0.94	Adequate

Given that the measurement model consisted of a single latent construct, the assessment of discriminant validity was not applicable and therefore not conducted.

Discussion

This study aimed to examine the validity and reliability of an emotion regulation measurement instrument designed for university students who use artificial intelligence (AI) in scientific writing. The findings from the CFA (Confirmatory Factor Analysis) provide strong empirical support for the psychometric adequacy of the proposed instrument.

Regarding construct validity, the successful convergence of the CFA model and the acceptable χ^2/df ratio (1.31) indicate that the measurement structure was appropriately specified and statistically stable, even with a moderate sample size of $N = 100$. This result is consistent with prior validation studies of adapted emotion regulation instruments, which have similarly demonstrated adequate fit under comparable sample conditions (Radde et al., 2021; Messina et al., 2022). The significant and substantial factor loadings across all 18 indicators (0.57–0.76) demonstrate that each item contributes meaningfully to the measurement of emotion regulation, confirming convergent validity (Hair et al., 2010). The successful convergence of the model and the acceptable ratio of Chi-square to degrees of freedom indicate that the measurement structure was appropriately specified and statistically stable, even with a moderate sample size. These results suggest that emotion regulation in the context of AI-assisted academic writing can be represented as a coherent latent construct measured through multiple observable indicators.

An important theoretical consideration in this study concerns the conceptualization of emotion regulation strategies. The original Emotion Regulation Questionnaire (ERQ) developed by Gross & John (2003) conceptualizes emotion regulation as comprising two

primary strategies: cognitive reappraisal and expressive suppression. In the present study, this theoretical framework is maintained, and the instrument is specified as a two-strategy measurement model rather than a strictly unidimensional construct.

The expansion of the instrument from 10 to 18 items does not introduce new dimensions but aims to provide a more comprehensive representation of these two regulation strategies within the context of AI-assisted scientific writing. Specifically, additional items were developed to capture how students employ cognitive reappraisal and expressive suppression across different stages of the writing process. This approach allows for a more context-sensitive assessment while preserving the original theoretical structure of the ERQ.

Thus, emotion regulation in this study is understood as a higher-order construct reflected through two interrelated strategies, rather than as a single undifferentiated factor. Future research may further examine the structural relationships between these strategies, including testing alternative models such as higher-order or bifactor structures, to determine the most appropriate representation for this context

The RMSEA and incremental fit indices (CFI, TLI, IFI) all exceeded commonly accepted thresholds, indicating that the measurement model adequately reflects the observed data structure. These indices provide convergent evidence for the structural validity of the instrument, consistent with recommendations from methodologists in the CB-SEM framework (Hair Jr et al., 2010; Hu & Bentler, 1999).

With respect to reliability, the Construct Reliability (CR) value met the recommended threshold of ≥ 0.70 , confirming acceptable internal consistency. This finding aligns with the psychometric evidence reported for the original ERQ and its Indonesian adaptation (Radde et al., 2021). The stability of parameter estimates and the absence of estimation problems further corroborate the instrument's reliability for use in future research contexts.

The AVE value exceeding 0.50 not only confirms convergent validity but also suggests that the latent construct of emotion regulation meaningfully captures the shared variance among its indicators. This indicates that the adapted instrument represents a coherent psychological construct, thereby providing a robust psychometric foundation for examining emotion regulation in the evolving context of AI-assisted scientific writing.

These results are consistent with prior validation studies of adapted emotion regulation instruments in non-Western academic contexts (Radde et al., 2021; Messina et al., 2022), reinforcing the cross-cultural applicability of the ERQ framework when adapted systematically. Importantly, the retention of the two-strategy structure (Cognitive Reappraisal

and Expressive Suppression) in the context of AI-assisted writing aligns with theoretical arguments that emotion regulation is inherently situation-specific (Gross, 2015; Bonanno & Burton, 2013). In AI-assisted writing contexts, cognitive reappraisal may enable students to reinterpret challenges arising from AI-generated outputs—such as inaccuracies, ambiguity, or lack of originality—as opportunities for revision and learning. In contrast, expressive suppression may reflect students' tendency to regulate visible emotional reactions, such as frustration or dissatisfaction when AI tools do not meet expectations. This suggests that both strategies play complementary roles in managing emotional responses during human–AI interaction, highlighting the ecological relevance of the ERQ framework in technology-mediated academic environments. Future studies should examine whether the relative effectiveness of these two strategies differs across writing stages, as this may have implications for tailored psychoeducational interventions in technology-integrated higher education.

Beyond psychometric evidence, this study also raises important methodological considerations related to instrument use in multilingual research contexts. Maintaining the original language of administration (Indonesian) was essential to preserve the established validity and reliability of the instrument, as linguistic translation may alter item meaning and introduce construct-irrelevant variance. English translations were provided solely for explanatory purposes in this manuscript, consistent with international recommendations emphasizing transparency and replicability in educational measurement research.

In summary, the adapted emotion regulation instrument demonstrates adequate validity and reliability for use in the specific context of AI-assisted scientific writing among TEI students. These findings extend the existing literature by providing a psychometrically grounded tool for assessing context-specific emotional regulation, contributing to both measurement science and the practical evaluation of students' psychological well-being in technology-integrated educational environments.

Beyond the psychometric evidence supporting the instrument, this study also raises important methodological considerations related to instrument use and reporting in multilingual research contexts. This study highlighted methodological considerations in the use of research instruments within non-English-speaking contexts. Maintaining the original language of administration was essential to preserve the established validity and reliability of the instrument, as linguistic translation may alter item meaning and introduce construct-irrelevant variance. Yet, to balance psychometric rigor and transparency, English translations of questionnaire items were provided solely for explanatory purposes. This strategy aligns with

international recommendations emphasizing transparency and replicability in educational measurement research. The findings therefore offer methodological insights applicable to researchers working across diverse linguistic and cultural settings, particularly in contexts where preserving measurement quality is a primary concern. Although grounded in a local context, the study's methodological implications extend to broader theoretical and practical discussions in educational assessment.

Recommendation

The results indicate that the developed instrument demonstrates a stable factor structure, with all indicators contributing substantially to the underlying latent construct. Accordingly, this instrument is recommended as an initial screening tool in courses that involve intensive AI-assisted academic writing, such as research methodology, proposal seminars, and thesis or scholarly article writing. Periodic administration of the instrument enables lecturers to identify students' emotion regulation patterns at an early stage, particularly related to the anxiety arising from ethical uncertainty, concerns about the validity of academic work, and dependence on AI tools in the academic writing process.

Future research is encouraged to develop a more adaptive measurement model by incorporating contextual indicators that reflect diverse patterns of AI use (e.g., AI as a language support tool, idea-generation aid, or data analysis assistant). In addition, re-examining the indicators through item parcelling or a second-order factor approach may enhance model efficiency and reduce potential error correlations among indicators identified in the modification indices analysis. Such refinements are essential to ensure that the instrument remains sensitive to the evolving dynamics of AI use while maintaining precision in measuring academic anxiety.

At the institutional level, the findings suggest the need for policy approaches that move beyond restrictive regulations on AI use toward data-driven strategies grounded in psychometric assessment of students. Higher education institutions are therefore advised to integrate the results of academic emotion regulation measurement into academic support services and student guidance programs, such as scientific writing clinics or AI literacy initiatives. This approach enables institutional decision-making regarding AI use in academic writing to be informed by empirical evidence on students' psychological conditions rather than normative assumptions or institutional concerns.

References

- Aldao, A., Nolen-Hoeksema, S., & Schweizer, S. (2010). Emotion-regulation strategies across psychopathology: A meta-analytic review. *Clinical Psychology Review, 30*(2), 217–237. <https://doi.org/10.1016/j.cpr.2009.11.004>
- Andrés, M. L., Stelzer, F., Juric, L. C., Introzzi, I., Rodríguez-Carvajal, R., & Guzmán, J. I. N. (2017). Emotion regulation and academic performance: A systematic review of empirical relationships. *Psicologia em Estudo, 22*(3), 299–311. <https://doi.org/10.4025/psicolestud.v22i3.34360>
- Aryansah, J. E., & Sari, S. P. (2021). Analisis peran regulasi emosi mahasiswa terhadap kebijakan school from home di masa pandemi Covid-19 [Analysis of the role of student emotion regulation toward the school-from-home policy during the Covid-19 pandemic]. *Jurnal Pemerintahan dan Politik, 6*(1), 8–14. <https://doi.org/10.36982/jpg.v6i1.1312>
- Balabdaoui, F., Dittmann-Domenichini, N., Grosse, H., Schlienger, C., & Kortemeyer, G. (2024). A survey on students' use of AI at a technical university. *Discover Education, 3* Article 51. <https://doi.org/10.1007/s44217-024-00136-4>
- Bonanno, G. A., & Burton, C. L. (2013). Regulatory flexibility: An individual differences perspective on coping and emotion regulation. *Perspectives on Psychological Science, 8*(6), 591–612. <https://doi.org/10.1177/1745691613504116>
- Chubb, J., Cowling, P., & Reed, D. (2022). Speeding up to keep up: Exploring the use of AI in the research process. *AI & Society, 37*(4), 1439–1457. <https://doi.org/10.1007/s00146-021-01259-0>
- Costa, K., Ntsobi, P. M., & Mfolo, L. (2024). Challenges, benefits and recommendations for using generative artificial intelligence in academic writing: A case of ChatGPT. *Medicon Engineering Themes, 7*(4), 3–38. <https://doi.org/10.55162/MCET.07.236>
- Črček, N., & Patekar, J. (2023). Writing with AI: University students' use of ChatGPT. *Journal of Language and Education, 9*(4), 128–138. <https://doi.org/10.17323/jle.2023.17379>
- Dan-Glauser, E. S., & Scherer, K. R. (2013). The difficulties in emotion regulation scale (DERS). *Swiss Journal of Psychology, 72*(1), 5–11. <https://doi.org/10.1024/1421-0185/a000093>
- Goodwin, C. J. (1995). *Research in psychology: Methods and design* (2nd ed.). John Wiley & Sons.

- Gratz, K. L., & Roemer, L. (2004). Multidimensional assessment of emotion regulation and dysregulation: Development, factor structure, and initial validation of the Difficulties in Emotion Regulation Scale. *Journal of Psychopathology and Behavioral Assessment*, 26(1), 41–54. <https://doi.org/10.1023/B:JOBA.0000007455.08539.94>
- Gross, J. J. (1998). The emerging field of emotion regulation: An integrative review. *Review of General Psychology*, 2(3), 271–299. <https://doi.org/10.1037/1089-2680.2.3.271>
- Gross, J. J. (Ed.). (2014). *Handbook of emotion regulation* (2nd ed.). Guilford Press.
- Gross, J. J. (2015). Emotion regulation: Current status and future prospects. *Psychological Inquiry*, 26(1), 1–26. <https://doi.org/10.1080/1047840X.2014.940781>
- Gross, J. J., & John, O. P. (2003). Individual differences in two emotion regulation processes: Implications for affect, relationships, and well-being. *Journal of Personality and Social Psychology*, 85(2), 348–362. <https://doi.org/10.1037/0022-3514.85.2.348>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Pearson.
- Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- John, O. P., & Gross, J. J. (2004). Healthy and unhealthy emotion regulation: Personality processes, individual differences, and life span development. *Journal of Personality*, 72(6), 1301–1334. <https://doi.org/10.1111/j.1467-6494.2004.00298.x>
- Kaewta, R., Witchayawat, W., Seehamat, L., & Sapphasuk, W. (2025). An analysis of factors and indicators of research competencies of basic education teachers. *Journal of Educational Measurement, Mahasarakham University*, 31(2), 22–38.
- Kumala, K. H., & Darmawanti, I. (2022). Strategi regulasi emosi pada mahasiswa dengan banyak peran [Emotion regulation strategies among students with multiple roles]. *Character: Jurnal Penelitian Psikologi*, 9(3), 19–29. https://ejournal.unesa.ac.id/index.php/character/article/view/46151?utm_source=chatgpt.com (in Indonesian)
- Larantukan, A. M. F., & Yudiarso, A. (2025). Validity of the Zarit Burden Interview (ZBI) Measurement Tool Using the Rasch Model in the Indonesian Population. *Humanitas*, 9(1), 18–40. <https://doi.org/10.28932/humanitas.v9i1.11007>

- López González, J., Fernández Espinosa, V., & Crespi, P. (2025). Integrating emotional regulation in higher education: An Aristotelian-Thomistic approach. *Journal of Human Behavior in the Social Environment*, 36(5), 328-249.
<https://doi.org/10.1080/10911359.2025.2457966>
- Meliala, A. K., & Ahman, A. (2024). Systematic literature review: Validitas konstruk skala psychological well-being Ryff [Systematic literature review: Construct validity of Ryff's psychological well-being scale]. *Ideguru: Jurnal Karya Ilmiah Guru*, 9(2), 950-955.
<https://doi.org/10.51169/ideguru.v9i2.995>
- Messina, I., Spataro, P., Grecucci, A., Moskow, D. M., Marogna, C., & Hofmann, S. G. (2022). Interpersonal emotion regulation questionnaire: Psychometric properties of the Italian version and associations with psychopathology. *Research in Psychotherapy: Psychopathology, Process and Outcome*, 25(2), 239-246.
<https://doi.org/10.4081/ripppo.2022.616>
- Radde, H. A., Nurrahmah, N., Nurhikmah, N., & Saudi, A. N. A. (2021). Uji validitas konstruk dari Emotion Regulation Questionnaire versi bahasa Indonesia dengan menggunakan confirmatory factor analysis [Construct validity test of the Indonesian version of the Emotion Regulation Questionnaire using confirmatory factor analysis]. *Jurnal Psikologi Karakter*, 1(2), 152-160. <https://doi.org/10.56326/jpk.v1i2.1284>
- Retnawati, H. (2016). *Analisis kuantitatif instrumen penelitian* [Quantitative analysis of research instruments]. Parama Publishing.
- Sheppes, G., Suri, G., & Gross, J. J. (2015). Emotion regulation and psychopathology. *Annual Review of Clinical Psychology*, 11, 379-405. <https://doi.org/10.1146/annurev-clinpsy-032814-112739>
- Weiss, N. H., Kiefer, R., Goncharenko, S., Raudales, A. M., Forkus, S. R., Schick, M. R., & Contractor, A. A. (2022). Emotion regulation and substance use: A meta-analysis. *Drug and Alcohol Dependence*, 230, Article 109131.
<https://doi.org/10.1016/j.drugalcdep.2021.109131>
- Zeidner, M., Matthews, G., & Roberts, R. D. (2009). Emotional intelligence and academic achievement: A meta-analytic review. *Educational Psychologist*, 44(4), 124-136.
<https://doi.org/10.1080/00461520903174853>

Development and Validation of an Academic Engagement Scale for Undergraduate Students in Distance Learning

Supunsa Langprasert¹ and Pisutthipa Metheekul^{2*}

(Received: March 2, 2026; Revised: April 24, 2026; Accepted: June 11, 2026)

Abstract

Distance education at open universities faces significant challenges, including high attrition rates driven by social isolation and the demands of self-directed learning. A context-specific instrument is essential for understanding the mechanisms that support student persistence. This study aimed to develop and validate an academic engagement scale, with development referring to the refinement of an existing framework. Drawing on constructivist and motivational theories, academic engagement was conceptualized as a second-order construct comprising cognitive, emotional, and behavioral dimensions. Data were collected from 440 students at Sukhothai Thammathirat Open University using cluster sampling based on academic disciplines (Social Sciences, Health Sciences, and Science and Technology). Confirmatory Factor Analysis (CFA) was conducted to examine the hierarchical structure, while Cronbach's alpha, Average Variance Extracted (AVE), and the Heterotrait-Monotrait (HTMT) ratio were used to assess reliability and construct validity.

The model fit indices indicated an acceptable fit to the data ($\chi^2/df = 3.22$, RMSEA = .07, CFI = .92, NFI = .91, and SRMR = .04). All dimensions demonstrated satisfactory reliability and validity, with HTMT values below .85, supporting the distinctiveness of the constructs. Cognitive engagement showed strong associations within the model, highlighting the role of self-regulated learning in distance education. The findings provide empirical support for a multidimensional structure of academic engagement and contribute a context-specific measurement tool that can be applied to support student engagement and retention in an open university distance learning context.

Keywords: academic engagement, open university, distance learning, scale validation

¹ Lecturer, Office of Registration, Records and Evaluation, Sukhothai Thammathirat Open University, Thailand

² Assistant Professor, School of Human Ecology, Sukhothai Thammathirat Open University, Thailand

* Corresponding Author E-mail: Pisutthipa.Met@stou.ac.th

Introduction

Academic engagement is widely recognized as a key predictor of students' academic success, encompassing cognitive, emotional, and behavioral dimensions (Fredricks et al., 2004; Fredricks, 2012). These dimensions reflect students' active involvement in learning processes, including their effort, emotional connection, and use of self-regulated learning strategies. Students who demonstrate high levels of engagement are more likely to persist in their studies, employ adaptive learning strategies, and achieve positive academic outcomes. From a theoretical perspective, academic engagement can be understood through the lens of constructivist theory. Rooted in the work of Vygotsky (1978) and Piaget (1970), constructivism posits that learning is an active process in which individuals construct knowledge through interaction with their environment. In this context, academic engagement represents the driving force that enables learners to actively participate in and make meaning from learning experiences. This perspective is particularly relevant in distance learning environments, where learners must rely on their own initiative and self-regulation to sustain meaningful learning.

Previous studies have consistently demonstrated that academic engagement plays an important role in student achievement and persistence across educational contexts (Fredricks et al., 2004; Fredricks, 2012). In distance learning environments, engagement has been closely associated with self-regulated learning processes, as learners are required to take greater responsibility for managing their own learning (Lerdpornkulrat, 2023). These findings suggest that academic engagement is not only influenced by individual motivation but also by learners' ability to regulate their learning behaviors.

Sukhothai Thammathirat Open University (STOU) represents a key institution in Thailand's open and distance learning system, providing flexible access to higher education for diverse student populations. While distance education offers advantages such as flexibility and accessibility (Al-Fraihat et al., 2020), it also introduces challenges, including social isolation and the need for strong self-directed learning skills. According to Moore's (1993) theory of transactional distance, the separation between learners and institutions may create psychological distance, which can negatively influence engagement and persistence. Empirical studies in the STOU context have shown that student success is influenced by both individual factors, such as motivation and self-regulated learning, and institutional support mechanisms (Sungsri, 2018; Metheekul, 2022). These findings highlight the importance of academic engagement as a key mechanism linking learners to their educational environment.

Another important factor closely associated with academic engagement in distance learning is self-regulated learning. In technology-mediated environments, learners who actively regulate their learning—by setting goals, monitoring progress, and applying appropriate strategies—tend to exhibit higher levels of engagement and academic performance (Zimmerman, 2000; Lerdpornkulrat, 2023). These characteristics align with constructivist principles and underscore the importance of internal learning processes in sustaining engagement. However, despite the recognized importance of academic engagement, the applicability of existing measurement instruments in the Thai higher education context remains limited. Existing research in Thailand has primarily focused on factors influencing learning success rather than the validation of context-specific measurement tools. As a result, there remains a lack of rigorously validated instruments specifically designed to assess academic engagement in open and distance learning environments.

Although academic engagement has been widely studied, several measurement instruments have been developed in different educational contexts, such as the Utrecht Work Engagement Scale for Students (UWES-S; Schaufeli et al., 2002) and the National Survey of Student Engagement (NSSE; Kuh, 2001). However, these instruments were primarily designed for traditional or Western higher education settings and may not fully capture the characteristics of learners in open and distance learning environments. In particular, they often overlook the flexible learning structures, reduced face-to-face interaction, and the central role of self-regulated learning in such contexts. In the Thai higher education context, existing studies have focused more on factors influencing learning success rather than on the validation of context-specific engagement instruments. As a result, there remains a lack of empirically validated tools specifically tailored to measure academic engagement in open university systems. This gap highlights the need to refine and validate an academic engagement scale that is theoretically grounded and contextually appropriate for distance learning environments.

Therefore, there is a need to refine and validate an academic engagement scale that is appropriate for open university settings. In this study, the term “development” refers to the refinement and contextual adaptation of an existing instrument rather than the creation of a completely new scale. Accordingly, the purpose of this study is to refine and validate an Academic Engagement Scale for undergraduate students in a distance learning context at Sukhothai Thammathirat Open University.

Objective

To examine the factorial structure and assess the reliability and construct validity of the academic engagement scale using Confirmatory Factor Analysis (CFA).

Methods

Research Design

This study employed a quantitative research design to examine the factorial structure and validate the psychometric properties of the Academic Engagement Scale for undergraduate students in a distance learning context.

Population

The population comprised undergraduate students at Sukhothai Thammathirat Open University (STOU) who were actively enrolled during the study period. The sample consisted of 440 undergraduate students who voluntarily participated in the online survey. The use of active enrollment status, rather than a fixed cohort year, reflects the flexible enrollment structure of open universities, where students may enroll intermittently over extended periods.

Sampling and Sample Procedures

The sample size was determined based on recommendations for structural equation modeling. According to Tabachnick and Fidell (2013), a sample size greater than 300 is considered adequate for factor analysis. A total of 500 questionnaires were distributed, and 440 valid responses were received, representing a response rate of 88%.

The sample was obtained using a cluster sampling technique based on three academic disciplines: Social Sciences, Health Sciences, and Science and Technology. This approach was adopted to ensure representation across the major academic fields within the university. Due to the flexible enrollment structure of open universities, students at Sukhothai Thammathirat Open University (STOU) do not follow a conventional year-level progression. In addition, students may enroll in courses across different academic levels simultaneously. As a result, year level does not accurately reflect students' academic standing or learning experience and is therefore less appropriate for grouping purposes in this context.

Table 1 Sample size

Academic field group	Target Sample Size (n)	Actual Responses Collected (n)
Social Sciences	250	285
Health Sciences	150	103
Science and Technology	100	52
Total	500	440

Instrument

The Academic Engagement Scale was adapted from the frameworks of Fredricks et al. (2004) and Wangrunkit (2018). The instrument consists of 27 items measured on a 5-point Likert scale, ranging from 1 (Strongly Disagree) to 5 (Strongly Agree). The instrument consists of three dimensions: Cognitive Engagement (12 items), Emotional Engagement (11 items), and Behavioral Engagement (4 items). The instrument was originally developed and validated in a traditional university context (Wangrunkit, 2018). In the present study, the scale was refined to align with the characteristics of students in an open and distance learning environment at STOU. Minor modifications were made to item wording to reflect the learning context while preserving the original conceptual structure. Given that the construct and dimensional structure had been previously established (Fredricks et al., 2004), a theory-driven approach was adopted. Therefore, Confirmatory Factor Analysis (CFA) was employed to validate the hypothesized structure in the new context. Exploratory Factor Analysis (EFA) was not conducted, as the study aimed to confirm an existing theoretical model rather than to explore a new factor structure. Conducting both EFA and CFA on the same dataset may also lead to overfitting and is generally not recommended (Kline, 2016).

Data Collection

Data were collected through an online survey from undergraduate students at Sukhothai Thammathirat Open University (STOU) who were currently enrolled and willing to participate voluntarily. Participants were informed about the purpose of the study and their rights, and informed consent was obtained prior to participation. The questionnaire was administered across the three academic clusters, and responses were screened for completeness and consistency. Cases with missing or invalid data were excluded.

Data Analysis

Confirmatory Factor Analysis (CFA) was conducted using the Maximum Likelihood (ML) estimation method to examine the factorial structure of the Academic Engagement Scale (Kline, 2016). Model fit was evaluated using multiple indices, including chi-square (χ^2), the Root Mean Square Error of Approximation (RMSEA), the Standardized Root Mean Square Residual (SRMR), the Comparative Fit Index (CFI), and the Normed Fit Index (NFI). Acceptable model fit was determined based on the following criteria: RMSEA < .08, SRMR < .08, and CFI and NFI > .90 (Hu & Bentler, 1999). A χ^2/df value less than 5 was considered indicative of acceptable model fit (Kline, 2016). Given the sensitivity of the chi-square statistic to sample size, a significant χ^2 value does not necessarily indicate poor model fit (Kline, 2016). Reliability and construct validity were assessed using Cronbach's alpha, Composite Reliability (CR), and Average Variance Extracted (AVE). An AVE value greater than .50 and a CR value greater than .70 were considered acceptable (Hair et al., 2010).

Ethical Considerations

This study received approval from the Human Research Ethics Committee of Nakhon Ratchasima Rajabhat University (Ethics Number: HE-RDI-NRRU 139/2024). All participants were informed about the study and provided informed consent prior to data collection.

Results

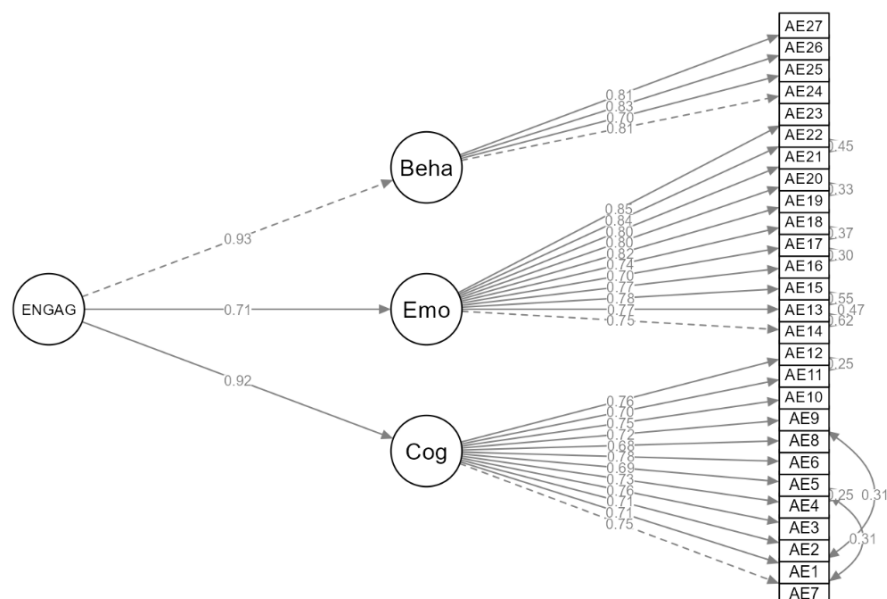
Demographic Characteristics of the Sample

The demographic characteristics of the 440 undergraduate students from STOU are presented in Table 2. The majority of the participants were female ($n = 338$, 76.8%), while male participants accounted for 23.2% ($n = 102$). In terms of age distribution, the largest group consisted of individuals aged 20–25 years ($n = 270$, 61.4%). This was followed by the 36–40 age group ($n = 40$, 9.1%), the 31–35 age group ($n = 39$, 8.9%), and the 26–30 age group ($n = 36$, 8.2%), respectively. Regarding academic clusters, most participants were from Social Sciences ($n = 285$, 64.8%), followed by Health Sciences ($n = 103$, 23.4%) and Science and Technology ($n = 52$, 11.8%).

Table 2 Frequency and Percentage of General Characteristics of the Sample

Demographic Variables	Frequency (n)	Percentage (%)
1. Gender		
Male	102	23.2
Female	338	76.8
2. Age		
20 – 25 years	270	61.4
26 – 30 years	36	8.2
31 – 35 years	39	8.9
36 – 40 years	40	9.1
41 – 45 years	16	3.6
46 – 50 years	16	3.6
51 – 55 years	10	2.3
56 – 60 years	10	2.3
61 years and above	3	0.7
3. Academic Cluster		
Social Sciences	285	64.8
Health Sciences	103	23.4
Science and Technology	52	11.8
Total	440	100

The hypothesized second-order model of academic engagement is presented in Figure 1.

**Figure 1** Second-order CFA model of academic engagement

Confirmatory Factor Analysis

The academic engagement scale items met acceptable thresholds for univariate normality, with skewness and kurtosis within Kline's (2016) recommended range (skewness < 3; kurtosis < 8). The highest skewness was 1.67 (Item 20) and the highest kurtosis was 3.47 (Item 19). The model fit indices indicated an acceptable fit to the data: $\chi^2/df = 3.22$, RMSEA = .07, CFI = .92, NFI = .91, and SRMR = .04. The chi-square statistic was significant ($p < .05$), which is common in large samples. All second-order constructs were significantly associated with their respective first-order subdimensions. Specifically, the behavioral domain, comprising 4 items across three subdimensions, showed standardized factor loadings ranging from .70 to .83 ($p < .001$). The cognitive domain, consisting of 12 items across three subdimensions, yielded loadings between .68 and .78 ($p < .001$). Lastly, the emotional domain, comprising 11 items across three subdimensions, demonstrated loadings ranging from .70 to .85 ($p < .001$). These findings support the proposed second-order structure of academic engagement (Table 3).

Table 3 Validation results of the academic engagement scale

No	Item	Loading	SE	AVE	CR
First-order constructs					
Cognitive Engagement				.53	.93
AE7	I ask myself questions to review my understanding of what I have read.	.75**	-		
AE1	I devote sufficient time to reading the course materials.	.71**	.06		
AE2	Before completing end-of-unit activities, I make sure to understand the questions or issues involved.	.71**	.06		
AE3	I connect what I read in the course materials with my prior knowledge to enhance my understanding.	.76**	.06		
AE4	When reading the course materials, I create examples to help me better understand the content.	.73**	.05		
AE5	I summarize the key points from the course materials I have read.	.69**	.05		
AE6	Whenever I read course materials or view learning media, I try to fully understand the content of the lessons.	.78**	.06		
AE8	I seek additional knowledge from other learning resources beyond the course materials.	.68**	.06		
AE9	I review key points before examinations.	.72**	.05		

Table 3 (Continued)

No	Item	Loading	SE	AVE	CR
AE10	I set goals for reading the course materials and strive to accomplish them every time.	.75**	.06		
AE11	I check the accuracy of my work after completing end-of-unit activities.	.70**	.05		
AE12	I reassess my understanding after completing end-of-unit activities.	.76**	.05		
Emotional Engagement				0.62	0.95
AE14	I feel comfortable and happy when I am at the university.	.75**	-		
AE13	I am happy to be studying at this university.	.77**	.03		
AE15	I feel proud to be a student of this university.	.78**	.04		
AE16	I enjoy learning new things.	.77**	.06		
AE17	I am aware of the importance of my studies.	.70**	.05		
AE18	I am determined to successfully complete my studies so that I can apply the knowledge gained.	.74**	.05		
AE19	I feel that instructors provide academic support.	.82**	.05		
AE20	I feel that instructors care about students.	.80**	.05		
AE21	I feel that university personnel (peers, instructors, and staff) listen to opinions and suggestions.	.80**	.06		
AE22	I feel that university personnel (peers, instructors, and staff) treat students fairly.	.83**	.05		
AE23	I feel that I receive academic support from university personnel.	.85**	.05		
Behavioral Engagement				0.62	0.87
AE24	I allocate my time effectively to complete reading the course materials before examinations.	.81**	-		
AE25	I regularly keep up with university-related information concerning my studies.	.70**	.05		
AE26	I regulate myself to follow the designated study schedule.	.83**	.05		
AE27	I manage distractions in my surroundings that interfere with reading the course materials.	.81**	.05		
Second-order constructs					
Behavioral Engagement		.93	-		
Cognitive Engagement		.92	.07		
Emotional Engagement		.71	.05		

Note: SE = standard error, AVE = average variance extracted, CR = composite reliability.

Convergent Validity

Convergent validity was evaluated using average variance extracted (AVE) and composite reliability (CR). All first-order constructs showed significant standardized loadings (.68–.85, $p < .001$), indicating strong item-construct relationships. AVE values ranged from .53 to .62 and CR from .87 to .95, surpassing recommended thresholds. These results confirm that the constructs captured substantial variance from their indicators.

Discriminant Validity

Discriminant validity was assessed using the Heterotrait–Monotrait ratio (HTMT), which is considered a more rigorous criterion than traditional methods. As shown in Table 4, all HTMT values ranged from .66 to .83. These values remain strictly below the conservative threshold of .85 (Henseler et al., 2015). The results confirm that the cognitive, emotional, and behavioral dimensions are empirically distinct.

Table 4 The heterotrait–monotrait ratio of correlations

	Cognitive	Emotional	Behavioral
Cognitive	1.00		
Emotional	.66	1.00	
Behavioral	.83	.68	1.00

Internal Consistency

The reliability of the scores was assessed in terms of internal consistency. Cronbach's alpha for the 27-item academic engagement scale was .96, indicating excellent internal consistency reliability. When considering each domain, the Cronbach's alpha values for the cognitive, emotional, and behavioral dimensions were .93, .95, and .87, respectively. In addition, the analysis of inter-item and item–total correlations indicated that the items were not redundant. The average inter-item correlation was below .74, well under the .80 threshold typically considered indicative of redundancy (Streiner et al., 2015). Furthermore, all corrected item–total correlations exceeded .30, the generally accepted minimum for item retention and evidence of meaningful contribution to the overall construct (Boateng et al., 2018). These results suggest that the scale was internally consistent while maintaining dimensional distinctiveness across its subdomains.

Discussion

This study provides empirical support for a multidimensional, second-order model of academic engagement in a distance learning context. The findings indicate that academic engagement can be conceptualized as a hierarchical construct comprising cognitive, emotional, and behavioral dimensions. This result is consistent with prior theoretical frameworks (Fredricks et al., 2004), suggesting that engagement is a complex and integrated construct rather than a set of isolated components. In addition, the findings are consistent with empirical evidence reported by Wang and Eccles (2012), which demonstrated that multiple dimensions of engagement interact to influence students' learning processes and academic outcomes. Compared to widely used engagement instruments such as the Utrecht Work Engagement Scale for Students (UWES-S) and the National Survey of Student Engagement (NSSE), the present findings highlight the importance of context-specific measurement in open university settings, where learning structures and student experiences differ substantially from traditional higher education environments.

At the first-order level, cognitive engagement demonstrated relatively strong associations with its indicators, highlighting the importance of self-regulated learning processes in distance education. In line with Zimmerman's (2000) theory of self-regulated learning, students in open university systems are required to actively manage their own learning, including planning, monitoring, and evaluating their academic progress. This finding is consistent with previous empirical studies showing that self-regulated learners tend to exhibit higher levels of engagement and persistence in technology-mediated learning environments (Lerdpornkulrat, 2023). These results suggest that cognitive engagement appears to play an important role in supporting students' adaptation to autonomous learning environments.

Emotional engagement was also found to be an important dimension, particularly in relation to perceived institutional support and sense of belonging. This finding aligns with prior research indicating that emotional connection to the institution contributes to student persistence in distance learning contexts (Wang & Eccles, 2012). While learners in open universities may experience limited face-to-face interaction, their perceived connection to instructors and the institution remains an important aspect of sustained engagement. From the researcher's perspective, this may reflect the role of perceived support as a psychological mechanism that helps reduce feelings of isolation commonly associated with distance learning environments.

Behavioral engagement, reflected through observable study-related actions such as time management and adherence to study schedules, further supports the multidimensional nature of academic engagement. The results indicate that behavioral engagement complements cognitive and emotional dimensions, contributing to a more comprehensive understanding of student participation in distance learning environments. This finding is consistent with previous studies that emphasize the role of active participation and study behaviors in maintaining engagement in online learning contexts (Fredricks et al., 2004).

An additional point worth noting is the relatively lower second-order loading of the emotional engagement dimension (.71) compared to the cognitive and behavioral components. This may reflect contextual characteristics of distance learning environments, where emotional attachment to the institution is less directly reinforced due to limited face-to-face interaction. It is also possible that some items in this dimension capture perceived institutional support rather than emotional engagement itself, which may have influenced the strength of the loading. Furthermore, although the model demonstrated acceptable fit according to established criteria, the fit indices suggest that there is still room for improvement. Therefore, the findings should be interpreted with caution, and future research is encouraged to refine the measurement model and explore alternative model specifications.

From a practical perspective, the findings suggest that the validated scale may serve as a useful tool for assessing student engagement in open university settings. By identifying patterns of engagement across cognitive, emotional, and behavioral dimensions, educators and institutions may better understand areas where students require additional support. However, further research is needed to examine how these engagement profiles relate to academic outcomes and retention across different learner groups. In addition, the study relies on a well-established theoretical framework, which may not fully capture emerging forms of engagement in rapidly evolving digital learning environments.

Limitations and Future Research

Despite the acceptable psychometric properties of the scale, several limitations should be acknowledged to provide a balanced interpretation of the results. First, the data were collected exclusively through self-report measures, which may be subject to social desirability bias—where students might overstate their level of engagement. Future research should triangulate these findings with objective behavioral data from Learning Management Systems (LMS), such as login frequency, time spent on modules, and assignment submission

patterns, to provide a more comprehensive behavioral profile. Second, the disproportionate sample size in the Science and Technology group ($n=52$) suggests that subsequent studies should employ Multi-Group Confirmatory Factor Analysis (MGCFA) to establish measurement invariance across disciplines. Finally, as this cross-sectional study captured engagement at a single point in time, longitudinal designs are recommended to monitor the dynamic fluctuations of engagement and determine its long-term causal impact on academic persistence from enrollment to graduation. Although the instrument was adapted from previously validated sources, content validity was not re-evaluated using the Index of Item-Objective Congruence (IOC) in the current study. Future research should consider conducting expert-based content validation to further strengthen the instrument and ensure its suitability across different educational contexts.

Conclusion

This study provides empirical support for an academic engagement scale conceptualized as a multidimensional, second-order construct comprising cognitive, emotional, and behavioral dimensions. The results of the confirmatory factor analysis indicate that the proposed hierarchical model demonstrates acceptable fit, with all dimensions contributing to the overall construct. Evidence of convergent validity, discriminant validity, and internal consistency suggests that the scale possesses satisfactory psychometric properties.

The findings highlight that academic engagement in distance learning environments is a multidimensional construct involving both behavioral actions and underlying psychological processes. The validated scale may provide a useful foundation for future research and practice in open university contexts.

Declarations

Funding: This research was supported by the Sukhothai Thammathirat Open University (STOU) Academic Research Grant (Innovation Research Category) for the fiscal year 2024.

Conflicts of Interest: The authors declare no conflicts of interest.

Ethical Approval Statement: The study was conducted in accordance with the Declaration of Helsinki, and approved by the Institutional Review Board from the Human Research Ethics Committee at the Research and Development Institute Nakhon Ratchasima Rajabhat University (Ethics Number: HE-RDI-NRRU 139/2024). The approval was granted on 27 September 2024 and is valid until 26 September 2025.

References

- Al-Fraihat, D., Joy, M., Masa'deh, R., & Sinclair, J. (2020). Evaluating e-learning systems success: An empirical study. *Computers in Human Behavior, 102*, 67–86.
<https://doi.org/10.1016/j.chb.2019.08.004>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quinonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health, 6*, 149.
<https://doi.org/10.3389/fpubh.2018.00149>
- Fredricks, J. A. (2012). School engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 1–11). Springer.
- Fredricks, J. A., Blumenfeld, P. C., & Paris, A. H. (2004). School engagement: Potential of the concept, state of the evidence. *Review of Educational Research, 74*(1), 59–109.
<https://doi.org/10.3102/00346543074001059>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Pearson.
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135. <https://doi.org/10.1007/s11747-014-0403-8>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. <https://doi.org/10.1080/10705519909540118>
- Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Publications.
- Kuh, G. D. (2001). Assessing what really matters to student learning: Inside the National Survey of Student Engagement. *Change: The Magazine of Higher Learning, 33*(3), 10–17.
<https://doi.org/10.1080/00091380109601795>
- Lerdpornkulrat, T. (2023). Analyzing student learning behavior in a blended-learning environment in Thai higher education. *Journal of New Horizon in Higher Education, 1*(1), 38–46. <https://so09.tci-thaijo.org/index.php/NHHE/article/view/2147>
- Moore, M. G. (1993). Theory of transactional distance. In D. Keegan (Ed.), *Theoretical principles of distance education* (pp. 22–38). Routledge.
- Piaget, J. (1970). *Science of education and the psychology of the child* (D. Coltman, Trans.). Orion Press.

- Schaufeli, W. B., Martínez, I. M., Pinto, A. M., Salanova, M., & Bakker, A. B. (2002). Burnout and engagement in university students: A cross-national study. *Journal of Cross-Cultural Psychology, 33*(5), 464–481. <https://doi.org/10.1177/0022022102033005003>
- Streiner, D. L., Norman, G. R., & Cairney, J. (2015). *Health measurement scales: A practical guide to their development and use* (5th ed.). Oxford University Press.
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Pearson Education.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.
- Wang, M. T., & Eccles, J. S. (2012). Social support matters: Longitudinal effects of social support on three dimensions of school engagement from middle to high school. *Child Development, 83*(3), 877–895. <https://doi.org/10.1111/j.1467-8624.2012.01745.x>
- Zimmerman, B. J. (2000). Attaining self-regulation: A social cognitive perspective. In M. Boekaerts, P. R. Pintrich, & M. Zeidner (Eds.), *Handbook of self-regulation* (pp. 13–39). Academic Press. <https://doi.org/10.1016/B978-012109890-2/50031-7>

Translate Thai References

- Metheekul, P. (2022). Factors contributing to learning success in distance education: A case study of undergraduate students at Sukhothai Thammathirat Open University. *Silpakorn University Journal of Education, 20*(1), 85–101. (In Thai)
- Sungsri, S. (2018). A model of educational services to enable graduate students of Sukhothai Thammathirat Open University to achieve learning success in the distance education system. *STOU Education Journal, 10*(2), 65–78. (In Thai)
- Wangrunakit, P. (2018). *Development of indicators and program to enhance academic engagement of undergraduate students with personal best goals as a mediator* [Doctoral dissertation]. Srinakharinwirot University. (In Thai)

Development Guidelines for Lifelong Learning Among Retirees in Guangdong Province, People's Republic of China: A Practitioner-oriented Inquiry

Zhou Hong¹, Teerawat Monthaisong^{2*}, and Wacharapatr Techawattanasiridumrong³

(Received: May 24, 2026; Revised: June 25, 2026; Accepted: June 28, 2026)

Abstract

China's accelerating population ageing, the State Council's 2024 Silver Economy initiative, and the expansion of the Seniors University network have repositioned community elder learning centers as a strategic site for late-life development. This study examined how such centers can convert empirical evidence on the determinants of lifelong learning into culturally embedded guidance for practice. The study pursued four objectives: to examine the level of the factors influencing lifelong learning, to examine the level of lifelong learning across its five dimensions, to analyze the predictive contribution of the institutional, family, and learner domains, and to propose practitioner-oriented development guidelines. Using a sequential explanatory mixed-methods design, the research surveyed 384 retirees aged 60 years and above in Guangdong Province, selected through multistage random sampling, with a 60-item Likert questionnaire (overall reliability $\alpha = .985$) measuring three antecedent domains—institutional, family, and learner—and five outcome dimensions of lifelong learning. Quantitative data were analyzed using means, standard deviations, and stepwise multiple regression, and the interview data were analyzed thematically. The three antecedent domains and overall lifelong learning were all rated at a high level. Stepwise multiple regression confirmed all three domains as significant positive predictors, jointly explaining 68.0% of the variance ($R^2 = .680$, $p < .01$), with the learner domain dominant ($\beta = .312$). Semi-structured interviews with seven purposively selected experts were analyzed thematically, and the draft

¹ Ph.D. Student, Doctor of Education Program, Faculty of Education, Phranakhon Si Ayutthaya Rajabhat University

² Assistant Professor, Doctor of Education Program, Faculty of Education, Phranakhon Si Ayutthaya Rajabhat University

³ Lecturer, Doctor of Education Program, Faculty of Education, Phranakhon Si Ayutthaya Rajabhat University

* Corresponding Author E-mail: teerawatmontaisong@gmail.com

guidance was appraised for feasibility and utility by 12 senior learning-center administrators. The study delivered 24 practitioner-oriented development guidelines distributed across the three domains, with exemplar guidance illustrating their operation within Guangdong's 4-2-1 family structure and WeChat-saturated communication environment. Contributions include a relational reconception of the elder learning center, an implementation sequence aligned with the Silver Economy policy architecture, and a feasibility profile flagging the Silver Credit accumulation system as the guideline most dependent on cross-sectoral coordination.

Keywords: development guidelines, lifelong learning, retirees, community elder learning centers, Guangdong Province

Introduction

The accelerated demographic transition unfolding across East Asia has reframed the policy conversation about later life, shifting attention from care provision toward sustained learning, social participation, and civic contribution (United Nations, 2020). By the close of 2024, the Chinese population aged 60 and above had reached approximately 310 million, or roughly 22% of the country's residents, and the Ministry of Education had recorded more than 76,000 senior universities and grassroots learning centers serving over 22 million older students (Xinhua News Agency, 2024). The promulgation in January 2024 of the State Council's guidelines on developing the Silver Economy formally elevated elder education to a national strategic agenda and earmarked dedicated funding for community-based learning centers as part of an inclusive ageing infrastructure (UNDP China, 2025).

Guangdong Province occupies a particularly significant position within this transition. As the most populous and economically developed province in southern China and a core constituent of the Greater Bay Area, Guangdong simultaneously confronts the steepest population ageing curve outside the country's northeast and the most concentrated exposure to mobile payments, telehealth, e-commerce, and short-video platforms (Wu & Jiang, 2025). Retirees in this region must therefore adapt to rapidly changing digital infrastructures even as they navigate the relational expectations of the prevailing 4-2-1 family structure, in which a single adult child is responsible for two parents and four grandparents (Zhang & Cheng, 2024). This combination of demographic pressure, accelerated digital transition, and contracted family caregiving capacity makes Guangdong a critical site for examining how community elder learning centers can mobilize lifelong learning at scale.

Lifelong learning is understood here as purposeful learning undertaken throughout life to enlarge knowledge, skills, and competencies (European Commission, 2001), operationalized through five interlocking dimensions of the adult learner: attitude toward learning, learning motivation, learning competency, self-directed knowledge-seeking, and communication ability (Kirby et al., 2010). Active ageing literature confirms that participation in such learning is positively associated with cognitive vitality, psychological well-being, reduced depression, and stronger social integration in later life (Narushima et al., 2018; World Health Organization, 2020). The Chinese context introduces additional complexity. Confucian self-cultivation supplies a culturally resonant rationale for late-life study (Tan, 2017), but contemporary fieldwork in Guangzhou suggests that traditional reciprocity is fading and that retirees increasingly construct themselves as self-reliant elders rather than as dependents (Zhang & Cheng, 2024). The pandemic period further exposed a stratified digital divide that constrained the most isolated retirees from accessing online learning opportunities (Sun et al., 2024).

Although the determinants of lifelong learning among older adults are well theorized, translation from determinant identification to operational guidance remains uneven. Regression evidence consistently points to a triadic structure of institutional, family, and learner factors (Mu, 2021; Wu & Jiang, 2025), yet community learning centers continue to rely on generic curricula imported from formal education rather than on guidance calibrated to the cultural and infrastructural conditions of their constituents. Recent systematic reviews of digital-literacy programming for older adults reach a similar conclusion: instructional content has multiplied, but design choices have not yet caught up with the relational, motivational, and physiological needs of late-life learners (Xu et al, 2024; Xu et al., 2025). The present study addresses this translation gap. It pairs a regression analysis of the determinants of lifelong learning with an expert-informed qualitative synthesis to produce a practitioner-oriented set of development guidelines tailored to the elder learning centers of Guangdong Province.

Objectives

The study pursued four interrelated objectives:

1. to examine the level of factors influencing lifelong learning among retirees aged 60 years and above in Guangdong Province;
2. to examine the level of lifelong learning among the same population across its five operational dimensions;
3. to analyze the predictive contribution of institutional, family, and learner factors to lifelong learning; and

4. to propose a practitioner-oriented set of development guidelines, derived from expert interviews and appraised by senior learning-center administrators, for advancing lifelong learning in this population.

Research Methodology

1. Research Design

A sequential explanatory mixed-methods design was adopted. Phase 1 used a cross-sectional survey to quantify the level and predictive structure of the determinants. Phase 2 used expert interviews and an administrator feasibility-utility appraisal to translate those determinants into context-appropriate guidance. The two phases were integrated through the elder learning center as the unit of implementation: each guideline articulated in Phase 2 had to address a determinant validated in Phase 1 and had to be exactable within the operational conditions of community senior education in Guangdong. The overall two-phase design is summarized in Figure 1.

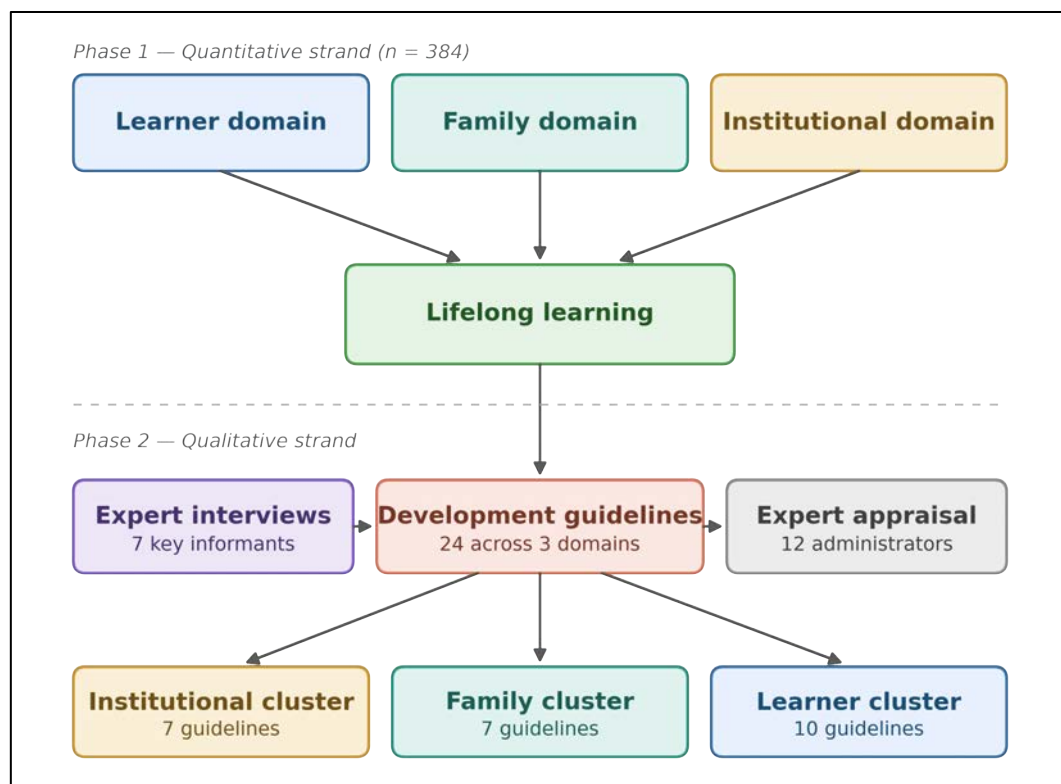


Figure 1: Research Framework of the Two-Phase Mixed-Methods Design

Note. Phase 1 examined the predictive contribution of the institutional, family, and learner domains to lifelong learning; Phase 2 translated the findings into 24 development guidelines that were appraised by senior administrators.

2. Phase 1: Quantitative Strand

2.1 Population and Sample

The population comprised retirees aged 60 years and above who were currently enrolled in at least one organized learning activity at a community senior center, an elder learning institution, or a community college in Guangdong Province. The sample consisted of 384 retirees, a size determined using the Krejcie and Morgan (1970) table for a population exceeding 100,000.

2.2 Sampling Method

Respondents were selected through multistage random sampling. First, three cities representing urban, suburban, and rural contexts within Guangdong were selected. Second, eligible learning centers were identified within each city. Third, individual respondents were recruited through systematic random selection from the registers of those centers.

2.3 Research Instrument

Data were collected with a structured questionnaire organized in three parts. Part 1 collected demographic information (sex, age band, completed education). Part 2 measured the three antecedent domains across 60 items: the institutional domain (classroom climate, learner–teacher relationships, learner–peer relationships, and learner-centered instruction), the family domain (family support, learner–family relationships), and the learner domain (locus of control, future orientation, self-esteem, self-knowledge). Part 3 measured the five lifelong learning dimensions. All items used a five-point Likert format from 1 (strongly disagree) to 5 (strongly agree). Content validity was established through expert review, yielding an Index of Item-Objective Congruence of 1.00. A pilot test with 30 retirees in Sichuan Province—selected to ensure comparability without contaminating the main sample—produced Cronbach's alpha coefficients of .990 for the antecedent scale and .980 for the lifelong learning scale, indicating high reliability.

2.4 Data Collection

Phase 1 data were collected over a four-month period by trained research assistants. Informed consent was obtained from each respondent before completion of the questionnaire.

2.5 Data Analysis

Demographic data were summarized using frequencies and percentages. The levels of the antecedent factors and of lifelong learning were reported as means and standard deviations and interpreted against a conventional five-band rubric. Prior to the regression analysis, the relationships among the three predictor domains were examined using Pearson product-moment correlation, and multicollinearity was assessed through tolerance and variance inflation factor (VIF) statistics. The predictive contribution of the three domains was then estimated using stepwise multiple regression analysis, which was preferred over simultaneous entry because the study sought the smallest set of predictors capable of explaining a substantial proportion of the variance in lifelong learning.

3. Phase 2: Qualitative Strand

3.1 Key Informants and Sampling

Phase 2 used purposive sampling. Seven key informants were selected on the basis of documented expertise in lifelong education, gerontology, senior learning-center administration, or policy formulation for older adult education. A separate panel of 12 senior administrators—each responsible for the operational management of an elder learning center in Guangdong—subsequently appraised the draft guidelines for feasibility and utility.

3.2 Research Instrument

A semi-structured interview schedule was used, organized around six themes derived from the regression results: pedagogical climate, physical environment, family activation, motivational architecture, digital communication, and recognition systems. The schedule was reviewed by three experts for content validity before use. The administrator appraisal used a five-point feasibility-and-utility rating form.

3.3 Data Collection

The interviews were audio-recorded and transcribed verbatim, with informed consent obtained from each key informant. The draft guidelines derived from the analysis were subsequently distributed to the 12-member administrator panel for rating.

3.4 Data Analysis

The interview transcripts were analyzed using thematic analysis. Two researchers first read the transcripts independently and generated initial codes through open coding. The codes

were then compiled into a shared codebook and grouped into candidate themes aligned with the six interview domains. Inter-coder agreement was examined, and the discrepancies that arose were resolved through discussion until consensus was reached, after which the themes were reviewed, defined, and named. The administrator panel's mean feasibility and utility ratings for each guideline informed the final wording and prioritization of the guidance.

Results

In line with the study's four objectives, the findings are reported in four parts: the levels of the factors influencing lifelong learning (Objective 1), the level of lifelong learning across its five dimensions (Objective 2), the predictive contribution of the three domains (Objective 3), and the expert-informed development guidelines with their feasibility and utility appraisal (Objective 4).

1. Levels of Factors Influencing Lifelong Learning

Descriptive analysis confirmed that all three antecedent domains were rated at a high level. Table 1 presents the overall pattern.

Table 1 Levels of Factors Influencing Lifelong Learning Among Retirees in Guangdong (n = 384)

Antecedent Domain	Mean (<i>M</i>)	<i>SD</i>	Level
Learner domain	4.48	0.46	High
Family domain	4.45	0.50	High
Institutional domain	4.01	0.49	High
Overall	4.31	0.48	High

The learner domain recorded the highest mean ($M = 4.48$, $SD = 0.46$), followed closely by the family domain ($M = 4.45$, $SD = 0.50$), with the institutional domain registering the lowest mean among the three ($M = 4.01$, $SD = 0.49$). The proximity of the learner and family means is consistent with ethnographic reports that retirees in Guangdong rely on personal initiative reinforced by intergenerational encouragement (Sun et al., 2024; Zhang & Cheng, 2024).

2. Levels of Lifelong Learning

The composite level of lifelong learning was rated high ($M = 4.04$, $SD = 0.47$). Among the five dimensions, attitude toward learning recorded the highest mean ($M = 4.22$), while communication ability recorded the lowest ($M = 3.92$). Self-directed knowledge-seeking and learning motivation occupied an intermediate position. The gap between strong learning

attitudes and weaker communication competence—observed across all three sampling sites—signals an asymmetry that any practitioner-oriented guidance must address.

Table 2 Levels of Lifelong Learning by Dimension (n = 384)

Lifelong Learning Dimension	Mean (<i>M</i>)	<i>SD</i>	Level
Attitude toward learning	4.22	0.45	High
Learning competency	4.11	0.47	High
Self-directed knowledge-seeking	4.05	0.49	High
Learning motivation	3.99	0.48	High
Communication ability	3.92	0.51	High
Overall	4.04	0.47	High

3. Predictive Contribution of the Three Domains

Before estimating the regression model, the relationships among the three predictor domains were examined. The domains were positively and moderately intercorrelated (all $p < .01$), and the collinearity diagnostics were within acceptable limits (tolerance $> .60$ and VIF < 2.0 for every domain), indicating that multicollinearity did not threaten the stability of the regression estimates.

Stepwise multiple regression identified all three antecedent domains as statistically significant positive predictors of lifelong learning at the .01 level. The combined model explained 68.0% of the variance in the outcome ($R^2 = .680$). Standardized regression coefficients indicated that the learner domain contributed most strongly ($\beta = .312$), followed by the family domain ($\beta = .262$) and the institutional domain ($\beta = .238$). Table 3 summarizes the model.

Table 3 Stepwise Multiple Regression Predicting Lifelong Learning (n = 384)

Predictor Domain	B	SE	β	t	p
Learner	0.298	0.041	.312	7.27	$< .01$
Family	0.241	0.038	.262	6.34	$< .01$
Institutional	0.214	0.040	.238	5.35	$< .01$
Constant	0.612	0.105	—	5.83	$< .01$

Note. $R^2 = .680$; Adjusted $R^2 = .678$; $F(3, 380) = 269.16$, $p < .01$. Predictors entered using stepwise selection.

The fitted model can be expressed in both raw-score and standardized-score forms. In raw-score (unstandardized) form: $\hat{Y} = 0.612 + 0.298(\text{Learner}) + 0.241(\text{Family}) + 0.214(\text{Institutional})$. In standardized-score form: $Z\hat{Y} = 0.312(Z\text{Learner}) + 0.262(Z\text{Family}) + 0.238(Z\text{Institutional})$. Both equations confirm that the learner domain exerts the strongest unique contribution to lifelong learning, followed by the family and institutional domains.

4. Expert-Informed Development Guidelines

Thematic analysis of the seven expert interviews produced 24 development guidelines distributed across the three predictor domains: seven institutional, seven family, and ten learner guidelines. Rather than enumerating the full set, this section presents the guiding logic of each cluster together with exemplar guidance, since the Phase 2 evaluation revealed that practitioners absorbed and adopted the guidelines more readily when their underlying rationale was visible. The complete set of all 24 guidelines is provided in the Appendix (Appendix Table).

4.1 Institutional Cluster: From Classroom to Relational Space

Experts converged on the position that elder learning centers in Guangdong should reduce reliance on lecture-format pedagogy and reorganize space, time, and ritual to make sociality the primary medium of learning. Exemplar guidance includes reconfiguring seating into circular or U-shaped arrangements to flatten the symbolic hierarchy between instructor and learner; designating communal pre-class areas where retirees can socialize informally; and replacing terminal examinations with public recognition ceremonies that honor sustained participation. This cluster operationalizes Oldenburg's (1989) conception of the third place by treating the learning center as the relational anchor between household and former workplace, and it aligns with active ageing evidence that social embeddedness sustains psychological well-being in later life (Narushima et al., 2018; World Health Organization, 2020).

4.2 Family Cluster: Activating the 4-2-1 Household as a Learning Partner

The family guidelines reposition adult children and grandchildren from caregivers to co-participants. Exemplar guidance includes an intergenerational classroom in which grandchildren tutor grandparents in digital tasks while grandparents transmit local cultural practices in return; a WeChat-based progress notification channel through which the learning center shares photos and achievements with family group chats; and a structured family advisory service that supports households in articulating supportive expectations for late-life learning. These guidelines respond to the contemporary 4-2-1 family structure in which

intergenerational caregiving capacity is contracting and traditional reciprocity is being renegotiated (Wang et al., 2024; Zhang & Cheng, 2024). They also anchor the implementation logic in the existing communication infrastructure of Chinese households, where WeChat functions as the de facto channel of family interaction for older adults (Yu et al., 2023).

4.3 Learner Cluster: A Strength-Based, Empowerment Pathway

The learner cluster takes a strength-based stance, treating retirees as carriers of intellectual capital rather than as remedial students. Exemplar guidance includes a Wisdom Inventory administered at course entry to surface each learner's prior expertise, which is then documented in an Honour Portfolio; a Silver Credit accumulation system through which sustained participation is rewarded with locally redeemable benefits; and a reverse-mentoring arrangement in which younger volunteers coach retirees on contemporary digital tasks while learning vocational or cultural skills in return. The Silver Credit system is the most ambitious component, drawing on Skinnerian reinforcement principles, Bandura's (1997) self-efficacy theory, and Dweck's (2006) growth mindset framework to convert episodic motivation into a durable learning trajectory. Reverse mentoring and storytelling-based pedagogy further address the attitude–communication gap observed in Phase 1 by giving retirees authentic communicative roles that draw on their accumulated life experience.

5. Feasibility and Utility Appraisal

The 12 senior administrators rated each domain on a five-point scale. Table 4 summarizes their ratings.

Table 4 Feasibility and Utility Ratings of the Development Guidelines (n = 12 Administrators)

Domain	Mean (M)		Level
	Feasibility	Utility	
Institutional cluster	4.42	4.58	Highest utility; high feasibility
Family cluster	4.39	4.61	Highest utility; high feasibility
Learner cluster	4.18	4.36	High utility; high feasibility

Institutional and family guidance received the strongest utility scores, reflecting administrator confidence that physical reconfiguration and family activation address the daily reality of community learning centers in Guangdong. The learner cluster received slightly lower feasibility ratings, an outcome traceable to the Silver Credit system: while administrators recognized its motivational logic, several expressed concern that operationalizing a digital

points infrastructure would require coordination across education bureaus, social welfare departments, and local commercial partners (UNDP China, 2024). This finding is methodologically important. It indicates that the guidelines vary in implementation horizon, and that adoption should follow a sequenced pathway in which the institutional and family clusters are deployed first, with the more infrastructure-dependent learner mechanisms phased in as cross-sectoral coordination matures. Figure 2 summarizes the predictive model together with the resulting development guidelines and their appraisal.

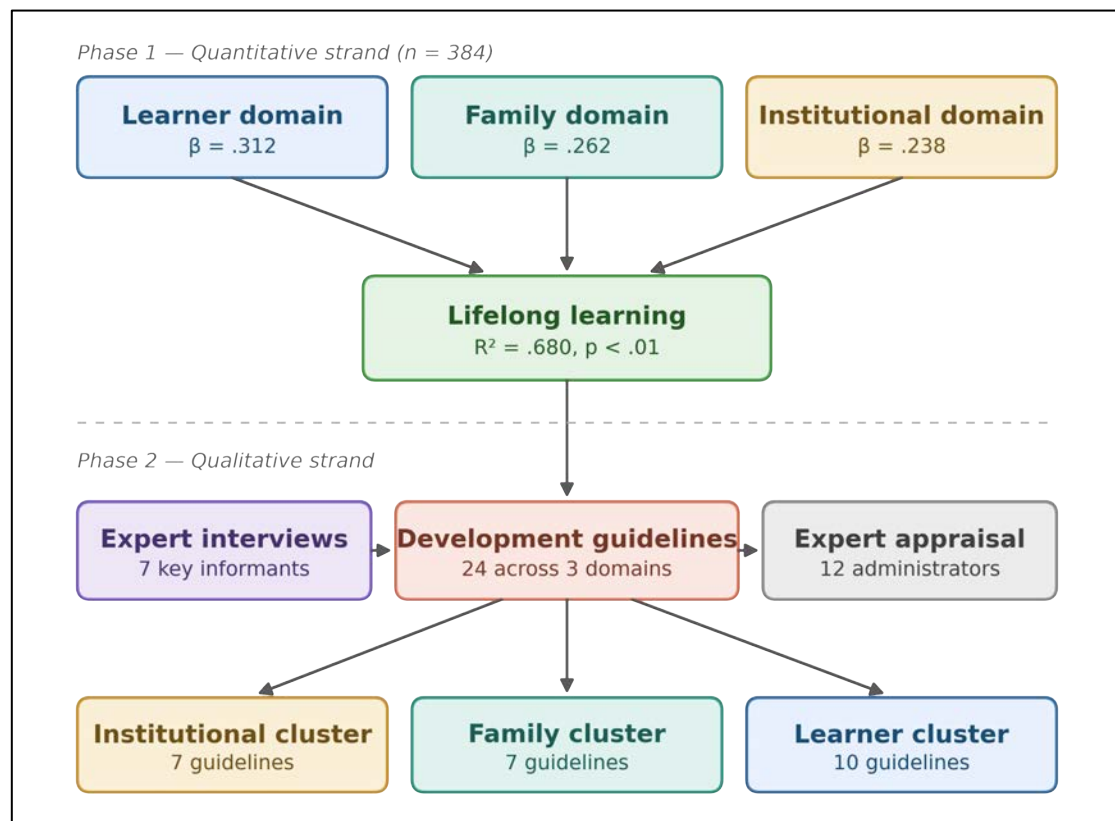


Figure 2: Summary of the Predictive Model and Development Guidelines

Note. β = standardized regression coefficient; R^2 = proportion of variance in lifelong learning explained by the three antecedent domains. All predictors were significant at $p < .01$.

Discussion

The discussion is organized to correspond with the study's four objectives. It interprets in turn the high levels of the factors influencing lifelong learning (Objective 1) and of lifelong learning itself (Objective 2), the predictive structure of the three domains (Objective 3), and the development guidelines derived from the expert interviews and appraised by administrators (Objective 4).

1. Why the Learner Domain Leads the Regression

The primacy of the learner domain ($\beta = .312$) is interpretively significant. It indicates that internal motivation, future orientation, and self-esteem are the conditions under which family and institutional supports become productive rather than merely available. Knowles (1980) anticipated this dynamic in his andragogical theory, which positioned adult learners as self-directed agents whose readiness is the foundation upon which external scaffolds operate. The finding also resonates with recent Chinese evidence that retirees increasingly construct themselves as self-reliant elders who pursue learning for personal meaning rather than as a duty to the household (Zhang & Cheng, 2024). The practitioner implication is direct: guidance that begins by surfacing each learner's intellectual capital—rather than diagnosing deficits—has a stronger theoretical and empirical mandate than remedial pedagogies imported from formal schooling.

2. The Attitude–Communication Asymmetry

Phase 1 produced one analytic surprise: the highest-rated learning dimension was attitude toward learning ($M = 4.22$), and the lowest was communication ability ($M = 3.92$). This asymmetry, replicated across the three sampling cities, suggests that retirees in Guangdong are motivated but communicationally constrained, particularly within the digital channels through which contemporary social participation is increasingly mediated. Recent systematic reviews of digital-literacy education converge on the same conclusion (Xu et al, 2024; Xu et al, 2025). The Family Cluster guidelines respond to this asymmetry by anchoring digital learning in the WeChat infrastructure that already mediates family life (Yu et al., 2023), and the Learner Cluster's reverse-mentoring guidance gives retirees authentic communicative roles that recruit their accumulated experience as a resource rather than positioning them as remedial digital users.

3. Cultural Fit With the 4-2-1 Family Structure

The strength of the family domain ($\beta = .262$) reflects a specifically Chinese cultural configuration in which familial relationships continue to anchor daily life even as caregiving capacity contracts. Li (2018) argues that Confucian self-cultivation supplies a culturally resonant rationale for late-life study, while Wang et al. (2024) document the operational pressure placed on the single adult child within the contemporary 4-2-1 household. The Family Cluster guidelines were designed to reduce this pressure by recasting children and grandchildren as learning partners rather than as exclusive caregivers, and the administrator panel's high utility rating for this cluster indicates that practitioners recognize the strategic value of distributing relational labor across the household.

4. Implementation Sequence and the Silver Economy Context

The feasibility-utility profile suggests a three-stage implementation pathway. Stage 1 implements the institutional guidelines, which lie within direct managerial control and require modest capital expenditure. Stage 2 activates the family guidelines through low-cost digital channels and intergenerational programming. Stage 3 develops the learner guidelines, including the Silver Credit system, in concert with provincial education bureaus and social welfare departments. This sequencing is consistent with the architecture of the State Council's Silver Economy guidance (UNDP China, 2024), which envisions a layered ecosystem of institutional services, family engagement, and individual benefits, and it offers community elder learning centers a tractable migration path from the present to the policy horizon.

Conclusion and Recommendations

1. Conclusion

This study addressed the translation gap between determinant identification and practitioner guidance for lifelong learning among retirees in Guangdong Province. The institutional, family, and learner domains jointly explained 68.0% of the variance in lifelong learning, with the learner domain emerging as the strongest predictor. Expert interviews translated this evidence into 24 development guidelines whose underlying logic was organized around three propositions: the elder learning center is a relational space rather than a transmission classroom; the 4-2-1 household is an active learning partner rather than a passive support; and the retiree is a carrier of intellectual capital rather than a remedial learner. Administrators appraised the institutional and family guidance as having the highest utility and rated the learner guidance highly but with sequencing caveats tied to the cross-sectoral architecture of the Silver Credit system.

2. Recommendations for Policy and Practice

Three recommendations follow directly from the findings. First, provincial and municipal authorities should adopt the development guidelines as a reference standard for the accreditation and evaluation of community elder learning centers in Guangdong, with dedicated funding for Stage 1 modifications—seating reconfiguration, communal areas, accessibility upgrades, and recognition ceremonies—treated as a minimum quality threshold. Second, accreditation criteria should require a structured family engagement component, including a WeChat-based notification channel and at least one intergenerational learning event per academic year. Third, the Silver Credit accumulation system should be piloted within

selected centers in collaboration with education bureaus and social welfare departments, with redemption mechanisms designed to connect sustained learning to locally meaningful benefits.

3. Recommendations for Future Research

Four research directions merit attention. Comparative work between urban and rural learning centers within Guangdong would identify necessary adaptations to the guidelines for less digitally saturated environments. Longitudinal designs are needed to assess whether the guidelines, once adopted, produce sustained changes in attendance, motivation, and well-being over a three-year horizon. The Silver Credit system itself should be evaluated through quasi-experimental designs that estimate its causal effects on the attitude–communication asymmetry observed in this study. Finally, cross-context studies in other Confucian-heritage societies, including South Korea, Vietnam, and Singapore, would clarify which components of the guidance travel and which require local adaptation.

References

- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W. H. Freeman.
- Dweck, C. S. (2006). *Mindset: The new psychology of success*. Random House.
- European Commission. (2001). *Making a European area of lifelong learning a reality*. Office for Official Publications of the European Communities.
[https://www.europarl.europa.eu/meetdocs/committees/cult/20020122/com\(2001\)678_en.pdf](https://www.europarl.europa.eu/meetdocs/committees/cult/20020122/com(2001)678_en.pdf)
- Kirby, J. R., Knapper, C., Lamon, P., & Egnatoff, W. J. (2010). Development of a scale to measure lifelong learning. *International Journal of Lifelong Education*, 29(3), 291–302.
<https://doi.org/10.1080/02601371003751713>
- Knowles, M. S. (1980). *The modern practice of adult education: From pedagogy to andragogy* (2nd ed.). Cambridge Adult Education.
- Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities. *Educational and Psychological Measurement*, 30(3), 607–610.
<https://doi.org/10.1177/001316447003000308>
- Narushima, M., Liu, J., & Diestelkamp, N. (2018). Lifelong learning in active ageing discourse: Its conserving effect on wellbeing, health and vulnerability. *Ageing and Society*, 38(4), 651–675. <https://doi.org/10.1017/S0144686X1600122X>
- Oldenburg, R. (1989). *The great good place*. Paragon House.

- Sun, Y., Wang, H., & Chen, L. (2024). Linking personal initiative and family help as well as social support: A case study of learning challenges and solutions for older adults in rural China during the COVID-19 pandemic. *Frontiers in Public Health*, 12, Article 1384527. <https://doi.org/10.3389/fpubh.2024.1384527>
- Tan, C. (2017). A Confucian perspective of self-cultivation in learning: Its implications for self-directed learning. *Journal of Adult and Continuing Education*, 23(2), 250–262. <https://doi.org/10.1177/1477971417721719>
- UNDP China. (2025). *Financing the silver economy in China: Empowering an inclusive and sustainable ageing society*. https://www.undp.org/sites/g/files/zskgke326/files/2026-02/financing_the_silver_economy_in_china-empowering_an_inclusive_and_sustainable_ageing_society-en.pdf
- United Nations. (2020). *World population ageing 2020*. Department of Economic and Social Affairs. https://www.un.org/development/desa/pd/sites/www.un.org.development.desa.pd/files/undesd_pd_2020_world_population_ageing_highlights.pdf
- Wang, R., Liu, T., & Zhao, Q. (2024). The impact of intergenerational support on older adults' care expectations in rural areas in China. *Frontiers in Public Health*, 12, Article 1423173. <https://doi.org/10.3389/fpubh.2024.1423173>
- World Health Organization. (2020). *Decade of healthy ageing 2021–2030: Baseline report*. <https://www.who.int/publications/i/item/9789240017900>
- Wu, F., & Jiang, F. (2025). Lifelong learning among older adults in China: Characteristics, influencing factors, and heterogeneity analysis. *Educational Gerontology*, 51(7), 821–839. <https://doi.org/10.1080/03601277.2025.2515565>
- Xinhua News Agency. (2024, October 11). China focus: Silver-haired seniors take on university life. *Xinhua News Agency*. <https://english.news.cn/20241011/c080191912cc4205bb05b867ef432089/c.html>
- Xu, L., Lam, Y. Y., Ng, T. K., Peng, P., Suen, W. L., Lee, C. S. Y., & Chu, S. K. W. (2024). Digital literacy education for older adults: A scoping review. *Proceedings of the Association for Information Science and Technology*, 61(1), 691–696. <https://doi.org/10.1002/pa2.1081>
- Xu, L., Ng, D. T. K., Lee, C. S. Y., Peng, P., & Chu, S. K. W. (2025). A systematic review of digital literacy in lifelong learning for older adults: Challenges, strategies, and learning outcomes. *Educational Technology Research and Development*, 73, 3627–3674. <https://doi.org/10.1007/s11423-025-10530-w>

Yu, J., Hu, Z., & Wang, Y. (2023). Chinese older adults' social media use: A study of WeChat from a social practice perspective. *Information Technology & People*, 36(8), 102–124. <https://doi.org/10.1108/ITP-10-2022-0787>

Zhang, M., & Cheng, Q. (2024). Becoming a self-reliant elder: Fading intergenerational reciprocity and pursuing active aging among middle-class seniors in urban China. *SAGE Open*, 14(2), 1–14. <https://doi.org/10.1177/2158244024125916>

Appendix

Appendix Table: The 24 Development Guidelines for Lifelong Learning Among Retirees in Guangdong

Cluster	No.	Development Guideline
Institutional	I1	Reconfigure classroom seating into circular or U-shaped arrangements to flatten the symbolic hierarchy between instructor and learner.
Institutional	I2	Designate communal pre-class areas where retirees can socialize informally before and after sessions.
Institutional	I3	Replace terminal examinations with public recognition ceremonies that honor sustained participation.
Institutional	I4	Train instructors in facilitative, dialogue-based pedagogy in place of lecture-format transmission.
Institutional	I5	Schedule regular small-group and peer-learning activities to strengthen learner–peer relationships.
Institutional	I6	Improve the physical accessibility of learning spaces (lighting, seating, signage, step-free access) for older learners.
Institutional	I7	Operate the learning center as a “third place” open beyond class hours for informal social and cultural activity.
Family	F1	Establish an intergenerational classroom in which grandchildren tutor grandparents in digital tasks while grandparents transmit local cultural practices in return.
Family	F2	Create a WeChat-based progress-notification channel that shares learners' photos and achievements with family group chats.
Family	F3	Provide a structured family advisory service that helps households articulate supportive expectations for late-life learning.
Family	F4	Offer family orientation sessions explaining the goals and benefits of the elder learning center.
Family	F5	Invite family members to periodic open days and learner showcase events.
Family	F6	Develop simple guidance materials that help adult children encourage and sustain their parents' participation.

Appendix Table (Continued)

Cluster	No.	Development Guideline
Family	F7	Recognize and celebrate family involvement to reinforce the household as a learning partner.
Learner	L1	Administer a Wisdom Inventory at course entry to surface each learner's prior expertise.
Learner	L2	Document each learner's expertise and achievements in an Honour Portfolio.
Learner	L3	Introduce a Silver Credit accumulation system that rewards sustained participation with locally redeemable benefits.
Learner	L4	Implement reverse-mentoring in which younger volunteers coach retirees on digital tasks while learning cultural or vocational skills in return.
Learner	L5	Use storytelling-based pedagogy that draws on learners' accumulated life experience.
Learner	L6	Support learners in setting personal learning goals to strengthen future orientation and self-direction.
Learner	L7	Provide structured opportunities for self-directed knowledge-seeking and project work.
Learner	L8	Embed authentic communicative roles (presenting, peer-teaching) to address the attitude–communication gap.
Learner	L9	Offer guided digital-communication practice anchored in everyday WeChat use.
Learner	L10	Provide affirmation and progress feedback that reinforce self-esteem and a growth mindset.

บทบาทการส่งผ่านของความไว้วางใจระหว่างภาวะผู้นำของผู้บริหารกับความพึงพอใจ
ในการทำงานและการสอนของครู : หลักฐานเชิงประจักษ์จากข้อมูล PISA 2022 ของฮ่องกง
The Mediating Role of Trust in the Relationship between Principal Leadership and
Teachers' Job and Teaching Satisfaction: Evidence from Hong Kong's 2022 PISA Data

ธัญชนก ทองนพเก้า¹ สุภชาติ ผดุงผล¹ ธีรนนท์ ศรีวิทัศน์¹ และธมรัตน์ ศิริภาพ^{2*}

Thanyachanok Thongnoppakaow¹ Suphachit Phadungphon¹ Thiranan Sriwithas¹ and Thomrat Siriparp^{2*}

(Received: December 8, 2025; Revised: February 12, 2026; Accepted: February 15, 2026)

บทคัดย่อ

การวิจัยครั้งนี้ เป็นการศึกษาข้อมูลแบบสำรวจ โดยใช้ข้อมูลจากโปรแกรมประเมินสมรรถนะนักเรียนมาตรฐานสากล (Programme for International Student Assessment: PISA) มีวัตถุประสงค์ 1) เพื่อตรวจสอบความสอดคล้องของโมเดลสมการโครงสร้างความพึงพอใจในการทำงานและการสอนของครู 2) เพื่อวิเคราะห์อิทธิพลทางตรงและทางอ้อมของภาวะผู้นำของผู้บริหารที่มีต่อความพึงพอใจในการทำงานและการสอนของครู และ 3) เพื่อตรวจสอบอิทธิพลการส่งผ่านของความไว้วางใจที่มีต่อภาวะผู้นำของผู้บริหารและความพึงพอใจในการทำงานและการสอนของครู กลุ่มตัวอย่างประกอบด้วยครูผู้สอน จำนวน 2,335 คน จากโรงเรียน 155 แห่ง ซึ่งได้มาจากการสุ่มตัวอย่างแบบสองขั้นตอนตามกรอบการสุ่มของโครงการ PISA เครื่องมือที่ใช้ ได้แก่ แบบสอบถามที่มีลักษณะเป็นมาตราส่วนประมาณค่าแบบลิเคิร์ต จำนวน 19 ข้อ มีค่าความเที่ยงรวมของแบบสอบถาม 0.885 งานวิจัยนี้ใช้การวิเคราะห์โมเดลสมการเชิงโครงสร้างและใช้เทคนิค Bootstrap ในการทดสอบอิทธิพลการส่งผ่าน ผลการวิจัยพบว่า โมเดลสมการเชิงโครงสร้างมีความสอดคล้องกับข้อมูลเชิงประจักษ์ ภาวะผู้นำของผู้บริหารมีอิทธิพลเชิงบวกโดยตรงต่อความไว้วางใจของครู ($\beta = 0.50$) ความไว้วางใจของครูมีอิทธิพลเชิงบวกต่อความพึงพอใจในการทำงาน ($\beta = 0.76$) และต่อความพึงพอใจในการสอน ($\beta = 0.43$) ภาวะผู้นำของผู้บริหารไม่มีอิทธิพลทางตรงอย่างมีนัยสำคัญต่อความพึงพอใจทั้งสองมิติ แต่มีอิทธิพลทางอ้อมผ่านความไว้วางใจต่อความพึงพอใจในการทำงาน ($\beta = 0.38$) และต่อความพึงพอใจในการสอน ($\beta = 0.21$) อย่างมีนัยสำคัญทางสถิติที่ .05

สรุปได้ว่าความไว้วางใจของครูต่อผู้บริหารทำหน้าที่เป็นตัวแปรส่งผ่านอย่างสมบูรณ์ ในความสัมพันธ์ระหว่างภาวะผู้นำของผู้บริหารกับความพึงพอใจของครูทั้งสองมิติ ผลการวิจัยชี้ให้เห็นถึงความสำคัญของ

¹ นิสิตดุขุภินันธิต สาขาวิชาการศึกษาบริหารการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

² อาจารย์ สาขาวิชาการศึกษาบริหารการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

¹ Doctoral student, Educational Administration, Faculty of Education, Kasetsart University

² Lecturer, Educational Administration, Faculty of Education, Kasetsart University

* Corresponding Author E-mail: thomrat@gmail.com

การสร้างควมไว้วางใจภายในองค์กรการศึกษา โดยเฉพาะในบริบทโรงเรียนของฮ่องกงที่มีการแข่งขันสูงและมีโครงสร้างการบริหารแบบลำดับชั้น

คำสำคัญ ความไว้วางใจ ภาวะผู้นำ ความพึงพอใจในการทำงาน ความพึงพอใจในการสอน

Abstract

This study employed a survey research design utilizing data from the Programme for International Student Assessment (PISA). The purposes of this study were: 1) to validate the proposed structural equation model of teachers' job and teaching satisfaction 2) to analyze the direct and indirect effects of principal leadership on teachers' job and teaching satisfaction, and 3) to investigate the mediating role of teachers' trust in the relationship between principal leadership and teacher satisfaction. The sample consisted of 2,335 teachers from 155 schools, derived from the two-stage sampling design employed in the PISA framework. The instrument used was a 19-item Likert-scale questionnaire, with an overall reliability coefficient of 0.885. Data were analyzed using structural equation modeling (SEM) and bootstrapping techniques to test the significance of mediation effects. The results revealed that the proposed model demonstrated a good fit with the empirical data. Principal leadership had a significant direct effect on teachers' trust ($\beta = 0.50$). Teachers' trust exerted significant positive effects on job satisfaction ($\beta = 0.76$) and teaching satisfaction ($\beta = 0.43$). However, principal leadership had no significant direct effects on either dimension of satisfaction. Instead, it exerted indirect effects through trust on job satisfaction ($\beta = 0.38$) and teaching satisfaction ($\beta = 0.21$), both of which were statistically significant at the 0.05 level.

In conclusion, teachers' trust served as a full mediator in the relationship between principal leadership and teachers' satisfaction. The findings highlight the pivotal role of trust in educational organizations, particularly in Hong Kong's highly competitive and hierarchically structured school system.

Keywords: trust, principal leadership, job satisfaction, teaching satisfaction

บทนำ

ในยุคปัจจุบันที่ระบบการศึกษาทั่วโลกกำลังเผชิญกับความท้าทายครั้งใหญ่ ซึ่งสะท้อนผ่านผลของโครงการประเมินนักเรียนนานาชาติ (Programme for International Student Assessment: PISA 2022) ขององค์การเพื่อความร่วมมือทางเศรษฐกิจและการพัฒนา (OECD, 2023a) ที่เผยให้เห็นถึงวิกฤตการณ์เรียนรู้ระดับโลก (global learning crisis) อย่างชัดเจน หลังการแพร่ระบาดของไวรัสโควิด-19 โดยคะแนนเฉลี่ยของ

นักเรียนทั่วโลกลดลงในทุกวิชา ทั้งคณิตศาสตร์ วิทยาศาสตร์ และการอ่าน ซึ่งชี้ให้เห็นถึงปัญหาความเหลื่อมล้ำและความยั่งยืนของคุณภาพการศึกษาในหลายประเทศ แต่อย่างไรก็ตามยังมีบางประเทศสามารถรักษาคุณภาพการศึกษาได้ภายใต้สภาพการณ์ที่ท้าทายนี้ โดยฮ่องกงเป็นหนึ่งในไม่กี่ประเทศที่สามารถรักษาคุณภาพการศึกษาไว้ได้อย่างต่อเนื่อง จากผลการทดสอบ PISA 2022 นักเรียนฮ่องกงมีคะแนนเฉลี่ยอยู่ในลำดับที่ 4 ของโลกในวิชาคณิตศาสตร์ ลำดับที่ 5 ในวิชาวิทยาศาสตร์ และลำดับที่ 11 ในด้านการอ่าน แสดงถึงมาตรฐานทางวิชาการที่สูงและคุณภาพการศึกษาที่ดี อย่างไรก็ตาม ภายใต้ความสำเร็จทางวิชาการนี้ ระบบการศึกษาของฮ่องกงกลับเผชิญกับแรงกดดันภายในเป็นอย่างมาก ความสำเร็จเชิงประจักษ์ในระดับโลกจึงเป็นเหมือนเพียงสิ่งที่กำลังบดบังปัญหาความตึงเครียดของบุคลากรภายใน ซึ่งเป็นรากฐานสำคัญของระบบที่กำลังรอการแก้ไข

จากสภาวะความกดดันภายในของฮ่องกง บ่งบอกว่าโจทย์ของการบริหารการศึกษาจึงไม่ได้จำกัดอยู่เพียงการมุ่งพัฒนาผลสัมฤทธิ์ของผู้เรียนเท่านั้น จากโครงสร้างการบริหารที่เข้มงวดที่สร้างภาระงานและความกดดันให้กับบุคลากรครู ส่งผลต่อเนื่องให้ครูจำนวนมากเกิดความเครียดในการทำงาน (Hallinger & Walker, 2017) จนเกิดคณะกรรมการว่าด้วยงานของครู (The Committee on Teachers' Work) ที่จัดตั้งขึ้นโดยรัฐบาลเพื่อศึกษาและเสนอแนะแนวทางในการจัดการกับภาระงานและความเครียดของครูในโรงเรียน เพื่อลดอัตราการลาออกของครูที่เพิ่มขึ้นในช่วงไม่กี่ปีที่ผ่านมา (Committee on Teachers' Work, 2007) จากสถานการณ์นี้สะท้อนถึงความจำเป็นที่ผู้บริหารจะต้องเข้าใจกลไกการบริหาร ที่จะลดระดับความกดดันและความเครียดของครูแต่ไปยกระดับความพึงพอใจในการทำงาน (job satisfaction) และความพึงพอใจในการสอน (teaching satisfaction) ซึ่งเป็นกลไกขับเคลื่อนระบบการศึกษาที่ยั่งยืน (OECD, 2023b)

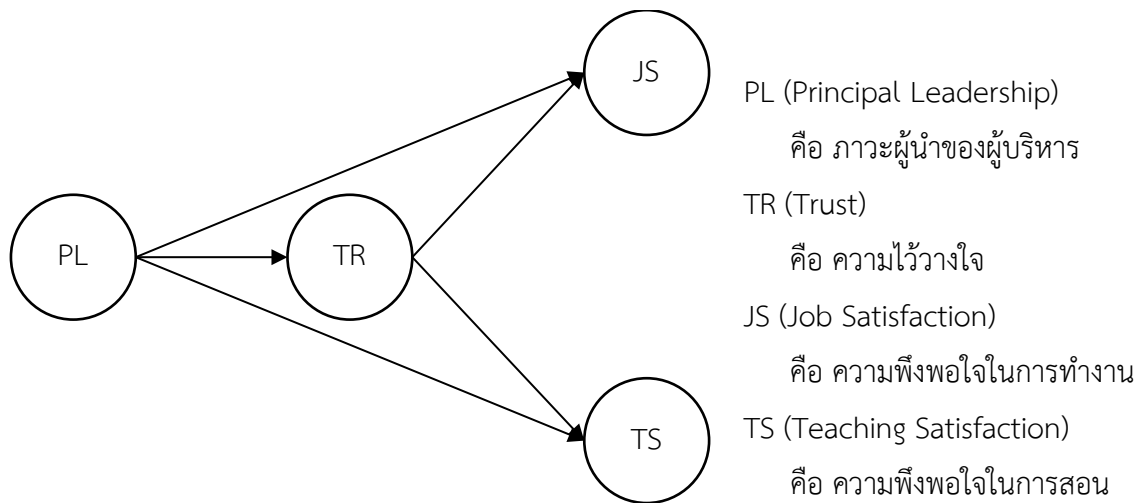
ดังนั้นจึงต้องพิจารณาว่าอะไรคือปัจจัยที่ส่งผลโดยตรงต่อระดับความพึงพอใจในการทำงานและความพึงพอใจในการสอน ซึ่งถือเป็นเป้าหมายสูงสุดในรักษามูลค่าของครูที่มีคุณภาพในระบบการศึกษา โดยความพึงพอใจในที่นี้ไม่ได้หมายถึงเพียงความรู้สึกส่วนตัวเท่านั้น แต่หมายถึงทัศนคติเชิงบวกต่อบทบาทและสภาพแวดล้อมในการทำงานที่สะท้อนถึงแรงจูงใจภายในและความผูกพันต่อวิชาชีพ ความพึงพอใจนี้เป็นตัวชี้วัดความยั่งยืนของระบบการศึกษา (OECD, 2023b) และเชื่อมโยงกับคุณภาพของครู (teacher quality) เนื่องจากครูที่มีความพึงพอใจในการทำงานและการสอนสูงจะมีแนวโน้มที่จะมีความผูกพันในงานสูงขึ้นและแสดงประสิทธิภาพในการสอนที่ดีกว่า โดยเฉพาะอย่างยิ่งในบริบทของฮ่องกงที่ระบบการศึกษามีโครงสร้างลำดับขั้นที่ชัดเจน (Xu & Shen, 2007) ความสัมพันธ์ระหว่างผู้บริหารและครู จึงเป็นแกนกลางสำคัญที่กำหนดทิศทางของแรงจูงใจและความพึงพอใจในการทำสิ่งต่าง ๆ (Hallinger & Walker, 2017) ด้วยเหตุนี้ ปัจจัยเชิงเหตุที่ถูกรับรู้กันอย่างเข้มข้นเพื่อขับเคลื่อนคุณภาพครู จึงหนีไม่พ้นบทบาทของภาวะผู้นำของผู้บริหารโรงเรียน ซึ่งถูกพิจารณาในฐานะตัวแปรต้นทางที่มีอิทธิพลต่อความพึงพอใจของครูโดยตรง

ภาวะผู้นำของผู้บริหาร (principal leadership) จึงเป็นตัวแปรสำคัญที่ถูกพิจารณาในฐานะตัวขับเคลื่อนความพึงพอใจและคุณภาพของบุคลากร โดยเฉพาะอย่างยิ่ง ภาวะผู้นำแบบกระจายอำนาจ (distributed leadership) เป็นแนวคิดที่มองว่าการนำองค์กรไม่ควรถูกจำกัดอยู่ที่ผู้บริหารเพียงคนเดียว แต่ครูและบุคลากรมีส่วนร่วมในการตัดสินใจ แบ่งปันความรับผิดชอบ และร่วมกันพัฒนาองค์กร (Spillane, 2006) แนวคิดนี้ยังได้รับการสนับสนุนอย่างยิ่งจากองค์การเพื่อความร่วมมือทางเศรษฐกิจและการพัฒนา

(OECD) ในฐานะยุทธศาสตร์หลักในการสร้างการบริหารโรงเรียนที่ยั่งยืนและมีประสิทธิภาพสูงสุด ซึ่งแนวคิดนี้สะท้อนคำถามว่า ผู้บริหารจะอย่างไรให้ครูพอใจในทุกมิติ โดยการเปลี่ยนบทบาทเป็นผู้เอื้ออำนวยความสะดวกมากกว่าผู้ควบคุม ซึ่งเน้นความยืดหยุ่น ความร่วมมือ และการสร้างความเป็นเจ้าของในงานของครู (Harris, 2014) สอดคล้องกับงานวิจัยของ Baptiste (2019) ที่ชี้ว่า ผู้บริหารที่สนับสนุนครูและให้ความยุติธรรม จะเสริมสร้างแรงจูงใจและความพึงพอใจได้อย่างมีนัยสำคัญ นอกจากนี้ ผู้บริหารที่สามารถสร้างแรงบันดาลใจ ถ่ายทอดวิสัยทัศน์ และให้ความสำคัญกับการพัฒนาบุคลากร ย่อมสร้างความไว้วางใจและแรงจูงใจภายในแก่ครูได้โดยตรง (Bass & Riggio, 2006) อย่างไรก็ตาม เมื่อมองในบริบทของประเทศไทยซึ่งอ่องกงที่ความสัมพันธ์ระหว่างภาวะผู้นำของผู้บริหารกับความพึงพอใจของครูอาจไม่ได้เกิดขึ้นโดยตรงเสมอไป ซึ่งนำไปสู่การค้นพบกลไกสำคัญที่ขาดไม่ได้ในการทำให้ภาวะผู้นำสัมฤทธิ์ผล นั่นคือความไว้วางใจ

แนวคิดเชิงกลไกที่ว่าด้วยความไว้วางใจ (trust) ได้รับการอธิบายไว้ใน ทฤษฎีทุนทางสังคม (social capital theory) ของ Bryk & Schneider (2002) ซึ่งมองว่าความไว้วางใจในโรงเรียนเป็นทุนเชิงสังคมที่สะสมจากการแลกเปลี่ยนความร่วมมือ ความจริงใจ และความเคารพระหว่างสมาชิกในองค์กร งานวิจัยเชิงประจักษ์แสดงให้เห็นว่าความไว้วางใจเป็นตัวแปรส่งผ่านที่มีบทบาทสำคัญที่สามารถเชื่อมความสัมพันธ์ระหว่างความพึงพอใจของครูและภาวะผู้นำของผู้บริหารภายใต้ระบบโครงสร้างลำดับชั้นได้ (Balan & Jamaluddin, 2024; Balyer, 2017) ภายใต้บริบทการบริหารของอ่องกงที่เน้นการบริหารจัดการโรงเรียนโดยใช้โรงเรียนเป็นฐาน (School Based Management: SBM) ความไว้วางใจยิ่งทวีความสำคัญสูงสุด เนื่องจาก SBM เรียกร้องให้ผู้บริหารถูกคาดหวังให้เปิดโอกาสให้ครูมีส่วนร่วมในการตัดสินใจและมีความรู้สึกเป็นเจ้าขององค์กรทางการศึกษา (Wong & Liu, 2018) การมอบอำนาจและกระจายความรับผิดชอบนี้ ไม่อาจเกิดขึ้นได้อย่างมีประสิทธิภาพหากขาดพื้นฐานของความไว้วางใจ ความไว้วางใจดังกล่าวจึงทำหน้าที่เป็นกลไกทางสังคมที่เปลี่ยนพฤติกรรมเชิงบวกของผู้บริหารให้กลายเป็น ความปลอดภัยทางจิตวิทยา (psychological safety) และความเป็นอิสระทางวิชาชีพของครู ซึ่งเป็นปัจจัยหลักที่ส่งผลต่อความผูกพันและความพึงพอใจในการทำงาน (Tschannen-Moran, 2014) ดังนั้น ความไว้วางใจระหว่างผู้บริหารและครูจึงเป็นเงื่อนไขสำคัญในการสร้างความพึงพอใจให้กับครู และเป็นกลไกที่ตอบเจตยในการบริหารว่าผู้บริหารควรดำเนินการอย่างไรเพื่อส่งเสริมความพึงพอใจของบุคลากรครูให้ได้ในทุกมิติ

งานวิจัยนี้จึงมุ่งศึกษากลไกการส่งผ่านของความไว้วางใจของครู ที่เชื่อมโยงระหว่างภาวะผู้นำของผู้บริหารกับความพึงพอใจของครู ทั้งในมิติความพึงพอใจในการทำงานและความพึงพอใจในการสอน โดยอาศัยกรอบแนวคิดของการวิเคราะห์ตัวแปรส่งผ่าน (mediation analysis) เพื่ออธิบายว่าผู้บริหารที่มีประสิทธิภาพสามารถสร้างความสัมพันธ์ภายในองค์กรที่โปร่งใสและมีส่วนร่วม ส่งผลให้เกิดความไว้วางใจ ซึ่งเป็นกลไกสำคัญที่ส่งเสริมแรงจูงใจและความมั่นคงทางอารมณ์ของครู จนนำไปสู่ความพึงพอใจในการทำงานและการสอนในระดับสูงขึ้น การศึกษานี้จึงมีความสำคัญในการยืนยันบทบาทของความไว้วางใจในฐานะตัวแปรส่งผ่านที่เชื่อมโยงภาวะผู้นำกับความพึงพอใจของครู โดยใช้ข้อมูลเชิงประจักษ์จาก PISA 2022 ของอ่องกง เพื่อสร้างองค์ความรู้เชิงนโยบายในการพัฒนาภาวะผู้นำทางการศึกษาในศตวรรษที่ 21 อย่างมีประสิทธิภาพ



ภาพ 1 กรอบแนวคิดการวิจัย

ความมุ่งหมายของการวิจัย

1. เพื่อตรวจสอบความสอดคล้องของโมเดลสมการเชิงโครงสร้างความพึงพอใจในการทำงานและการสอนของครู
2. เพื่อวิเคราะห์อิทธิพลทางตรงและทางอ้อมของภาวะผู้นำของผู้บริหารที่มีต่อความพึงพอใจในการทำงานและการสอนของครู
3. เพื่อตรวจสอบอิทธิพลการส่งผ่านของความไว้วางใจที่มีต่อภาวะผู้นำของผู้บริหารและความพึงพอใจในการทำงานและการสอนของครู

สมมติฐานของการวิจัย

- H1 ภาวะผู้นำของผู้บริหาร (PL) มีอิทธิพลทางตรงในเชิงบวกอย่างมีนัยสำคัญต่อระดับความไว้วางใจ (TR)
- H2 ความไว้วางใจ (TR) มีอิทธิพลทางตรงในเชิงบวกต่อความพึงพอใจในการทำงาน (JS)
- H3 ความไว้วางใจ (TR) มีอิทธิพลทางตรงในเชิงบวกต่อความพึงพอใจในการสอน (TS)
- H4 ภาวะผู้นำของผู้บริหาร (PL) มีอิทธิพลทางตรงในเชิงบวกต่อความพึงพอใจในการทำงาน (JS)
- H5 ภาวะผู้นำของผู้บริหาร (PL) มีอิทธิพลทางตรงในเชิงบวกต่อความพึงพอใจในการสอน (TS)
- H6 ความไว้วางใจ (TR) ทำหน้าที่เป็นตัวแปรส่งผ่าน (Mediator) ความสัมพันธ์ระหว่างภาวะผู้นำของผู้บริหาร (PL) กับความพึงพอใจทั้งสองมิติในการทำงาน (JS) และการสอน (TS)

วิธีดำเนินการวิจัย

1. ประชากรและตัวอย่าง

การวิจัยครั้งนี้ เป็นการศึกษาข้อมูลทัศนคติของครูในประเทศฮ่องกง จากฐานข้อมูล PISA 2022 ประชากรของการวิจัย คือ ครูผู้สอนในประเทศฮ่องกง ตัวอย่างวิจัย คือ ครูผู้สอนชาวฮ่องกงจำนวน 2,335 คน จาก 155 โรงเรียน ได้มาโดยวิธีการสุ่มตัวอย่างแบบสองขั้นตอน กล่าวคือ สุ่มโรงเรียน และสุ่มครูในโรงเรียน

2. เครื่องมือที่ใช้ในการวิจัย

ตัวแปรภายใต้กรอบแนวคิดการวิจัยทั้ง 4 ตัวแปร ได้แก่ ภาวะผู้นำของผู้บริหาร (Principal Leadership) ความไว้วางใจ (Trust) ความพึงพอใจในการทำงาน (Job Satisfaction) และความพึงพอใจในการสอน (Teaching Satisfaction) รวบรวมจากข้อคำถามในแบบสอบถามสำหรับครู จำนวนทั้งหมด 19 ข้อ พัฒนาขึ้นโดย OECD นำมาตรวจสอบความเที่ยง (reliability) ด้วยค่าสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's alpha coefficient) และตรวจสอบความตรงเชิงโครงสร้าง (construct validity) ด้วยการวิเคราะห์องค์ประกอบเชิงยืนยัน (confirmatory factor analysis) ผลการตรวจสอบ มีดังนี้

2.1 แบบวัดภาวะผู้นำของผู้บริหาร (Principal Leadership: PL) ประกอบด้วยข้อคำถามจำนวน 7 ข้อ (TC253Q1 – TC253Q7) มีลักษณะเป็นมาตราประมาณค่า 4 ระดับ (ไม่เคยเลย - เป็นประจำ) ตัวอย่างข้อคำถาม ได้แก่ ผู้อำนวยการโรงเรียนดำเนินการเพื่อสนับสนุนความร่วมมือระหว่างครูในการพัฒนาวิธีการสอนใหม่ ๆ (TC253Q4) ผู้อำนวยการสังเกตการณ์การเรียนการสอนในห้องเรียน (TC253Q2) ผู้อำนวยการดำเนินการเพื่อให้แน่ใจว่าครูมีความรับผิดชอบในการพัฒนาทักษะการสอนของตนเอง (TC253Q5) มีค่าความเที่ยงเท่ากับ 0.90 มีความตรงเชิงโครงสร้างพิจารณาได้จาก $\chi^2(10, N = 2,335) = 16.42, p > .05$, CFI = 1.00, TLI = 1.00, RMSEA = 0.00, SRMR = 0.01

2.2 แบบวัดความไว้วางใจ (Trust: TR) ประกอบด้วยข้อคำถามจำนวน 5 ข้อ (TC241Q1 – TC241Q5) มีลักษณะเป็นมาตราประมาณค่า 4 ระดับ (ไม่เห็นด้วยอย่างยิ่ง - เห็นด้วยอย่างยิ่ง) ตัวอย่างข้อคำถาม ได้แก่ ครูสามารถพึ่งพาฝ่ายบริหารของโรงเรียนในการสนับสนุนทางวิชาชีพได้ (TC241Q1) ผู้อำนวยการมีความมั่นใจในความเชี่ยวชาญของครู (TC241Q2) แม้ในสถานการณ์ที่ยากลำบาก เพื่อนร่วมงานของฉันก็รู้ว่าพวกเขาสามารถไว้วางใจฉันได้ (TC241Q3) มีค่าความเที่ยงเท่ากับ 0.81 มีความตรงเชิงโครงสร้างพิจารณาได้จาก $\chi^2(3, N = 2,335) = 7.31, p > .05$, CFI = 1.00, TLI = 0.99, RMSEA = 0.03, SRMR = 0.01

2.3 แบบวัดความพึงพอใจในการทำงาน (Job Satisfaction: JS) ประกอบด้วยข้อคำถามจำนวน 4 ข้อ มีลักษณะเป็นมาตราประมาณค่า 4 ระดับ (ไม่เห็นด้วยอย่างยิ่ง - เห็นด้วยอย่างยิ่ง) ตัวอย่างข้อคำถาม ได้แก่ ฉันมีความสุขกับการทำงานที่โรงเรียนนี้ (TC198Q5) ฉันอยากจะแนะนำโรงเรียนของฉันว่าเป็นสถานที่ทำงานที่ดี (TC198Q7) ฉันพอใจกับผลการปฏิบัติงานของฉันในโรงเรียนแห่งนี้ (TC198Q9) โดยรวมแล้วฉันพอใจกับงานของฉัน (TC198Q10) มีค่าความเที่ยงเท่ากับ 0.80 มีความตรงเชิงโครงสร้างพิจารณาได้จาก $\chi^2(1, N = 2,335) = 1.49, p > .05$, CFI = 1.00, TLI = 1.00, RMSEA = 0.02, SRMR = 0.00

2.4 แบบวัดความพึงพอใจในการสอน (Teaching Satisfaction: TS) ประกอบด้วยข้อคำถามจำนวน 3 ข้อ มีลักษณะเป็นมาตราประมาณค่า 4 ระดับ (ไม่เห็นด้วยอย่างยิ่ง - เห็นด้วยอย่างยิ่ง) ตัวอย่างข้อคำถาม ได้แก่ ข้อดีของการเป็นครูมีมากกว่าข้อเสียอย่างชัดเจน (TC198Q1) หากฉันสามารถตัดสินใจอีกครั้งฉันก็ยังคงเลือกที่จะเป็นครู (TC198Q2) ฉันคิดว่าอาชีพครูมีคุณค่าในสังคม (TC198Q8) มีค่าความเที่ยงเท่ากับ 0.68 มีความตรงเชิงโครงสร้างพิจารณาได้จาก $\chi^2(1, N = 2,335) = 0.00, p > .05$, CFI = 1.00, TLI = 1.00, RMSEA = 0.00, SRMR = 0.00

3. การเก็บรวบรวมข้อมูล

งานวิจัยนี้เป็นการวิจัยเชิงปริมาณโดยใช้ข้อมูลทุติยภูมิ (secondary data) จากฐานข้อมูลในโครงการประเมินสมรรถนะนักเรียนมาตรฐานสากล PISA 2022 ที่ดำเนินการโดย OECD ใช้ข้อมูลจากกลุ่มตัวอย่างครูในประเทศฮ่องกง ซึ่งเป็นส่วนหนึ่งของการสำรวจ PISA 2022 ในส่วนของแบบสอบถามครู (teacher questionnaire data file)

4. การวิเคราะห์ข้อมูล

ชุดข้อมูลของประเทศฮ่องกงจากฐานข้อมูลข้างต้น พบว่ามีตัวอย่างวิจัยที่มีข้อมูลในตัวแปรหลักครบถ้วนจำนวน 2,335 คน วิเคราะห์ข้อมูลพื้นฐานของตัวอย่างการวิจัย (sample) ด้วยสถิติเชิงพรรณนา ได้แก่ ค่าเฉลี่ย (Mean) ส่วนเบี่ยงเบนมาตรฐาน (SD) และค่าการกระจายสัมพัทธ์ (%CV) การวิเคราะห์เพื่อตอบวัตถุประสงค์ของการวิจัย ใช้การวิเคราะห์โมเดลสมการโครงสร้าง (SEM) และทดสอบอหิทธิพลการส่งผ่านด้วยวิธี Bootstrap 5,000 รอบ ด้วยโปรแกรม Mplus Version 7.3 (Muthén & Muthén, 2012) ทั้งนี้ ผู้วิจัยเลือกใช้เทคนิค Bootstrap ในการทดสอบอหิทธิพลทางอ้อม เนื่องจากการแจกแจงของผลคูณสัมประสิทธิ์ (product of coefficients) มักมีลักษณะเบ้และไม่เป็นปกติ โดยการใช้การประมาณค่าช่วงความเชื่อมั่นแบบปรับแก้ หรือ Bias-corrected Bootstrap ที่มีความแม่นยำและน่าเชื่อถือสูงกว่า การทดสอบแบบดั้งเดิมอย่างการทดสอบ Sobel (MacKinnon et al., 2004; Preacher & Hayes, 2008)

สำหรับการวิจัยครั้งนี้ ใช้เกณฑ์การประเมินความตรงสำหรับกลุ่มตัวอย่างขนาดใหญ่ Hair et al. (2010) เสนอว่า กรณีโมเดลวิจัยประกอบด้วย ตัวแปรสังเกตได้ระหว่าง 12 ถึง 20 ตัว ขนาดตัวอย่างมากกว่า 200 คน การทดสอบด้วยสถิติไคสแควร์มีแนวโน้มที่จะปฏิเสธสมมติฐานหลัก ($p < .05$) ให้พิจารณาค่าดัชนีวัดความสอดคล้องอื่น ๆ ร่วมด้วย ได้แก่ CFI และ TLI มีค่าตั้งแต่ 0.92, SRMR น้อยกว่า 0.08 และ RMSEA น้อยกว่า 0.07

ผลการวิจัย

ข้อมูลทั่วไปของตัวอย่าง

ตัวอย่างครูชาวฮ่องกงจากฐานข้อมูล PISA 2022 จำนวน 2,335 คน ประกอบด้วยครูเพศหญิง 1,169 คน (ร้อยละ 50.07) ครูเพศชาย 1,137 คน (ร้อยละ 48.69) และไม่ระบุ 29 คน (ร้อยละ 1.24) ส่วนใหญ่มีระดับการศึกษาปริญญาโท หรือเทียบเท่า 1,221 คน (ร้อยละ 52.30) รองลงมา คือ ปริญญาตรี หรือเทียบเท่า 872 คน (ร้อยละ 37.34) และไม่ระบุ 212 คน (ร้อยละ 9.08) มีเพียงส่วนน้อยจบระดับปริญญาเอกหรือเทียบเท่า 30 คน (ร้อยละ 1.28) รายละเอียดดังตาราง 1

ตาราง 1 ข้อมูลพื้นฐานตัวอย่างวิจัย

เพศ	ความถี่	ร้อยละ	ระดับการศึกษา	ความถี่	ร้อยละ
ชาย	1,137	48.69	ปริญญาตรี หรือเทียบเท่า	872	37.34
หญิง	1,169	50.07	ปริญญาโท หรือเทียบเท่า	1,221	52.30
ไม่ระบุ	29	1.24	ปริญญาเอก หรือเทียบเท่า	30	1.28
รวม	2,335	100.00	ไม่ระบุ	212	9.08
			รวม	2,335	100.00

เมื่อพิจารณาค่าสถิติพื้นฐานของตัวแปรในการวิจัยทั้ง 4 ตัวแปร ได้แก่ ภาวะผู้นำของผู้บริหาร (PL) ความไว้วางใจ (TR) ความพึงพอใจในการทำงาน (JS) และ ความพึงพอใจในการสอน (TS) พบว่า มีการกระจายคิดเป็นร้อยละ 28.37 – 15.57 โดยภาวะผู้นำของผู้บริหารมีการกระจายสูงสุด (ร้อยละ 28.37) รองลงมา คือ ความพึงพอใจในการการสอน (ร้อยละ 17.39) ความพึงพอใจในการทำงาน (ร้อยละ 16.44) และความไว้วางใจมีการกระจายต่ำสุด (ร้อยละ 15.57) นอกจากนี้ เมื่อวิเคราะห์ค่าสหสัมพันธ์ภายในของตัวแปรพบว่าตัวแปรทุกตัวแปรมีความสัมพันธ์กันอย่างมีนัยสำคัญทางสถิติที่ .05 มีค่าอยู่ระหว่าง .15 – .54 คู่ความสัมพันธ์ที่มากที่สุด คือ ความไว้วางใจกับความพึงพอใจในการทำงาน และ ความพึงพอใจในการทำงานกับความพึงพอใจในการสอน ($r=.54$) ส่วนคู่ความสัมพันธ์ที่น้อยที่สุด คือ ภาวะผู้นำของผู้บริหารกับความพึงพอใจในการสอน รายละเอียดดังตาราง 2

ตาราง 2 ค่าสถิติพื้นฐาน และค่าสหสัมพันธ์ภายในของตัวแปรที่ใช้ในการวิจัย

ตัวแปร	Mean	SD	%CV	PL	TR	JS	TS
PL	2.15	0.61	28.37	–			
TR	2.89	0.45	15.57	.37*	–		
JS	2.98	0.49	16.44	.31*	.54*	–	
TS	2.99	0.52	17.39	.15*	.32*	.54*	–

Note: * $p < .05$

การวิเคราะห์เพื่อตอบวัตถุประสงค์การวิจัย

การตอบวัตถุประสงค์การวิจัยข้อที่ 1

ผลการวิเคราะห์โมเดลสมการโครงสร้าง พบว่า โมเดลความพึงพอใจในการทำงานและการสอน มีความสอดคล้องกับข้อมูลเชิงประจักษ์ พิจารณาได้จาก $\chi^2(143, N = 2,335) = 1471.98, p < .001, CFI = 0.94, TLI = 0.93, RMSEA = 0.07$ และ $SRMR = 0.05$ ค่าดัชนีวัดความสอดคล้องที่ได้เป็นไปตามเกณฑ์ของ Hair et al. (2010) กรณีตัวอย่างวิจัยมีขนาดใหญ่ (ภาพ 2)

การตอบวัตถุประสงค์การวิจัยข้อที่ 2

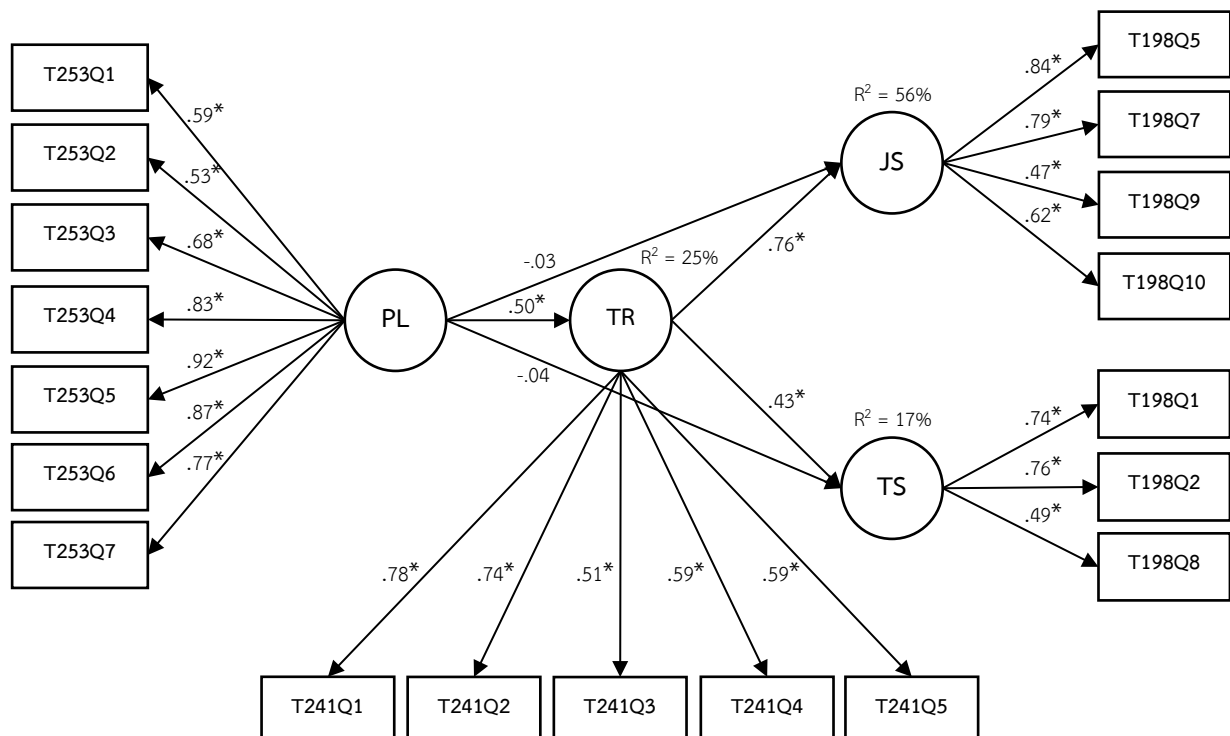
เมื่อพิจารณาขนาดอิทธิพลรวม (total effect) ของตัวแปรภายในโมเดลวิจัย พบว่า ภาวะผู้นำของผู้บริหาร (PL) ส่งผลต่อความพึงพอใจในการทำงาน (JS) ด้วยขนาดอิทธิพลรวมเท่ากับ .35 อย่างมีนัยสำคัญทางสถิติที่ .05 เป็นอิทธิพลทางอ้อม .38 และเป็นอิทธิพลทางตรง -.03 และ ภาวะผู้นำของผู้บริหาร (PL) ส่งผลต่อความพึงพอใจในการสอน (TS) ด้วยขนาดอิทธิพลรวมเท่ากับ .17 อย่างมีนัยสำคัญทางสถิติที่ .05 เป็นอิทธิพลทางอ้อม .21 และเป็นอิทธิพลทางตรง -.04 สำหรับอิทธิพลทางตรง (direct effect) ภาวะผู้นำของผู้บริหารส่งผลต่อความไว้วางใจมากที่สุด ($\beta = .50$) รองลงมาคือ ความพึงพอใจในการทำงาน ($\beta = -.03$) และความพึงพอใจในการสอน ($\beta = -.04$) ตัวแปรในโมเดลร่วมกันอธิบายความแปรปรวนของความพึงพอใจในการทำงาน (JS) ร้อยละ 56 และความพึงพอใจในการสอน (TS) ร้อยละ 17 (ตาราง 3)

ตาราง 3 อิทธิพลทางตรง อิทธิพลทางอ้อม และอิทธิพลรวม

ผล	TR			JS			TS		
เหตุ	DE	IE	TE	DE	IE	TE	DE	IE	TE
PL	.50*	–	.50*	-.03	.38*	.35*	-.04	.21*	.17*
	(0.04)		(0.04)	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.03)
TR				.76*	-	.76*	.43*	-	.43*
				(0.04)		(0.04)	(0.04)		(0.04)
R ²	T198Q1	T198Q2	T198Q5	T198Q7	T198Q8	T198Q9	T198Q1	T241Q1	T241Q2
	.55	.58	.71	.63	.24	.22	0 .38	.61	.54
	T241Q3	T241Q4	T241Q5	T253Q1	T253Q2	253Q3	T253Q4	T253Q5	T253Q6
	.26	.35	.34	.35	.28	.46	.69	.84	.76
	T253Q7								
	.60								
สมการโครงสร้าง	R ²			TR	JS	TS			
				.25	.56	.17			

Note: ตัวเลขในวงเล็บ คือ ค่าความคลาดเคลื่อนมาตรฐาน (standard error); DE =อิทธิพลทางตรง,

IE = อิทธิพลทางอ้อม, TE= อิทธิพลรวม, * p< .05



$$\chi^2 (143, N = 2,335) = 1471.98, P < .05, CFI = 0.94, TLI = 0.93, RMSEA = 0.07, SRMR = 0.05$$

ภาพ 2 ผลการวิเคราะห์โมเดลสมการโครงสร้างความพึงพอใจในการทำงานและการสอนของครู (นำเสนอด้วยค่ามาตรฐาน)

การตอบวัตถุประสงค์การวิจัยข้อที่ 3

จากการทดสอบอิทธิพลการส่งผ่านของความไว้วางใจ (TR) โดยทำการเปรียบเทียบ 2 เส้นทาง สมมติฐาน ด้วยวิธีการ Bootstrap 5000 รอบ ผลการทดสอบ พบว่า ภายใต้วงความเชื่อมั่น 95% มี 2 เส้นทางการส่งผ่านตามสมมติฐานพบนัยสำคัญทางสถิติ คือ $PL \rightarrow TR \rightarrow JS$ ($\beta = 0.38$) และ $PL \rightarrow TR \rightarrow TS$ ($\beta = 0.21$) (ตาราง 4)

ตาราง 4 การทดสอบอิทธิพลส่งผ่านในโมเดลวิจัย ด้วยวิธี Bootstrap 5000 รอบ

เส้นทาง อิทธิพลทางอ้อม	b (β)	SE	95% CI	
			Lower	Upper
$PL \rightarrow TR \rightarrow JS$	0.47 (.38)	0.02	0.34	0.42
$PL \rightarrow TR \rightarrow TS$	0.21 (.21)	0.02	0.17	0.26

ผลการทดสอบสมมติฐานของการวิจัย

H1 ภาวะผู้นำของผู้บริหาร (PL) มีอิทธิพลทางตรงในเชิงบวกต่อความไว้วางใจ (TR) อย่างมีนัยสำคัญทางสถิติ ($p < 0.05$) โดยมีค่าสัมประสิทธิ์เท่ากับ 0.50

H2 ความไว้วางใจ (TR) มีอิทธิพลทางตรงในเชิงบวกต่อความพึงพอใจในการทำงาน (JS) อย่างมีนัยสำคัญทางสถิติ ($p < 0.05$) โดยมีค่าสัมประสิทธิ์เท่ากับ 0.76

H3 ความไว้วางใจ (TR) มีอิทธิพลทางตรงในเชิงบวกต่อความพึงพอใจในการสอน (TS) อย่างมีนัยสำคัญทางสถิติ ($p < 0.05$) โดยมีค่าสัมประสิทธิ์เท่ากับ 0.43

H4 ภาวะผู้นำของผู้บริหาร (PL) มีอิทธิพลทางตรงต่อความพึงพอใจในการทำงาน (JS) อย่างไม่มีนัยสำคัญทางสถิติ โดยมีค่าสัมประสิทธิ์เท่ากับ - 0.03

H5 ภาวะผู้นำของผู้บริหาร (PL) มีอิทธิพลทางตรงต่อความพึงพอใจในการสอน (TS) อย่างไม่มีนัยสำคัญทางสถิติ โดยมีค่าสัมประสิทธิ์เท่ากับ - 0.04

H6 ความไว้วางใจ (TR) ทำหน้าที่เป็นตัวแปรส่งผ่าน (Mediator) ความสัมพันธ์ระหว่างภาวะผู้นำของผู้บริหาร (PL) กับความพึงพอใจทั้งสองมิติในการทำงาน (JS) และการสอน (TS) รวมกันเป็นความพึงพอใจในการทำงานของครู (Teacher Satisfaction)

1. ความไว้วางใจ (TR) ทำหน้าที่เป็นตัวแปรส่งผ่าน (Mediator) ความสัมพันธ์ระหว่างภาวะผู้นำของผู้บริหาร (PL) กับความพึงพอใจในการทำงาน (JS) (ผลกระทบทางอ้อม = .38, $p < .05$)

2. ความไว้วางใจ (TR) ทำหน้าที่เป็นตัวแปรส่งผ่าน (Mediator) ความสัมพันธ์ระหว่างภาวะผู้นำของผู้บริหาร (PL) กับความพึงพอใจในการสอน (TS) (ผลกระทบทางอ้อม = .21, $p < .05$)

จากการที่อิทธิพลทางตรงจากภาวะผู้นำของผู้บริหาร (PL) ไปยังความพึงพอใจในการทำงาน (JS) และความพึงพอใจในการสอน (TS) ไม่มีนัยสำคัญ แต่อิทธิพลทางอ้อมผ่านความไว้วางใจ (TR) ไปยังความพึงพอใจในการทำงาน (JS) และความพึงพอใจในการสอน (TS) มีนัยสำคัญ แสดงว่าความไว้วางใจของครูต่อผู้บริหาร (TR) ทำหน้าที่เป็นตัวแปรส่งผ่านแบบสมบูรณ์ (Full Mediation) ในความสัมพันธ์ระหว่างภาวะผู้นำของผู้บริหาร (PL) กับความพึงพอใจในการทำงาน (JS) และความพึงพอใจในการสอน (TS) สรุปได้ว่าภาวะผู้นำของผู้บริหารส่งผลต่อความพึงพอใจในการทำงานและความพึงพอใจในการสอนผ่านทางความไว้วางใจต่อผู้บริหารเท่านั้น

อภิปรายผล

การวิเคราะห์โมเดลสมการเชิงโครงสร้าง (SEM) สะท้อนกลไกความสัมพันธ์ระหว่างตัวแปรหลักในบริบทของระบบการศึกษาที่มีมาตรฐานสูงและมีแรงคาดหวังทางวิชาชีพเข้มข้นเช่นฮ่องกง

สมมติฐาน H1 ได้รับการสนับสนุน โดยผลการวิจัยพบว่า ภาวะผู้นำของผู้บริหารโรงเรียนมีอิทธิพลทางตรงในเชิงบวกอย่างมีนัยสำคัญต่อความไว้วางใจ ($\beta = 0.50$, $p < .05$) ซึ่งสอดคล้องกับแนวคิดที่ว่า พฤติกรรมของผู้นำมีบทบาทสำคัญในการก่อรูปและธำรงความไว้วางใจในองค์กร (Balan & Jamaluddin, 2024; Baptiste, 2019) ในบริบทของฮ่องกง ซึ่งดำเนินการจัดการศึกษาตามแนวทาง School-Based Management (SBM)

(Education Commission, 2019) การเป็นผู้นำที่มีประสิทธิผลไม่ได้จำกัดอยู่เพียงการบริหารจัดการเชิงโครงสร้าง แต่ยังต้องแสดงบทบาทเชิงสัมพันธ์ที่โปร่งใส มีความรับผิดชอบ และส่งเสริมความเป็นมืออาชีพของครู ความเชื่อมั่นในศักดิ์ศรีและศักยภาพของผู้บริหารจึงเป็นปัจจัยสำคัญที่ก่อให้เกิดความไว้วางใจ ซึ่งสอดคล้องกับแกนหลักของทฤษฎีทุนทางสังคม (social capital theory) ที่มองว่าความไว้วางใจเป็นทุนทางสังคมสำคัญในการพัฒนาโรงเรียนอย่างยั่งยืน (Bryk & Schneider, 2002)

สมมติฐาน H2 และ H3 ได้รับการสนับสนุน ผลการวิจัยแสดงให้เห็นว่า ความไว้วางใจมีอิทธิพลทางตรงในเชิงบวกต่อทั้งความพึงพอใจในการทำงาน ($\beta = 0.76$) และความพึงพอใจในการสอน ($\beta = 0.43$) ความสัมพันธ์ดังกล่าวสะท้อนให้เห็นว่า ความไว้วางใจระหว่างครูและผู้บริหารทำหน้าที่เป็นกลไกป้องกันเชิงบวก (buffer mechanism) ต่อแรงกดดันจากภาระงานและความคาดหวังที่สูงในระบบการศึกษา (LegCo, 2007) ความไว้วางใจจึงเปรียบเสมือนสะพานเชื่อมจากความตึงเครียดไปสู่ความรู้สึกปลอดภัยทางจิตวิทยา (psychological safety) และความเป็นอิสระทางวิชาชีพของครู (Tschannen-Moran, 2014) ซึ่งเป็นพื้นฐานสำคัญของความผูกพัน และความพึงพอใจในวิชาชีพ (Tobias, 2017)

ในทางกลับกัน สมมติฐาน H4 และ H5 ไม่ได้รับการสนับสนุน ผลการวิเคราะห์พบว่า ภาวะผู้นำของผู้บริหาร ไม่มีผลโดยตรงอย่างมีนัยสำคัญต่อทั้งความพึงพอใจในการทำงานและความพึงพอใจในการสอน โดยค่าสัมประสิทธิ์มีทิศทางลบเล็กน้อย ($\beta = -0.03$ และ -0.04 ตามลำดับ) ซึ่งแตกต่างจากงานวิจัยก่อนหน้านี้ที่มักพบความสัมพันธ์เชิงบวกระหว่างภาวะผู้นำกับความพึงพอใจของครู (Baptiste, 2019; Tobias, 2017) ความแตกต่างนี้อาจสะท้อนความซับซ้อนของระบบการบริหารที่มีความคาดหวังสูง เช่นในบริบทของ SBM ของฮ่องกง ที่ผู้บริหารมีบทบาทเชิงรุกในการขับเคลื่อนนโยบายและพัฒนาโครงการต่าง ๆ ซึ่งอาจสร้างแรงกดดันทางอ้อมให้ครูรู้สึกมีภาระงานเพิ่มขึ้น โดยเฉพาะเมื่อทรัพยากรมีจำกัด (LegCo, 2007) ดังนั้น ผลลัพธ์นี้ไม่ควรถูกตีความว่าเป็นภาวะผู้นำเชิงลบ แต่ควรมองว่าเป็น “ความไม่สมดุลเชิงบริหาร” ระหว่างเป้าหมายขององค์กรกับขีดจำกัดของทรัพยากรบุคลากร

สมมติฐาน H6 ได้รับการสนับสนุนในลักษณะของ การส่งผ่านแบบสมบูรณ์ (Full Mediation) กล่าวคือ ภาวะผู้นำของผู้บริหารส่งผลทางตรงต่อความพึงพอใจอย่างไม่มีนัยสำคัญทางสถิติ (JS: $\beta = -0.03$; TS: $\beta = -0.04$) อย่างไรก็ตาม พบว่า ภาวะผู้นำของผู้บริหารมีอิทธิพลทางอ้อมผ่านความไว้วางใจอย่างมีนัยสำคัญทางสถิติ (JS: $\beta = 0.38$; TS: $\beta = 0.21$) ผลดังกล่าวสะท้อนว่า “ความไว้วางใจ” ทำหน้าที่เป็นกลไกสำคัญในการถ่ายทอดอิทธิพลของภาวะผู้นำไปสู่ความพึงพอใจของครู ซึ่งสอดคล้องกับแนวคิดด้านการวิเคราะห์อิทธิพลส่งผ่านที่ชี้ให้เห็นบทบาทของตัวแปรส่งผ่านในการอธิบายความสัมพันธ์เชิงสาเหตุทางอ้อม (Hayes, 2009) กล่าวคือ กล่าวอีกนัยหนึ่ง ภาวะผู้นำเพียงลำพังอาจไม่เพียงพอในการสร้างความพึงพอใจ หากไม่ก่อให้เกิดความไว้วางใจในระดับบุคคล

ในเชิงนโยบาย การสร้างความไว้วางใจในองค์กรการศึกษาจึงควรถูกจัดลำดับเป็นเป้าหมายเชิงยุทธศาสตร์ของการพัฒนาโรงเรียน ภาวะผู้นำที่มีประสิทธิภาพต้องอาศัยการสื่อสารแบบเปิดเผย การยอมรับความคิดเห็น และการกระจายอำนาจอย่างมีส่วนร่วมเพื่อเสริมสร้างทุนทางสังคมในระดับโรงเรียน ทั้งนี้ควรส่งเสริมความร่วมมือระหว่างผู้บริหาร ครู และชุมชนในลักษณะระบบนิเวศทางการศึกษา เพื่อเสริมพลังซึ่งกันและกันในมิติของความไว้วางใจ

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

ผลการวิจัยเชิงประจักษ์จากข้อมูล PISA 2022 ของฮ่องกงได้เผยให้เห็นข้อสรุปสำคัญว่า ความไว้วางใจ (TR) ไม่เพียงแต่เป็นตัวแปรที่สำคัญ แต่ยังทำหน้าที่เป็น ตัวแปรส่งผ่านแบบสมบูรณ์ (full mediator) ในความสัมพันธ์ระหว่าง ภาวะผู้นำของผู้บริหาร (PL) กับความพึงพอใจในการทำงาน (JS) และความพึงพอใจในการสอน (TS) ข้อค้นพบนี้ชี้ให้เห็นว่า การเสริมสร้างความไว้วางใจภายในโรงเรียนมีความสำคัญต่อการขับเคลื่อนภาวะผู้นำไปสู่ผลลัพธ์เชิงบวกด้านความเป็นอยู่ทางวิชาชีพของครูอย่างแท้จริง ดังนั้น ข้อเสนอแนะเชิงนโยบายและการพัฒนามีดังนี้

ระดับประเทศ

ผลการวิจัยชี้ให้เห็นว่า “ความไว้วางใจ” เป็นกลไกสำคัญที่เชื่อมโยงภาวะผู้นำของผู้บริหารกับความพึงพอใจของครู ดังนั้น การพัฒนานโยบายระดับประเทศควรมุ่งเสริมสร้างวัฒนธรรมองค์กรที่ส่งเสริมความไว้วางใจในสถานศึกษาอย่างเป็นระบบ โดยสนับสนุนให้โรงเรียนเป็นพื้นที่ที่เปิดโอกาสให้เกิดการสื่อสารอย่างโปร่งใส การมีส่วนร่วมในการตัดสินใจ และการเคารพในความเป็นมืออาชีพของครู

แนวทางหนึ่งคือ การบรรจุ “การสร้างและธำรงรักษาความไว้วางใจ” เป็นองค์ประกอบสำคัญของหลักสูตรพัฒนาผู้บริหารสถานศึกษา โดยเน้นภาวะผู้นำแบบมีส่วนร่วม (participatory/distributed leadership) ที่เอื้อต่อความร่วมมือ ความสม่ำเสมอในการสื่อสาร และความเป็นธรรมในการบริหารจัดการ ซึ่งจะช่วยเสริมสร้างบรรยากาศองค์กรที่เอื้อต่อการปฏิบัติงานของครูอย่างยั่งยืน

นอกจากนี้ ภาครัฐอาจพิจารณานำ “ระดับความไว้วางใจภายในโรงเรียน” มาเป็นตัวชี้วัดเชิงคุณภาพประกอบการประเมินประสิทธิภาพการบริหารสถานศึกษา การติดตามตัวชี้วัดดังกล่าวจะช่วยให้หน่วยงานกำกับดูแลด้านการศึกษา (EDB) สามารถประเมินความเข้มแข็งของวัฒนธรรมองค์กร และกำหนดมาตรการสนับสนุนที่สอดคล้องกับบริบทของแต่ละโรงเรียนได้อย่างเหมาะสม

ในเชิงโครงสร้างการทำงาน แม้งานวิจัยนี้ได้ศึกษาภาระงานของครูโดยตรง แต่ผลการวิเคราะห์ที่สะท้อนบทบาทสำคัญของความไว้วางใจ ชี้ให้เห็นว่าการจัดสภาพแวดล้อมการทำงานที่เอื้อต่อการปฏิบัติหน้าที่หลักของครู อาจเป็นปัจจัยสนับสนุนทางอ้อมต่อการเสริมสร้างความไว้วางใจระหว่างครูกับผู้บริหาร อย่างไรก็ตาม ประเด็นดังกล่าวควรได้รับการศึกษาเพิ่มเติมในอนาคตเพื่อยืนยันกลไกเชิงประจักษ์

ระดับสถานศึกษา

ผลการวิจัยสะท้อนอย่างชัดเจนว่า ภาวะผู้นำของผู้บริหารส่งผลกระทบต่อความพึงพอใจของครูผ่านกลไก “ความไว้วางใจ” เท่านั้น ดังนั้น การพัฒนาระดับสถานศึกษาควรมุ่งเสริมสร้างความไว้วางใจในฐานะทุนทางสังคมขององค์กร โดยให้ความสำคัญกับทั้งบทบาทของผู้บริหารและครูในฐานะผู้มีส่วนร่วมในกระบวนการเดียวกัน

ในส่วนของผู้บริหาร โรงเรียนควรปรับแนวทางการบริหารจากการสั่งการเชิงลำดับชั้นไปสู่การบริหารแบบมีส่วนร่วมที่เปิดโอกาสให้ครูเข้ามามีบทบาทในการตัดสินใจอย่างแท้จริง โดยเฉพาะในประเด็น

ที่เกี่ยวข้องกับวิชาชีพและการจัดสรรทรัพยากร ความโปร่งใส ความสม่ำเสมอ และความเป็นธรรมในการบริหารจัดการเป็นองค์ประกอบสำคัญของการสร้างความไว้วางใจ นอกจากนี้ การสนับสนุนอิสระทางวิชาชีพ และการทำหน้าที่เป็น “ผู้สนับสนุนทางวิชาการ” มากกว่าผู้ควบคุม จะช่วยเสริมสร้างความไว้วางใจในระดับบุคคล (interpersonal trust) ซึ่งเป็นกลไกสำคัญที่เชื่อมโยงไปสู่ความพึงพอใจในการทำงานและการสอนของครู

ในส่วนของครู ในฐานะผู้ขับเคลื่อนคุณภาพการศึกษา ควรมีบทบาทเชิงรุกในการร่วมสร้างวัฒนธรรมความไว้วางใจ ผ่านการมีส่วนร่วมอย่างสร้างสรรค์ การแลกเปลี่ยนเรียนรู้ทางวิชาชีพ (Professional Learning Community: PLC) และการสื่อสารอย่างตรงไปตรงมา ความไว้วางใจจึงมิใช่เพียงผลลัพธ์จากการบริหารของผู้บริหาร แต่เป็นกระบวนการที่เกิดจากการตอบรับและการมีส่วนร่วมของครูด้วยเช่นกัน การแสดงความรับผิดชอบร่วมกันและการสนับสนุนบรรยากาศเชิงบวกภายในองค์กร จะช่วยให้ความไว้วางใจทำหน้าที่เป็นกลไกเสริมสร้างความพึงพอใจและความยั่งยืนของโรงเรียนในระยะยาว

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

ผลการวิจัยนี้ได้ยืนยันบทบาทของความไว้วางใจ (trust) ในฐานะตัวแปรส่งผ่านระหว่างภาวะผู้นำของผู้บริหารและความพึงพอใจของครู ซึ่งถือเป็นพื้นฐานสำคัญในการทำความเข้าใจกลไกในระดับบุคคลภายในองค์กรการศึกษา อย่างไรก็ตาม เพื่อเป็นการยกระดับองค์ความรู้ให้ครอบคลุมมิติเชิงบริบทและโครงสร้างที่ซับซ้อนยิ่งขึ้น ผู้วิจัยจึงมีข้อเสนอแนะสำหรับการวิจัยครั้งต่อไป ดังนี้

ประการแรก ต่อยอดสู่การวิเคราะห์เชิงลึกด้วย รูปแบบ Multi-Level Structural Equation Modeling (ML-SEM) แม้ว่าการวิจัยครั้งนี้จะให้ผลสรุปที่ชัดเจนเกี่ยวกับกลไกทางจิตวิทยาและปฏิสัมพันธ์ระหว่างบุคคล ซึ่งเป็นหัวใจสำคัญของความสัมพันธ์ระหว่างผู้บริหารกับครู แต่เนื่องจากข้อมูล PISA มีโครงสร้างลำดับชั้น โดยครูถูกรวมกลุ่มอยู่ภายในโรงเรียน การใช้โมเดลหลายระดับจะช่วยแยกความแปรปรวนของความไว้วางใจในระดับบุคคลออกจากความไว้วางใจร่วมของโรงเรียน ซึ่งจะนำไปสู่การทำความเข้าใจที่ชัดเจนยิ่งขึ้นว่า ความไว้วางใจควรได้รับการส่งเสริมผ่านมาตรการรายบุคคล หรือผ่านการพัฒนาบรรยากาศของโรงเรียนในภาพรวม

ประการที่สอง ควรมีการศึกษาความคงเส้นคงวาของกลไกการส่งผ่านแบบสมบูรณ (full mediation) ในกลุ่มประชากรย่อยที่แตกต่างกัน โดยใช้กลยุทธ์กลุ่มพหุ (multiple group strategy) เช่น การเปรียบเทียบครูตามประสบการณ์สอน หรือตามระดับชั้นที่รับผิดชอบ การวิเคราะห์ดังกล่าวจะช่วยให้ผู้บริหารเข้าใจว่าครูแต่ละกลุ่มอาจต้องการรูปแบบการสร้าง ความไว้วางใจที่แตกต่างกัน ทำให้สามารถออกแบบกลยุทธ์การบริหารที่ตอบสนองต่อความต้องการของครูได้อย่างเหมาะสมยิ่งขึ้น

ประการสุดท้าย ตรวจสอบเพิ่มเติมกับข้อมูลจากประเทศหรือเขตปกครองที่มีโครงสร้างระบบการศึกษาใกล้เคียงกัน เช่น สิงคโปร์หรือไต้หวัน อาจช่วยเพิ่มความเข้าใจเกี่ยวกับความคงทนของโมเดลในบริบทที่หลากหลาย การวิเคราะห์เชิงเปรียบเทียบเช่นนี้มีได้มีเป้าหมายเพื่อสรุปผลในเชิงสากล หากแต่เพื่อพิจารณาว่า บทบาทของความไว้วางใจในฐานะกลไกระหว่างภาวะผู้นำและความพึงพอใจของครู เป็นปรากฏการณ์เฉพาะบริบทของฮ่องกง หรือปรากฏในระบบการศึกษาที่มีลักษณะคล้ายคลึงกันด้วย ซึ่งจะช่วยสนับสนุนความเข้าใจด้านกลไกเชิงทฤษฎีให้มีความรอบด้านมากขึ้น

References

- Balan, A., & Jamaluddin, R. (2024). Leadership practices and teacher satisfaction: The critical role of trust and professional development in mid-career stages. *International Journal of Academic Research in Business and Social Sciences*, 14(9), 488–503.
<https://doi.org/10.6007/IJARBS/v14-i9/22635>
- Balyer, A. (2017). Trust in school principals: Teachers' opinions. *Journal of Education and Learning*, 6(2), 317–325. <https://doi.org/10.5539/jel.v6n2p317>
- Baptiste, M. (2019). No teacher left behind: The impact of principal leadership styles on teacher job satisfaction and student success. *Journal of International Education and Leadership*, 9(1). 1-11.
- Bass, B. M., & Riggio, R. E. (2006). *Transformational leadership*. Psychology Press.
- Bryk, A. S., & Schneider, B. (2002). *Trust in schools: A core resource for improvement*. Russell Sage Foundation.
- Committee on Teachers' Work. (2007). *The Committee on Teachers' Work final report*. Legislative Council of the Hong Kong Special Administrative Region.
<https://www.legco.gov.hk/yr06-07/english/panels/ed/papers/ed0212cb2-1041-6-e.pdf>
- Education Commission. (2019). *Review report of the Task Force on School-based Management Policy*. https://www.e-c.edu.hk/wp-content/uploads/2019/07/Report_SBM_2019.pdf
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Pearson.
- Hallinger, P., & Walker, A. (2017). Leading learning in Asia—Emerging empirical insights from five societies. *Journal of Educational Administration*, 55(2), 130–146.
<https://doi.org/10.1108/JEA-02-2017-0015>
- Harris, A. (2014). *Distributed leadership matters: Perspectives, practicalities, and potential*. Corwin Press.
- Hayes, A. F. (2009). Beyond Baron and Kenny: Statistical mediation analysis in the new millennium. *Communication Monographs*, 76(4), 408–420.
<https://doi.org/10.1080/03637750903310360>
- MacKinnon, D. P., Lockwood, C. M., & Williams, J. (2004). Confidence limits for the indirect effect: Distribution of the product and resampling methods. *Multivariate Behavioral Research*, 39(1), 99–128. https://doi.org/10.1207/s15327906mbr3901_4
- Muthén, L. K., & Muthén, B. O. (2012). *Mplus user's guide* (7th ed.). Muthén & Muthén.

- OECD. (2023a). *PISA 2022 results (Volume I): The state of learning and equity in education*. OECD Publishing. <https://doi.org/10.1787/53f23881-en>
- OECD. (2023b). *PISA 2022 results (Volume II): Learning during and after the COVID-19 pandemic*. OECD Publishing. <https://doi.org/10.1787/a77a4129-en>
- Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects in multiple mediator models. *Behavior Research Methods*, 40(3), 879–891. <https://doi.org/10.3758/BRM.40.3.879>
- Spillane, J. P. (2006). *Distributed leadership*. Jossey-Bass.
- Tobias, L. J. (2017). *A study of teacher job satisfaction, teacher preferred leadership behaviors, and the impact of the leadership behaviors on teacher job satisfaction* [Doctoral dissertation]. Gardner-Webb University. https://digitalcommons.gardner-webb.edu/education_etd/263/
- Tschannen-Moran, M. (2014). *Trust matters: Leadership for successful schools* (2nd ed.). Jossey-Bass.
- Wong, T. Y., & Liu, P. (2018). Hong Kong principal professional development: Context, challenges, and opportunities. *Chinese Education & Society*, 51(5), 359–371. <https://doi.org/10.1080/10611932.2018.1510689>
- Xu, F., & Shen, J. (2007). Research on job satisfaction of elementary and high school teachers and strategies to increase job satisfaction. *Chinese Education and Society*, 40(5), 86–96. <https://doi.org/10.2753/CED1061-1932400509>

การพัฒนาดัชนีความเสี่ยงทางการศึกษาที่เชื่อมโยงกับสภาพภูมิอากาศของประเทศไทย Development of a Climate-Mapped Educational Risk Index for Thailand

สิวะโชติ ศรีสุทธิยากร¹ กนิษฐ์ ศรีเคลือบ^{1*} ประภาศิริ รัชชประภาพรกุล¹ และ วาทีนี อมรไพศาลเลิศ¹

Siwachoat Srisuttiyarkorn¹ Kanit Sriklaub^{1*} Prapasiri Ratchaprapapornkul¹

and Watinee Amornpaisarnler¹

(Received: December 26, 2025; Revised: February 26, 2026; Accepted: March 2, 2026)

บทคัดย่อ

ระบบการศึกษาของโลกและประเทศไทยกำลังเผชิญกับความเปราะบางจากภัยพิบัติที่เกี่ยวข้องกับการเปลี่ยนแปลงภูมิอากาศ โดยเฉพาะในพื้นที่ชนบทและพื้นที่ที่มีข้อจำกัดทางเศรษฐกิจและสังคม แม้ว่าการรอบนโยบายระดับโลกจะให้ความสำคัญกับการปกป้องเด็กจากความเสี่ยงทางสิ่งแวดล้อม แต่ดัชนีชี้วัดที่มีอยู่ยังขาดความละเอียดเชิงพื้นที่และไม่ได้ผนวกผลลัพธ์ทางการศึกษาเข้าไปในโครงสร้างโดยตรง งานวิจัยนี้มีวัตถุประสงค์เพื่อ (1) พัฒนาดัชนีเชิงพื้นที่ของความเสี่ยงทางการศึกษาที่เชื่อมโยงกับสภาพภูมิอากาศ (Climate-Mapped Educational Risk Index: CMER) และ (2) วิเคราะห์ความสัมพันธ์ระหว่างดัชนีดังกล่าวกับปัจจัยด้านภูมิศาสตร์ โรงเรียน และนักเรียน การวิจัยใช้ข้อมูลคะแนน O-NET ระดับโรงเรียนจาก 4 รายวิชาหลักของโรงเรียน 30,112 แห่งในช่วง พ.ศ. 2560–2564 ร่วมกับข้อมูลภูมิอากาศความละเอียดสูงจากโครงการ SEACLID/CORDEX-SEA โดยเปรียบเทียบกับระยะฐาน พ.ศ. 2513–2550 ดัชนี CMER พัฒนาขึ้นด้วยการวิเคราะห์องค์ประกอบหลักระหว่างตัวแปรภูมิอากาศสุดขีดกับผลสัมฤทธิ์ทางการเรียน และจัดกลุ่มโรงเรียนตามระดับความเสี่ยงด้วยเกณฑ์ ROC จากนั้น ใช้การถดถอยแบบ Elastic Net วิเคราะห์ความสัมพันธ์ระหว่างค่า CMER กับปัจจัยเสี่ยงรายเขตพื้นที่การศึกษา และสร้างเมทริกซ์ความเสี่ยง-ความสำคัญ (Risk-Importance Matrix: RIM) เพื่อสนับสนุนการตัดสินใจเชิงนโยบาย

ผลการวิจัยพบว่า ระยะเวลาของปรากฏการณ์ฝนตกสุดขั้วเป็นมิติเดียวที่มีความสัมพันธ์เชิงโครงสร้างกับผลสัมฤทธิ์ทางการเรียนอย่างมีนัยสำคัญ โดยโรงเรียนในพื้นที่ภาคเหนือ ภาคตะวันออกเฉียงเหนือ ภาคกลางและภาคใต้มีแนวโน้มเสี่ยงสูง และผลการวิเคราะห์ RIM พบว่าปัจจัยทางการศึกษามีส่วนขยายขนาดความเสี่ยงแตกต่างกันในแต่ละพื้นที่ การออกแบบมาตรการที่จำเพาะต่อบริบทจึงมีความจำเป็น ดัชนี CMER และ RIM เป็นกรอบข้อมูลที่สามารถใช้ระบุความเสี่ยงและชี้เป้าการจัดสรรทรัพยากรเพื่อสร้างระบบการศึกษาที่เสมอภาคและยืดหยุ่นต่อการเปลี่ยนแปลงภูมิอากาศ

คำสำคัญ การเปลี่ยนแปลงสภาพภูมิอากาศ ความเสี่ยงทางการศึกษา

ดัชนีความเสี่ยงทางการศึกษาที่เชื่อมโยงกับสภาพภูมิอากาศ เมทริกซ์ความเสี่ยง-ความสำคัญ

¹ หน่วยการวิจัยทางการศึกษาพิเศษและการเรียนรวม คณะครุศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย

¹ Special and Inclusive Education Research Unit, Faculty of Education, Chulalongkorn University

งานวิจัยนี้ได้รับทุนสนับสนุนงานวิจัยจากองค์การยูนิเซฟ (UNICEF) ประเทศไทย

This research was funded by UNICEF Thailand

* Corresponding Author E-mail: kanit.s@chula.ac.th

Abstract

Education systems worldwide, including Thailand's, face increasing vulnerability to climate change-related disasters, particularly in rural and socioeconomically disadvantaged areas. Although global policy frameworks emphasize protecting children from environmental risks, existing indices lack spatial granularity and do not directly incorporate educational outcomes into their structure. This study aims to (1) develop a spatial index of educational risk linked to climate conditions—the Climate-Mapped Educational Risk Index (CMER)—and (2) analyze the relationships between this index and geographic, school, and student factors. The study integrated school-level O-NET scores from four core subjects across 30,112 schools during 2017–2021 with high-resolution climate data from the SEACLID/CORDEX-SEA project covering the same period relative to a baseline of 1970–2007. The CMER Index was developed using Principal Component Analysis of extreme climate variables and student achievement, with schools classified by risk levels using ROC-based criteria. Elastic Net regression was then applied to analyze relationships between CMER scores and risk factors across educational service areas, followed by the construction of a Risk-Importance Matrix (RIM) to support policy decision-making.

The results revealed that the duration of extreme precipitation events was the only dimension showing a significant structural relationship with student achievement. Schools in the Northern, Northeastern, Central, and Southern regions exhibited higher risk levels. RIM analysis indicated that educational factors amplify risk to varying degrees across different areas, underscoring the need for context-specific interventions. The CMER Index and RIM provide a data-driven framework for identifying climate-linked educational risks and targeting resource allocation to build a more equitable and climate-resilient education system.

Keywords: climate change, educational risk, Climate-Mapped Educational Risk Index, Risk-Importance Matrix

บทนำ

เหตุการณ์สภาพอากาศสุดขั้ว เช่น ฝนตกหนักยืดยาว น้ำท่วมรุนแรง และคลื่นความร้อน กำลังทวีความถี่และความรุนแรงขึ้นทั่วโลก (Intergovernmental Panel on Climate Change [IPCC], 2023) ก่อให้เกิดผลกระทบอย่างรุนแรงและซับซ้อนต่อความต่อเนื่องทางการเรียนรู้ โดยเฉพาะกับเด็กซึ่งเป็นกลุ่มเปราะบางที่สุด (Auer-Frege et al., 2024; UNICEF, 2021) สำหรับประเทศไทยซึ่งถูกจัดอยู่ในกลุ่มประเทศที่มีความเสี่ยงสูง (Eckstein et al., 2021) ภัยพิบัติทางสภาพภูมิอากาศได้ส่งผลกระทบเป็นวงกว้าง เช่น การสร้างความเสียหายต่อโครงสร้างพื้นฐานของโรงเรียน การหยุดชะงักของการเดินทาง ไปจนถึงการปิดสถานศึกษาชั่วคราว ซึ่งทำให้เกิดการสูญเสียโอกาสในการเรียนรู้ และยิ่งซ้ำเติมความเหลื่อมล้ำด้านการศึกษาที่มีอยู่เดิมให้รุนแรงยิ่งขึ้น (Auer-

Frege et al., 2023; Thampanishvong et al., 2022; UNICEF, 2021) ถึงแม้จะมีการตระหนักเกี่ยวกับปัญหาดังกล่าว แต่ปัจจุบันยังขาดเครื่องมือหรือดัชนีที่สามารถวัดและประเมินระดับผลกระทบทางการศึกษาในแต่ละพื้นที่ได้อย่างเป็นระบบ โดยดัชนีชี้วัดระดับโลกที่มีอยู่ยังขาดความละเอียดเชิงพื้นที่และไม่ได้ถูกออกแบบมาเพื่อจับความแปรปรวนของระบบการศึกษาโดยตรง

ช่องว่างในการประเมินความเสี่ยงนี้อาจมีสาเหตุมาจากการข้อจำกัดในการเข้าถึงข้อมูลสภาพภูมิอากาศของประเทศในระดับพื้นที่ย่อยที่มีความสม่ำเสมอและครอบคลุมได้เพียงพอ ส่งผลให้การประเมินความเสี่ยงของโรงเรียนในพื้นที่ห่างไกล พื้นที่ภูเขา หรือชายฝั่ง ซึ่งมักมีความเสี่ยงเฉพาะถิ่นที่แตกต่างกันและเป็นไปอย่างจำกัด (Auer-Frege et al., 2023; Auer-Frege et al., 2024; Adler et al., 2022) ขณะเดียวกัน แผนนโยบายการศึกษาเชิงพื้นที่จะให้ความสำคัญกับการใช้ข้อมูลเพื่อการพัฒนา (area-based education) แต่ฐานข้อมูลที่ใช้วิเคราะห์ยังไม่พบว่ามีกรณีผลพวงภูมิอากาศเข้ามาในกรอบการวิเคราะห์อย่างเป็นระบบ ทำให้การชี้เป้าเพื่อจัดสรรทรัพยากรและการสนับสนุนยังคงพึ่งพาการรับรู้ทั่วไป มากกว่าการอ้างอิงหลักฐานเชิงประจักษ์ที่เชื่อมโยงความเสี่ยงจากสภาพภูมิอากาศเข้ากับบริบทของโรงเรียนได้อย่างแท้จริง

จากสภาพข้างต้นประเด็นปัญหาวิจัยจึงไม่ใช่การระบุพื้นที่ที่เผชิญเหตุการณ์สภาพอากาศสุดขั้ว แต่คือการสร้างกรอบวิเคราะห์ที่สามารถเชื่อมโยงข้อมูลสภาพภูมิอากาศความละเอียดสูงเข้ากับผลลัพธ์ทางการศึกษาในระดับโรงเรียน เพื่อใช้เป็นสารสนเทศเชิงลึกที่สามารถอธิบายพลวัตของความเสี่ยงในแต่ละบริบทพื้นที่ การวิเคราะห์ในระดับพื้นที่ย่อยนี้มีความสำคัญอย่างยิ่ง เนื่องจากกลไกของความเสี่ยงทางการศึกษานั้นมีความแตกต่างกันไปตามบริบท ซึ่งสนับสนุนโดยงานวิจัยของ Srisuttiyakorn (2019) ที่พบว่าความสัมพันธ์ระหว่างปัจจัยด้านประสิทธิภาพการบริหารจัดการของโรงเรียนและระดับฐานะทางเศรษฐกิจของครอบครัวกับความเหลื่อมล้ำทางการศึกษานั้นมีความแตกต่างกันไปตามเขตพื้นที่การศึกษา สะท้อนให้เห็นถึงความหลากหลายของบริบทที่ไม่สามารถใช้มาตรการเดียวกันครอบคลุมทุกพื้นที่ได้อย่างมีประสิทธิภาพ การทำความเข้าใจความแตกต่างนี้จึงจะนำไปสู่การจัดลำดับความสำคัญของการแทรกแซงเชิงนโยบายที่มีเป้าหมายชัดเจน อย่างไรก็ตาม การวางแผนการแทรกแซงเชิงนโยบายที่มีประสิทธิภาพควรพิจารณาจากปัจจัยที่ส่งผลหรือเป็นตัวขับเคลื่อนความเสี่ยงดังกล่าวโดยเฉพาะปัจจัยเชิงโครงสร้างหรือปัจจัยทางการศึกษาในพื้นที่

จากสภาพข้างต้นและการทบทวนวรรณกรรมที่เกี่ยวข้อง สามารถวิเคราะห์ช่องว่างทางวิชาการที่สำคัญสามประการ ประการแรก ดัชนีความเสี่ยงที่มีอยู่ในระดับสากลขาดความละเอียดเชิงพื้นที่ในระดับโรงเรียนหรือเขตพื้นที่การศึกษา ประการที่สอง ยังไม่มีดัชนีที่ผนวกผลลัพธ์ทางการเรียนรู้ที่วัดได้จริงเข้าไปในโครงสร้างของดัชนีโดยตรง ดัชนีที่มีอยู่มักใช้ตัวแปรแทน (proxy variables) เช่น อัตราการเข้าเรียนหรือสัดส่วนเด็กด้อยโอกาส เพื่อบ่งชี้ความเปราะบางทางการศึกษา และรวมคะแนนของตัวแปรแทนดังกล่าวเข้ากับมิติการเผชิญภัยด้วยโมเดลเชิงบวก (additive model) โดยมีการกำหนดน้ำหนักของคะแนนแต่ละมิติจากกรอบแนวคิดทฤษฎีหรือสมมุติฐานของผู้วิจัยที่กำหนดไว้ล่วงหน้า วิธีการนี้ทำให้ไม่สามารถตรวจจับได้ว่ามิติใดของการเผชิญภัยภูมิอากาศมีความสัมพันธ์เชิงโครงสร้างกับการลดลงของผลสัมฤทธิ์ที่วัดได้จริง และยังไม่สามารถสำรวจมิติแฝงของความเสี่ยงทางการศึกษาที่ปรากฏการณ์ในทั้งสองมิติแปรผันร่วมกันเชิงประจักษ์ได้ ประการที่สาม แผนนโยบายการจัดการศึกษาเชิงพื้นที่ (area-based education) จะให้ความสำคัญกับการใช้ข้อมูลเชิงพื้นที่ แต่ยังไม่มีการอบวิเคราะห์ที่บูรณาการผลพวงภูมิอากาศเข้ากับบริบทของโรงเรียนอย่างเป็นระบบ

ทำให้การชี้เป้าเพื่อจัดสรรทรัพยากรยังคงอิงการรับรู้ทั่วไปมากกว่าหลักฐานเชิงประจักษ์ ช่องว่างทั้งสามประการนี้ชี้ให้เห็นถึงความจำเป็นในการพัฒนาดัชนีที่มีความละเอียดเชิงพื้นที่ ผนวกผลลัพธ์ทางการเรียนรู้ที่วัดได้จริง และสามารถค้นพบความสัมพันธ์เชิงโครงสร้างระหว่างภูมิอากาศกับการศึกษาจากข้อมูลเชิงประจักษ์ซึ่งนำไปสู่วัตถุประสงค์หลักของงานวิจัยนี้

วัตถุประสงค์ของการวิจัย

1. พัฒนาดัชนีเชิงพื้นที่ของความเสี่ยงทางการศึกษาที่เชื่อมโยงกับการเปลี่ยนแปลงสภาพภูมิอากาศ
2. วิเคราะห์ความสัมพันธ์ระหว่างดัชนีเชิงพื้นที่ของความเสี่ยงทางการศึกษาที่เชื่อมโยงกับการเปลี่ยนแปลงสภาพภูมิอากาศ กับปัจจัยด้านภูมิศาสตร์ ลักษณะของโรงเรียน และลักษณะของนักเรียน

วิธีดำเนินการวิจัย

1. ข้อมูลที่ใช้ในการวิจัย

ข้อมูลในการวิจัยจำแนกตามวัตถุประสงค์การใช้งานเป็น 2 ส่วน ส่วนแรกเป็นข้อมูลสำหรับสร้างดัชนี CMER ประกอบด้วยสองมิติ ได้แก่ มิติการเผชิญภัยภูมิอากาศ (climate hazard exposure) ซึ่งวัดจากตัวชี้วัดสภาพภูมิอากาศสุดขีดครอบคลุมด้านความเข้ม ความถี่ และระยะเวลา และมิติความเปราะบางทางการศึกษา (educational vulnerability) ซึ่งวัดจากผลสัมฤทธิ์ทางการเรียน (O-NET) ในฐานะตัวชี้วัดที่สะท้อนว่าการเรียนรู้สะสมได้รับผลกระทบจนปรากฏเป็นผลลัพธ์ที่วัดได้จริงหรือไม่ ส่วนที่สองเป็นข้อมูลลักษณะโรงเรียนและนักเรียน ซึ่งใช้วิเคราะห์ปัจจัยที่ขยายหรือลดทอนความเปราะบางดังกล่าวในแต่ละบริบทพื้นที่ตามวัตถุประสงค์ที่ 2 รายละเอียดมีดังนี้

1.1 ข้อมูลภูมิอากาศ (climate data) ข้อมูลภูมิอากาศที่ใช้ในงานวิจัยครอบคลุม 2 ช่วงเวลา คือ ระยะเวลา (พ.ศ. 2513–2550) และระยะศึกษา (พ.ศ. 2560–2564) จากโครงการ SEACLID/CORDEX Southeast Asia (Phase II) ซึ่งได้จากการจำลองภูมิอากาศระดับภูมิภาคและ downscaling ให้มีความละเอียด 25×25 กม. ครอบคลุมพื้นที่ประเทศไทยทั้งหมด 2,619 กริด ตัวชี้วัดแบ่งเป็น 2 กลุ่ม คือ ปรากฏการณ์ฝนตกสุดขั้ว 7 ตัว และปรากฏการณ์อุณหภูมิสุดขั้ว 11 ตัว ครอบคลุม 3 มิติการวัด ได้แก่ ความเข้ม ความถี่ และระยะเวลา รายละเอียดแสดงในตารางที่ 1 (Ngo-Duc et al., 2016)

1.2 ข้อมูลทางการศึกษา (educational data) ข้อมูลทางการศึกษา หรือข้อมูลที่ไม่ใช่สภาพภูมิอากาศ (non-climate data) ที่ใช้ในงานวิจัย ประกอบด้วย ข้อมูลคุณลักษณะของโรงเรียน และคุณลักษณะของนักเรียน

ข้อมูลคุณลักษณะของโรงเรียนเป็นข้อมูลของโรงเรียนในสังกัดสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานในปีการศึกษา 2560 และ 2564 ครอบคลุมโรงเรียน 30,112 แห่งทั่วประเทศ จำแนกเป็นโรงเรียนในเขตพื้นที่การศึกษาประถมศึกษา (สพป.) และเขตพื้นที่การศึกษามัธยมศึกษา (สพม.) ทั้งนี้ในการวิเคราะห์เลือกใช้เฉพาะโรงเรียนที่มีข้อมูลครบถ้วนทุกรายวิชาและทุกปีในช่วง พ.ศ. 2560–2564 ซึ่งทำให้จำนวนโรงเรียนที่ใช้ในการวิเคราะห์จริงมี 26,311 แห่ง (ร้อยละ 87.38)

ข้อมูลคุณลักษณะของโรงเรียนที่ใช้ในการวิเคราะห์ ประกอบด้วย (1) ข้อมูลทั่วไป ได้แก่ ชื่อโรงเรียน ขนาดโรงเรียน (ขนาดเล็ก กลาง ใหญ่ และใหญ่พิเศษ) สังกัดของโรงเรียน และที่อยู่ของโรงเรียน (2) พิกัด

ภูมิศาสตร์ของโรงเรียน ซึ่งได้จากการดึงข้อมูลผ่าน Google Maps API (Google, 2024) โดยค้นหาพิกัดอิงจากชื่อและที่อยู่ของโรงเรียน (3) ระยะห่างจากศูนย์กลางอำเภอ (กิโลเมตร) ซึ่งคำนวณจากพิกัดของโรงเรียนไปยังจุดอ้างอิงที่ว่าการอำเภอในแต่ละพื้นที่ โดยใช้สูตร Haversine เพื่อให้ได้ระยะทางทางภูมิศาสตร์ที่มีความแม่นยำ และ (4) ระดับความยากลำบากในการเข้าถึงโรงเรียน ซึ่งได้จากประกาศของ สพฐ. ในปีงบประมาณ 2565–2566 เกี่ยวกับการกำหนดโรงเรียนในพื้นที่พิเศษ เช่น พื้นที่สูงหรือพื้นที่เกาะ โดยจำแนกความยากลำบากออกเป็น 4 ระดับ ได้แก่ ไม่ยากลำบาก (not difficult) จนถึงยากลำบากมากที่สุด (most difficult) และ (5) ข้อมูลโครงสร้างพื้นฐานและการเชื่อมต่อ เกี่ยวข้องกับการมีและลักษณะของสาธารณูปโภคของโรงเรียน ได้แก่ ไฟฟ้า น้ำประปา และการเชื่อมต่ออินเทอร์เน็ตของโรงเรียน ข้อมูลส่วนนี้เก็บรวบรวมด้วยเทคนิค web scraping จากเว็บไซต์ระบบข้อมูลสินทรัพย์ สพฐ. ในเดือนกันยายน 2567

ข้อมูลคุณลักษณะนักเรียน เป็นข้อมูลระดับโรงเรียน ประกอบด้วย (6) ร้อยละของนักเรียนในโรงเรียนที่ครอบครัวยากจน โดยตัวแปรนี้ถูกจัดกระทำให้เป็นตัวแปรดัมมี่โดยมีค่าเป็น 1 เมื่อโรงเรียนจำนวนนักเรียนยากจนมากกว่าร้อยละ 50 และ (7) ผลสัมฤทธิ์ทางการเรียนของนักเรียน ใช้คะแนน O-NET ระดับโรงเรียนของนักเรียนระหว่างปี พ.ศ. 2560–2564 จำแนกตามระดับชั้น ป.6 ม.3 และ ม.6 ในรายวิชาคณิตศาสตร์ ภาษาไทย ภาษาอังกฤษ และวิทยาศาสตร์ ข้อมูลนี้ได้จากฐานข้อมูลผู้สอบ O-NET ของสถาบันทดสอบทางการศึกษาแห่งชาติ (สทศ.) โดยคะแนน O-NET ถูกนำไปใช้ในสองบทบาทในการวิเคราะห์ บทบาทแรกคือเป็นข้อมูลนำเข้าร่วมกับตัวชี้วัดภูมิอากาศสุดขีดเพื่อสร้างดัชนี CMER และบทบาทที่สองใช้เป็นตัวแปรอ้างอิงในการกำหนดจุดตัดด้วย ROC เพื่อปรับแต่งให้ระดับคะแนน CMER ที่ใช้จำแนกความเสี่ยงมีความสอดคล้องกับผลสัมฤทธิ์ทางการเรียนที่ต่ำกว่าเกณฑ์ ซึ่งทำให้สามารถรับประกันความสอดคล้องกันระหว่างสิ่งที่ดัชนี CMER วัดกับเกณฑ์ที่ใช้ตัดสินใจระดับความเสี่ยง

2. การประมวลผลและการวิเคราะห์ข้อมูล

การวิเคราะห์ข้อมูลจำแนกออกเป็น 2 ส่วนหลัก ได้แก่ การวิเคราะห์ข้อมูลเพื่อพัฒนาดัชนีความเสี่ยงทางการศึกษาที่เชื่อมโยงกับสภาพภูมิอากาศ (Climate-Mapped Educational Risk Index: CMER Index) ผลการวิเคราะห์ส่วนนี้ใช้ตอบวัตถุประสงค์ข้อที่ 1 และการวิเคราะห์ความสัมพันธ์ระหว่างปัจจัยทางการศึกษากับดัชนี CMER ที่พัฒนาขึ้นภายในแต่ละเขตพื้นที่การศึกษา ตอบวัตถุประสงค์ข้อที่ 2

ตาราง 1 ข้อมูลภูมิอากาศด้านปรากฏการณ์ฝนตกสุดขีดและอุณหภูมิสุดขีด

ตัวชี้วัด	คำจำกัดความ	หน่วย	หมวดหมู่
ตัวชี้วัดปรากฏการณ์ฝนตกสุดขีด (extreme precipitation events)			
Rx1day	ปริมาณฝนสะสมสูงสุดใน 1 วัน	มิลลิเมตร (mm)	ความเข้ม (intensity)
Rx5day	ปริมาณฝนสะสมสูงสุดใน 5 วัน	มิลลิเมตร (mm)	
R95p	ผลรวมรายปีของปริมาณฝนประจำวันที่มีมากกว่าค่าเปอร์เซ็นต์ไทล์ที่ 95	มิลลิเมตร (mm)	
R99p	ผลรวมรายปีของปริมาณฝนประจำวันที่มีมากกว่าค่าเปอร์เซ็นต์ไทล์ที่ 99	มิลลิเมตร (mm)	
R10mm	จำนวนวันที่มีปริมาณฝน ≥ 10 มม.	วัน (days)	ความถี่
R20mm	จำนวนวันที่มีปริมาณฝน ≥ 20 มม.	วัน (days)	
CWD	จำนวนวันที่ฝนตกต่อเนื่องสูงสุดต่อปี (ฝน ≥ 1 มม.)	วัน (days)	ระยะเวลา

ตัวชี้วัด	คำจำกัดความ	หน่วย	หมวดหมู่
ตัวชี้วัดปรากฏการณ์อุณหภูมิสุดขั้ว (extreme temperature events)			
TXx	ค่าอุณหภูมิสูงสุดรายวันที่สูงสุดในแต่ละเดือน	องศาเซลเซียส (°C)	ความเข้ม (Intensity)
TNx	ค่าอุณหภูมิต่ำสุดรายวันที่สูงสุดในแต่ละเดือน	°C	
TXn	ค่าอุณหภูมิสูงสุดรายวันที่ต่ำสุดในแต่ละเดือน	°C	
TNn	ค่าอุณหภูมิต่ำสุดรายวันที่ต่ำสุดในแต่ละเดือน	°C	
DTR	ค่าผลต่างเฉลี่ยระหว่างอุณหภูมิสูงสุดและต่ำสุดในแต่ละเดือน	°C	
TX10p	ร้อยละของวันที่ Tmax < เปอร์เซ็นไทล์ที่ 10	% ของวัน	ความถี่
TN10p	ร้อยละของวันที่ Tmin < เปอร์เซ็นไทล์ที่ 10	% ของวัน	
TX90p	ร้อยละของวันที่ Tmax > เปอร์เซ็นไทล์ที่ 90	% ของวัน	
TN90p	ร้อยละของวันที่ Tmin > เปอร์เซ็นไทล์ที่ 90	% ของวัน	
CSD	จำนวนวันที่ Tmin < เปอร์เซ็นไทล์ที่ 10 ต่อเนื่องกันอย่างน้อย 6 วัน	วัน	ระยะเวลา
WSD	จำนวนวันที่ Tmax > เปอร์เซ็นไทล์ที่ 90 ต่อเนื่องกันอย่างน้อย 6 วัน	วัน	(Duration)

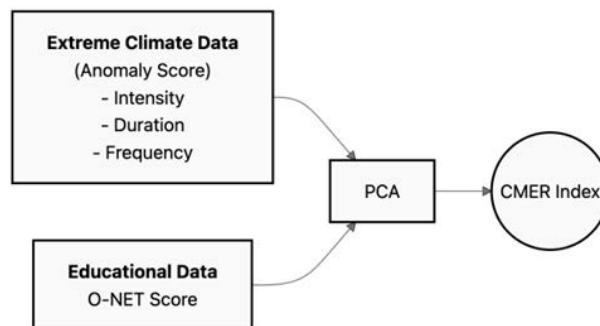
2.1 การพัฒนาดัชนี CMER ดัชนีเชิงพื้นที่ของความเสี่ยงทางการศึกษาที่เชื่อมโยงกับการเปลี่ยนแปลงสภาพภูมิอากาศที่ใช้ในการวิจัยนี้ ถูกออกแบบให้เติมเต็มช่องว่างของการสร้างดัชนีวัดความเสี่ยงแบบเดิมที่ใช้กันในปัจจุบัน เช่น Children's Climate Risk Index (CCRI) ของ UNICEF (2021) และ WorldRiskIndex (Bündnis Entwicklung Hilft, 2024) ที่มักสร้างขึ้นจากการนิยามตัวชี้วัดออกเป็น 2 มิติ ได้แก่ มิติการเผชิญภัย (exposure) และความเปราะบาง (vulnerability) แยกกันตามกรอบทฤษฎี จากนั้นรวมคะแนนของตัวชี้วัดดังกล่าวเข้าด้วยกันด้วยด้วยโมเดลเชิงบวก (additive model) วิธีการนี้มีข้อจำกัดสำคัญ 2 ประการ ประการแรกคือ ไม่สามารถระบุได้จากข้อมูลว่าการเผชิญภัยทางภูมิอากาศดังกล่าวมีความสัมพันธ์กับผลลัพธ์ทางการศึกษาอย่างไร ประการที่สองที่ผ่านมามีดัชนีดังกล่าวมักใช้ข้อมูลทางการศึกษาในมิติที่เน้นการบ่งชี้ความเปราะบางทางสังคม เช่น อัตราการเข้าเรียน หรือสัดส่วนเด็กด้อยโอกาส โดยไม่ได้ผนวกผลลัพธ์ทางการเรียนเข้าไปประเด็นนี้เป็นช่องว่างให้ดัชนีดังกล่าวไม่สามารถตรวจจับได้ว่าความเสี่ยงจากสภาพภูมิอากาศที่เกิดขึ้นมีความสัมพันธ์กับการลดลงของผลลัพธ์การเรียนรู้ในระดับใด และพื้นที่ใด

งานวิจัยนี้นำเสนอกรอบแนวคิดของการวัดความเสี่ยงที่แตกต่างออกไป โดยพิจารณาความเสี่ยงทางการศึกษาในฐานะปรากฏการณ์ที่เกิดจากปฏิสัมพันธ์เชิงโครงสร้างระหว่างการเผชิญภัยภูมิอากาศกับความเปราะบางของระบบการศึกษา ซึ่งสอดคล้องกับกรอบการประเมินความเสี่ยงภัยพิบัติของ UNDRR (2015) อย่างไรก็ตามงานวิจัยนี้ปรับแนวคิดดังกล่าวใน 2 ประการ ประการแรก แทนที่จะวัดความเปราะบางทางการศึกษาด้วยตัวแปรแทน (proxy variable) เช่น อัตราการเข้าเรียน หรือสัดส่วนเด็กด้อยโอกาส ซึ่งเป็นแนวทางที่ดัชนีความเสี่ยงระดับโลกส่วนใหญ่ใช้ งานวิจัยนี้นิยามความเปราะบางทางการศึกษาเป็นผลสัมฤทธิ์ทางการเรียนผ่านผลการสอบ O-NET ที่วัดได้จริงในระดับโรงเรียน ประการที่สอง งานวิจัยนี้มองว่า ความเสี่ยงทางการศึกษาที่เชื่อมโยงกับสภาพภูมิอากาศเป็นสภาพที่ปรากฏขึ้นในมิติแฝง (latent dimension) ที่การเผชิญภัยทางภูมิอากาศและการด้อยผลสัมฤทธิ์ทางการเรียนแปรผันร่วมกันในทิศทางตรงข้าม และมีนัยสำคัญในเชิงปฏิบัติ เรียกดัชนีที่สร้างขึ้นจากแนวคิดนี้ว่าดัชนีความเสี่ยงทางการศึกษาที่เชื่อมโยงกับสภาพภูมิอากาศ (CMER)

การสร้างดัชนี CMER เริ่มจากการใช้ข้อมูลภูมิอากาศ (ตัวชี้วัดฝนตก และอุณหภูมิสุดขีด) มาคำนวณเป็นคะแนนความผิดปกติ (anomaly scores) ที่สะท้อนความเบี่ยงเบนของสภาพภูมิอากาศในช่วงเวลาที่ทำการศึกษาจากการเปรียบเทียบข้อมูลภูมิอากาศ 2 ระยะ ได้แก่ ระยะศึกษา (พ.ศ. 2560 - 2564) กับระยะฐาน (พ.ศ. 2513 - 2550) คะแนนความผิดปกตินี้ถูกนำมาผสานกับข้อมูลผลสัมฤทธิ์ทางการเรียน (คะแนน O-NET ระดับโรงเรียน จำนวน 4 รายวิชา) จำแนกตามระดับชั้น ภาพ 1 แสดงกรอบการวิเคราะห์ที่ใช้สร้างดัชนี CMER ด้วยการวิเคราะห์องค์ประกอบหลัก (principal component analysis) วิธีการนี้เป็นเทคนิคหนึ่งในกลุ่มการเรียนรู้ของเครื่องแบบไม่มีการกำกับ (unsupervised learning) ที่ช่วยลดมิติข้อมูลและเน้นรักษาความแปรปรวนสำคัญของข้อมูลต้นฉบับเอาไว้ให้มากที่สุด ผลลัพธ์ที่ได้จากการวิเคราะห์นี้คือ คะแนนองค์ประกอบหลักที่สะท้อนความเสี่ยงทางการศึกษาที่สัมพันธ์กับการเปลี่ยนแปลงของสภาพภูมิอากาศในระดับโรงเรียนอย่างเป็นระบบ

คะแนนองค์ประกอบหลักถูกแปลงเป็นคะแนนมาตรฐานที่มีค่าอยู่ในช่วง 0 ถึง 1 ด้วยฟังก์ชันการแจกแจงสะสมแบบปกติ (normal cumulative distribution function: CDF) การแปลงนี้ช่วยให้มีเกณฑ์ที่ชัดเจนในเชิงตำแหน่งเปอร์เซ็นต์และเพิ่มความแกร่งต่อค่าผิดปกติ (outlier) ซึ่งช่วยให้การวิเคราะห์เพื่อจำแนกกลุ่มเสี่ยงด้วย ROC และการวิเคราะห์ความสัมพันธ์ สามารถดำเนินการได้อย่างมีความหมายและมีความแม่นยำมากขึ้น

การกำหนดคะแนนจุดตัดให้กับดัชนี CMER ผู้วิจัยจำแนกระดับความเสี่ยงเป็น 3 ระดับ (ต่ำ ปานกลาง สูง) โดยใช้การวิเคราะห์ Receiver Operating Characteristic (ROC) เพื่อหาจุดตัดที่มีประสิทธิภาพสูงสุด โดยคะแนนผลสัมฤทธิ์ทางการเรียนถูกแปลงเป็น Z-score ($Z < -0.5$ = ต่ำ, $Z < -2.5$ = ต่ำมาก) และใช้เป็น ตัวแปรอ้างอิง (ground truth) ในการคำนวณค่าความไว (sensitivity) และค่าความจำเพาะ (specificity) เพื่อเลือกจุดตัดที่รักษาสมดุลระหว่างสองค่าดังกล่าว (Unal, 2017)



ภาพ 1 กรอบการวิเคราะห์เพื่อพัฒนาดัชนี CMER

2.2 การวิเคราะห์ความสัมพันธ์ระหว่างปัจจัยทางการศึกษากับดัชนี CMER ผู้วิจัยวิเคราะห์ความสัมพันธ์ระหว่างปัจจัยทางการศึกษาและดัชนี CMER ที่พัฒนาขึ้นตามกรอบดังภาพ 2 โดยปัจจัยทางการศึกษาที่ใช้วิเคราะห์มี 6 ตัวแปร ได้แก่ (1) การเป็นโรงเรียนขนาดเล็ก (small school) เป็นตัวแปรดัมมี่ = 1 เมื่อเป็นโรงเรียนขนาดเล็ก ซึ่งสะท้อนข้อจำกัดด้านทรัพยากรการบริหารจัดการ (2) ระยะทางถึงศูนย์กลางอำเภอ (school distance) คำนวณจากพิกัดโรงเรียนกับที่ว่าการอำเภอใกล้ที่สุด แปลงเป็นคะแนนมาตรฐานแบบ Max-Min (0-1) (3) ความยากลำบากในการเข้าถึง (hard to reach) ดัมมี่ = 1 เมื่อมีระดับความยากลำบากตั้งแต่ “ยากมาก” ขึ้นไป ซึ่งทั้ง (2) และ (3) สะท้อนความท้าทายในการเดินทาง โดยเฉพาะช่วงเกิดสภาพภูมิอากาศสุดขีด (4) การขาดแคลนโครงสร้างพื้นฐาน (lack of infrastructure) ดัมมี่ = 1 เมื่อโรงเรียนขาดไฟฟ้า น้ำประปา หรืออินเทอร์เน็ตที่มี

คุณภาพอย่างน้อย 1 ด้าน (5) ร้อยละนักเรียนยากจนสูง (poor family) ดัชนี = 1 เมื่อมีนักเรียนจากครอบครัวยากจนเกิน 50% (ใช้เฉพาะโรงเรียนประถมศึกษาเนื่องจากข้อจำกัดข้อมูล) (6) ความเป็นนักเรียนอายุน้อย (younger student) ดัชนี = 1 เมื่อนักเรียนอยู่ระดับชั้น ป.6

ในการวิเคราะห์ ผู้วิจัยใช้โมเดลการวิเคราะห์การถดถอยแบบ Elastic Net เพื่อคัดเลือกตัวแปรอิสระภายใต้ปัญหาความสัมพันธ์กันสูงระหว่างตัวแปรทำนาย (multicollinearity) โดยทำการปรับแต่งไฮเปอร์พารามิเตอร์ของโมเดลด้วยเทคนิค 5-fold cross-validation ที่ทำซ้ำ 3 ครั้งจำแนกตามเขตพื้นที่การศึกษาจำนวน 243 เขตพื้นที่ จากนั้นคัดเลือกตัวแปรทำนายที่มีค่า Variable Importance $\geq 10\%$ ไปวิเคราะห์ด้วยการถดถอยเชิงเส้นปกติ (OLS) อีกครั้งหนึ่ง เพื่อให้ตีความเชิงอัตราการเปลี่ยนแปลงของดัชนี CMER ได้ชัดเจน โดยผลค่าสัมประสิทธิ์จาก OLS จากแต่ละเขตพื้นที่การศึกษาได้ถูกนำมาวิเคราะห์เพื่อใช้ประโยชน์ต่อไป

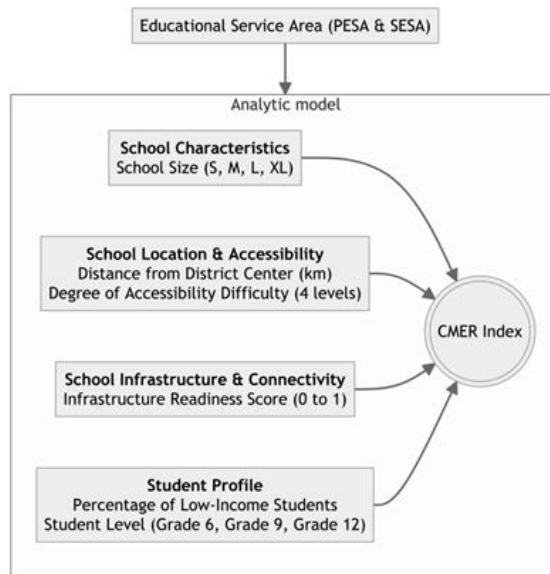
การเลือกใช้ Elastic Net มีเหตุผลเฉพาะจากลักษณะข้อมูลในงานวิจัยนี้ จากตัวแปรทางการศึกษาทั้ง 6 ตัว มีแนวโน้มสัมพันธ์เชิงเส้นแบบพหุกันสูงโดยโครงสร้าง (multicollinearity) เนื่องจากธรรมชาติของโรงเรียนในประเทศไทยที่มีคุณลักษณะคล้ายกันมักมีแนวโน้มที่จะมีสภาพเชิงบริบทและภูมิหลังที่ใกล้เคียงกันด้วยการวิเคราะห์การถดถอยแบบริดจ์ (ridge regression) สามารถจัดการปัญหานี้ได้ดีผ่าน L2 penalty อย่างไรก็ตาม L2 penalty ยังขาดกลไกคัดตัวแปรที่ไม่เกี่ยวข้องออก ซึ่งการวิเคราะห์การถดถอยแบบลาสโซ (lasso regression) สามารถแก้ไขได้ผ่าน L1 penalty ที่ช่วยให้โมเดลกระชับและตีความได้ง่าย อย่างไรก็ตามภายใต้ความสัมพันธ์เชิงเส้นแบบพหุ การถดถอยแบบลาสโซมีแนวโน้มเลือกเพียงตัวแปรเดียวในกลุ่มที่สัมพันธ์กันและกำหนดให้ตัวแปรที่เหลือมีค่าสัมประสิทธิ์เป็นศูนย์ ซึ่งไม่สอดคล้องกับวัตถุประสงค์ของงานวิจัยนี้ที่ต้องการระบุว่าปัจจัยใดบ้างขับเคลื่อนความเสี่ยงในแต่ละพื้นที่ การสูญเสียตัวแปรที่มีนัยสำคัญเชิงเนื้อหาย่อมกระทบต่อความสมบูรณ์ของสารสนเทศเชิงลึกที่ได้การวิเคราะห์ การถดถอยแบบ Elastic Net จึงเป็นตัวเลือกที่เหมาะสมที่สุดในบริบทนี้ เนื่องจากรวมจุดเด่นของทั้งสองวิธีเข้าด้วยกัน กล่าวคือรักษาตัวแปรในกลุ่มที่สัมพันธ์กันไว้ในโมเดลพร้อมกันผ่าน L2 penalty ขณะที่ยังคงกลไกการคัดตัวแปรที่ไม่จำเป็นออกของ Lasso ไว้ได้

ผลลัพธ์ที่ได้ถูกนำเสนอในรูปแบบเมทริกซ์ความเสี่ยงและความสำคัญ (Risk-Importance Matrix: RIM) เพื่อให้สามารถจำแนกเขตพื้นที่ตามระดับความเสี่ยงและความสำคัญเชิงบริบท และประเมินว่าปัจจัยทางการศึกษาใดบ้าง ที่มีความสัมพันธ์กับระดับความรุนแรงหรือความเสี่ยงทางการศึกษาที่เชื่อมโยงกับสภาพภูมิอากาศของแต่ละเขตพื้นที่ การนำเสนอ RIM นี้ประยุกต์แนวคิดจาก Risk-Impact Matrix และ Importance-Performance Matrix (Landell, 2016; Rivadeneira Moreira, 2025) ประกอบด้วย (1) คำนวณค่า Relative Importance (RI) จากอัตราส่วนของค่าสัมบูรณ์ Variable Importance (VIP) ต่อผลรวม VIP ทั้งหมดในพื้นที่ และแปลงเป็นร้อยละ และ (2) จำแนกพื้นที่ด้วยแผนภาพ RIM โดยใช้ค่า RI (แกน X) และค่ามัธยฐานของดัชนี CMER (แกน Y) เพื่อจัดพื้นที่ออกเป็น 4 กลุ่ม ได้แก่ กลุ่มเป้าหมายเชิงกลยุทธ์ กลุ่มเร่งพัฒนา กลุ่มเฝ้าระวัง และกลุ่มความเสี่ยงต่ำ

(1) **กลุ่มเป้าหมายเชิงกลยุทธ์ (strategic focus)** คือกลุ่มเขตพื้นที่การศึกษาที่มีมัธยฐาน CMER $\geq .60$ และ RI $\geq 40\%$ หรือมี RI เด่นแม้ CMER ต่ำกว่า .60 ขอบเขตของกลุ่มจำแนกด้วยเส้นโค้งจาก (RI=40%, CMER=1.00) ถึง (RI=100%, CMER=.50) เพื่อครอบคลุมพื้นที่ที่มีความเสี่ยงสัมพันธ์กับปัจจัยเด่นอย่างชัดเจน

(2) **กลุ่มเร่งพัฒนา (development priority)** เป็นกลุ่มของเขตพื้นที่การศึกษาที่มีมัธยฐาน CMER ระหว่าง .20-.60 และ RI $\geq 20\%$ โดยใช้เส้นทแยงจาก (RI=0%, CMER=.60) ถึง (RI=100%, CMER=.20) เพื่อระบุพื้นที่

ที่เสี่ยงปานกลางและมีปัจจัยเด่นที่เหมาะสมกับการแทรกแซงเชิงป้องกัน (3) กลุ่มเฝ้าระวัง (monitoring) คือเขตพื้นที่ที่ตกอยู่บริเวณใต้เส้นทแยงมุมแต่ยังมีค่ามัธยฐานของ CMER $> .20$ บ่งชี้ว่ามีความเสี่ยงต่ำถึงปานกลางแต่ยังไม่พบปัจจัยเด่น และ (4) กลุ่มความเสี่ยงต่ำ (low-risk) คือเขตพื้นที่ที่ตกอยู่ในบริเวณพื้นที่ใต้เส้นโค้งจาก (RI=0%, CMER=.20) ถึง (RI=50%, CMER=0.00) แสดงถึงพื้นที่ที่มีทั้งความเสี่ยงและความสำคัญต่ำ



ภาพ 2 กรอบการวิเคราะห์ความสัมพันธ์ระหว่างปัจจัยทางการศึกษากับดัชนี CMER

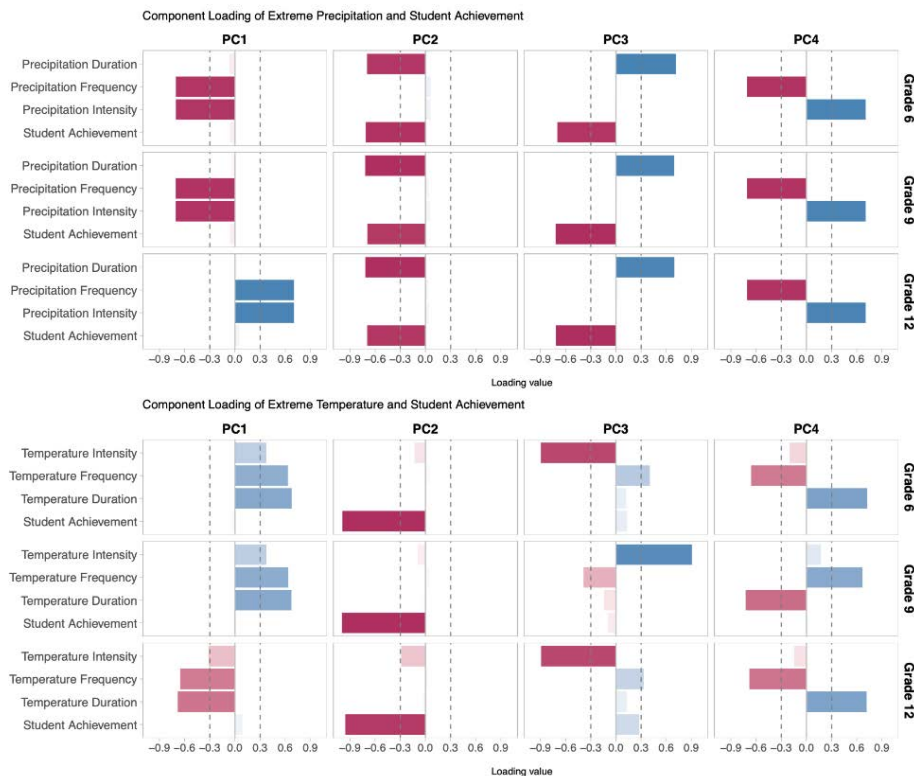
ผลการวิจัย

1. ผลการพัฒนาดัชนี CMER

การพัฒนาดัชนี CMER ใช้การวิเคราะห์องค์ประกอบหลัก (PCA) ที่แยกเป็น 2 ชุด ได้แก่ ปรากฏการณ์ฝนตกสุดซัวและอุณหภูมิสุดซัวกับผลสัมฤทธิ์ทางการเรียน โดยวิเคราะห์แยกตามระดับชั้น ป.6 (Grade 6) ม.3 (Grade 9) และ ม.6 (Grade 12) ผลการวิเคราะห์พบว่า องค์ประกอบหลักที่ 1 (PC1) ของแต่ละการวิเคราะห์สามารถอธิบายความแปรปรวนได้มากที่สุดประมาณ 40% ขณะที่ PC2 และ PC3 อธิบายได้ประมาณ 25% แต่ละองค์ประกอบและ PC4 สามารถอธิบายความแปรปรวนได้ประมาณ 7-8% อย่างไรก็ตามในการวิเคราะห์นี้เกณฑ์การเลือกองค์ประกอบหลักมุ่งค้นหาองค์ประกอบหลักที่ตัวแปรภูมิอากาศสุดซัวและผลสัมฤทธิ์ทางการเรียนมีน้ำหนักองค์ประกอบหลักในทิศทางตรงข้ามอย่างมีนัยสำคัญ ซึ่งสะท้อนการเกิดร่วมกันเชิงโครงสร้างที่มีความหมายเป็นความเสี่ยงทางการศึกษาที่เชื่อมโยงกับสภาพภูมิอากาศสุดซัว ทั้งนี้ ความเสี่ยงดังกล่าวโดยธรรมชาติเป็นปรากฏการณ์ที่เกิดขึ้นเฉพาะบางพื้นที่และบางบริบท ไม่ใช่ปรากฏการณ์ที่ครอบคลุมระบบการศึกษาทั้งหมด ค่าความแปรปรวนที่ PC3 อธิบายได้ประมาณ 25% จึงสะท้อนว่าการเกิดร่วมกันเชิงโครงสร้างระหว่างการเผชิญภัยภูมิอากาศและการด้อยผลสัมฤทธิ์ทางการเรียนมีอยู่จริงในระบบการศึกษา แต่อาจเกิดขึ้นเพียงบางส่วนและกระจายอยู่ในบางพื้นที่และบางบริบท ซึ่งสอดคล้องกับลักษณะของความเสี่ยงเชิงพื้นที่ที่งานวิจัยนี้มุ่งตรวจจับ

แผนภูมิแท่งในภาพ 3 แสดงค่าน้ำหนักองค์ประกอบ (component loadings) จากการวิเคราะห์ PCA จำแนกตามประเภทตัวแปร ได้แก่ ตัวชี้วัดความชื้น ความถี่ และระยะเวลาของปรากฏการณ์ภูมิอากาศสุดซัว (ฝนตก

และอุณหภูมิ) และผลสัมฤทธิ์ทางการเรียน โดยแกน X แสดงค่าน้ำหนักในช่วง -1.00 ถึง 1.00 ซึ่งสะท้อนทิศทางและขนาดของความสัมพันธ์กับองค์ประกอบหลัก



ภาพ 3 น้ำหนักองค์ประกอบหลักของปรากฏการณ์ภูมิอากาศสุดขั้วกับผลสัมฤทธิ์ทางการเรียนของนักเรียน

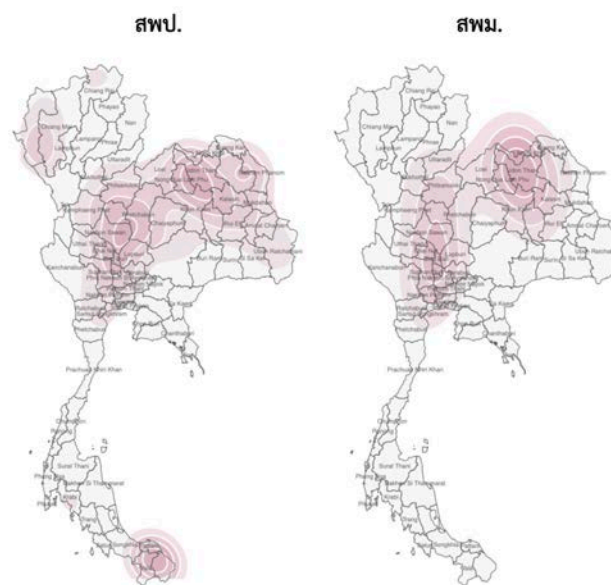
จากภาพ 3 พบว่า ในชุดข้อมูลฝนตกสุดขั้วและผลสัมฤทธิ์ทางการเรียน (แผนภูมิชุดบน) องค์ประกอบหลักที่ 3 (PC3) มีรูปแบบความสัมพันธ์ที่โดดเด่น กล่าวคือ ระยะเวลาของปรากฏการณ์ฝนตกสุดขั้ว (Precipitation Duration) มีค่าน้ำหนักเป็นบวก ในขณะที่ผลสัมฤทธิ์ทางการเรียนมีค่าน้ำหนักลบ และทั้งสองมีค่าสัมบูรณ์ของน้ำหนักมากกว่า 0.3 (Hair et al., 2019; Jolliffe & Cadima, 2016) สะท้อนความสัมพันธ์เชิงผกผันระหว่างตัวแปรทั้งสองในองค์ประกอบนี้ แสดงให้เห็นว่า PC3 บ่งชี้ถึงความเสี่ยงทางการศึกษาที่มีความเชื่อมโยงกับระยะเวลาของปรากฏการณ์ฝนตกสุดขั้ว ในทางตรงกันข้าม ชุดข้อมูลอุณหภูมิสุดขั้วไม่พบองค์ประกอบที่สะท้อนความสัมพันธ์เชิงโครงสร้างระหว่างตัวแปรทั้งสองกลุ่มอย่างชัดเจน ผลการวิเคราะห์นี้ชี้ว่า PC3 ที่ใช้ชุดข้อมูลฝนตกสุดขั้วและผลสัมฤทธิ์ทางการเรียน เป็นองค์ประกอบเดียวที่เชื่อมโยงสภาพภูมิอากาศกับผลกระทบทางการศึกษาในเชิงความเสี่ยงได้อย่างชัดเจน จึงถูกใช้เป็นฐานในการสร้างดัชนี CMER โดยนำคะแนน PC3 มาแปลงเป็นคะแนนมาตรฐานแบบ CDF ในช่วง 0–1 เพื่อให้การตีความและกำหนดเกณฑ์ทำได้อย่างเป็นปรนัย

ทั้งนี้แม้ว่า PC1 และ PC2 จะเป็นองค์ประกอบที่อธิบายความแปรปรวนของสภาพภูมิอากาศสุดขั้วได้มากกว่า PC3 แต่เมื่อพิจารณารูปแบบของน้ำหนักองค์ประกอบพบว่า น้ำหนักองค์ประกอบของผลสัมฤทธิ์ทางการเรียนมีค่าใกล้ศูนย์ทั้งหมด ซึ่งสะท้อนว่าความแปรปรวนของสภาพภูมิอากาศสุดขั้วในองค์ประกอบทั้งสองนั้นเกิดขึ้นอย่างไม่สัมพันธ์กับผลสัมฤทธิ์ทางการเรียน ไม่สามารถใช้เป็นฐานของดัชนีความเสี่ยงทางการศึกษาได้

จากการเลือกองค์ประกอบหลัก PC3 เพื่อสร้างดัชนี CMER ผู้วิจัยกำหนดเกณฑ์การแปลความหมายของค่าดัชนีให้เป็นปรนัยด้วยการวิเคราะห์ ROC โดยใช้คะแนน O-NET เป็นตัวแปรอ้างอิงในการคำนวณค่าความไวและ

ค่าความจำเพาะ เนื่องจากผลสัมฤทธิ์ทางการเรียนเป็นตัวชี้วัดในมิติด้านความเปราะบางทางการศึกษาในกรอบแนวคิดของงานวิจัยนี้ การใช้ตัวแปรเดียวกันเป็นเกณฑ์อ้างอิงจึงสอดคล้องกับหลักการกำหนดจุดตัดเชิงเกณฑ์ (criterion-referenced cut-off) ผลการวิเคราะห์ทำให้สามารถจำแนกโรงเรียนเป็น 3 ระดับ ได้แก่ (1) ความเสี่ยงต่ำ ($CMER < 0.60$) (2) ความเสี่ยงปานกลาง ($0.60 \leq CMER \leq 0.80$) โดยจุดตัด 0.60 ให้ Sensitivity = .80, Specificity = .79 และ AUC = .88 และ (3) ความเสี่ยงสูง ($CMER > 0.80$) โดยจุดตัด 0.80 ให้ Sensitivity = 1.00, Specificity = .95 และ AUC = .98 เกณฑ์ดังกล่าวช่วยให้การประเมินและติดตามความเสี่ยงดำเนินได้อย่างเป็นระบบ และใช้ประกอบการจัดลำดับความสำคัญเชิงนโยบายเพื่อจัดสรรทรัพยากรลดผลกระทบทางการศึกษาจากการเปลี่ยนแปลงสภาพภูมิอากาศได้อย่างมีประสิทธิภาพ

การวิเคราะห์ตามเกณฑ์จุดตัดข้างต้น พบว่า โรงเรียนส่วนใหญ่ (68.24%, 17,954 แห่ง) อยู่ในกลุ่มความเสี่ยงต่ำ โดยมีค่าเฉลี่ยดัชนี .422 (SD = .135, ช่วง .010–.598) รองลงมาเป็นกลุ่มความเสี่ยงปานกลาง (28.81%, 7,580 แห่ง) มีค่าเฉลี่ย .683 (SD = .051) และกลุ่มความเสี่ยงสูง (2.95%, 777 แห่ง) มีค่าเฉลี่ย .843 (SD = .048) สะท้อนการกระจายของความเสี่ยงทางการศึกษาที่เชื่อมโยงกับสภาพภูมิอากาศอย่างชัดเจนระหว่างกลุ่ม



ภาพ 4 แผนที่เส้นชั้นความหนาแน่น (contour map) ของพื้นที่ที่มีดัชนี CMER อยู่ในระดับสูง

ภาพ 4 แสดงแผนที่เส้นชั้นความหนาแน่นของโรงเรียนที่มีความเสี่ยง CMER สูง ($CMER > 0.80$) ทั่วประเทศไทย จำแนกตามเขตพื้นที่การศึกษาประถมศึกษา (สพป.) และมัธยมศึกษา (สพม.) ผลการแสดงผลพบว่าโรงเรียนกลุ่มเสี่ยงสูงมีการกระจายตัวครอบคลุมหลายภูมิภาค โดยมีบริเวณที่หนาแน่นเด่นชัดในภาคใต้ ภาคตะวันออกเฉยงเหนือ ภาคกลาง และบางส่วนของภาคเหนือ ขณะเดียวกัน ลักษณะการกระจายตัวแตกต่างกันตามระดับเขตพื้นที่การศึกษา โดยโรงเรียนในเขตพื้นที่ สพป. มีการกระจายตัวกว้างกว่าในหลายภูมิภาค ในขณะที่โรงเรียนในเขตพื้นที่ สพม. มีแนวโน้มกระจุกตัวมากกว่า โดยเฉพาะในภาคกลางและภาคตะวันออกเฉยงเหนือ

2. ผลการวิเคราะห์ความสัมพันธ์ระหว่างดัชนี CMER กับ ปัจจัยด้านภูมิศาสตร์ ลักษณะของโรงเรียน และลักษณะของนักเรียน

การวิเคราะห์ในวัตถุประสงค์ที่ 2 มุ่งระบุปัจจัยทางการศึกษาที่ส่งผลต่อดัชนี CMER ในแต่ละภูมิภาค ผลการวิเคราะห์ พบว่า ในกลุ่มเขตพื้นที่ สพป. ปัจจัยที่เด่นตามเกณฑ์นี้มี 2 ปัจจัย คือ การเป็นโรงเรียนขนาดเล็ก (small school) และ ความยากลำบากในการเข้าถึงโรงเรียน (hard to reach) โดยปัจจัย small school ถูกเลือกใน 181 เขต (98.90%) มีค่าเฉลี่ย .14 (SD = .12) ในจำนวนนี้พบว่ามีเขตพื้นที่จำนวน 30.4% ที่สัมพันธ์กับการถดถอยดังกล่าวมีนัยสำคัญทางสถิติ ($p < .05$) ส่วนปัจจัย Hard to Reach ถูกเลือกใน 33 เขตพื้นที่ (18.03%) แต่มีค่าเฉลี่ยของสัมประสิทธิ์การถดถอยเท่ากับ .11 (SD = .08) โดยในจำนวนนี้มีเขตพื้นที่จำนวน 77.9% ที่พบมีนัยสำคัญทางสถิติ ($p < .05$) สำหรับเขตพื้นที่ สพม. พบปัจจัยเด่น 3 ปัจจัย ได้แก่ Small School ซึ่งถูกเลือกในทุกเขตพื้นที่ (61 เขตพื้นที่) มีค่าสัมประสิทธิ์เฉลี่ย .18 (SD = .10) ในจำนวนนี้มีเขตพื้นที่จำนวน 77.0% ที่สัมพันธ์กับการถดถอยมีนัยสำคัญทางสถิติ ($p < .05$) รองลงมาคือ Hard to Reach ซึ่งถูกเลือกใน 5 เขตพื้นที่ (ร้อยละ 8.20) มีค่าสัมประสิทธิ์เฉลี่ย .13 (SD = .07) อย่างไรก็ตามในจำนวนนี้ไม่มีเขตพื้นที่ใดที่ค่าสัมประสิทธิ์ดังกล่าวมีนัยสำคัญทางสถิติ และปัจจัย School Distance ซึ่งถูกเลือกใน 61 เขตพื้นที่ มีค่าสัมประสิทธิ์เฉลี่ย .11 (SD = .12) และในจำนวนนี้มีเขตพื้นที่ร้อยละ 19.7 ที่มีค่าสัมประสิทธิ์มีนัยสำคัญทางสถิติ ($p < .05$)

ภาพ 5 แสดงผลการวิเคราะห์ด้วย Risk-Importance Matrix (RIM) โดยบูรณาการคะแนนความเสี่ยง CMER กับค่าความสำคัญสัมพัทธ์ (Relative Importance: RI) ของปัจจัยทางการศึกษา เพื่อจำแนกเขตพื้นที่ตามระดับความเสี่ยงและบทบาทของปัจจัยขับเคลื่อนหลัก โดยภาพรวมสามารถจัดกลุ่มพื้นที่ออกเป็น 4 กลุ่มสำคัญ ได้แก่ (1) กลุ่มมุ่งเน้นเชิงกลยุทธ์ ซึ่งมีค่า CMER สูงร่วมกับค่า RI สูง แสดงพื้นที่ที่มีความเสี่ยงทางการศึกษาเชื่อมโยงกับปัจจัยเชิงโครงสร้างอย่างเด่นชัด (2) กลุ่มเร่งพัฒนา ที่มีค่า CMER สูงแต่ค่า RI ยังไม่เด่นชัด สะท้อนพื้นที่ที่ควรพัฒนาและค้นหาปัจจัยเพิ่มเติม (3) กลุ่มเฝ้าระวัง ซึ่งมีค่า CMER ระดับปานกลางหรือค่า RI เริ่มสูงขึ้น แสดงถึงพื้นที่ที่ควรติดตามแนวโน้มความเสี่ยง และ (4) กลุ่มความเสี่ยงต่ำ ที่มีค่า CMER และ RI ต่ำ โดยในที่นี้นำเสนอผลแยกตามระดับเขตพื้นที่การศึกษา ได้แก่ สพป. และ สพม.

เมื่อพิจารณาในกลุ่มเขตพื้นที่ สพป. (ภาพ 5 ก.) พบว่าเขตพื้นที่ที่ถูกจัดให้อยู่ในกลุ่มเป้าหมายเชิงกลยุทธ์ (Strategic Focus) มีจำนวน 19 เขตพื้นที่ โดยมีค่า CMER อยู่ในช่วง 0.52–0.73 และมีค่า RI อยู่ในช่วง 43.88–100% ปัจจัยที่ปรากฏเด่นในกลุ่มนี้ได้แก่ ความยากจนของครัวเรือน ระยะทาง/การเข้าถึงโรงเรียน ความเป็นโรงเรียนขนาดเล็ก และการขาดแคลนโครงสร้างพื้นฐาน สะท้อนความเปราะบางที่เกี่ยวข้องกับข้อจำกัดด้านเศรษฐกิจ การเข้าถึงบริการทางการศึกษา และทรัพยากรของโรงเรียน

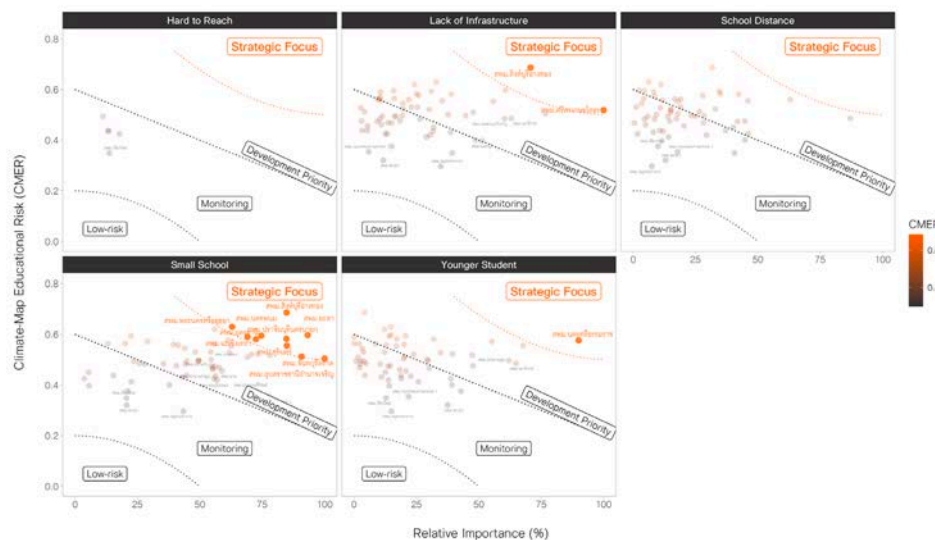
สำหรับเขตพื้นที่ สพม. (ภาพ 5 ข.) พบว่าเขตพื้นที่กลุ่มเป้าหมายเชิงกลยุทธ์มีจำนวน 12 เขตพื้นที่ โดยมีค่า CMER อยู่ในช่วง 0.50–0.69 และปัจจัยที่พบเด่นชัดในหลายเขตพื้นที่คือความเป็นโรงเรียนขนาดเล็ก ขณะที่บางเขตพื้นที่ยังปรากฏปัจจัยด้านการขาดแคลนโครงสร้างพื้นฐานและลักษณะของผู้เรียนเป็นปัจจัยเด่นร่วมกัน ผลลัพธ์ดังกล่าวชี้ว่าการจำแนกด้วย RIM ตามระดับการศึกษาช่วยเพิ่มความละเอียดเชิงพื้นที่สำหรับการกำหนดเป้าหมายการสนับสนุนเชิงนโยบาย

ตาราง 2 สถิติพื้นฐานของค่าสัมประสิทธิ์การถดถอยรายเขตพื้นที่ของปัจจัยทางการศึกษาที่ส่งผลต่อ CMER

Feature	N	M	SD	Q1	Med	Q3	Max	%Sig.
เขตพื้นที่การศึกษาประถมศึกษา (PESA)								
Small School	181	.14	.12	.04	.12	.20	.49	30.4%
Hard to Reach	33	.11	.08	.06	.08	.19	.30	39.4%
School Distance	183	.09	.08	.04	.07	.13	.42	27.9%
Lack of Infrastructure	176	.06	.05	.02	.05	.08	.23	15.9%
Young Student	183	.04	.04	.01	.02	.05	.37	16.4%
Poor Family	176	.03	.03	.01	.02	.04	.31	14.2%
เขตพื้นที่ศึกษามัธยมศึกษา (SESA)								
Small School	61	.18	.10	.12	.18	.22	.61	77.0%
Hard to Reach	5	.13	.07	.09	.11	.11	.26	0.0%
School Distance	61	.11	.12	.04	.09	.12	.61	19.7%
Lack of Infrastructure	61	.07	.06	.03	.07	.10	.27	18.0%
Young Student	61	.06	.06	.02	.04	.08	.32	24.6%

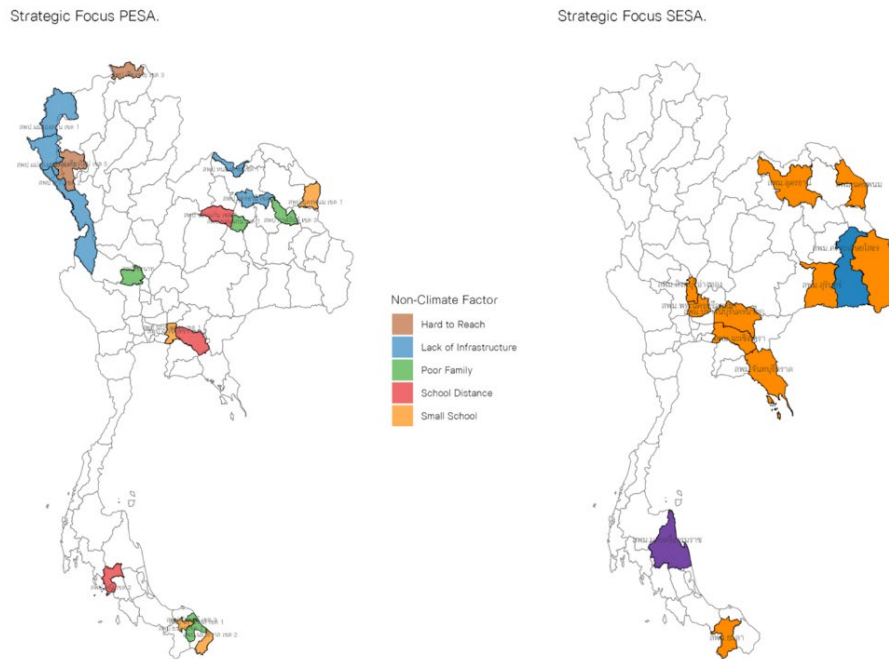


ก. เขตพื้นที่การศึกษาประถมศึกษา



ข. เขตพื้นที่ศึกษามัธยมศึกษา

ภาพ 5 แสดงเมทริกซ์ความเสี่ยงและความสำคัญ (RIM) ของโรงเรียนระดับเขตพื้นที่การศึกษา



ภาพ 6 ปัจจัยทางการศึกษาที่มีความสำคัญมากที่สุดในแต่ละเขตพื้นที่การศึกษาในกลุ่มมุ่งเน้นเชิงกลยุทธ์

ภาพ 6 แสดงแผนที่การกระจายความเสี่ยงเชิงพื้นที่ของเขตพื้นที่การศึกษา สพป. และ สพม. ที่ถูกจัดให้อยู่ในกลุ่มมุ่งเน้นเชิงกลยุทธ์ พร้อมทั้งระบุปัจจัยทางการศึกษาที่เป็นปัจจัยเด่นของแต่ละเขตพื้นที่ กล่าวคือ ปัจจัยที่มีค่า RI สูงสุดเมื่อเทียบกับปัจจัยอื่นในเขตพื้นที่เดียวกัน จากรูปสีที่แตกต่างกันใช้แทนประเภทของปัจจัยเด่นที่เชื่อมโยงกับความเสี่ยง CMER ในแต่ละบริบทพื้นที่ เมื่อพิจารณาแผนที่การกระจายความเสี่ยงของเขตพื้นที่ สพป. พบว่ามีความหลากหลายของปัจจัยเด่น ได้แก่ ความยากลำบากในการเข้าถึงพื้นที่ การขาดแคลนโครงสร้างพื้นฐาน ความยากจนของครัวเรือน และความเป็นโรงเรียนขนาดเล็ก บ่งชี้ว่าปัจจัยที่สัมพันธ์กับความเปราะบางของโรงเรียนประถมศึกษาเกิดจากข้อจำกัดที่มีความหลากหลายและแตกต่างกันไปตามบริบทของพื้นที่ ขณะที่ในระดับมัธยมศึกษา ปัจจัยเด่นในเขตพื้นที่ส่วนใหญ่คือ ความเป็นโรงเรียนขนาดเล็ก ซึ่งกระจุกตัวในหลายพื้นที่ แสดงให้เห็นว่าความเสี่ยงทางการศึกษาในระดับมัศึกษามีความเชื่อมโยงกับข้อจำกัดเชิงโครงสร้างของโรงเรียนอย่างเด่นชัด

ผลการวิเคราะห์จากภาพ 5 และ 6 ยังบ่งชี้ให้เห็นอย่างชัดเจนว่า แม้หลายเขตพื้นที่จะถูกจัดอยู่ในกลุ่มที่ต้องให้ความสำคัญเชิงนโยบายเหมือนกัน แต่ปัจจัยเด่นทางการศึกษาที่มีเชื่อมโยงกับความเสี่ยง CMER และควรได้รับการจัดการกลับมีความแตกต่างกันไปตามบริบทพื้นที่ ดังนั้น ผลลัพธ์ดังกล่าวจึงควรถูกนำไปใช้ประกอบการออกแบบมาตรการสนับสนุนเชิงพื้นที่ที่มีความจำเพาะ มุ่งแก้ปัญหาได้ตรงจุด และสอดคล้องกับความเปราะบางของแต่ละเขตพื้นที่มากขึ้น

อภิปรายผลการวิจัย

1. จากผลการสร้างดัชนี CMER ในงานวิจัยนี้พบว่า ดัชนี CMER สร้างขึ้นจากความสัมพันธ์ระหว่างระยะเวลาของปรากฏการณ์ฝนตกสุดชั้วกับผลสัมฤทธิ์ทางการเรียนของนักเรียน โดยไม่พบความสัมพันธ์เชิงโครงสร้างจากตัวชี้วัดของสภาพภูมิอากาศสุดชั้วในด้านอื่น ผลการวิจัยนี้สนับสนุนสมมุติฐานว่าที่ผ่านมา

ผลกระทบของภูมิอากาศต่อการศึกษาของประเทศไทยมีแนวโน้มเป็นไปแบบสะสมต่อเนื่องมากกว่าผลกระทบแบบฉับพลันรุนแรง โดยผลกระทบแบบสะสมนี้เป็นปัจจัยที่ลดทอนเวลาเรียนและโมเมนตัมการเรียนรู้ของนักเรียนอย่างค่อยเป็นค่อยไป หลักการเดียวกันนี้ปรากฏในงานวิจัยด้านเศรษฐศาสตร์การศึกษา (Gottfried, 2014; Liu et al., 2021) ที่พบว่าการขาดเรียนของนักเรียนแบบกระจายส่งผลเสียต่อผลสัมฤทธิ์ทางการเรียนมากกว่าการขาดเรียนแบบกระจุกในจำนวนวันที่เท่ากัน เพราะทำให้ขาดความต่อเนื่องและเพิ่มภาระในการติดตามเนื้อหา ซึ่งสภาพปรากฏการณ์ที่ฝนตกหนักยึดเยื้อในบางพื้นที่ของประเทศไทยก็มีแนวโน้มสร้างเงื่อนไขการขาดเรียนในลักษณะเดียวกัน

นอกจากนี้การไม่พบความสัมพันธ์เชิงโครงสร้างระหว่างอุณหภูมิสุดขีดกับผลสัมฤทธิ์ทางการเรียน แม้จะมีหลักฐานจากงานวิจัยในประเทศว่าความร้อนส่งผลต่อสมรรถนะทางปัญญา (Park et al., 2020; Zivin et al., 2014) สะท้อนถึงความเฉพาะผลกระทบที่มีต่อผลลัพธ์ กล่าวคือ งานวิจัยที่พบผลของอุณหภูมิส่วนใหญ่ใช้ตัวชี้วัดด้านสมรรถนะทางปัญญาในขณะนั้น เช่น คะแนนสอบในวันที่อากาศร้อน แต่คะแนน O-NET เป็นผลลัพธ์เชิงสะสมที่สะท้อนการเรียนรู้ตลอดช่วงชั้นของการศึกษา ซึ่งนักเรียนส่วนใหญ่ของประเทศไทยมีความคุ้นเคยกับสภาพอากาศร้อนอยู่แล้ว นอกจากนี้ปัญหาสภาพอากาศร้อนสามารถแก้ไขหรือปรับตัวได้ง่ายกว่า เช่น การเลื่อนเวลาเรียน การใช้พัดลมหรือเครื่องปรับอากาศ หรือการเรียนชดเชยในช่วงที่อากาศเย็นลงในทางกลับกันสภาพฝนตกสุดขั้วอย่างยึดเยื้อสร้างผลกระทบที่ชดเชยได้ยากกว่า

2. ดัชนีความเสี่ยงด้านภูมิอากาศที่ใช้กันในระดับสากล เช่น Children's Climate Risk Index (CCRI) ของ UNICEF (2021) และ WorldRiskIndex (Bündnis Entwicklung Hilft, 2024) มีข้อจำกัดสำคัญ 2 ประการ สำหรับการนำมาประยุกต์ใช้ทางด้านการศึกษา ประการแรก ดัชนีดังกล่าววัดการเผชิญภัย (exposure) และความเปราะบาง (vulnerability) ในระดับประเทศหรือจังหวัด ซึ่งขาดรายละเอียดเชิงพื้นที่สำหรับการตัดสินใจระดับโรงเรียนหรือเขตพื้นที่การศึกษา ประการที่สอง ไม่ได้ผนวกรวมตัวชี้วัดด้านผลลัพธ์ทางการศึกษาเข้าไปในกรอบแนวคิดของดัชนี ทำให้ไม่สามารถระบุได้ว่าความเสี่ยงที่พบในดัชนีดังกล่าวจะสะท้อนเป็นความเสี่ยงทางการศึกษาได้จริง หรืออย่างน้อยเพียงใด

ดัชนี CMER ที่พัฒนาขึ้นในงานวิจัยนี้ได้เสนอแนวทางที่แตกต่าง โดยใช้การสร้างดัชนีที่ใช้ข้อมูลเป็นฐาน (data-driven approach) ผ่านการวิเคราะห์องค์ประกอบหลักแทนการกำหนดน้ำหนักหรือโครงสร้างทางทฤษฎีเอาไว้ล่วงหน้า วิธีการนี้มีข้อดีคือช่วยให้ค้นพบรูปแบบความสัมพันธ์ที่ไม่ถูกอธิบายได้ด้วยทฤษฎีเดิม ดังที่พบว่า ความยาวนานของฝนตกสุดขั้วมีความสำคัญมากกว่าความรุนแรงของฝนตกสุดขั้ว ซึ่งขัดกับสัญชาตญาณทั่วไปที่มักให้ความสำคัญกับความรุนแรงของภัยพิบัติ อย่างไรก็ตาม ข้อจำกัดของวิธีการนี้คือผลลัพธ์ที่ได้อาจมีความจำเพาะต่อบริบทและต้องระมัดระวังหากจะนำไปใช้หรือขยายผลต่อในบริบทอื่น

3. ผลการวิจัยพบว่าความเป็นโรงเรียนขนาดเล็ก เป็นปัจจัยที่มีความสัมพันธ์กับ CMER อย่างมีนัยสำคัญในหลายเขตพื้นที่ ทั้งใน สพป. และ สพม. ข้อค้นพบนี้สามารถอธิบายได้ผ่านกรอบแนวคิดความสามารถในการฟื้นตัวขององค์กร (Sutcliffe & Vogus, 2003; Duchek, 2020) ที่ระบุว่าความสามารถในการรับมือและฟื้นตัวจากวิกฤตขององค์กรขึ้นอยู่กับทรัพยากรสำรอง ความหลากหลายของบุคลากร และความยืดหยุ่นในการปรับโครงสร้าง ซึ่งโรงเรียนที่มีขนาดเล็ก (รวมทั้งขนาดปานกลาง) ของประเทศไทยมีแนวโน้มพบข้อจำกัดในทุกมิติดังกล่าว กล่าวคือมีงบประมาณ ทรัพยากร และบุคลากรที่จำกัด ขาดครูสำรอง

เมื่อครูไม่สามารถเดินทางมาสอนได้ และมีความยืดหยุ่นต่ำในการปรับตารางเรียนหรือรูปแบบการจัดการเรียนรู้ ดังนั้นเมื่อพบกับปัญหาทางสภาพภูมิอากาศ โรงเรียนขนาดเล็กจึงมีแนวโน้มที่ได้รับผลกระทบรุนแรงกว่า และฟื้นตัวช้ากว่า ทั้งนี้ขนาดโรงเรียนไม่ใช่ปัจจัยเสี่ยงในตัวเอง แต่ทำหน้าที่เป็นตัวขยายความเปราะบางที่ขยายขนาดความเสี่ยง CMER ในรุนแรงยิ่งขึ้น

ข้อค้นพบนี้มีความสำคัญเชิงนโยบาย กล่าวคือ การลดความเสี่ยงทางการศึกษาจากสภาพภูมิอากาศ ไม่จำเป็นต้องมุ่งเน้นการลดการเผชิญภัยโดยตรง แต่สามารถดำเนินการผ่านการเสริมสร้างความสามารถในการปรับตัว (adaptive capacity) ของโรงเรียนที่เปราะบาง เช่น การสร้างเครือข่ายความร่วมมือระหว่างโรงเรียน การพัฒนาระบบการเรียนรู้ทางไกลสำรองที่มีคุณภาพและเท่าเทียม หรือการจัดสรรทรัพยากรแบบยืดหยุ่นที่สามารถนำมาใช้ได้ภาวะฉุกเฉิน

4. งานวิจัยนี้เสนอเมทริกซ์ความเสี่ยง-ความสำคัญ (RIM) เป็นเครื่องมือสนับสนุนการตัดสินใจที่ผสมผสานข้อมูล 2 มิติ ได้แก่ ระดับความเสี่ยง (มัธยฐานของ CMER) และความสำคัญเชิงสัมพัทธ์ของปัจจัยทางการศึกษา แนวคิดนี้ประยุกต์จาก Importance-Performance Analysis (Martilla & James, 1977) และ Risk-Impact Matrix (International Organization for Standardization [ISO], 2018) แต่ปรับให้เหมาะกับบริบทการจัดการความเสี่ยงทางการศึกษา คุณค่าของ RIM คือช่วยสนับสนุนผู้กำหนดนโยบาย หลีกเลียงการตัดสินใจแบบ one-size-fit-all ที่ผิดพลาด กล่าวคือแม้หลายพื้นที่จะมีความเสี่ยงที่อยู่ในระดับใกล้เคียงกัน แต่ปัจจัยทางการศึกษาที่ขยายความเสี่ยงดังกล่าวมีความแตกต่างกัน ดังที่เห็นจากผลการวิจัยว่าในกลุ่มเขตพื้นที่ทางการศึกษาที่ต้องให้ความสำคัญเชิงกลยุทธ์ บางพื้นที่มีปัจจัยด้านความยากจนเป็นปัจจัยเด่น บางพื้นที่เป็นความห่างไกล และบางพื้นที่เป็นโครงสร้างพื้นฐาน ดังนั้นการใช้มาตรการเดียวกันทุกพื้นที่จึงไม่เพียงไม่มีประสิทธิภาพ แต่อาจสิ้นเปลืองทรัพยากรโดยไม่จำเป็น

ข้อเสนอแนะ

1. ดัชนี CMER ที่พัฒนาขึ้นสามารถบ่งชี้โรงเรียนหรือพื้นที่ที่มีความเสี่ยงจากสภาพภูมิอากาศได้แม่นยำกว่าวิธีเดิมที่อิงเพียงข้อมูลน้ำท่วมหรือปริมาณฝน เนื่องจากผสมผสานข้อมูลภูมิอากาศเข้ากับผลสัมฤทธิ์ทางการเรียนและบริบทของโรงเรียนโดยตรง หน่วยงานระดับนโยบาย เช่น สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (สพฐ.) หรือองค์กรระหว่างประเทศที่ทำงานด้านการศึกษาและการเปลี่ยนแปลงสภาพภูมิอากาศ สามารถนำดัชนี CMER ร่วมกับเมทริกซ์ความเสี่ยง-ความสำคัญ (RIM) ไปใช้ประโยชน์ใน 2 ระดับ ดังนี้ (1) **ระดับเขตพื้นที่การศึกษา** ใช้ RIM จำแนกเขตพื้นที่ตามระดับความเสี่ยงเชิงโครงสร้าง (จากค่า CMER) และความสำคัญเชิงสัมพัทธ์ของปัจจัยทางการศึกษา (จากค่า RI) เพื่อจัดลำดับความเร่งด่วนและระบุปัจจัยหลักที่ขับเคลื่อนความเสี่ยงในแต่ละพื้นที่ ไม่ว่าจะเป็นโรงเรียนขนาดเล็ก ความยากจน โครงสร้างพื้นฐาน หรือความห่างไกล RIM จึงทำหน้าที่เป็นกลไกชี้เป้าเชิงพื้นที่ที่ช่วยกำหนดทิศทางนโยบายให้สอดคล้องกับปัญหาของแต่ละพื้นที่อย่างเฉพาะเจาะจง และ (2) **ระดับโรงเรียนภายในเขตพื้นที่** หลังจากระบุเขตพื้นที่เป้าหมายแล้ว ให้จัดลำดับโรงเรียนด้วยค่า CMER ระดับโรงเรียน และจำแนกเป็นกลุ่มตามระดับความเสี่ยง (เช่น กลุ่มปานกลาง 0.60–0.79 และกลุ่มสูง ≥ 0.80) เพื่อกำหนดระดับการสนับสนุนที่แตกต่างกัน โดยออกแบบมาตรการให้สอดคล้องกับปัจจัยเสี่ยงเฉพาะของเขตพื้นที่นั้น ๆ ตามที่ RIM ระบุ

2. จากผลการวิจัยที่พบว่าลักษณะของปัจจัยทางการศึกษาที่มีผลต่อความเสี่ยง CMER มีความแตกต่างกันในแต่ละเขตพื้นที่ การออกแบบแนวทางหรือมาตรฐานสนับสนุนจึงควรยึดหลักความจำเพาะของบริบท โดยจำแนกเป็น 2 แนวทางได้แก่ (1) **พื้นที่ที่มีปัจจัยเสี่ยงเด่นเพียงปัจจัยเดียว** ควรใช้มาตรการมุ่งเป้า เพื่อแก้ไขปัญหาอย่างตรงจุด เช่น กรณีที่ความเป็นโรงเรียนขนาดเล็กเป็นปัจจัยเด่น ควรเน้นการจัดสรรทรัพยากรที่จำเป็นทั้งด้านงบประมาณ เทคโนโลยีดิจิทัล และบุคลากร ควบคู่กับการสร้างเครือข่ายความร่วมมือกับโรงเรียนแม่ข่ายขนาดใหญ่เพื่อแบ่งปันทรัพยากรและพัฒนาวิชาชีพครู (2) **พื้นที่ซึ่งเผชิญกับความเสี่ยงซับซ้อนจากหลายปัจจัยร่วมกัน** จำเป็นต้องวิเคราะห์เพื่อระบุปัจจัยขับเคลื่อนหลักและปัจจัยเสริมก่อนออกแบบแนวทางหรือมาตรฐาน เช่น ในพื้นที่ที่ความยากจนเป็นปัจจัยขับเคลื่อนหลัก (ดังกรณี สปป. นราธิวาส เขต 1) มาตรการหลักควรมุ่งเน้นการสนับสนุนทางเศรษฐกิจแก่นักเรียนและครอบครัวเป็นลำดับแรก โดยมีมาตรการด้านโครงสร้างพื้นฐานเป็นส่วนเสริม ในทางกลับกัน พื้นที่ที่ความห่างไกลเป็นปัจจัยหลัก (ดังกรณี สปป. เชียงใหม่ เขต 5) ควรให้ความสำคัญกับการพัฒนาโครงสร้างพื้นฐานและระบบคมนาคมเป็นลำดับแรก

3. การวิจัยในอนาคตควรขยายขอบเขตของตัวชี้วัดในด้านผลลัพธ์ทางการศึกษาให้ครอบคลุมมากกว่าคะแนน O-NET เช่น อัตราการขาดเรียน อัตราการออกกลางคัน และสุขภาพทางกายและทางจิตของนักเรียน การขยายตัวชี้วัดในลักษณะนี้จะช่วยให้เข้าใจกลไกเชิงลึกกว่าฝนตกยึดเหี่ยวส่งผลต่อการเรียนรู้ผ่านช่องทางใดบ้าง อันจะนำไปสู่การออกแบบมาตรการป้องกันและเยียวยาที่ตรงจุดยิ่งขึ้น นอกจากนี้การวิจัยในอนาคตควรพิจารณาออกแบบการศึกษาในเชิงตามยาว (longitudinal study) โดยติดตามการเปลี่ยนแปลงของดัชนี CMER และปัจจัยทางการศึกษาในแต่ละเขตพื้นที่ข้ามช่วงเวลา เพื่อตรวจจับว่าความเสี่ยงทางการศึกษาที่เชื่อมโยงกับสภาพภูมิอากาศมีแนวโน้มสะสมหรือรุนแรงขึ้นตามการเปลี่ยนแปลงของสภาพภูมิอากาศในระยะยาวหรือไม่ และมาตรการที่ดำเนินการในพื้นที่เป้าหมายสามารถลดค่าดัชนีได้จริงในเชิงประจักษ์หรือไม่ ซึ่งจะช่วยให้ CMER เปลี่ยนจากเครื่องมือวินิจฉัยเป็นเครื่องมือติดตามผลเชิงนโยบายได้

กิตติกรรมประกาศ

คณะผู้วิจัยขอขอบพระคุณองค์การยูนิเซฟ (UNICEF) ประเทศไทย ที่ได้ให้การสนับสนุนทุนวิจัยและอำนวยความสะดวกในการดำเนินงานจนกระทั่งงานวิจัยฉบับนี้สำเร็จลุล่วงด้วยดี และขอขอบพระคุณ ศูนย์วิจัยการเปลี่ยนแปลงสภาพภูมิอากาศและพลังงานหมุนเวียนระดับภูมิภาค มหาวิทยาลัยรามคำแหง (Ramkhamhaeng University Center of Regional Climate Change and Renewable Energy: RU-CORE) ที่อนุเคราะห์ชุดข้อมูลสภาพภูมิอากาศซึ่งมีความสำคัญยิ่งต่อการพัฒนาดัชนี CMER และการวิเคราะห์ความเสี่ยงทางการศึกษาภายใต้บริบทการเปลี่ยนแปลงสภาพภูมิอากาศในงานวิจัยฉบับนี้

References

- Adler, C., Wester, P., Bhatt, I., Huggel, C., Insarov, G. E., Morecroft, M. D., Muccione, V., & Prakash, A. (2022). *Cross-Chapter Paper 5: Mountains*. In H. O. Pörtner, D. C. Roberts, M. Tignor, E. S. Poloczanska, K. Mintenbeck, A. Alegria, M. Craig, S. Langsdorf, S. Löschke, V. Möller, A. Okem, & B. Rama (Eds.), *Climate change 2022: Impacts, adaptation and vulnerability. Contribution of Working Group II to the Sixth Assessment Report of the Intergovernmental Panel on Climate Change* (pp.2273-2318). Cambridge University Press. <https://doi.org/10.1017/9781009325844.022>

- Auer-Frege, I., Befus, E., Bih, K. B., Fawwad, M., Ide, T., Kiefer, M., Kienzl, P., Klüppel, D., Korte, L., Kortendieck, M., Moloinyane, S., Peters, K., Radtke, K., Schneider, S., & Weller, D. (2024). *WorldRiskReport 2024: Focus on education and vulnerability to climate risks*. Bündnis Entwicklung Hilft & Institute for International Law of Peace and Armed Conflict. <https://weltrisikobericht.de/worldriskreport/>
- Auer-Frege, I., Bliecke, V., Bradshaw, S., Dijkzeul, D., Funke, C., Kienzl, P., Küsters, K., Masuch, L., Nasreen, M., Radtke, K., Schneider, S., Thielbörger, P., Weller, D., Wiegers, O., Yaman, D. K., & Zennig, K. (2023). *WorldRiskReport 2023: Focus on diversity*. Bündnis Entwicklung Hilft & Institute for International Law of Peace and Armed Conflict. <https://weltrisikobericht.de/worldriskreport/>
- Bündnis Entwicklung Hilft & Ruhr-Universität Bochum, Institut für Internationales Friedens- und Konfliktrecht (IFHV). (2024). *WorldRiskReport 2024: Multiple crises*. https://securesustain.org/wp-content/uploads/2024/10/WorldRiskReport_2024.pdf
- Duchek, S. (2020). Organizational resilience: A capability-based conceptualization. *Business Research*, 13(1), 215–246. <https://doi.org/10.1007/s40685-019-0085-7>
- Eckstein, D., Künzel, V., & Schäfer, L. (2021). *Global climate risk index 2021: Who suffers most from extreme weather events? Weather-related loss events in 2019 and 2000-2019*. Germanwatch. <https://germanwatch.org/en/cr>
- Google. (2024). Google Maps Platform documentation. <https://developers.google.com/maps>
- Gottfried, M. A. (2014). Chronic absenteeism and its effects on students' academic and socioemotional outcomes. *Journal of Education for Students Placed at Risk*, 19(1–2), 53–75. <https://doi.org/10.1080/10824669.2014.962696>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2021). *Multivariate data analysis* (9th ed.). Cengage.
- International Organization for Standardization. (2018). *ISO 31000:2018 Risk management—Guidelines*. Author.
- Intergovernmental Panel on Climate Change. (2023). *Climate change 2023: Synthesis report. Summary for policymakers*. IPCC. <https://doi.org/10.59327/IPCC/AR6-9789291691647.001>
- Jolliffe, I. T., & Cadima, J. (2016). Principal component analysis: A review and recent developments. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 374(2065), Article 20150202. <https://doi.org/10.1098/rsta.2015.0202>
- Landell, H. (2016). *The risk matrix as a tool for risk analysis: How to apply existing theories in practice in order to overcome its limitations* [Master's thesis, University of Gävle]. DiVA Portal. <https://www.diva-portal.org/smash/get/diva2:944825/FULLTEXT01.pdf>

- Liu, J., Lee, M., & Gershenson, S. (2021). The short- and long-run impacts of secondary school absences. *Journal of Public Economics*, 199, Article 104441. <https://doi.org/10.1016/j.jpubeco.2021.104441>
- Martilla, J. A., & James, J. C. (1977). Importance-Performance Analysis. *Journal of Marketing*, 41(1), 77–79. <https://doi.org/10.2307/1250495>
- Ngo-Duc, T., Tangang, F. T., Santisirisomboon, J., Cruz, F., Trinh-Tuan, L., Nguyen-Xuan, T., & Aldrian, E. (2016). Performance evaluation of RegCM4 in simulating extreme rainfall and temperature indices over the CORDEX-Southeast Asia region. *International Journal of Climatology*, 37(3), 163-182. <https://doi.org/10.1002/joc.4803>
- Park, J., Goodman, J., Hurwitz, M., & Smith, J. (2020). Heat and learning. *American Economic Journal: Economic Policy*, 12(2), 306-339. <https://doi.org/10.1257/pol.20180612>
- Rivadeneira Moreira, J. C. (2025). Risk management in education: A systematic literature review of the last five years. *Revista de Gestión Social y Ambiental*, 19(1), e010977. <https://doi.org/10.24857/rgsa.v19n1-127>
- Sutcliffe, K. M., & Vogus, T. J. (2003). *Organizing for resilience*. In K. S. Cameron, J. E. Dutton, & R. E. Quinn (Eds.), *Positive organizational scholarship: Foundations of a new discipline* (pp. 94–110). Berrett-Koehler Publishers.
- Thampanishvong, K., Santisirisomboon, J., Sumano, B., Channuntapipat, C., Ngaosri, K., Witoonchart, P., Kaewbuadee, N., & Pinkaew, L. (2022). *Impact assessment of climate change and environmental degradation on children in Thailand*. Thailand Development Research Institute. <https://www.unicef.org/thailand/media/10361/file/Climate%20Change%20Impact%20Assessment%20-%20Full%20report.pdf>
- Unal, I. (2017). Defining an optimal cut-point value in ROC analysis: An alternative approach. *Computational and Mathematical Methods in Medicine*, 2017, Article 3762651. <https://doi.org/10.1155/2017/3762651>
- UNICEF. (2021). *The climate crisis is a child-rights crisis: Introducing the Children's Climate Risk Index*. <https://www.unicef.org/reports/climate-crisis-child-rights-crisis>
- United Nations Office for Disaster Risk Reduction. (2015). Sendai Framework for Disaster Risk Reduction 2015–2030. UNDRR.
- Zivin, J. G., & Neidell, M. (2014). Temperature and the allocation of time: Implications for climate change. *Journal of Labor Economics*, 32(1), 1-26. <https://doi.org/10.1086/671766>

Translate Thai Reference

- Srisuttiyakorn, S. (2019). Educational inequality and its factors: Multilevel analysis integrated with median-based class of generalized entropy inequality index. *Journal of Research Methodology*, 32(3), 342–379. (in Thai)

การประเมินอภิมานรายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย

Meta-evaluation of Final Evaluation Reports by Save the Children Australia

วรวรรณ จิรธนาพิวัฒน์¹ ศิวะพร ภูพันธ์^{2*} และ อุทัยทิพย์ เจียวิวรรณ์กุล³

Worrawan Jirathanapiwat¹ Siwaporn Phupan^{2*} and Uthaithip Jiawiwatkul³

(Received: March 12, 2026; Revised: May 11, 2026; Accepted: May 12, 2026)

บทคัดย่อ

การวิจัยครั้งนี้ มีวัตถุประสงค์เพื่อประเมินคุณภาพรายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ตามหลักการประเมินอภิมาน (meta-evaluation) การวิจัยครั้งนี้ เป็นการวิจัยเอกสาร โดยทำการประเมินคุณภาพรายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ระหว่าง ปี พ.ศ. 2558–2566 จำนวน 14 ฉบับ เครื่องมือที่ใช้ในการเก็บรวบรวมข้อมูล คือ แบบตรวจสอบรายการสำหรับการประเมินอภิมาน วิเคราะห์ข้อมูลโดยใช้โปรแกรม Jamovi สถิติที่ใช้ ได้แก่ ความถี่ ร้อยละ คะแนนรวม ผลการวิจัยพบว่า รายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย จำนวน 14 ฉบับ มีการระบุวัตถุประสงค์ ตัวชี้วัดไว้อย่างชัดเจน ใช้ทั้งข้อมูลเชิงปริมาณและเชิงคุณภาพ และนำผลการวิเคราะห์ข้อมูลทั้งสองประเภทมาตีความสรุปผลร่วมกัน เมื่อพิจารณาคุณภาพรายงานการประเมินฯ จำนวน 4 มาตรฐาน พบว่า มาตรฐานด้านการใช้ประโยชน์ รายงานการประเมินฯ ส่วนใหญ่มีคุณภาพอยู่ในระดับดีมาก คิดเป็นร้อยละ 71.43 มาตรฐานด้านความเป็นไปได้ รายงานการประเมินฯ มีคุณภาพอยู่ในระดับดีมาก คิดเป็นร้อยละ 64.29 มาตรฐานด้านความเหมาะสม รายงานการประเมินฯ มีคุณภาพเท่ากัน คือ ระดับดีมาก และยอดเยี่ยม คิดเป็นร้อยละ 35.71 และมาตรฐานด้านความถูกต้องแม่นยำ รายงานการประเมินฯ ส่วนใหญ่มีคุณภาพอยู่ในระดับดีมาก คิดเป็นร้อยละ 57.14

คำสำคัญ การประเมินอภิมาน การประเมินอภิมานแบบสรุปรวม การประเมินแบบผสมวิธี รายงานการประเมินระยะสิ้นสุดโครงการ

¹ นักศึกษาปริญญาโท สาขาวิชาการประเมินโครงการทางสังคม คณะสังคมศาสตร์และมนุษยศาสตร์ มหาวิทยาลัยมหิดล

² ผู้ช่วยศาสตราจารย์ ภาควิชาศึกษาศาสตร์ คณะสังคมศาสตร์และมนุษยศาสตร์ มหาวิทยาลัยมหิดล

³ รองศาสตราจารย์ ภาควิชาศึกษาศาสตร์ คณะสังคมศาสตร์และมนุษยศาสตร์ มหาวิทยาลัยมหิดล

¹ Master's degree student, Social Programs Evaluation Program, Faculty of Social Sciences and Humanities, Mahidol University

² Assistant Professor, Department of Education, Faculty of Social Sciences and Humanities, Mahidol University

³ Associate Professor, Department of Education, Faculty of Social Sciences and Humanities, Mahidol University

* Corresponding Author E-mail: siwaporn.phu@mahidol.ac.th

Abstract

The purpose of this study was to evaluate the quality of final evaluation reports supported by Save the Children Australia using the principles of meta-evaluation. Documentary research was employed with the intent to assess 14 final evaluation reports, published during 2015-2023. The research instrument was a meta-evaluation checklist. The data were analyzed with frequency, percentage, and total scores using Jamovi. The findings revealed that all 14 final evaluation reports clearly specified their objectives and indicators, utilized both quantitative and qualitative data, and integrated findings from both types of data in the interpretation and conclusion process. Regarding the quality assessment based on four evaluation standards, the results showed that, under the utility standard, most evaluation reports were evaluated at a very good level (71.43%). Under the feasibility standard, 64.29% of the reports were evaluated at a very good level. For the propriety standard, the reports were equally evaluated at very good and excellent levels (35.71% each). Under the accuracy standard, most reports were evaluated at a very good level (57.14%).

Keywords: meta-evaluation, summative meta-evaluation, mixed-methods evaluation, final evaluation reports

บทนำ

การดำเนินงานโครงการพัฒนาต่าง ๆ เพื่อช่วยเหลือเด็ก พัฒนาคุณภาพชีวิตเด็ก การคุ้มครองเด็ก การปกป้องสิทธิเด็ก นับเป็นประเด็นที่ทุกประเทศทั่วโลก ต่างให้ความสำคัญ สำหรับการดำเนินการสำคัญที่สะท้อนถึงการให้ความสำคัญกับการช่วยเหลือเด็ก คือ การดำเนินโครงการพัฒนาของเครือข่ายองค์กรช่วยเหลือเด็ก ระดับโลก ซึ่งเป็นเครือข่ายการทำงานที่มีเป้าหมายเพื่อพัฒนาคุณภาพชีวิตเด็กในภูมิภาคต่าง ๆ ทั่วโลก สำหรับองค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย นับเป็นส่วนหนึ่งของเครือข่ายองค์กรช่วยเหลือเด็ก ระดับโลก ซึ่งดำเนินการโครงการที่มุ่งยกระดับคุณภาพชีวิตของเด็กทั้งในประเทศออสเตรเลียและในประเทศต่าง ๆ สำหรับการดำเนินโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ครอบคลุมด้านต่าง ๆ ได้แก่ ด้านการศึกษา สุขภาพและโภชนาการ การคุ้มครองเด็ก การบรรเทาทุกข์ฉุกเฉินและการตอบสนองด้านมนุษยธรรม อย่างไรก็ตาม องค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ยังมีบทบาทสำคัญในการพัฒนาระหว่างประเทศ ซึ่งมีการกึ่งให้การสนับสนุนการดำเนินโครงการในประเทศกำลังพัฒนา ประมาณ 20 ประเทศ ครอบคลุมภูมิภาคเอเชีย แปซิฟิก แอฟริกา และตะวันออกกลาง (Save the Children Australia, 2024)

ในการดำเนินโครงการต่าง ๆ ขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลียนั้น ทุกโครงการได้กำหนดวัตถุประสงค์ ผลลัพธ์ที่คาดหวัง กลุ่มเป้าหมาย พื้นที่ดำเนินงาน และกิจกรรมอย่างเป็นรูปธรรม ซึ่งจะมีการติดตามและประเมินผลเมื่อสิ้นสุดการดำเนินการเพื่อสรุปความสำเร็จ ประสิทธิภาพ และบทเรียนที่ได้รับ

รวมถึง การจัดทำรายงานการประเมินระยะสิ้นสุดโครงการ อย่างไรก็ตาม เมื่อโครงการต่าง ๆ ได้ดำเนินโครงการเสร็จสิ้นลง ผู้รับผิดชอบโครงการจะต้องจัดทำรายงานการประเมินระยะสิ้นสุดโครงการ เพื่อส่งมอบให้กับหน่วยงานเจ้าของโครงการและผู้ให้ทุนหลัก ซึ่งในกรณีส่วนใหญ่ก็คือ องค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย เพื่อใช้เป็นเอกสารประกอบการสรุปผลการดำเนินงานตามแผนงานประจำปีขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย รวมถึง รายงานต่อผู้ให้ทุนระดับสากล เช่น รัฐบาลออสเตรเลีย องค์กรเอกชน ผู้สนับสนุน กองทุนระหว่างประเทศอื่น ๆ เป็นต้น นอกจากนี้ รายงานการประเมินระยะสิ้นสุดโครงการ จะถูกส่งมอบให้กับหน่วยงานภาคีในประเทศที่มีส่วนร่วมในการดำเนินโครงการ เช่น องค์กรภาคประชาสังคม มหาวิทยาลัย หน่วยงานรัฐในพื้นที่ หรือบริษัทที่ปรึกษา เป็นต้น เพื่อนำผลการประเมินไปใช้สำหรับการปรับปรุงการดำเนินงานในระยะต่อไป และในบางกรณี รายงานการประเมินระยะสิ้นสุดโครงการ สามารถนำไปใช้เพื่อเสนอต่อองค์กรช่วยเหลือเด็กในประเทศที่ดำเนินงาน เพื่อใช้ในการสรุปบทเรียน (lessons learned) และกำหนดยุทธศาสตร์ใหม่ขององค์กรต่าง ๆ ในภูมิภาค

ข้อมูลสารสนเทศจากรายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย จึงอาจจะถูกนำไปใช้ในการพัฒนานโยบายต่าง ๆ ที่มาจากหลักฐาน (evidence-driven policy) ดังนั้น รายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ควรจะเป็นรายงานที่มีคุณภาพ เพื่อเป็นต้นทุนข้อมูลสารสนเทศที่มีคุณภาพ นำไปสู่การพัฒนาคุณภาพชีวิตของเด็กทั้งในประเทศออสเตรเลีย และประเทศต่าง ๆ ในแถบภูมิภาค กล่าวคือ หากมีวิธีการสรุปรายงานผลการประเมินระยะสิ้นสุดโครงการต่าง ๆ เหล่านี้ ต่อองค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย และหน่วยงานที่เกี่ยวข้อง นับเป็นสิ่งที่สำคัญต่อการขับเคลื่อนองค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย และยังเป็นองค์ความรู้ที่ทำให้ทราบถึงคุณภาพรายงานผลการประเมินระยะสิ้นสุดโครงการ

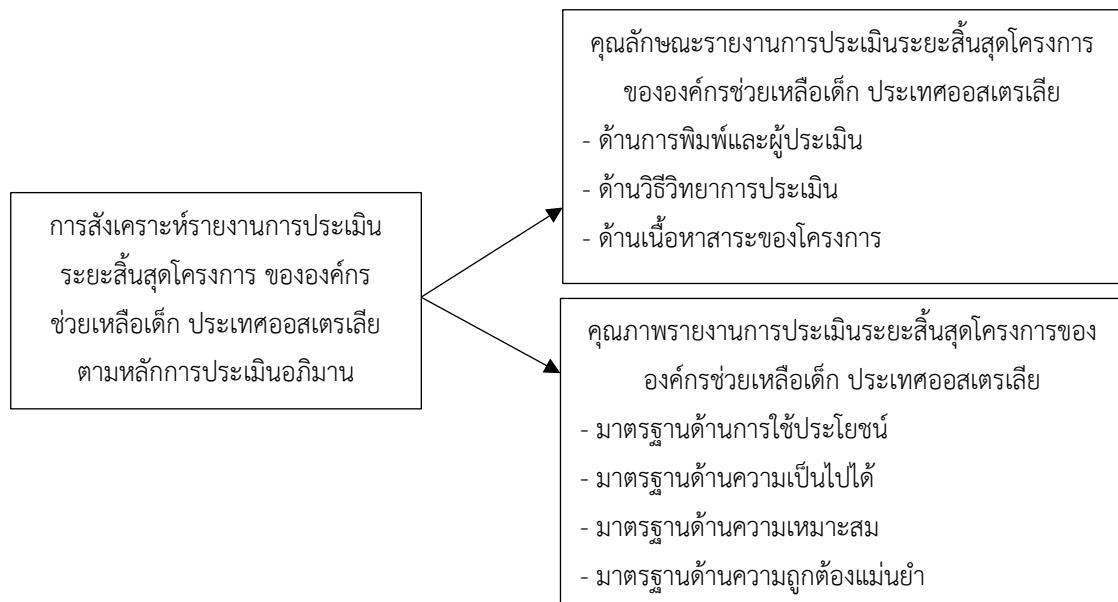
ด้วยเหตุนี้ คณะผู้วิจัยจึงมีความสนใจในการประยุกต์ใช้หลักการประเมินอภิมาน (meta-evaluation) ในลักษณะการประเมินอภิมานแบบสรุปรวม (summative meta evaluation) ซึ่งมีจุดมุ่งหมายเพื่อประเมินคุณภาพของการดำเนินการประเมินผลโครงการ ในระยะสิ้นสุดโครงการ ดังนั้น ในการศึกษาครั้งนี้ คณะผู้วิจัยจึงทำการประเมินคุณภาพรายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ซึ่งผลการวิจัยที่ได้จากการศึกษาครั้งนี้ จะเป็นข้อค้นพบเกี่ยวกับคุณภาพรายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ซึ่งจะเป็นประโยชน์ต่อการบริหารโครงการต่าง ๆ ขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย รวมไปถึง การดำเนินโครงการของหน่วยงานหรือองค์กรต่าง ๆ ที่ได้รับการสนับสนุนจากองค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย เพื่อยกระดับคุณภาพการดำเนินโครงการต่าง ๆ นำไปสู่การพัฒนาคุณภาพชีวิตของเด็กทั้งในประเทศออสเตรเลีย และประเทศต่าง ๆ ในแถบภูมิภาคต่อไป

ความมุ่งหมายของการวิจัย

เพื่อประเมินคุณภาพรายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ตามหลักการประเมินอภิมาน

กรอบแนวคิดในการวิจัย

จากการศึกษาเอกสาร และแนวคิดทฤษฎีที่เกี่ยวข้องกับการวิจัยครั้งนี้ คณะผู้วิจัยได้พัฒนากกรอบแนวคิดในการวิจัย เพื่อสังเคราะห์รายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ตามหลักการประเมินอภิมาน ซึ่งแบ่งออกเป็น 2 ส่วนหลัก ๆ คือ *ส่วนแรก* เป็นการวิเคราะห์คุณลักษณะของรายงานการประเมินฯ ประกอบด้วย ด้านการพิมพ์และผู้ประเมิน ด้านวิธีวิทยาการประเมิน และด้านเนื้อหาสาระของโครงการ สำหรับ*ส่วนที่สอง* เป็นการประเมินคุณภาพรายงานการประเมินฯ ในส่วนนี้ คณะผู้วิจัยได้นำแนวคิดมาตรฐานการประเมินโครงการ (The Program Evaluation Standards) ของ Stufflebeam (1999) ซึ่งแบ่งออกเป็น 4 มาตรฐาน คือ มาตรฐานด้านการใช้ประโยชน์ มาตรฐานด้านความเป็นไปได้ มาตรฐานด้านความเหมาะสม และมาตรฐานด้านความถูกต้องแม่นยำ มาใช้เป็นกรอบในการประเมินคุณภาพรายงานการประเมินฯ ซึ่งกรอบแนวคิดในการวิจัยมีรายละเอียดดังภาพ 1



ภาพ 1 กรอบแนวคิดในการวิจัย

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ เป็นการวิจัยเอกสาร (documentary research) มีจุดมุ่งหมายเพื่อสังเคราะห์รายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ตามหลักการประเมินอภิมาน (meta-evaluation) คณะผู้วิจัยได้เลือกใช้การประเมินอภิมานแบบสรุปรวม (summative meta-evaluation) โดย Stufflebeam (2001) ได้อธิบายว่ามีจุดมุ่งหมายเพื่อสรุปตัดสินคุณภาพโดยรวมของการประเมินว่ามีความเที่ยงตรง น่าเชื่อถือ และเป็นประโยชน์ต่อผู้ใช้ผลการประเมิน ซึ่งมีรายละเอียดวิธีดำเนินการวิจัย ดังนี้

1. ประชากรและตัวอย่างวิจัย

ประชากร คือ รายงานการประเมินระยะสิ้นสุดโครงการ (final evaluation report) ฉบับสมบูรณ์ ขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ที่ได้รับการรับรองผลการประเมิน และเผยแพร่ผ่านทางเว็บไซต์ขององค์กร ในช่วงปี พ.ศ. 2558–2566 (ค.ศ. 2015–2023) จำนวน 23 เล่ม

ตัวอย่างวิจัย คือ รายงานการประเมินระยะสิ้นสุดโครงการฉบับสมบูรณ์ ขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ที่ได้รับการรับรองผลการประเมินและเผยแพร่ผ่านทางเว็บไซต์ขององค์กร ในช่วงปี พ.ศ. 2558–2566 (ค.ศ. 2015–2023) จำนวน 14 เรื่อง ซึ่งผ่านเกณฑ์การคัดเลือกที่คณะผู้วิจัยได้กำหนดขึ้น คือ เป็นรายงานการประเมินระยะสิ้นสุดโครงการ ที่ได้ดำเนินการประเมินโดยเลือกใช้การประเมินแบบผสมวิธี (mixed methods evaluation) กล่าวคือ ใช้ทั้งวิธีการเชิงปริมาณและเชิงคุณภาพ ในการเก็บรวบรวมข้อมูลและวิเคราะห์ข้อมูล เพื่อตอบคำถามการประเมิน

2. เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัย คือ แบบบันทึกเอกสาร (documentary protocol) ซึ่งแบ่งออกเป็น 2 ตอน ดังนี้

ตอนที่ 1 แบบบันทึกข้อมูลเบื้องต้นเกี่ยวกับลักษณะของรายงาน/โครงการ มีลักษณะเป็นแบบตรวจสอบรายการและแบบเติมข้อความ มีจำนวน 17 ข้อ ซึ่งแบ่งออกเป็น 3 ส่วน คือ ด้านการพิมพ์และผู้ประเมิน (จำนวน 5 ข้อ) ด้านวิธีวิทยาการประเมิน (จำนวน 6 ข้อ) และด้านเนื้อหาของสาระของโครงการ (จำนวน 6 ข้อ)

ตอนที่ 2 แบบบันทึกข้อมูลคุณภาพรายงานการประเมินฯ ซึ่งคณะผู้วิจัยได้กำหนดเนื้อหาของการประเมิน ตามแนวคิดมาตรฐานการประเมินโครงการ (The Program Evaluation Standards) ของ Stufflebeam (1999) มีลักษณะการประเมินแบบรูบริก (rubric measurement) (ระดับ 1-4) ซึ่งแบ่งออกเป็น 4 มาตรฐาน มีจำนวน 30 มาตรฐานย่อย คือ (1) มาตรฐานด้านการใช้ประโยชน์ (utility) มีจำนวน 7 มาตรฐานย่อย (2) มาตรฐานด้านความเป็นไปได้ (feasibility) มีจำนวน 3 มาตรฐานย่อย (3) มาตรฐานด้านความเหมาะสม (propriety) มีจำนวน 8 มาตรฐานย่อย และ (4) มาตรฐานด้านความถูกต้องแม่นยำ (accuracy) มีจำนวน 12 มาตรฐานย่อย

ในการศึกษาครั้งนี้ คณะผู้วิจัยได้ดำเนินการตรวจสอบคุณภาพด้านความเที่ยง โดยนำแบบบันทึกเอกสารที่ได้พัฒนาขึ้น มาทดลองใช้ โดยนักวิจัย จำนวน 2 คน เพื่อตรวจสอบความเที่ยงระหว่างผู้ประเมิน (inter-rater reliability) ทั้งนี้ คณะผู้วิจัยได้สุ่มรายงานการประเมินฯ จำนวน 1 ฉบับ และกำหนดให้นักวิจัย จำนวน 2 คน ได้ทดลองบันทึกข้อมูลเบื้องต้นเกี่ยวกับลักษณะของรายงาน/โครงการ และประเมินคุณภาพรายงานการประเมินฯ ซึ่งเป็นข้อมูลประเภทไม่ต่อเนื่อง คือ 0 (ไม่ระบุ) และ 1 (ระบุ) ภายหลังจากนั้น จึงนำคะแนนการประเมินรายงานการประเมินฯ ของผู้ประเมินคนที่ 1 และผู้ประเมินคนที่ 2 มาวิเคราะห์ความเที่ยงระหว่างผู้ประเมิน ทั้ง 2 คน โดยใช้โปรแกรม Jamovi โดยเลือกวิธีของ Kappa สำหรับผลการวิเคราะห์ข้อมูล พบว่า ค่า Kappa เท่ากับ .96 ($p < .0001$) และค่าความสอดคล้อง เท่ากับ 98% สรุปได้ว่า การประเมินคุณภาพรายงานการประเมินฯ ของผู้ประเมินทั้ง 2 คน มีความใกล้เคียงกันสูงมาก

3. การเก็บรวบรวมข้อมูล

คณะผู้วิจัยดำเนินการเก็บรวบรวมข้อมูล ซึ่งมีขั้นตอน คือ (1) สืบค้นรายงานการประเมินระยะสิ้นสุดโครงการฉบับสมบูรณ์ ขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ที่ได้รับการรับรองผลการประเมิน และเผยแพร่ผ่านทางเว็บไซต์ขององค์กร ในช่วงปี พ.ศ. 2558–2566 (ค.ศ. 2015–2023) โดยคัดเลือกเฉพาะรายงานการประเมินฯ ที่สามารถเข้าถึงรายงานการประเมินฯ ฉบับเต็มได้ มีจำนวน 23 เรื่อง (2) อ่านและตรวจสอบเนื้อหา รายงานการประเมินฯ ว่ามีคุณสมบัติเป็นไปตามเกณฑ์ที่ได้กำหนดไว้ คือ เป็นรายงานการประเมินระยะสิ้นสุดโครงการ ที่ได้ดำเนินการประเมินโดยเลือกใช้การประเมินแบบผสมวิธี ทั้งนี้ มีรายงานการประเมินฯ จำนวน 14 เรื่อง ที่ผ่านเกณฑ์การคัดเลือก และ (3) อ่านรายงานการประเมินฯ โดยละเอียด และบันทึกข้อมูลเบื้องต้นของรายงาน/โครงการ รวมถึง ให้คะแนนคุณภาพรายงานการประเมินฯ ทั้ง 14 เรื่อง

4. การวิเคราะห์ข้อมูล

คณะผู้วิจัยดำเนินการวิเคราะห์ข้อมูลโดยใช้โปรแกรม Jamovi ซึ่งแบ่งการดำเนินการออกเป็น 2 ส่วน คือ (1) การวิเคราะห์ข้อมูลเบื้องต้นของรายงานการประเมินฯ ที่นำมาสังเคราะห์ เป็นการวิเคราะห์ค่าสถิติพื้นฐานของคุณลักษณะรายงานการประเมินฯ สถิติที่ใช้ ได้แก่ ค่าความถี่ ร้อยละ และ (2) การวิเคราะห์ข้อมูลคุณภาพรายงานการประเมินฯ เป็นการนำผลการประเมินคุณภาพ ในแต่ละรายการ และนำมาคำนวณเป็นคะแนนรวม และร้อยละ

ผลการวิจัย

1. ผลการวิเคราะห์ข้อมูลเบื้องต้นของรายงานการประเมินระยะสิ้นสุดโครงการ ขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย

1.1 ผลการวิเคราะห์ข้อมูลเบื้องต้นของรายงานการประเมินระยะสิ้นสุดโครงการ ด้านการพิมพ์รายงานการประเมินฯ และผู้ประเมิน พบว่า รายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ที่นำมาสังเคราะห์ทั้งหมด จำนวน 14 เรื่อง เป็นรายงานการประเมินระยะสิ้นสุดโครงการ ฉบับสมบูรณ์ ที่องค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ให้การรับรองรายงานผลการประเมินแล้ว และเผยแพร่รายงานการประเมินฯ ผ่านทางเว็บไซต์ขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ในระหว่างปี พ.ศ. 2558–2566 (ค.ศ. 2015–2023) เมื่อพิจารณาปีที่เผยแพร่รายงานการประเมินระยะสิ้นสุดโครงการ พบว่า รายงานการประเมินฯ ที่นำมาสังเคราะห์ ได้รับการเผยแพร่ผ่านทางเว็บไซต์ขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ปี พ.ศ. 2558 (ปี ค.ศ. 2015) จำนวน 3 เรื่อง (ร้อยละ 21.00) ปี พ.ศ. 2559 (ปี ค.ศ. 2016) จำนวน 2 เรื่อง (ร้อยละ 14.00) ปี พ.ศ. 2560 (ปี ค.ศ. 2017) จำนวน 3 เรื่อง (ร้อยละ 21.00) ปี พ.ศ. 2563 (ปี ค.ศ. 2020) จำนวน 1 เรื่อง (ร้อยละ 7.00) ปี พ.ศ. 2564 (ปี ค.ศ. 2021) จำนวน 3 เรื่อง (ร้อยละ 21.00) และปี พ.ศ. 2566 (ปี ค.ศ. 2023) จำนวน 2 เรื่อง (ร้อยละ 14.00) เมื่อพิจารณาภูมิภาคที่ดำเนินโครงการ พบว่า รายงานการประเมินระยะสิ้นสุดโครงการ ครอบคลุมพื้นที่ต่าง ๆ ได้แก่ ภูมิภาคแปซิฟิก จำนวน 5 เรื่อง (ร้อยละ 35.00) ภูมิภาคเอเชียตะวันออกเฉียงใต้ จำนวน 3 เรื่อง (ร้อยละ 21.00) ประเทศออสเตรเลีย จำนวน

2 เรื่อง (ร้อยละ 14.00) ภูมิภาคเอเชียกลาง จำนวน 1 เรื่อง (ร้อยละ 7.00) ภูมิภาคเอเชียใต้ จำนวน 1 เรื่อง (ร้อยละ 7.00) รวมถึง การดำเนินโครงการ ระดับโลก จำนวน 2 เรื่อง (ร้อยละ 14.00) เมื่อพิจารณา**ประเภทของผู้ประเมินโครงการ** พบว่า โครงการส่วนใหญ่ออกแบบให้มีผู้ประเมินภายนอก จำนวน 10 เรื่อง (ร้อยละ 71.00) ในขณะที่โครงการเลือกใช้ระบบผู้ประเมินภายใน มีจำนวน 4 เรื่อง (ร้อยละ 29.00) อย่างไรก็ตาม เมื่อพิจารณา **สังกัดของผู้ประเมินและแหล่งทุนของโครงการ** พบว่า มีสังกัดของผู้ประเมินและแหล่งทุนของโครงการที่หลากหลาย ได้แก่ การดำเนินการประเมินโดยองค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย จำนวน 3 เรื่อง (ร้อยละ 21.00) การดำเนินการประเมินโดยองค์กรประเมินเอกชนร่วมกับองค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย จำนวน 3 เรื่อง (ร้อยละ 21.00) การดำเนินการประเมินโดยผู้ประเมินอิสระร่วมกับองค์กรช่วยเหลือเด็ก จำนวน 3 เรื่อง (ร้อยละ 21.00) การดำเนินการประเมินโดยองค์กรประเมินเอกชนที่ได้รับทุนจากกระทรวงการต่างประเทศและการค้าออสเตรเลีย (Department of Foreign Affairs and Trade : DFAT) จำนวน 3 เรื่อง (ร้อยละ 21.00) และการดำเนินการประเมินซึ่งได้รับการสนับสนุนกระทรวงการต่างประเทศและการค้าออสเตรเลีย (Department of Foreign Affairs and Trade : DFAT) ร่วมกับองค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย จำนวน 2 เรื่อง (ร้อยละ 14.00)

1.2 ผลการวิเคราะห์ข้อมูลเบื้องต้นของรายงานการประเมินระยะสิ้นสุดโครงการ **ด้านวิธีวิทยาการประเมิน** พบว่า ในภาพรวมรายงานการประเมินฯ ส่วนใหญ่มีการออกแบบและนำเสนอวิธีวิทยาการประเมินอย่างเป็นระบบและครบถ้วนในหลายมิติ โดยรายงานการประเมินฯ ทั้ง 14 เรื่อง (ร้อยละ 100.00) ได้อธิบายรายละเอียด 3 ประเด็น คือ (1) การเลือกใช้การประเมินแบบผสมวิธี กล่าวคือ ใช้ทั้งวิธีการเชิงปริมาณและเชิงคุณภาพในการเก็บรวบรวมข้อมูลและการวิเคราะห์ข้อมูล เพื่อตอบคำถามการประเมิน (2) การระบุจำนวนกลุ่มตัวอย่างหรือผู้ให้ข้อมูลอย่างครบถ้วน และ (3) การอธิบายข้อจำกัดของการประเมินและการตรวจสอบความน่าเชื่อถือของผลการประเมิน ทั้งนี้ รายงานการประเมินฯ จำนวน 13 เรื่อง จากรายงานการประเมินฯ 14 เรื่อง (ร้อยละ 93.00) มีการอธิบายเทคนิคการเก็บข้อมูลและวิธีการคัดเลือกกลุ่มตัวอย่างอย่างชัดเจน เช่น การสัมภาษณ์เชิงลึก การสนทนากลุ่ม และการสำรวจ เป็นต้น นอกจากนี้ ยังพบว่า รายงานการประเมินฯ ส่วนใหญ่มีการระบุประเภทของการประเมินแบบผสมวิธี ไว้อย่างชัดเจน จำนวน 11 เรื่อง (ร้อยละ 79.00) รายละเอียดดังตาราง 1

1.3 ผลการวิเคราะห์ข้อมูลเบื้องต้นของรายงานการประเมินระยะสิ้นสุดโครงการ **ด้านเนื้อหาสาระของโครงการ** พบว่า มีจำนวน 3 ประเด็น ที่รายงานการประเมินฯ ทั้ง 14 เรื่อง (ร้อยละ 100.00) มีการอธิบายอย่างชัดเจน คือ (1) ระบุขอบเขตเนื้อหาของโครงการ ได้แก่ ด้านการศึกษา ด้านสุขภาพ ด้านการคุ้มครองเด็ก (2) กำหนดวัตถุประสงค์และผลลัพธ์ที่คาดหวังของโครงการ และ (3) นำเสนอข้อค้นพบและข้อเสนอแนะเชิงวิเคราะห์ ทั้งนี้ ในประเด็นเนื้อหาสาระของโครงการ ยังพบว่า รายงานการประเมินฯ ส่วนใหญ่ได้อธิบายรายละเอียดประเด็นอื่น ๆ คือ อธิบายตัวชี้วัดผลลัพธ์ (outcome indicators) และผลกระทบ (impact) จำนวน 12 เรื่อง (ร้อยละ 86.00) และระบุการนำผลการประเมินไปใช้ (การนำไปใช้เชิงนโยบาย การออกแบบโครงการต่อเนื่อง หรือการเผยแพร่สู่สาธารณะ) จำนวน 11 เรื่อง (ร้อยละ 79.00) นอกจากนี้ ยังพบว่า มีรายงานการประเมินฯ เพียง 8 เรื่อง (ร้อยละ 57.00) ที่สามารถตอบคำถามการประเมินตามกรอบ

OECD/DAC ครบทั้ง 5 มิติ ได้แก่ ความสอดคล้อง (relevance) ประสิทธิภาพ (effectiveness) ประสิทธิภาพ (efficiency) ผลกระทบ (impact) และความยั่งยืน (sustainability) รายละเอียดดังตาราง 1

ตาราง 1 จำนวนและร้อยละของรายงานการประเมินระยะสิ้นสุดโครงการ ขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ที่นำมาสังเคราะห์

ข้อมูลเบื้องต้นของรายงานการประเมินระยะสิ้นสุดโครงการ	การระบุรายละเอียด	
	ไม่ระบุ n (%)	ระบุ n (%)
1. ด้านวิวิธนาการประเมิน		
1) ประเภทของการประเมินแบบผสมวิธี	3 (21.00%)	11 (79.00%)
2) การประเมินแบบผสมวิธี (ใช้ทั้งวิธีการเชิงปริมาณและเชิงคุณภาพ)	-	14 (100.00%)
3) เทคนิคการเก็บข้อมูลและวิธีเลือกกลุ่มตัวอย่าง	1 (7.00%)	13 (93.00%)
4) ขนาดของกลุ่มตัวอย่างหรือจำนวนผู้ให้ข้อมูล	-	14 (100.00%)
5) การวิเคราะห์ข้อมูล (สถิติ/เนื้อหา/เปรียบเทียบ/OECD-DAC)	-	14 (100.00%)
6) ข้อจำกัดของการประเมินและการตรวจสอบความน่าเชื่อถือของผลการประเมิน	-	14 (100.00%)
2. ด้านเนื้อหาของโครงการ		
1) เนื้อหาของโครงการ ได้แก่ ด้านการศึกษา ด้านสุขภาพ ด้านการคุ้มครองเด็ก	-	14 (100.00%)
2) วัตถุประสงค์และผลลัพธ์ที่คาดหวังของโครงการ	-	14 (100.00%)
3). ตัวชี้วัดผลลัพธ์ (outcome indicators) และผลกระทบ (impact)	2 (14.00%)	12 (86.00%)
4) ตอบคำถามการประเมินตามกรอบ OECD/DAC ครบทั้ง 5 มิติ คือ ความสอดคล้อง (relevance) ประสิทธิภาพ (effectiveness) ประสิทธิภาพ (efficiency) ผลกระทบ (impact) และความยั่งยืน (sustainability)	6 (43.00%)	8 (57.00%)
5). การนำเสนอข้อค้นพบและข้อเสนอแนะเชิงวิเคราะห์	-	14 (100.00%)
6) ระบุการนำผลการประเมินไปใช้ ได้แก่ การนำไปใช้เชิงนโยบาย การออกแบบโครงการต่อเนื่อง การเผยแพร่สู่สาธารณะ	3 (21.00%)	11 (79.00%)

2. ผลการสังเคราะห์รายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ตามหลักการประเมินอภิมาน

ผลการสังเคราะห์รายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ตามหลักการประเมินอภิมาน (meta-evaluation) เมื่อพิจารณาเป็นรายด้าน มีรายละเอียดดังนี้

มาตรฐานการใช้ประโยชน์ (utility) มีจำนวน 7 มาตรฐานย่อย พบว่า รายงานการประเมินฯ มีคะแนนรวมอยู่ระหว่าง 53 – 96 คะแนน ซึ่งเป็นคุณภาพใน 3 ระดับ คือ ระดับดี ดีมาก และยอดเยี่ยม ทั้งนี้ รายงานการประเมินฯ ส่วนใหญ่ มีคุณภาพอยู่ในระดับดีมาก (จำนวน 10 เรื่อง คิดเป็นร้อยละ 71.43)

มาตรฐานด้านความเป็นไปได้ (feasibility) มีจำนวน 3 มาตรฐานย่อย พบว่า รายงานการประเมินฯ มีคะแนนรวมอยู่ระหว่าง 58 – 100 คะแนน ซึ่งเป็นคุณภาพใน 3 ระดับ คือ ระดับดี ดีมาก และยอดเยี่ยม ทั้งนี้ รายงานการประเมินฯ ส่วนใหญ่มีคุณภาพอยู่ในระดับดีมาก (จำนวน 9 เรื่อง คิดเป็นร้อยละ 64.29)

มาตรฐานด้านความเหมาะสม (propriety) มีจำนวน 8 มาตรฐานย่อย พบว่า รายงานการประเมินฯ มีคะแนนรวมอยู่ระหว่าง 59 – 96 คะแนน ซึ่งเป็นคุณภาพใน 3 ระดับ คือ ระดับดี ดีมาก และยอดเยี่ยม ทั้งนี้ รายงานการประเมินฯ มีคุณภาพใกล้เคียงกัน จำนวน 3 ระดับ คือ ระดับยอดเยี่ยม (จำนวน 5 เรื่อง คิดเป็นร้อยละ 35.71) ระดับดีมาก (จำนวน 5 เรื่อง คิดเป็นร้อยละ 35.71) และระดับดี (จำนวน 4 เรื่อง คิดเป็นร้อยละ 28.57)

มาตรฐานด้านความถูกต้องแม่นยำ (accuracy) มีจำนวน 12 มาตรฐานย่อย พบว่า รายงานการประเมินฯ มีคะแนนรวมอยู่ระหว่าง 31 – 97 คะแนน ซึ่งเป็นคุณภาพใน 4 ระดับ คือ ระดับพอใช้ ดี ดีมาก และยอดเยี่ยม ทั้งนี้ รายงานการประเมินฯ ส่วนใหญ่ มีคุณภาพอยู่ในระดับดีมาก (จำนวน 8 เรื่อง คิดเป็นร้อยละ 57.14)

สำหรับผลการสังเคราะห์รายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็กประเทศออสเตรเลีย จำนวน 14 เรื่อง มีรายละเอียดดังตาราง 2 และภาพ 2

ตาราง 2 ผลการสังเคราะห์รายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก

ประเทศออสเตรเลีย ตามหลักการประเมินอภิमान

รายการประเมิน	รายงานการประเมินระยะสิ้นสุดโครงการ ขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย													
	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14
มาตรฐาน														
ด้านการใช้ประโยชน์														
1) การระบุผู้มีส่วนได้ส่วนเสีย	4	3	3	4	4	4	4	1	4	4	3	4	4	4
2) ความน่าเชื่อถือของ ผู้ประเมิน	4	4	4	4	4	4	4	3	4	4	4	4	4	4
3) ขอบเขตและการคัดเลือก ข้อมูล	4	4	4	4	4	4	4	3	4	4	4	4	4	4
4) การระบุคุณค่า	4	4	3	4	4	4	4	2	4	4	2	4	4	4
5) ความชัดเจนของรายงาน	4	4	3	4	4	4	4	4	4	4	3	4	4	4
6) ความตรงเวลาและการ เผยแพร่ผล	2	2	2	2	3	3	2	2	3	3	2	2	3	3
7) ผลกระทบของการ ประเมิน	3	2	0	1	3	3	2	0	4	4	2	3	2	2
รวมคะแนน (จากคะแนน เต็ม 28 คะแนน)	25	23	19	23	26	26	24	15	27	27	20	25	25	25
สรุปผล (%)	89	82	67	82	92	92	85	53	96	96	71	89	89	89
แปลผล	ดี มาก	ดี มาก	ดี	ดี มาก	ดี มาก	ดี มาก	ดี มาก	ดี	ยอด เยี่ยม	ยอด เยี่ยม	ดี มาก	ดี มาก	ดี มาก	ดี มาก

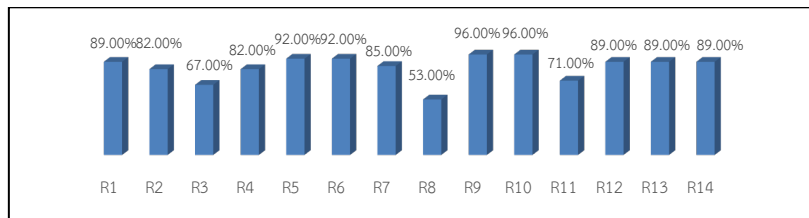
ตาราง 2 (ต่อ)

รายการประเมิน	รายงานการประเมินระยะสิ้นสุดโครงการ ขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย													
	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14
มาตรฐานด้านความเป็นไปได้														
1) ขั้นตอนปฏิบัติได้จริง	4	2	4	4	4	4	4	4	4	4	3	4	4	4
2) ความเป็นไปได้ทางการเมือง	3	3	1	3	3	3	3	2	4	4	2	3	3	2
3) ความคุ้มค่าของต้นทุน	4	2	2	4	4	4	4	4	4	4	3	4	4	4
รวมคะแนน (จากคะแนนเต็ม 12 คะแนน)	11	7	7	11	11	11	11	10	12	12	8	11	11	10
สรุปผล (%)	91	58	58	91	91	91	91	83	100	100	66	91	91	83
แปลผล	ดี มาก	ดี	ดี	ดี มาก	ดี มาก	ดี มาก	ดี มาก	ดี มาก	ยอดเยี่ยม	ยอดเยี่ยม	ดี	ดี มาก	ดี มาก	ดี มาก
มาตรฐานด้านความเหมาะสม														
1) การมุ่งเน้นบริการ	4	4	3	4	4	4	4	3	4	4	4	4	4	4
2) ข้อตกลงอย่างเป็นทางการ ทำข้อตกลงล่วงหน้าเป็นลาย ลักษณ์อักษรเกี่ยวกับ	3	3	2	3	4	4	2	1	4	4	2	4	4	2
3) สิทธิของผู้เข้าร่วม	4	4	4	4	4	4	4	4	4	4	4	4	4	4
4) ปฏิสัมพันธ์กับผู้เกี่ยวข้อง	4	4	4	4	4	4	4	3	4	4	4	4	4	4
5) การประเมินที่ครบถ้วน และเป็นธรรม	3	4	3	3	4	4	3	3	4	4	3	4	4	3
6) การเปิดเผยผลการ ประเมิน	4	4	4	4	4	4	4	4	4	4	4	4	4	4
7) ความขัดแย้งทาง ผลประโยชน์	0	0	1	1	3	3	0	1	3	4	0	2	3	2
8) ความรับผิดชอบด้าน งบประมาณ	0	0	0	0	3	3	0	0	3	3	0	2	3	3
รวมคะแนน (จากคะแนนเต็ม 32 คะแนน)	22	23	21	23	30	30	21	19	30	31	21	28	30	26
สรุปผล (%)	68	71	65	71	93	93	65	59	93	96	65	87	93	81
แปลผล	ดี มาก	ดี มาก	ดี	ดี มาก	ยอดเยี่ยม	ยอดเยี่ยม	ดี	ดี	ยอดเยี่ยม	ยอดเยี่ยม	ดี	ดี มาก	ยอดเยี่ยม	ดี มาก
มาตรฐานด้านความถูกต้องแม่นยำ														
1) การบันทึกข้อมูลโครงการ	2	4	2	4	4	4	4	3	4	4	2	4	4	4
2) การวิเคราะห์บริบท	2	3	3	3	3	3	3	3	2	4	2	3	4	3
3) การอธิบายวัตถุประสงค์ และขั้นตอน	1	4	3	3	3	4	4	1	3	3	1	3	3	4
4) แหล่งข้อมูลเชื่อถือได้	3	4	4	4	4	4	4	3	4	4	3	4	4	4
5) ข้อมูลที่ถูกต้อง	1	4	4	4	4	4	4	4	4	4	4	4	4	4
6) ความเชื่อมั่นของข้อมูล	0	4	2	3	3	3	2	0	2	4	3	4	3	4
7) ความเป็นระบบของข้อมูล	0	4	1	4	3	4	2	0	2	4	2	2	4	2
8) การวิเคราะห์ข้อมูล เชิงปริมาณ	2	3	3	4	0	1	2	0	0	3	3	3	3	4

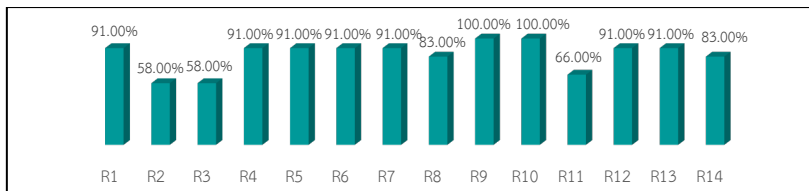
ตาราง 2 (ต่อ)

รายการประเมิน	รายงานการประเมินระยะสิ้นสุดโครงการ ขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย													
	R1	R2	R3	R4	R5	R6	R7	R8	R9	R10	R11	R12	R13	R14
9) การวิเคราะห์ข้อมูลเชิงคุณภาพ	4	4	4	4	4	4	4	4	4	4	4	4	4	4
10) ข้อเสนอแนะที่มีเหตุผล	0	4	3	3	4	4	3	3	4	4	4	4	4	4
11) รายงานที่เป็นกลาง	0	3	2	3	3	4	2	1	3	3	2	3	3	3
12) การประเมินอภิมาน	0	2	1	1	0	1	1	0	1	1	1	1	1	2
รวมคะแนน (จากคะแนนเต็ม 48 คะแนน)	15	47	32	40	35	40	35	22	33	46	31	39	41	42
สรุปผล (%)	31	97	66	83	72	83	72	45	68	95	64	81	85	87
แปลผล	พอใช้	ยอดเยี่ยม	ดี	ดีมาก	ดี	ดี	ดี	พอใช้	ดี	ยอดเยี่ยม	ดี	ดีมาก	ดี	ดีมาก

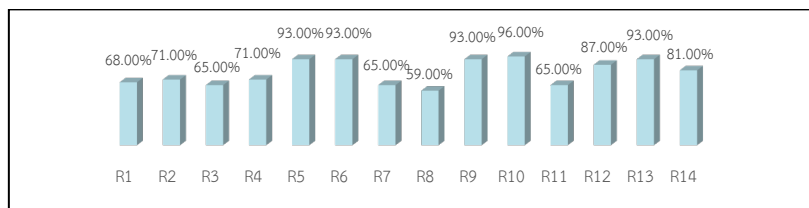
หมายเหตุ: เกณฑ์การจัดระดับคุณภาพ 4 ระดับ สำหรับการแปลความหมายคะแนนรวมในแต่ละมาตรฐาน คือ คะแนนรวม 0 – 24 อยู่ในระดับต่ำ คะแนนรวม 25 – 49 อยู่ในระดับพอใช้ คะแนนรวม 50 – 67 อยู่ในระดับดี คะแนนรวม 68 – 92 อยู่ในระดับดีมาก และคะแนนรวม 93 – 100 อยู่ในระดับยอดเยี่ยม



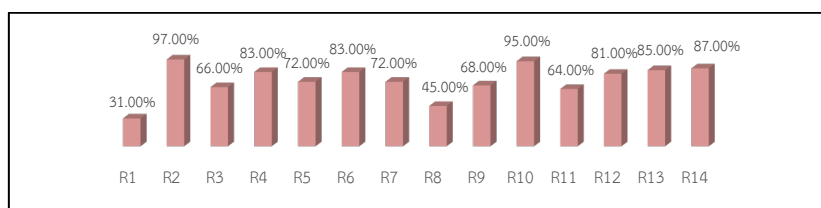
ภาพ 2 (ก) มาตรฐานด้านการใช้ประโยชน์



ภาพ 2 (ข) มาตรฐานด้านความเป็นไปได้



ภาพ 2 (ค) มาตรฐานด้านความเหมาะสม



ภาพ 2 (ง) มาตรฐานด้านความถูกต้องแม่นยำ

ภาพ 2 ผลการสังเคราะห์รายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก
ประเทศออสเตรเลีย

อภิปรายผล

จากการศึกษาครั้งนี้ คณะผู้วิจัยสรุปประเด็นการอภิปราย ดังนี้

1. จากผลการวิเคราะห์ข้อมูลเบื้องต้นของรายงานการประเมินระยะสิ้นสุดโครงการ ขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย มีประเด็นที่สำคัญดังนี้

1.1 ผลการวิเคราะห์ข้อมูลเบื้องต้น ด้านภูมิภาค พบว่า รายงานการประเมินฯ ที่นำมาสังเคราะห์ ในครั้งนี้ เป็นรายงานการประเมินฯ ที่ได้ดำเนินโครงการในประเทศออสเตรเลีย และในภูมิภาคต่าง ๆ ได้แก่ หมู่เกาะแปซิฟิก ภูมิภาคเอเชียกลาง ภูมิภาคเอเชียตะวันออกเฉียงใต้ ภูมิภาคเอเชียใต้ รวมถึง ในระดับโลก ประเด็นนี้กล่าวได้ว่าองค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ได้ดำเนินการโดยมุ่งเน้นการพัฒนาคุณภาพชีวิตเด็กและเยาวชนทั้งในประเทศออสเตรเลียและในภูมิภาคต่าง ๆ และยังสะท้อนให้เห็นถึงบทบาทสำคัญขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ในการพัฒนาระหว่างประเทศ (Save the Children Australia, 2024) และบทบาทเชิงยุทธศาสตร์ขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ที่มุ่งเน้นการทำงานในพื้นที่ที่ยังมีความต้องการความช่วยเหลือในด้านการพัฒนาเด็กและเยาวชน (UNESCO Institute for Statistics, 2025)

1.2 ผลการวิเคราะห์ข้อมูลเบื้องต้น ด้านเนื้อหาของสาระของโครงการ พบว่า รายงานการประเมินฯ ที่นำมาสังเคราะห์ในครั้งนี้ มีการดำเนินโครงการพัฒนา ในประเด็นที่เกี่ยวข้องกับการศึกษา สุขภาพ และการคุ้มครองเด็กในมิติต่าง ๆ ได้แก่ โครงการส่งเสริมความเท่าเทียมด้านเพศภาวะ (gender equality) การบริการด้านการศึกษาและการส่งเสริมสุขภาวะของเด็กในพื้นที่ห่างไกล ในที่นี้ ขอนำเสนอตัวอย่างโครงการพัฒนา ตัวอย่างเช่น โครงการของประเทศเวียดนามด้านการรู้รับปรับตัวกับสภาพแวดล้อม (Sterrett, 2015) และโครงการของประเทศไทยด้านการปกป้องและให้ความช่วยเหลือเด็กและเยาวชนข้ามชาติ (Ma, 2017) ซึ่งโครงการพัฒนาต่าง ๆ ที่ได้รับการสนับสนุนการดำเนินโครงการจากองค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ได้ออกแบบประเด็นเนื้อหาของโครงการและกิจกรรมภายใต้โครงการให้เหมาะสมกับความต้องการและข้อจำกัดของพื้นที่ที่ดำเนินโครงการเป็นสำคัญ เพื่อให้โครงการสามารถสร้างประโยชน์ให้กับเด็กและเยาวชนได้อย่างแท้จริง อย่างไรก็ตาม โครงการส่วนใหญ่ไม่ได้จำกัดเฉพาะการมุ่งพัฒนาเด็กและเยาวชนแต่เพียงอย่างเดียว แต่ยังครอบคลุมไปถึงครอบครัวและสังคมรอบตัวของเด็ก ๆ ด้วย เช่น โครงการให้ความช่วยเหลือด้วยเงินสด ประเทศฟิลิปปินส์ (Foster & Richards-Melamdir, 2023) โครงการหนึ่งพันวันแรก ประเทศวานูอาตู (Save the Children, 2021) เป็นต้น

1.3 ผลการวิเคราะห์ข้อมูลเบื้องต้น ด้านผู้ประเมิน พบว่า รายงานการประเมินฯ ที่นำมาสังเคราะห์ในครั้งนี้ สามารถจำแนกลักษณะของผู้ประเมินโครงการ ออกเป็น 2 ประเภท คือ ผู้ประเมินภายนอก (external evaluator) และผู้ประเมินภายใน (internal evaluator) โดยรายงานการประเมินฯ ที่นำมาสังเคราะห์ส่วนใหญ่ มีลักษณะการดำเนินการประเมินโดยผู้ประเมินภายนอกหรือองค์กรประเมินอิสระ ซึ่งสังกัดบริษัทที่ปรึกษาด้านการประเมินในระดับประเทศและระหว่างประเทศ ทั้งนี้ อาจเนื่องมาจากข้อกำหนดของแหล่งทุน ซึ่งสอดคล้องกับสถาบันวิจัย ARCO Research Center (2024) เป็นหน่วยงานที่มีภารกิจด้านการวิจัย การอบรมและการตรวจประเมิน ของประเทศอิตาลี ที่ได้อธิบายถึงจุดแข็งของการตรวจ

ประเมินโครงการพัฒนาโดยผู้ประเมินภายนอกไว้ว่า ผู้ประเมินภายนอกมีความเป็นอิสระจากองค์กรดำเนินงานหรือผู้ให้ทุน จึงสามารถประเมินผลได้อย่างเป็นกลาง ช่วยสะท้อนความสำเร็จที่แท้จริงของโครงการ และเอื้อต่อการเรียนรู้ขององค์กรอย่างถูกต้อง รวมถึง ความน่าเชื่อถือของผลการประเมิน อันเนื่องมาจากความเป็นกลางและความไม่เอนเอียงของผู้ประเมิน

อย่างไรก็ตาม ยังมีข้อค้นพบที่สำคัญ คือ มีรายงานการประเมินฯ บางส่วน ที่ดำเนินการประเมินโดยผู้ประเมินภายใน ซึ่งเป็นเจ้าหน้าที่ผู้ปฏิบัติงานขององค์กรช่วยเหลือเด็ก ซึ่งในบางโครงการอาจจะมีการบวนการที่ได้มอบหมายให้สมาชิกในองค์กรหรือโครงการ เป็นผู้รับผิดชอบในการประเมินโครงการ ซึ่งสอดคล้องกับแนวคิดของ Chai-Aroon (2022) ที่ได้อธิบายถึงบทบาทของผู้ประเมินภายในต่อการดำเนินโครงการ คือ ผู้ประเมินภายใน มักจะมีความสัมพันธ์ใกล้ชิดกับการขับเคลื่อนโครงการ มีความเข้าใจในแผนงาน/โครงการอย่างดี มีความสามารถในการเข้าถึงข้อมูลของแผนงาน/โครงการ รวมถึง มีทักษะสอดคล้องกับแนวคิด ยุทธศาสตร์ และคุณค่าของแผนงาน/โครงการ

1.4 ผลการวิเคราะห์ข้อมูลเบื้องต้น ด้านแหล่งทุนของโครงการ พบว่า รายงานการประเมินฯ ที่นำมาสังเคราะห์ ในครั้งนี้ มีลักษณะของแหล่งทุนของโครงการ แบ่งออกเป็น 3 ลักษณะ คือ (1) การได้รับงบประมาณสนับสนุนโดยตรง จากองค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย (2) การได้รับงบประมาณสนับสนุนจากกระทรวงการต่างประเทศและการค้า ประเทศออสเตรเลีย (Department of Foreign Affairs and Trade, DFAT) ซึ่งมีความรับผิดชอบด้านนโยบายต่างประเทศ ความสัมพันธ์ระหว่างประเทศ การค้า การลงทุน และการส่งมอบความช่วยเหลือเพื่อการพัฒนา เพื่อสร้างความมั่นคงและความมั่งคั่งให้แก่ประเทศ ทั้งนี้หนึ่งในภารกิจของกระทรวงการต่างประเทศและการค้า ประเทศออสเตรเลีย คือ การบริหารจัดการโครงการความช่วยเหลือของออสเตรเลียในภูมิภาคต่าง ๆ (Department of Foreign Affairs and Trade, n.d.) และ (3) การได้รับงบประมาณการสนับสนุนมาจากหลายแหล่ง ทั้งจากการประสานความร่วมมือกับองค์กรพัฒนาเอกชนอื่น ๆ ทั้งในระดับประเทศและระดับสากลที่ทำงานด้านการพัฒนาเด็กและเยาวชน

1.5 ผลการวิเคราะห์ข้อมูลเบื้องต้น ด้านวิธีวิทยาการประเมิน พบว่า รายงานการประเมินฯ ที่นำมาสังเคราะห์ ทุกเรื่อง ใช้วิธีการประเมินแบบผสมวิธี กล่าวคือ โครงการได้ดำเนินโครงการโดยใช้ทั้งวิธีการเชิงปริมาณและเชิงคุณภาพ และนำผลการวิเคราะห์ข้อมูลทั้งเชิงปริมาณและเชิงคุณภาพมาสรุปผลร่วมกัน นอกจากนี้ ยังพบว่า รายงานการประเมินฯ ส่วนใหญ่มีการระบุประเภทของการประเมินไว้ ว่าโครงการได้เลือกใช้ในการประเมินแบบผสมวิธี (mixed methods evaluation) ไว้อย่างชัดเจน ข้อค้นพบเช่นนี้ อาจเนื่องมาจากเป็นข้อกำหนดขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ซึ่งโครงการจะต้องใช้ทั้งวิธีการเชิงปริมาณและวิธีการเชิงคุณภาพ อย่างไรก็ตาม ยังมีข้อค้นพบที่สำคัญ คือ ถึงแม้ว่ารายงานการประเมินส่วนใหญ่ได้ระบุถึงการเลือกใช้ในการประเมินแบบผสมวิธี (mixed methods evaluation) แต่ยังไม่พบการระบุหรือการอธิบายรายละเอียดเกี่ยวกับรูปแบบประเมินแบบผสมวิธี (mixed methods designs) ตามหลักสากล เช่น รูปแบบประเมินแบบผสมวิธีในการประเมิน (Creswell & Plano Clark, 2018; Mertens, 2018; United States Agency for International Development, 2013) รูปแบบการประเมินแบบผสมวิธีในการประเมินโครงการ (Prosek, 2020) เป็นต้น

2. จากผลการสังเคราะห์รายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็กประเทศออสเตรเลีย ตามหลักการประเมินนิมิต มีประเด็นที่สำคัญดังนี้

2.1 ผลการวิเคราะห์ข้อมูล พบว่า ในภาพรวม รายงานการประเมินฯ ที่นำมาสังเคราะห์ในครั้งนี้ มีคุณภาพสอดคล้องตามมาตรฐานการประเมินโครงการ (The Program Evaluation Standards) ของ Stufflebeam (1999) ทั้ง 4 มาตรฐาน คือ มาตรฐานด้านการใช้ประโยชน์ มาตรฐานด้านความเป็นไปได้ มาตรฐานด้านความเหมาะสม และมาตรฐานด้านความถูกต้องแม่นยำ

อย่างไรก็ตาม ผลการวิเคราะห์ข้อมูลในส่วนนี้ มีข้อค้นพบที่สำคัญ คือ มีจำนวน 3 มาตรฐาน ที่มีคุณภาพ 3 ระดับ คือ มาตรฐานด้านการใช้ประโยชน์ (มีคุณภาพ 3 ระดับ คือ ระดับดี ดีมาก และยอดเยี่ยม) มาตรฐานด้านความเป็นไปได้ (มีคุณภาพ 3 ระดับ คือ ระดับดี ดีมาก และยอดเยี่ยม) มาตรฐานด้านความเหมาะสม (มีคุณภาพ 3 ระดับ คือ ระดับดี ดีมาก และยอดเยี่ยม) ในขณะที่มาตรฐานด้านความถูกต้องแม่นยำ มีคุณภาพ 4 ระดับ คือ ระดับพอใช้ ดี ดีมาก และยอดเยี่ยม ข้อมูลนี้สะท้อนให้เห็นว่า มาตรฐานด้านความถูกต้องแม่นยำ มีความแปรปรวนของคะแนนสูงที่สุด เมื่อเทียบกับมาตรฐานอื่น ๆ

2.2 เมื่อพิจารณาผลการวิเคราะห์ข้อมูล มาตรฐานด้านความถูกต้องแม่นยำ พบว่า ในมาตรฐานย่อย ด้านการประเมินนิมิต มีคะแนนต่ำกว่าด้านอื่น ๆ และเป็นคะแนนคุณภาพที่น้อยมาก มีระดับ 0 คะแนน และระดับ 1 คะแนน ในประเด็นนี้จึงสรุปได้ว่า รายงานการประเมินฯ ที่นำมาสังเคราะห์ในครั้งนี้ มีข้อจำกัดด้านการประเมินนิมิต เนื่องจากรายงานการประเมินฯ จำนวนมาก ไม่ได้อธิบายรายละเอียดของการประเมินนิมิต และอาจเนื่องมาจาก แหล่งทุนไม่ได้ระบุประเด็นการประเมินนิมิต ไว้เป็นข้อกำหนดของการกระบวนการประเมินโครงการที่ผู้ประเมินจะต้องดำเนินการ จึงทำให้มาตรฐานด้านความถูกต้องแม่นยำ มีคะแนนรวมต่ำกว่ามาตรฐานอื่น ๆ

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการวิจัยไปใช้

1.1 จากผลการสังเคราะห์รายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็กประเทศออสเตรเลีย ตามหลักการประเมินนิมิต ครั้งนี้ ซึ่งคณะผู้วิจัยได้เลือกใช้ในการประเมินนิมิตแบบสรุปรวม (summative meta-evaluation) โดยได้กำหนดเนื้อหาของการประเมิน ตามแนวคิดมาตรฐานการประเมินโครงการ ซึ่งแบ่งออกเป็น 4 มาตรฐาน คือ มาตรฐานด้านการใช้ประโยชน์ มาตรฐานด้านความเป็นไปได้ มาตรฐานด้านความเหมาะสม และมาตรฐานด้านความถูกต้องแม่นยำ แสดงให้เห็นว่า ผลการสังเคราะห์รายงานการประเมินระยะสิ้นสุดโครงการครั้งนี้ สามารถส่งมอบข้อมูลสารสนเทศเกี่ยวกับคุณภาพรายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ในระหว่างปี พ.ศ. 2558–2566 (ค.ศ. 2015–2023) ที่สามารถสร้างความเชื่อมั่นในการนำข้อมูลสารสนเทศจากรายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ไปใช้ประโยชน์ ทั้งนี้ องค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย สามารถนำข้อมูลสารสนเทศที่เป็นคุณค่ารวมของรายงานการประเมินฯ และมาตรฐานของรายงานการประเมินฯ ดังกล่าว ไปใช้ประโยชน์ต่อไป

1.2 จากผลการสังเคราะห์รายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ครั้งนี้ มีข้อสังเกตที่สำคัญ คือ มาตรฐานด้านความถูกต้องแม่นยำ เป็นมาตรฐานที่มีความแปรปรวนของคะแนนสูงที่สุด ซึ่งรายงานการประเมินฯ ส่วนใหญ่ ยังมีข้อจำกัดในมาตรฐานย่อยด้านการประเมินอภิमान ดังนั้น องค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ควรกำหนดให้ผู้ประเมิน/โครงการ ได้อธิบายรายละเอียดการประเมินอภิमानในรายงานการประเมิน เพื่อการดำเนินโครงการต่าง ๆ เป็นไปอย่างมีประสิทธิภาพ และยังทำให้รายงานการประเมินที่โครงการจะต้องส่งมอบให้กับองค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย มีข้อมูลสารสนเทศที่ครบถ้วนสมบูรณ์และมีคุณภาพ

1.3 จากผลการสังเคราะห์รายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ครั้งนี้ มีข้อสังเกตที่สำคัญ คือ รายงานการประเมินส่วนใหญ่ ได้ระบุถึงการเลือกใช้การประเมินแบบผสมวิธี (mixed methods evaluation) แต่ยังไม่พบการระบุรูปแบบประเมินแบบผสมวิธี (mixed methods evaluation designs) ซึ่งรายงานการประเมินฯ ส่วนใหญ่ ยังมีข้อจำกัดในการระบุรายละเอียดของรูปแบบประเมินแบบผสมวิธี ดังนั้น ผู้ประเมิน/โครงการ ที่ดำเนินโครงการโดยเลือกใช้การประเมินแบบผสมวิธี ควรอธิบายรายละเอียดของการประเมินแบบผสมวิธี และรูปแบบประเมินแบบผสมวิธี ตามหลักสากล ไว้อย่างชัดเจน พร้อมทั้ง ระบุเหตุผลในการเลือกใช้รูปแบบประเมินแบบผสมวิธีไว้ในรายงานการประเมินของโครงการ ซึ่งการอธิบายรายละเอียดเหล่านี้ จะก่อให้เกิดคุณประโยชน์ทางวิชาการ ด้านการประเมินแบบผสมวิธีต่อไป

1.4 ในการศึกษาครั้งนี้ คณะผู้วิจัยได้เลือกใช้การประเมินอภิमानแบบสรุปรวม เพื่อสังเคราะห์รายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย ดังนั้น สำหรับผู้ประเมิน หรือผู้ที่จะเริ่มสังเคราะห์รายงานการประเมินระยะสิ้นสุดโครงการ สามารถเลือกใช้วิธีการประเมินอภิमानแบบสรุปรวม เพื่อส่งมอบข้อมูลสารสนเทศที่เป็นคุณค่ารวมของรายงานการประเมิน ทำให้ผู้ที่เกี่ยวข้องกับโครงการ ได้ทราบถึงความเป็นมาตรฐานของโครงการประเมิน/รายงานการประเมิน และยังทำให้ผู้ใช้ผลการประเมิน มีความเชื่อมั่นต่อผลการประเมินและข้อเสนอแนะจากรายงานการประเมินระยะสิ้นสุดโครงการ

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

จากผลการสังเคราะห์รายงานการประเมินระยะสิ้นสุดโครงการขององค์กรช่วยเหลือเด็ก ประเทศออสเตรเลีย แสดงให้เห็นว่า วิธีการประเมินแบบผสมวิธี (mixed methods evaluation) ได้ถูกนำมาใช้ในการประเมินโครงการ โดยการใช้ทั้งวิธีการเชิงปริมาณและเชิงคุณภาพ รวมถึง การนำผลการวิเคราะห์ข้อมูล ทั้งเชิงปริมาณและเชิงคุณภาพมาสรุปผลร่วมกัน ดังนั้น จึงควรมีการสังเคราะห์รายงานการประเมินระยะสิ้นสุดโครงการเชิงคุณลักษณะ (เชิงคุณภาพ) ซึ่งเป็นการอธิบายเชิงเนื้อหา สามารถอธิบายปรากฏการณ์ทางสังคม บริบททางสังคมและวัฒนธรรมได้

จริยธรรมการวิจัย

การศึกษาครั้งนี้ได้ผ่านกระบวนการพิจารณาและเห็นชอบการรับรองจริยธรรมการวิจัยในคน จากคณะกรรมการจริยธรรมการวิจัยในคน สาขาสังคมศาสตร์และมนุษยศาสตร์ มหาวิทยาลัยมหิดล ตามเอกสารรับรองหมายเลข 2025/027.1111

References

- ARCO Research Center. (2024). *5 key benefits of external evaluation for cooperation and development projects*. <https://www.arcolab.org/en/5-vantaggi-valutazione-monitoraggio-cooperazione-sviluppo/>
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage.
- Department of Foreign Affairs and Trade. (n.d.). *What we do*. <https://www.dfat.gov.au/about-us/department/what-we-do>
- Foster, J., & Richards-Melamdir, M. (2023). *Final evaluation report: Fiji cash assistance project*. https://www.savethechildren.org.au/getmedia/461a7639-c2a9-459f-a46b-a8cc67c8e0bf/STC01914_FJ-Cash-Evaluation-Report-2023-DESIGNED_sml2-compressed.pdf.aspx
- Ma, M. (2017). *Project evaluation: Improving migrant protection and assistance for children in Thailand (IMPACT)*. Save the Children Australia.
- Mertens, D. M. (2018). *Mixed methods design in evaluation*. SAGE Publications. <https://doi.org/10.4135/9781506330631>
- Prosek, E. A. (2020). An introduction to mixed methods design in program evaluation. *Counseling Outcome Research and Evaluation*, 11(2), 60–70. <https://doi.org/10.1080/21501378.2019.1590688>
- Save the Children Australia. (2024). *Where we work*. <https://www.savethechildren.org.au/about-us/where-we-work>
- Save the Children. (2021). *First 1,000 days program endline evaluation report: Sanma & Shefa Provinces, Vanuatu*. <https://www.savethechildren.org.au/getmedia/330f635a-2c5b-475f-a44a-ce4cc2c729c6/2021-Vanuatu-First-1000-Days-Endline-Evaluation.pdf.aspx>
- Sterrett, C. L. (2015). *Final evaluation of the Vietnam child-centred climate resilience program*. <https://resourcecentre.savethechildren.net/pdf/vnm-climate-resilience.pdf>
- Stufflebeam, D. L. (1999). *Program evaluations meta-evaluation checklist*. https://files.wmich.edu/s3fs-public/attachments/u350/2014/program_metaeval_short.pdf

Stufflebeam, D. L. (2001). Evaluation models. *New Directions for Evaluation*, 2001(89), 7–98.

<https://doi.org/10.1002/ev.3>

UNESCO Institute for Statistics. (2025). *SDG 4 scorecard progress report on national benchmarks: Focus on the out-of-school rate*. <https://doi.org/10.54676/SKLD4888>

United States Agency for International Development. (2013). *Technical note: Conducting mixed-method evaluations*. <https://alnap.org/help-library/resources/technical-note-conducting-mixed-method-evaluations/>

Translate Thai Reference

Chai-Aroon, T. (2022). The significance of internal evaluation for health promotion programs. In Un-ob, P. (Ed.), *Internal evaluation: A tool for learning and health promotion program development* (2nd ed.). (pp.41-51). P.A. Living Co., Ltd. (in Thai)

การศึกษาสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชา
คณิตศาสตร์ชั้นประถมศึกษาปีที่ 6 โรงเรียนในสังกัดสำนักงานเขตพื้นที่
การศึกษาประถมศึกษากาฬสินธุ์

The Study of Competency-Based Learning Assessment Problems of
Grade 6 Mathematics Teachers in Schools under the Kalasin
Primary Educational Service Area Office

ศักดิ์สิทธิ์ ฤทธิสิน^{1*} อมร มะลาศรี¹ คมสันท์ ขจรปัญญาไพศาล¹

สุวรรณวัฒน์ เทียนยุทธกุล¹ และ สมใจ ภูครองทุ่ง¹

Saksit Rittilun,^{1*} Amorn Malasri,¹ Komsan Khajornpanyapaipan¹

Suwannawat Thienyuttakul¹ and Somjai Phukrongtung¹

(Received: March 12, 2026; Revised: April 23, 2026; Accepted: April 24, 2026)

บทคัดย่อ

การวิจัยนี้มีความมุ่งหมายเพื่อศึกษาและเปรียบเทียบสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ชั้นประถมศึกษาปีที่ 6 โรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษากาฬสินธุ์ กลุ่มตัวอย่างเป็นครูผู้สอนรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 จำนวน 222 คน ได้มาโดยวิธีการสุ่มแบบแบ่งชั้นภูมิ เครื่องมือที่ใช้ในการวิจัย ได้แก่ แบบสอบถามสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะ มีค่าความเชื่อมั่น เท่ากับ 0.99 และแบบบันทึกการสนทนากลุ่ม มีค่าดัชนีความสอดคล้อง เท่ากับ 1.00 ทุกข้อ วิเคราะห์ข้อมูลด้วยร้อยละ ค่าเฉลี่ย ส่วนเบี่ยงเบนมาตรฐาน การทดสอบเอฟ และการวิเคราะห์เนื้อหา

ผลการวิจัย พบว่า 1) สภาพปัญหาการประเมินผลการเรียนรู้ด้านสมรรถนะหลักและสมรรถนะเฉพาะอยู่ในระดับปานกลาง โดยสมรรถนะหลักสภาพปัญหามากที่สุด คือ “การเป็นพลเมืองที่เข้มแข็ง” สมรรถนะเฉพาะสภาพปัญหามากที่สุด คือ “การสื่อสารและนำเสนอทางคณิตศาสตร์” 2) ครูที่มีวิทยฐานะแตกต่างกันมีสภาพปัญหาการประเมินผลการเรียนรู้สมรรถนะเฉพาะ แตกต่างกันอย่างมีนัยสำคัญทางสถิติ

¹ สาขาวิชานวัตกรรมการจัดการเรียนรู้ คณะศึกษาศาสตร์และนวัตกรรมการศึกษา มหาวิทยาลัยกาฬสินธุ์

¹ Program in Learning Management Innovation; Faculty of Education and Educational Innovation; Kalasin University

งบประมาณสนับสนุนงานมูลฐาน (Fundamental Fund) ประจำปีงบประมาณ พ.ศ. 2568 จากกองทุนส่งเสริมวิทยาศาสตร์ วิจัยและนวัตกรรม

The budget for supporting fundamental work (Fundamental Fund) for the fiscal year 2025 from the National Science Research and Innovation Fund (NSRF)

* Corresponding Author E-mail: srittilun@gmail.com

ที่ระดับ .05 ครูที่มีประสบการณ์การทำงานแตกต่างกันมีสภาพปัญหาการประเมินผลการเรียนรู้สมรรถนะเฉพาะแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 และครูที่สังกัดโรงเรียนที่มีขนาดแตกต่างกันมีสภาพปัญหาการประเมินผลการเรียนรู้สมรรถนะหลักและสมรรถนะเฉพาะไม่แตกต่างกัน และ 3) จากการสนทนากลุ่ม พบว่า ครูขาดความเข้าใจเกี่ยวกับสมรรถนะและการออกแบบตัวชี้วัดที่ซับซ้อน ใช้เวลามากในการออกแบบการประเมิน และประสบปัญหาในการสร้างเครื่องมือประเมินให้ครอบคลุมและชัดเจน

คำสำคัญ การประเมินผลการเรียนรู้ฐานสมรรถนะ สมรรถนะหลัก สมรรถนะเฉพาะ

Abstract

This research investigated and compared the conditions of problems regarding competency-based learning assessment among Grade 6 mathematics teachers in schools under the Kalasin Primary Educational Service Area Office, Thailand. Samples comprised 222 teachers selected through stratified random sampling. Research instrument used was a questionnaire with reliability coefficient of 0.99 and a focus group discussion form (IOC = 1.00). Data were analyzed using descriptive statistics, F-tests, and content analysis.

Results showed that: 1) problems in assessing core and subject-specific competencies were at moderate levels, with the most problematic areas being active citizenship (core) and mathematical communication and presentation (subject-specific); 2) teachers with different academic ranks showed significant differences at the .05 level in assessing subject-specific competencies, teachers with different experience levels also differed significantly at the .05 level in assessing subject-specific competencies, and teachers in schools of different sizes did not differ significantly for either core or subject-specific competencies; and 3) focus group findings revealed that teachers lacked clear understanding of curriculum-defined competencies, struggled with designing complex indicators, and faced time constraints in documentation, making it difficult to create comprehensive assessment tools with clear criteria and descriptors, which resulted in assessments lacking adequate validity.

Keywords: competency-based Learning assessment, core competencies, specific competencies

บทนำ

หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 ได้กำหนดสมรรถนะที่สำคัญไว้ 5 ประการ ประกอบด้วย 1) ความสามารถในการสื่อสาร 2) ความสามารถในการคิด 3) ความสามารถในการแก้ปัญหา 4) ความสามารถในการใช้ทักษะชีวิต และ 5) ความสามารถในการใช้เทคโนโลยี โดยหลักสูตรได้กำหนดให้ครูพัฒนาสมรรถนะให้แก่ผู้เรียน (Ministry of Education, 2008) แต่ตลอดระยะเวลาของการใช้

หลักสูตรพบว่า ครูยังไม่เข้าใจสมรรถนะโดยขาดทั้งพื้นฐานความรู้และทักษะในการพัฒนาสมรรถนะของผู้เรียน ส่งผลให้ผู้เรียนไม่ได้เรียนรู้และไม่ได้ฝึกนำความรู้ความเข้าใจ และทักษะต่าง ๆ ไปใช้ในการปฏิบัติงานให้เกิดสมรรถนะ (Independent Commission for Education Reform, n.d.)

แนวทางที่จะช่วยให้คุณภาพของผู้เรียนบรรลุผลลัพธ์ตามที่ต้องการ คือการปฏิรูปหลักสูตร การจัดการเรียนการสอน การวัดและประเมินผลให้มีความยืดหยุ่น และเอื้อต่อการพัฒนาสมรรถนะที่จำเป็นต่อการใช้ชีวิต การทำงานและการแก้ปัญหาต่าง ๆ รวมทั้งการปรับตัวให้ทันต่อการเปลี่ยนแปลงและความต้องการใหม่ ๆ ของสังคมและโลก สามารถทำได้หลายวิธี การปรับหลักสูตรให้เป็นหลักสูตรฐานสมรรถนะเป็นทางเลือกหนึ่งที่มีศักยภาพที่จะตอบโจทย์ปัญหาของครูและผู้เรียนที่เกิดขึ้น รวมทั้งสามารถตอบสนองความต้องการใหม่ ๆ ของโลกในศตวรรษที่ 21 (Office of the Education Council, 2020) หลักสูตรฐานสมรรถนะเป็นหลักสูตรที่มุ่งเป้าหมายการพัฒนาสมรรถนะของผู้เรียนตามช่วงวัย เน้นการพัฒนารายบุคคลอย่างเป็นองค์รวม เพื่อการเป็นเจ้าของการเรียนรู้และพัฒนาตนเองอย่างต่อเนื่อง มีความเชื่อมโยงระหว่างสมรรถนะหลักและสมรรถนะเฉพาะในการกำหนดผลลัพธ์การเรียนรู้เพื่อพัฒนาผู้เรียนให้สามารถนำไปใช้ในชีวิตและการทำงาน สามารถจัดสภาพแวดล้อมและเส้นทางการเรียนรู้ที่หลากหลายและระบบสนับสนุนการเรียนรู้ที่สอดคล้องกับพหุปัญญาและธรรมชาติของผู้เรียน เน้นกระบวนการจัดการเรียนรู้เชิงรุก การใช้สื่อและสถานการณ์การเรียนรู้ร่วมสมัย มีความหลากหลาย และยืดหยุ่น ตามความสนใจ ความถนัดของผู้เรียน บริบท จุดเน้นของสถานศึกษา และชุมชนแวดล้อม และมุ่งใช้ผลการประเมินเพื่อพัฒนาและสะท้อนสมรรถนะของผู้เรียนตามเกณฑ์การปฏิบัติที่เป็นธรรม เชื่อถือได้ เอื้อต่อการถ่วงดุลการเรียนรู้และพัฒนาในระดับที่สูงขึ้นตามระดับความสามารถ (Office of the Basic Education Commission, 2022)

สำนักงานเลขาธิการสภาการศึกษา (Office of the Education Council, 2025) ได้ศึกษากระบวนการพัฒนาหลักสูตรสถานศึกษาฐานสมรรถนะสถานศึกษานำร่องในพื้นที่นวัตกรรม โดยศึกษาปัจจัยระบบสนับสนุน และปัญหาอุปสรรคในการทำงานพัฒนาหลักสูตรสถานศึกษาฐานสมรรถนะของสถานศึกษานำร่องในพื้นที่นวัตกรรม จำนวน 35 แห่ง และสถานศึกษาทั่วไป อีกจำนวน 5 แห่ง พบว่า 1) สถานศึกษามีการเตรียมความพร้อมในหลากหลายด้าน ได้แก่ การสร้างความเข้าใจ และทำให้ ผู้เกี่ยวข้องเห็นความสำคัญ การเตรียมความพร้อมและพัฒนาบุคลากร การปรึกษาผู้รู้หรือผู้เชี่ยวชาญ การจัดตั้งคณะกรรมการพัฒนาหลักสูตรสถานศึกษา การศึกษาเอกสารและข้อมูลพื้นฐานด้านต่าง ๆ การเข้ารับการอบรมเกี่ยวกับการพัฒนาหลักสูตรสถานศึกษาฐานสมรรถนะเพื่อให้ผู้บริหารและครูสอน เกิดความรู้ความเข้าใจจากนั้นมีการเตรียมบุคลากร และเตรียมปรับกรอบความคิดของครูสอนให้พร้อม ในการเปลี่ยนแปลงและการทำงาน มีการศึกษาดูงาน/ ตัวอย่างหลักสูตรสถานศึกษาจากโรงเรียนอื่น ๆ เพื่อเป็นตัวอย่างสำหรับการดำเนินการ 2) กระบวนการพัฒนาหลักสูตรสถานศึกษาฐานสมรรถนะ ในขั้นการดำเนินงานพัฒนาหลักสูตรของสถานศึกษา 7 ด้าน พบว่า ด้านที่ดำเนินการยากที่สุด คือ ด้านแนวทางการวัดและประเมินผลการเรียนรู้ฐานสมรรถนะ

ปัจจุบันมีพื้นที่ที่ได้รับการประกาศเป็นพื้นที่นวัตกรรมการศึกษาแล้วจำนวน รวมทั้งสิ้น 20 พื้นที่ ในระยะแรกที่มีพระราชบัญญัติพื้นที่นวัตกรรม พ.ศ. 2562 (Government Gazette, 2019) มีจังหวัดนำร่อง 8 จังหวัด ได้แก่ ศรีสะเกษ ระยอง สตูล เชียงใหม่ กาญจนบุรี และจังหวัดชายแดนภาคใต้ ซึ่งประกอบด้วย

ยะลา ปัตตานี และนราธิวาส ต่อมาคณะรัฐมนตรีได้ประกาศกำหนดพื้นที่นวัตกรรมการศึกษาเพิ่มอีก 11 จังหวัด ประกอบด้วย กรุงเทพมหานคร กระบี่ จันทบุรี ตราด ภูเก็ต แม่ฮ่องสอน สงขลา สระแก้ว สุโขทัย สุราษฎร์ธานี และอุบลราชธานี (Government Gazette, 2022) และต่อมาคณะรัฐมนตรี ได้ประกาศกำหนด จังหวัดบุรีรัมย์ เป็นพื้นที่นวัตกรรมการศึกษา (Government Gazette, 2024) แม้ว่าโรงเรียนในสังกัด สำนักงานเขตพื้นที่การศึกษาประถมศึกษากาฬสินธุ์ จะไม่อยู่ในพื้นที่นวัตกรรมการศึกษาและไม่ได้เป็นโรงเรียน นำร่องการใช้หลักสูตรสถานศึกษาฐานสมรรถนะก็ตาม แต่เพื่อเป็นการเตรียมการรองรับการใช้หลักสูตร สถานศึกษาฐานสมรรถนะในอนาคต การศึกษาสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของ ครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 โรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษา ประถมศึกษากาฬสินธุ์ สารสนเทศที่ได้จะเป็นประโยชน์ต่อสำนักงานเขตพื้นที่การศึกษาประถมศึกษากาฬสินธุ์ ในการกำหนดนโยบายและวางแผนการพัฒนาศักยภาพครูได้สอดคล้องกับสภาพปัญหาและบริบทที่แท้จริง เพื่อให้เกิดการพัฒนาอย่างมีประสิทธิภาพและประสิทธิผลสูงสุด

ความมุ่งหมายของการวิจัย

1. เพื่อศึกษาสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชา คณิตศาสตร์ชั้นประถมศึกษาปีที่ 6 โรงเรียนในสำนักงานเขตพื้นที่การศึกษาประถมศึกษากาฬสินธุ์
2. เพื่อเปรียบเทียบสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชา คณิตศาสตร์ชั้นประถมศึกษาปีที่ 6 โรงเรียนในสำนักงานเขตพื้นที่การศึกษาประถมศึกษากาฬสินธุ์ จำแนกตาม วิทยฐานะ ประสบการณ์การทำงาน และขนาดโรงเรียนที่สังกัด

สมมติฐานการวิจัย

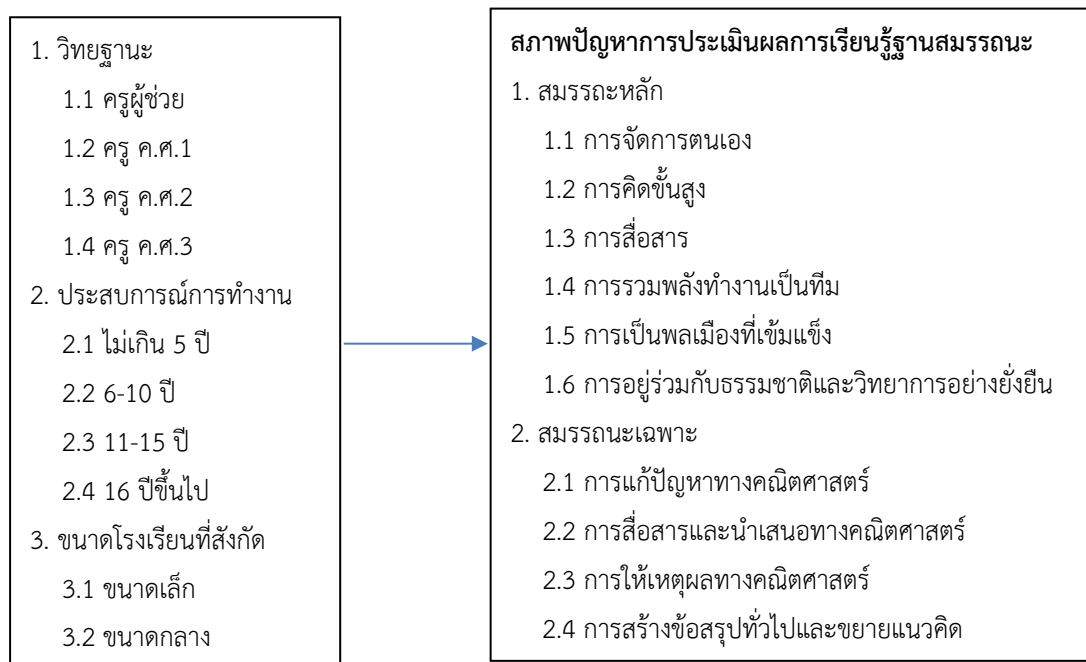
ครูที่มีวิทยฐานะ ประสบการณ์การทำงาน และขนาดโรงเรียนที่สังกัดต่างกัน มีสภาพปัญหา การประเมินผลการเรียนรู้ฐานสมรรถนะในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 แตกต่างกัน

กรอบแนวคิดการวิจัย

การศึกษาสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ใช้กรอบแนวคิดของสำนักงานคณะกรรมการการศึกษาขั้นพื้นฐาน (Office of the Basic Education Commission, 2022b) สมรรถนะหลัก 6 ด้าน ประกอบด้วย การจัดการตนเอง การคิดขั้นสูง การสื่อสาร การรวมพลังทำงานเป็นทีม การเป็นพลเมืองที่เข้มแข็ง และการอยู่ร่วมกับธรรมชาติและวิทยาการ อย่างยั่งยืน และสมรรถนะเฉพาะ 6 ด้าน ประกอบด้วย การแก้ปัญหาทางคณิตศาสตร์ การสื่อสารและนำเสนอ ทางคณิตศาสตร์ การให้เหตุผลทางคณิตศาสตร์ การสร้างข้อสรุปทั่วไปและขยายแนวคิด การคิดสร้างสรรค์ และ การใช้เครื่องมือในการเรียนรู้ ดังภาพ 1

ตัวแปรต้น

ตัวแปรตาม



ภาพ 1 กรอบแนวคิดการวิจัย

วิธีดำเนินการวิจัย

การศึกษาสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ของโรงเรียนสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษาพิจิตร มีวิธีดำเนินการดังต่อไปนี้

1. ประชากร กลุ่มตัวอย่าง และผู้ให้ข้อมูลหลัก

1.1 ประชากร ได้แก่ ครูผู้สอนคณิตศาสตร์ระดับชั้นประถมศึกษาปีที่ 6 โรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษาพิจิตร เขต 1 2 และ 3 ประจำปีภาคการศึกษาที่ 2 การศึกษา 2567 จำนวน 523 คน

1.2 กลุ่มตัวอย่าง ได้แก่ ครูผู้สอนคณิตศาสตร์ระดับชั้นประถมศึกษาปีที่ 6 โรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษาพิจิตร เขต 1 2 และ 3 ประจำปีภาคการศึกษาที่ 2 การศึกษา 2567 จำนวน 222 คน กำหนดขนาดกลุ่มตัวอย่างโดยใช้ตาราง Krejcie and Morgan (Krejcie and Morgan, 1970) และได้มาโดยวิธีการสุ่มแบบแบ่งชั้นภูมิ โดยใช้สำนักงานเขตพื้นที่การศึกษาและขนาดโรงเรียนเป็นชั้นภูมิ

1.3 กลุ่มผู้ให้ข้อมูลหลัก จำนวน 7 คน ประกอบด้วยคณาจารย์สังกัดสถาบันอุดมศึกษา จำนวน 3 คน ครูผู้สอนรายวิชาคณิตศาสตร์ระดับชั้นประถมศึกษาปีที่ 6 ในโรงเรียนนำร่องการใช้หลักสูตรฐานสมรรถนะ จำนวน 3 คน และศึกษานิเทศก์ที่รับผิดชอบโรงเรียนในสังกัดที่ใช้หลักสูตรฐานสมรรถนะ จำนวน 1 คน

2. เครื่องมือที่ใช้ในการวิจัย เครื่องมือที่ใช้ในขั้นตอนการศึกษาและสำรวจสภาพปัญหา ประกอบด้วย 2 ชนิด ดังต่อไปนี้

2.1 แบบสอบถามสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 โดยแบ่งออกเป็น 3 ตอน ได้แก่ ตอนที่ 1 ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม มีลักษณะเป็นแบบตรวจสอบรายการ (Check List) ประกอบด้วย วิทยฐานะ ประสบการณ์ในการทำงาน และขนาดโรงเรียนที่สังกัด และตอนที่ 2 แบบสอบถามสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะ ในรายวิชาคณิตศาสตร์ ของครูผู้สอนชั้นประถมศึกษาปีที่ 6 มีลักษณะเป็นแบบมาตราส่วนประมาณค่า 5 ระดับ จำนวน 44 ข้อ ประกอบด้วย 1) สมรรถนะหลัก 6 สมรรถนะ จำนวน 28 ข้อ และ 2) สมรรถนะเฉพาะ 6 สมรรถนะ จำนวน 16 ข้อ โดยมีค่าดัชนีความสอดคล้อง (IOC) ตั้งแต่ 0.80 ถึง 1.00 ค่าอำนาจจำแนกตามวิธี Item-Total Correlation ตั้งแต่ 0.75 ถึง 0.94 ค่าความเชื่อมั่นตามวิธีสัมประสิทธิ์แอลฟาของคอนบราค เท่ากับ 0.99

2.2 แบบบันทึกการสนทนากลุ่ม เกี่ยวกับสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 โดยกำหนดประเด็นการสนทนากลุ่มไว้ 3 ประเด็น ได้แก่ สภาพปัญหาด้านความรู้ความเข้าใจ ด้านการออกแบบการเรียนการสอนและการประเมินผลสมรรถนะผู้เรียน และด้านการสร้างเครื่องมือประเมินผลการเรียนรู้ฐานสมรรถนะ มีค่าดัชนีความสอดคล้อง เท่ากับ 1.00 ทุกข้อ

3. การเก็บรวบรวมข้อมูล ได้ดำเนินการเก็บรวบรวมข้อมูลทั้งข้อมูลเชิงปริมาณและข้อมูลเชิงคุณภาพ ดังต่อไปนี้

3.1 การสอบถามสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 โดยการส่งหนังสือขอความอนุเคราะห์ไปยังสำนักงานเขตพื้นที่การศึกษาประถมศึกษากาฬสินธุ์ทั้งสามเขตพื้นที่ ภายหลังเมื่อได้รับอนุญาตจากสำนักงานเขตพื้นที่การศึกษาแล้ว ผู้วิจัยได้ดำเนินการประสานกับโรงเรียนโดยตรง เพื่อขอความอนุเคราะห์และขอความร่วมมือผู้บริหารสถานศึกษาได้มอบหมายให้ครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ที่เป็นกลุ่มตัวอย่างเป็นผู้ตอบแบบสอบถาม ในช่วงวันที่ 25 กุมภาพันธ์ พ.ศ. 2568 ถึง วันที่ 31 มีนาคม พ.ศ. 2568 โดยใช้ช่องทางออนไลน์ผ่าน Google form ได้รับแบบสอบถามกลับคืนมาจำนวน 222 ชุด คิดเป็นร้อยละ 100

3.2 การสนทนากลุ่ม เกี่ยวกับสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 โดยผ่านระบบออนไลน์ ในวันที่ 20 มีนาคม พ.ศ. 2568

4. การวิเคราะห์ข้อมูล

4.1 การวิเคราะห์สภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 โดยการวิเคราะห์ค่าเฉลี่ย (M) ส่วนเบี่ยงเบนมาตรฐาน (SD) และการเปรียบเทียบค่าเฉลี่ยของสภาพปัญหาการประเมินการเรียนรู้ฐานสมรรถนะ จำแนกตามวิทยฐานะ ประสบการณ์การทำงาน และขนาดโรงเรียนที่สังกัด ด้วยการวิเคราะห์สถิติทดสอบเอฟ (F-test) กรณีที่พบความแตกต่าง ดำเนินการวิเคราะห์เปรียบเทียบรายคู่ด้วยวิธี LSD (Least Significant Difference)

การแปลความความหมายของค่าเฉลี่ย ใช้เกณฑ์การแปล ดังนี้ (Srisa-ard, 2017)
(1-1.50 น้อยที่สุด, 1.51-2.50 น้อย, 2.51-3.50 ปานกลาง, 3.51-4.50 มาก, 4.51-5.00 มากที่สุด)

4.2 การวิเคราะห์สภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 จากการสนทนากลุ่ม โดยการวิเคราะห์เนื้อหา จัดกลุ่มเนื้อหาตามประเด็นที่กำหนด ได้แก่ 1) ด้านความรู้ความเข้าใจ 2) ด้านการออกแบบการเรียนการสอนและการประเมินผลสมรรถนะผู้เรียน และ 3) ด้านการสร้างเครื่องมือประเมินผลการเรียนรู้ฐานสมรรถนะ

ผลการวิจัย

1. ผลการวิเคราะห์ข้อมูลทั่วไปของผู้ตอบแบบสอบถาม วิทยฐานะ พบว่า ครูผู้ช่วย 27 คน (12.16%) ครู คศ. 1 67 คน (30.18%) ครู คศ. 2 63 คน (28.38%) และ คศ. 3 65 (29.28%) ประสบการณ์การทำงาน พบว่า ไม่เกิน 5 ปี 75 คน (33.78%) 6-10 ปี 65 คน (29.28%) 11-15 ปี 26 คน (11.71%) และ 16 ปี ขึ้นไป 56 คน (25.23%) ขนาดสถานศึกษา พบว่า ขนาดเล็ก 150 คน (67.57%) ขนาดกลาง 66 คน (29.73%) ขนาดใหญ่ 5 คน (2.25%) และขนาดใหญ่พิเศษ 1 คน (0.45%)

2. สภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ปรากฏดังนี้

2.1 ด้านสมรรถนะหลัก ผลการวิเคราะห์สภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะ ด้านสมรรถนะหลัก ปรากฏ ดังตาราง 1

ตาราง 1 ผลการวิเคราะห์สภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะ ด้านสมรรถนะหลัก

รายการประเมิน	<i>M</i>	<i>SD</i>	แปลผล
การจัดการตนเอง	3.47	1.05	ปานกลาง
การคิดขั้นสูง	3.40	1.05	ปานกลาง
การสื่อสาร	3.37	1.13	ปานกลาง
การรวมพลังทำงานเป็นทีม	3.45	1.26	ปานกลาง
การเป็นพลเมืองที่เข้มแข็ง	3.48	1.41	ปานกลาง
การอยู่ร่วมกับธรรมชาติและวิทยาการอย่างยั่งยืน	3.42	1.12	ปานกลาง
โดยรวม	3.43	1.12	ปานกลาง

จากตาราง 1 พบว่า สภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะ ด้านสมรรถนะหลัก โดยรวมอยู่ในระดับปานกลาง ($M = 3.43$, $SD = 1.12$) เมื่อพิจารณาเป็นรายด้านพบว่าอยู่ในระดับปานกลางทุกด้าน โดยด้านที่มีสภาพปัญหามากที่สุด คือ ด้านการเป็นพลเมืองที่เข้มแข็ง ($M = 3.48$, $SD = 1.41$) และด้านที่มีสภาพปัญหาน้อยที่สุดคือ ด้านการสื่อสาร ($M = 3.37$, $SD = 1.13$)

2.2 ด้านสมรรถนะเฉพาะ ผลการวิเคราะห์สภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะ ด้านสมรรถนะเฉพาะ ปรากฏ ดังตาราง 2

ตาราง 2 ผลการวิเคราะห์สภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะ ด้านสมรรถนะเฉพาะ

รายการประเมิน	M	SD	แปลผล
การแก้ปัญหาทางคณิตศาสตร์	3.38	0.95	ปานกลาง
การสื่อสารและนำเสนอทางคณิตศาสตร์	3.41	0.96	ปานกลาง
การให้เหตุผลทางคณิตศาสตร์	3.35	1.01	ปานกลาง
การสร้างข้อสรุปทั่วไปและขยายแนวคิด	3.26	0.89	ปานกลาง
การคิดสร้างสรรค์	3.29	0.93	ปานกลาง
การใช้เครื่องมือในการเรียนรู้	3.30	0.95	ปานกลาง
โดยรวม	3.33	0.86	ปานกลาง

จากตาราง 2 สภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะ ด้านสมรรถนะเฉพาะ โดยรวมอยู่ในระดับปานกลาง ($M = 3.33$, $SD = 0.86$) เมื่อพิจารณาเป็นรายด้านพบว่าอยู่ในระดับปานกลางทุกด้าน ด้านที่มีสภาพปัญหามากที่สุดคือ ด้านการสื่อสารและนำเสนอทางคณิตศาสตร์ ($M = 3.41$, $SD = 0.96$) ส่วนด้านที่มีสภาพปัญหาน้อยที่สุด คือ การสร้างข้อสรุปทั่วไปและขยายแนวคิด ($M = 3.26$, $SD = 0.89$)

3. ผลการเปรียบเทียบสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ปรากฏดังนี้

3.1 ผลการเปรียบเทียบสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ที่มีวิทยฐานะแตกต่างกัน ปรากฏ ดังตาราง 3

ตาราง 3 ผลการวิเคราะห์เปรียบเทียบสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ที่มีวิทยฐานะแตกต่างกัน

องค์ประกอบ	แหล่งความแปรปรวน	SS	df	MS	F	p-value
การประเมินผล	ระหว่างกลุ่ม	7.92	3	2.640	2.157	.094
การเรียนรู้	ภายในกลุ่ม	266.84	218	1.224		
สมรรถนะหลัก	รวม	274.76	221			
การประเมินผล	ระหว่างกลุ่ม	13.47	3	4.490	6.475*	.000
การเรียนรู้	ภายในกลุ่ม	151.16	218	.0693		
สมรรถนะเฉพาะ	รวม	164.63	221			

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 3 พบว่า ครูที่มีวิทยฐานะแตกต่างกันมีสภาพปัญหาการประเมินผลการเรียนรู้สมรรถนะหลักไม่แตกต่างกัน ส่วนสภาพปัญหาการประเมินผลการเรียนรู้สมรรถนะเฉพาะมีสภาพปัญหาแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

สภาพปัญหาการประเมินผลสมรรถนะเฉพาะ มีความแตกต่างกัน 4 คู่ ได้แก่ ครูผู้ช่วยกับครู ค.ศ. 2 ครูผู้ช่วยกับครู ค.ศ. 3 ครู ค.ศ. 1 กับครู ค.ศ. 2 และครู ค.ศ. 1 กับครู ค.ศ. 3 นอกนั้นไม่แตกต่างกัน ปรากฏดังตาราง 4

ตาราง 4 ผลการวิเคราะห์เปรียบเทียบความแตกต่างค่าเฉลี่ยรายคู่จำแนกตามวิทยฐานะ

องค์ประกอบ	วิทยฐานะ	M	ครูผู้ช่วย	ครู ค.ศ. 1	ครู ค.ศ. 2	ครู ค.ศ. 3
การประเมิน	ครูผู้ช่วย	3.623	-	0.003	0.512*	0.489*
ผลสมรรถนะเฉพาะ	ครู ค.ศ. 1	3.620		-	0.509*	0.486*
	ครู ค.ศ. 2	3.110			-	-0.024
	ครู ค.ศ. 3	3.134				-

* มีนัยสำคัญทางสถิติที่ระดับ .05

Levene Statistic สมรรถนะเฉพาะ 2.175, p-value. 093

3.2 ผลการเปรียบเทียบสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอน ในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ที่มีประสบการณ์การทำงานแตกต่างกัน ปรากฏ ดังตาราง 5

ตาราง 5 ผลการวิเคราะห์เปรียบเทียบสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอน ในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ที่มีประสบการณ์การทำงานแตกต่างกัน

องค์ประกอบ	แหล่งความแปรปรวน	SS	df	MS	F	p-value
การประเมินผล	ระหว่างกลุ่ม	8.438	3	2.813	2.302	.078
การเรียนรู้	ภายในกลุ่ม	266.324	218	1.222		
สมรรถนะหลัก	รวม	274.762	221			
การประเมินผล	ระหว่างกลุ่ม	16.120	3	5.373	7.888*	.000
การเรียนรู้	ภายในกลุ่ม	148.507	218	.681		
สมรรถนะเฉพาะ	รวม	164.627	221			

* มีนัยสำคัญทางสถิติที่ระดับ .05

จากตาราง 5 พบว่า ครูที่มีประสบการณ์การทำงานแตกต่างกัน มีสภาพปัญหาการประเมินผลการเรียนรู้สมรรถนะหลักไม่แตกต่างกัน ส่วนสภาพปัญหาการประเมินผลการเรียนรู้สมรรถนะเฉพาะมีสภาพปัญหาแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

เมื่อพิจารณาผลการเปรียบเทียบรายคู่ พบว่า สภาพปัญหาการประเมินผลการเรียนรู้สมรรถนะเฉพาะแตกต่างกัน 3 คู่ ได้แก่ ไม่เกิน 5 ปี กับ 11-15 ปี ไม่เกิน 5 ปี กับ 16 ปีขึ้นไป และ 6-10 ปี กับ 16 ปีขึ้นไป นอกนั้นไม่แตกต่างกัน ปรากฏดังตาราง 6

ตาราง 6 ผลการวิเคราะห์เปรียบเทียบความแตกต่างค่าเฉลี่ยรายคู่จำแนกตามประสบการณ์การทำงาน

องค์ประกอบ	ประสบการณ์การทำงาน	M	ไม่เกิน 5 ปี	6-10 ปี	11-15 ปี	16 ปีขึ้นไป
การประเมิน	ไม่เกิน 5 ปี	3.615	-	.181	.543*	.654*
ผลสมรรถนะ	6-10 ปี	3.434		-	.361	.472*
เฉพาะ	11-15 ปี	3.072			-	.111
	16 ปีขึ้นไป	2.961				-

* มีนัยสำคัญทางสถิติที่ระดับ .05

Levene Statistic สมรรถนะเฉพาะ 2.077, p-value. 104

3.3 ผลการเปรียบเทียบสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ที่สังกัดโรงเรียนที่มีขนาดแตกต่างกัน ปรากฏ ดังตาราง 7

ตาราง 7 ผลการวิเคราะห์เปรียบเทียบสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ตามขนาดโรงเรียนที่สังกัดแตกต่างกัน

องค์ประกอบ	แหล่งความแปรปรวน	SS	df	MS	F	P-value
การประเมินผล	ระหว่างกลุ่ม	5.507	3	1.836	1.486	.219
การเรียนรู้	ภายในกลุ่ม	269.255	218	1.235		
สมรรถนะหลัก	รวม	274.762	221			
การประเมินผล	ระหว่างกลุ่ม	.582	3	.194	.258	.856
การเรียนรู้	ภายในกลุ่ม	164.045	218	.752		
สมรรถนะเฉพาะ	รวม	164.627	221			

จากตาราง 7 พบว่า ครูที่สังกัดโรงเรียนที่มีขนาดแตกต่างกัน 4 ขนาด ได้แก่ ขนาดใหญ่พิเศษ ขนาดใหญ่ ขนาดกลาง และขนาดเล็ก มีสภาพปัญหาการประเมินผลการเรียนรู้สมรรถนะหลักและสมรรถนะเฉพาะไม่แตกต่างกัน

4. สภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 จากการสนทนากลุ่มของผู้ให้ข้อมูลหลัก นำเสนอโดยไม่จัดเรียงตามระดับของสภาพปัญหา ปรากฏดังนี้

4.1 ด้านความรู้ความเข้าใจ พบว่า ครูผู้สอนยังขาดความเข้าใจเกี่ยวกับสมรรถนะต่าง ๆ ที่หลักสูตรกำหนด ทำให้เกิดความไม่ชัดเจนในการดำเนินการประเมินผลสมรรถนะหลักและสมรรถนะเฉพาะ ขาดความเข้าใจในการออกแบบตัวชี้วัดที่มีความซับซ้อน ส่งผลให้ครูผู้สอนไม่ได้ดำเนินการประเมินผลตามตัวชี้วัดดังกล่าว

4.2 ด้านการออกแบบการเรียนการสอนและการประเมินผลสมรรถนะผู้เรียน ครูผู้สอนต้องใช้เวลาและความละเอียดในการดำเนินการ จึงส่งผลให้ครูผู้สอนส่วนใหญ่ไม่ได้ดำเนินการจัดทำรายละเอียดดังกล่าวในแผนการจัดการเรียนรู้

4.3 ด้านการสร้างเครื่องมือประเมินผลการเรียนรู้ฐานสมรรถนะ การสร้างเครื่องมือประเมินผลให้ครอบคลุมทุกสมรรถนะและมีคุณภาพเป็นเรื่องยากสำหรับครู และการกำหนดประเด็นการประเมิน เกณฑ์การประเมิน และคำอธิบายในเครื่องมือประเมินไม่ครอบคลุมและไม่ชัดเจน ทำให้การประเมินผลไม่ครอบคลุมและขาดความเที่ยงตรงตามสมรรถนะที่ต้องการประเมิน

สรุปผลการวิจัย

1. สภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ทั้งด้านสมรรถนะหลักและสมรรถนะเฉพาะ โดยรวมอยู่ในระดับปานกลาง เมื่อพิจารณาเป็นรายด้านพบว่าอยู่ในระดับปานกลางทุกด้าน

2. ผลการเปรียบเทียบสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 จำแนกตามตัวแปรวิทยฐานะ ประสบการณ์การทำงาน และขนาดโรงเรียนที่สังกัด สรุปได้ดังนี้

2.1 ครูที่มีวิทยฐานะแตกต่างกันมีสภาพปัญหาการประเมินผลสมรรถนะหลักไม่แตกต่างกัน ส่วนสภาพปัญหาการประเมินผลสมรรถนะเฉพาะ มีสภาพปัญหาแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

2.2 ครูที่มีประสบการณ์การทำงานแตกต่างกัน มีสภาพปัญหาการประเมินผลสมรรถนะหลักไม่แตกต่างกัน ส่วนสภาพปัญหาการประเมินผลสมรรถนะเฉพาะ มีสภาพปัญหาแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05

2.3 ครูที่สังกัดโรงเรียนที่มีขนาดแตกต่างกัน มีสภาพปัญหาการประเมินผลสมรรถนะหลักและสมรรถนะเฉพาะไม่แตกต่างกัน

3. สภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 จากการสนทนากลุ่มผู้ให้ข้อมูลหลัก ปรากฏดังนี้

3.1 ด้านความรู้ความเข้าใจ พบว่า ครูผู้สอนยังขาดความเข้าใจเกี่ยวกับสมรรถนะต่าง ๆ ที่หลักสูตรกำหนด ทำให้เกิดความไม่ชัดเจนในการดำเนินการประเมินผลสมรรถนะหลักและสมรรถนะเฉพาะ ขาดความเข้าใจในการออกแบบตัวชี้วัดที่มีความซับซ้อน ส่งผลให้ครูผู้สอนไม่ได้ดำเนินการประเมินผลตามตัวชี้วัดดังกล่าว

3.2 ด้านการออกแบบการเรียนการสอนและการประเมินผลสมรรถนะผู้เรียน พบว่า ครูผู้สอนต้องใช้เวลาและความละเอียดในการดำเนินการ จึงส่งผลให้ครูผู้สอนส่วนใหญ่ไม่ได้ดำเนินการจัดทำรายละเอียดดังกล่าวในแผนการจัดการเรียนรู้

3.3 ด้านการสร้างเครื่องมือประเมินผลการเรียนรู้ฐานสมรรถนะ พบว่า การสร้างเครื่องมือประเมินผลให้ครอบคลุมทุกสมรรถนะและมีคุณภาพเป็นเรื่องยากสำหรับครู และการกำหนดประเด็น

การประเมิน เกณฑ์การประเมิน และคำอธิบายในเครื่องมือไม่ครอบคลุมและไม่ชัดเจน ทำให้การประเมินผลไม่ครอบคลุมและขาดความเที่ยงตรงตามสมรรถนะที่ต้องการประเมิน

อภิปรายผล

จากผลการศึกษาสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 มีประเด็นสำคัญที่ควรนำสู่การอภิปรายผล ดังต่อไปนี้

1. สภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ทั้งด้านสมรรถนะหลักและสมรรถนะเฉพาะ โดยรวมอยู่ในระดับปานกลาง เมื่อพิจารณาเป็นรายด้านพบว่าอยู่ในระดับปานกลางทุกด้าน ทั้งนี้อาจเนื่องมาจากหลักสูตรที่โรงเรียนใช้เป็นหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 เป็นหลักสูตรที่กำหนดมาตรฐานการเรียนรู้เป็นผลลัพธ์ที่ต้องการให้เกิดกับผู้เรียน และมีตัวชี้วัดขั้นปีไว้ มาตรฐานการเรียนรู้และตัวชี้วัดมีจำนวนมาก ซ้ำซ้อน และอิงเนื้อหาค่อนข้างมาก ดังที่สำนักงานคณะกรรมการการศึกษาขั้นพื้นฐานได้กล่าวไว้ (Office of the Basic Education Commission, n.d.) ซึ่งโรงเรียนที่เป็นกลุ่มตัวอย่างในการวิจัยครั้งนี้ยังไม่ได้ใช้หลักสูตรการศึกษาฐานสมรรถนะ ทั้งนี้หลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 มีการกล่าวถึงสมรรถนะไว้เพียงเล็กน้อย คือ “สมรรถนะที่สำคัญของผู้เรียน มุ่งพัฒนาผู้เรียนให้มีคุณภาพตามมาตรฐานการเรียนรู้ ซึ่งการพัฒนาให้ผู้เรียนบรรลุมาตรฐานการเรียนรู้ที่กำหนดนั้นจะช่วยให้ผู้เรียนเกิดสมรรถนะที่สำคัญ 5 ประการ...” ตลอดจนแนวทางการประเมินผลการเรียนรู้ก็มุ่งประเมินตามตัวชี้วัดการเรียนรู้เพื่อสะท้อนสมรรถนะที่สำคัญ และคุณลักษณะอันพึงประสงค์ของผู้เรียน (Ministry of Education, 2008) ซึ่งครูในโรงเรียนดังกล่าวยังไม่ได้ดำเนินการประเมินผลการเรียนรู้ฐานสมรรถนะตามแนวทางหลักสูตรการศึกษาฐานสมรรถนะที่อยู่ระหว่างการนำร่องการใช้หลักสูตร จึงอาจส่งผลให้ครูสะท้อนสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะในระดับปานกลาง นอกจากนี้ อาจขึ้นกับลักษณะของผู้ตอบแบบสอบถามที่มักเกิดจากความรู้สึกว่ามีสภาพปัญหามีอยู่จริง แต่ไม่รุนแรงมากจนกระทบต่อการทำหน้าที่ประจำวันมากนัก มีทั้งส่วนที่ดีและส่วนที่ต้องปรับปรุง ไม่อยากแสดงความคิดเห็นที่รุนแรงมากเกินไป หรือผู้ตอบยังไม่มีข้อมูลเพียงพอ ผู้ตอบอาจไม่เข้าใจปัญหาดีพอ หรือไม่มีประสบการณ์ตรงมากนัก ทำให้เลือกตอบกลาง ๆ เพื่อแสดงความสมดุลหรือเป็นทางเลือกที่ปลอดภัย ไม่สุดโต่งเกินไป

2. ครูที่มีวิทยฐานะ ประสบการณ์การทำงาน และขนาดโรงเรียนที่สังกัดแตกต่างกัน มีสภาพปัญหาการประเมินผลสมรรถนะหลักไม่แตกต่างกัน ทั้งนี้อาจเนื่องมาจาก สมรรถนะหลักเป็นสมรรถนะไม่อิงเนื้อหา แต่สอดแทรกและบูรณาการอยู่ในกระบวนการเรียนการสอนในทุกกลุ่มสาระการเรียนรู้ หรือสาระการเรียนรู้ต่าง ๆ การประเมินจึงเน้นกระบวนการและการปฏิบัติตามสภาพจริงของผู้เรียน ด้วยวิธีการประเมินที่หลากหลาย โดยอาศัยข้อมูลเชิงประจักษ์จากสังเกตพฤติกรรม การกระทำหรือการปฏิบัติจริงของผู้เรียน ดังนั้นครูทุกคนจึงสามารถดำเนินการพัฒนาสมรรถนะหลักไปด้วยกันได้ พร้อมประเมินผลเพื่อตรวจสอบความก้าวหน้าและสรุปผลการพัฒนา ดังนั้นครูที่มีวิทยฐานะ ประสบการณ์การทำงาน และขนาดโรงเรียน

ที่สังกัดต่างกัน จึงมีสภาพปัญหาการประเมินสมรรถนะหลักไม่แตกต่างกัน สอดคล้องกับแนวคิดของ Office of the Education Council (2021) Office of the Basic Education Commission (2022a) และ Kongpun and Meiang (2026) ที่กล่าวว่า สมรรถนะหลักเป็นสมรรถนะที่กำหนดให้เป็นพื้นฐานที่นักเรียนทุกคนต้องได้รับการพัฒนาให้เป็นความสามารถติดตัวเมื่อจบการศึกษา มีลักษณะเป็นสมรรถนะข้ามหรือคร่อมรายวิชา สามารถพัฒนาให้เกิดขึ้นแก่ผู้เรียนได้ในสาระการเรียนรู้ต่าง ๆ ที่หลากหลาย หรือสามารถนำไปประยุกต์ใช้ในการพัฒนาผู้เรียนให้เรียนรู้สาระต่าง ๆ ได้ดีขึ้น และมีลักษณะเป็น Content-Free คือ ไม่ขึ้นกับเนื้อหาสาระของศาสตร์ใด ๆ อย่างไรก็ตามสมรรถนะหลักโดยตัวมันเองไม่ได้ปราศจากความรู้ แต่ความรู้ที่เป็นองค์ประกอบของสมรรถนะหลักจะเป็นองค์ความรู้เชิงกระบวนการ (Procedural Knowledge) ซึ่งเป็นชุดของขั้นตอนหรือการปฏิบัติเพื่อดำเนินการให้บรรลุเป้าหมายของสมรรถนะนั้น ๆ เป็นได้ทั้งกระบวนการที่ใช้เฉพาะศาสตร์หรือบูรณาการข้ามศาสตร์ ครูจึงสามารถนำไปใช้ออกแบบการจัดการเรียนรู้ในกลุ่มสาระ/รายวิชา/กิจกรรมต่าง ๆ ได้อย่างหลากหลาย และสามารถช่วยกันพัฒนาสมรรถนะเหล่านี้ให้เกิดขึ้นกับผู้เรียนได้

3. ครูที่มีวิทยฐานะแตกต่างกัน และประสบการณ์การทำงานแตกต่างกัน มีสภาพปัญหาการประเมินผลสมรรถนะเฉพาะแตกต่างกันอย่างมีนัยสำคัญทางสถิติที่ระดับ .05 แต่พบว่ามีสภาพปัญหาการประเมินผลสมรรถนะหลักไม่แตกต่างกัน ดังประเด็นการอภิปรายข้อ 2 นั้น แสดงให้เห็นว่าคุณลักษณะของสมรรถนะหลักและสมรรถนะเฉพาะที่ต่างกันมีผลให้ครูมีสภาพปัญหาการประเมินผลแตกต่างกัน โดยเฉพาะคุณลักษณะของสมรรถนะเฉพาะที่อิงเนื้อหาทางคณิตศาสตร์ กล่าวคือเป็นการแสดงออกทางพฤติกรรมและการกระทำในการประยุกต์ใช้ความรู้ ทักษะ และคุณลักษณะต่าง ๆ ของนักเรียนในการทำงานหรือแก้ไขปัญหาทางคณิตศาสตร์ให้ประสบผลสำเร็จเกณฑ์ที่กำหนดไว้ ตัวอย่างเช่น สมรรถนะด้านการแก้ปัญหาทางคณิตศาสตร์ (mathematical problem solving) เป็นความสามารถของนักเรียนในการมองเห็นปัญหาและแก้ปัญหาในชีวิตจริงผ่านลงมือกระทำและสะท้อนผลจากประสบการณ์ รวมทั้งมีความมุ่งมั่นและเห็นคุณค่าของการใช้คณิตศาสตร์ในการแก้ปัญหาประกอบด้วยตัวบ่งชี้ได้แก่ 1) ความอยากรู้อยากเห็นสามารถมองเห็นปัญหาทางคณิตศาสตร์ ในชีวิตจริงด้วยมุมมองของตนเอง 2) แก้ปัญหาในชีวิตจริงผ่านการลงมือแก้ปัญหาในสถานการณ์ต่าง ๆ และ เรียนรู้คณิตศาสตร์ผ่านการสะท้อนความคิด (reflect) จากประสบการณ์ 3) มีความมุ่งมั่นในการทำความเข้าใจและแก้ปัญหาทางคณิตศาสตร์ และ 4) ตระหนักและเห็นคุณค่าของการใช้คณิตศาสตร์ในการแก้ปัญหา เป็นต้น ซึ่งสมรรถนะเฉพาะและตัวบ่งชี้เหล่านี้ยังไม่ปรากฏในหลักสูตรหรือเอกสารประกอบหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 และครูยังไม่เคยดำเนินการประเมินผลการเรียนรู้ตามสมรรถนะเฉพาะเหล่านี้

4. ครูที่สังกัดโรงเรียนที่มีขนาดแตกต่างกันมีสภาพปัญหาการประเมินผลสมรรถนะเฉพาะไม่แตกต่างกัน ทั้งนี้อาจเนื่องมาจากบริบทของโรงเรียนในสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษากาฬสินธุ์ทั้ง 3 เขตพื้นที่ ยังไม่ใช่เขตพื้นที่นวัตกรรมในการนำร่องการใช้หลักสูตรสถานศึกษาฐานสมรรถนะ ประกอบกับในการดำเนินการประเมินผลการเรียนรู้ของผู้เรียน ครูยึดหลักแนวทางการปฏิบัติของหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 โดยอิงมาตรฐานการเรียนรู้และตัวชี้วัดตามที่หลักสูตรกำหนด (Office of the Basic Education Commission, 2014) ส่งผลให้ครูที่สังกัดโรงเรียนที่มีขนาดแตกต่างกันมีสภาพปัญหาการประเมินผลสมรรถนะเฉพาะไม่แตกต่างกัน

5. สภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ที่เป็นข้อมูลเชิงคุณภาพจากการสนทนากลุ่ม พบว่า 1) ด้านความรู้ความเข้าใจ ครูผู้สอนยังขาดความเข้าใจเกี่ยวกับสมรรถนะต่าง ๆ ที่หลักสูตรกำหนด ทำให้เกิดความไม่ชัดเจนในการดำเนินการประเมินผลสมรรถนะหลักและสมรรถนะเฉพาะ ขาดความเข้าใจในการออกแบบตัวชี้วัดที่มีความซับซ้อน ส่งผลให้ครูผู้สอนไม่ได้ดำเนินการประเมินผลตามตัวชี้วัดดังกล่าว 2) ด้านการออกแบบการเรียนรู้และการสอนและการประเมินผลสมรรถนะผู้เรียน ครูผู้สอนต้องใช้เวลาและความละเอียดในการดำเนินการ จึงส่งผลให้ครูผู้สอนส่วนใหญ่ไม่ได้ดำเนินการจัดทำรายละเอียดดังกล่าวในแผนการจัดการเรียนรู้ และ 3) ด้านการสร้างเครื่องมือประเมินผลการเรียนรู้ฐานสมรรถนะ การสร้างเครื่องมือประเมินผลให้ครอบคลุมทุกสมรรถนะและมีคุณภาพเป็นเรื่องยากสำหรับครู และการกำหนดประเด็นการประเมิน เกณฑ์การประเมิน และคำอธิบายในเครื่องมือประเมินไม่ครอบคลุมและไม่ชัดเจน ทำให้การประเมินผลไม่ครอบคลุมและขาดความเที่ยงตรงตามสมรรถนะที่ต้องการวัด ทั้งนี้อาจเนื่องมาจาก การดำเนินการประเมินผลการเรียนรู้ตามเอกสารประกอบหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 แนวปฏิบัติการประเมินผลการเรียนรู้ได้กล่าวถึงการประเมินสมรรถนะสำคัญของผู้เรียนไว้เพียงเล็กน้อยว่า “การประเมินสมรรถนะที่สำคัญควรใช้วิธีการประเมินที่เน้นการปฏิบัติ และบูรณาการอยู่ในกระบวนการเรียนการสอน ไม่ควรแยกประเมินต่างหากอีก แต่ทั้งนี้สถานศึกษาต้องตรวจสอบว่าผู้เรียนเกิดสมรรถนะสำคัญตามที่หลักสูตรกำหนดหรือไม่” (Ministry of Education, 2010) ซึ่งยังไม่มีรายละเอียดเกี่ยวกับแนวปฏิบัติการประเมินสมรรถนะที่ชัดเจน ประกอบกับแผนพัฒนาการศึกษาขั้นพื้นฐาน ระยะ 5 ปี (พ.ศ. 2566-2570) และแผนปฏิบัติการประจำปีงบประมาณ พ.ศ. 2568 ของสำนักงานเขตพื้นที่การศึกษาประถมศึกษากาฬสินธุ์ ทั้ง 3 เขต พบว่า แผนงาน/โครงการ/กิจกรรม การพัฒนาศักยภาพครูและบุคลากรทางการศึกษา ด้านการประเมินผลการเรียนรู้ฐานสมรรถนะยังไม่ชัดเจน จะมีที่เกี่ยวข้องบ้าง ตัวอย่างเช่น การพัฒนากระบวนการจัดการเรียนรู้เชิงรุก (Active Learning) เพื่อส่งเสริมผู้เรียนให้มีคุณลักษณะและทักษะในศตวรรษที่ 21 การพัฒนาครูในด้านการประเมินผลเพื่อพัฒนาการเรียนรู้ของผู้เรียน (assessment for learning) ด้วยวิธีการที่หลากหลาย เพื่อส่งเสริมการเรียนรู้เป็นรายบุคคล (personalized learning) และการพัฒนาหลักสูตรเพื่อส่งเสริมสมรรถนะผู้เรียนในศตวรรษที่ 21 เป็นต้น (Kalasin Primary Educational Area Office 1, 2023; Kalasin Primary Educational Area Office 2, 2024; Kalasin Primary Educational Area Office 3, 2023) และอีกประการหนึ่งคือ การนำร่องการใช้หลักสูตรการศึกษาฐานสมรรถนะก็ยังอยู่ในวงจำกัดเฉพาะโรงเรียนเขตพื้นที่นวัตกรรม (Education Sandbox) เท่านั้น จึงอาจเป็น

สาเหตุที่ครูยังไม่ได้รับการพัฒนาด้านการประเมินผลการเรียนรู้ฐานสมรรถนะ ส่งผลให้ครูยังขาดความรู้ความเข้าใจและความสามารถในการออกแบบการประเมินผลการเรียนรู้ฐานสมรรถนะ

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

1.1 จากผลการวิจัยที่พบว่า ครูที่มีวิทยฐานะและประสบการณ์การทำงานแตกต่างกันมีสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะด้านสมรรถนะเฉพาะแตกต่างกัน ดังนั้น หน่วยงานที่รับผิดชอบในการพัฒนาศักยภาพครูควรนำสารสนเทศเกี่ยวกับตัวแปรวิทยฐานะและประสบการณ์การทำงานไปใช้ประกอบในการออกแบบการจัดกิจกรรมการพัฒนาครูด้านการประเมินผลการเรียนรู้ฐานสมรรถนะในรายวิชาคณิตศาสตร์ เพื่อเพิ่มประสิทธิภาพและประสิทธิผลของการพัฒนาครู

1.2 จากผลการวิจัยที่พบว่า ครูผู้สอนยังมีสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ทั้งด้านความรู้ความเข้าใจ ด้านการออกแบบการเรียนการสอน และการประเมินผลสมรรถนะผู้เรียน และด้านการสร้างเครื่องมือประเมินผลการเรียนรู้ฐานสมรรถนะ ดังนั้น หน่วยงานที่เกี่ยวข้องในระดับนโยบายควรกำหนดนโยบายและแผนพัฒนาครูอย่างครอบคลุมทุกด้านให้ชัดเจน และเร่งสื่อสารและถ่ายทอดลงสู่หน่วยงานระดับปฏิบัติเพื่อดำเนินการพัฒนาศักยภาพครูและบุคลากรที่เกี่ยวข้องในการรองรับการเปลี่ยนแปลงด้านหลักสูตรฐานสมรรถนะในอนาคตได้อย่างมีประสิทธิภาพ

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

การศึกษาสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูผู้สอนในรายวิชาคณิตศาสตร์ ชั้นประถมศึกษาปีที่ 6 ในครั้งนี้ ประชากรเป็นครูสังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษาภาพลันธุ์ทั้ง 3 เขต ซึ่งสังกัดโรงเรียนที่ไม่ใช่โรงเรียนนำร่องการใช้หลักสูตรสถานศึกษาฐานสมรรถนะ อาจเป็นเหตุผลหนึ่งที่ทำให้ครูสะท้อนปัญหาดังกล่าวอยู่ในระดับปานกลางทั้งโดยรวมและรายด้าน เนื่องจากยังไม่มี การนำสู่การปฏิบัติอย่างจริงจัง จึงยังไม่สะท้อนปัญหาไปในทิศทางใดทิศทางหนึ่งอย่างชัดเจน ดังนั้น จึงควรศึกษาสภาพปัญหาการประเมินผลการเรียนรู้ฐานสมรรถนะของครูในโรงเรียนนำร่องการใช้หลักสูตรสถานศึกษาฐานสมรรถนะ เพื่อให้ได้ข้อมูลสภาพปัญหาจากการปฏิบัติจริงของครู

Reference

Krejcie, R. V., & Morgan, D. W. (1970). Determining sample size for research activities.

Educational and Psychological Measurement, 30(3), 607–610.

<https://doi.org/10.1177/001316447003000308>

Translate Thai References

Government Gazette. (2019, April 30). *Education Sandbox Act, B.E. 2562 (2019)*. *Government Gazette*, 136(56A), pp. 102–120. (in Thai)

- Government Gazette. (2022, December 20). *Announcement of the Prime Minister's Office Regarding the Designation of Education Sandbox. Government Gazette, 139*(Special Issue 296), p. 8. (in Thai)
- Government Gazette. (2024, May 9). *Announcement of the Prime Minister's Office Regarding the Designation of Education Sandbox. Government Gazette, 141*(Special Issue 130), p. 1. (in Thai)
- Independent Commission for Education Reform. (n.d.). *Special report no. 12: Competency-based curriculum and instruction*. (in Thai)
- Kalasin Primary Educational Service Area Office 1. (2023). *Five-year basic education development plan (2023–2027)*. Policy and Planning Group, Kalasin Primary Educational Service Area Office 1. (in Thai)
- Kalasin Primary Educational Service Area Office 2. (2024). *Annual action plan for fiscal year 2025*. Policy and Planning Group, Kalasin Primary Educational Service Area Office 2. (in Thai)
- Kalasin Primary Educational Service Area Office 3. (2023). *Five-year education development plan (2023–2027)*. Policy and Planning Group, Kalasin Primary Educational Service Area Office 3. (in Thai)
- Kongpun, A., & Meiang, S. (2026). Assessment and evaluation of core competencies in the competency-based curriculum for primary school students. *Journal of Public Administration and Interdisciplinary Studies for Society, 2*(1), 1–15. (in Thai)
- Ministry of Education. (2008). *Basic Education Core Curriculum B.E. 2551 (A.D. 2008)*. Agricultural Cooperative Federation of Thailand Printing House. (in Thai)
- Ministry of Education. (2010). *Guidelines for learning measurement and assessment based on the Basic Education Core Curriculum B.E. 2551 (A.D. 2008)*. Agricultural Cooperative Federation of Thailand Printing House. (in Thai)
- Office of the Basic Education Commission. (n.d.). *(Draft) Manual for implementing the basic education curriculum framework B.E. 2565*. Ministry of Education. (in Thai)
- Office of the Basic Education Commission. (2014). *Guidelines for measuring and evaluating learning outcomes according to the Basic Education Core Curriculum B.E. 2551 (2008) (4th ed.)*. Agricultural Cooperative Federation of Thailand Printing House. (in Thai)

- Office of the Basic Education Commission. (2022a). *(Draft) Manual for implementing the basic education curriculum framework: Primary education level*. Documentation for the workshop on analysis, review, and revision of the draft Basic Education Curriculum Framework, August 28–31, 2022, Royal City Hotel, Bangkok. (in Thai)
- Office of the Basic Education Commission. (2022b). *(Draft) Basic Education Curriculum Framework B.E. 2565: Primary education level*. Ministry of Education. (in Thai)
- Office of the Education Council. (2020). *Active competency-based learning management*. 21st Century. (in Thai)
- Office of the Education Council. (2021). *The core competency framework of students at the basic education level*. Prikwan Graphic. (in Thai)
- Office of the Education Council. (2025). *Enhancing learner potential through competency-based school curriculum*. Office of the Education Council. (in Thai)
- Srisa-ard, B. (2017). *Basic Research* (10th ed.). Suviriyasarn. (in Thai)

การพัฒนาครูเชิงระบบเพื่อยกระดับสมรรถนะความฉลาดรู้ด้านวิทยาศาสตร์ของนักเรียน
ชั้นประถมศึกษาปีที่ 6 ตามกรอบการประเมิน PISA 2025 ด้วยรูปแบบ PCACK-4D: การวิจัย
เชิงวินิจฉัยและวิเคราะห์ความต้องการจำเป็น

Systemic Teacher Development for Enhancing Scientific Literacy Competency
of Grade 6 Students Based on the PISA 2025 Framework Using the PCACK-4D
Model: Diagnostic Research and Needs Assessment

จตุพล แสงสุข^{1*} และ สติติ คลาดโรค์²

Jatupol Saensuk^{1*} and Sathit Kladrok²

(Received: March 12, 2026; Revised: May 8, 2026; Accepted: May 11, 2026)

บทคัดย่อ

การวิจัยครั้งนี้มีวัตถุประสงค์เพื่อ 1) วินิจฉัยปัญหาการจัดการเรียนรู้เชิงวิกฤตจากผลการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) วิชาวิทยาศาสตร์ ชั้นประถมศึกษาปีที่ 6 ย้อนหลัง 3 ปีการศึกษา (พ.ศ. 2564–2566) 2) วิเคราะห์ข้อสอบจุดวิกฤตตามกรอบ Revised Bloom's Taxonomy และสมรรถนะความฉลาดรู้ด้านวิทยาศาสตร์ตามกรอบการประเมิน PISA 2025 และ 3) วิเคราะห์ความต้องการจำเป็นและนำเสนอแนวทางการพัฒนาครูเชิงระบบด้วยรูปแบบ PCACK-4D โดยใช้รูปแบบการวิจัยแบบผสมผสานประเภทการวิจัยเชิงวินิจฉัยและวิเคราะห์ความต้องการจำเป็น กลุ่มเป้าหมายคือผู้บริหารสถานศึกษาและครูผู้สอนวิชาวิทยาศาสตร์ระดับชั้นประถมศึกษาปีที่ 6 จำนวน 22 คน จาก 11 โรงเรียน ในศูนย์พัฒนาคุณภาพการศึกษาแห่งหนึ่ง สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษาชัยภูมิ เขต 1 คัดเลือกด้วยการสุ่มแบบเจาะจง เครื่องมือวิจัยประกอบด้วยโปรแกรม PROFILE แบบวิเคราะห์ Revised Bloom's Taxonomy × PISA 2025 Matrix และแบบสัมภาษณ์กึ่งโครงสร้าง (IOC = 0.67–1.00) วิเคราะห์ข้อมูลด้วยคะแนนมาตรฐานที่ร้อยละ การวิเคราะห์เนื้อหา และการตรวจสอบข้อมูลแบบสามเส้า

ผลการวิจัยพบว่า 1) ทุกมาตรฐานการเรียนรู้อยู่ในระดับ L1 โดยมาตรฐานที่เป็นปัญหาซ้ำซ้อน และมีคะแนนต่ำต่อเนื่อง ได้แก่ มาตรฐาน ว 1.2 (โครงสร้างและหน้าที่ของสิ่งมีชีวิต) ว 2.1 (สมบัติของสาร) และ ว 2.3 (แรงและพลังงาน) ซึ่งมีผลคุณน้าหนักสูงสุด (10 คะแนน) 2) จุดวิกฤตสูงสุดพบในมาตรฐาน ว 2.1 ตัวชี้วัด ป.4/1 ปีการศึกษา 2564 มีอัตราการตอบถูกเฉลี่ยเพียงร้อยละ 6.33 ซึ่งต่ำกว่าค่าความน่าจะเป็น

¹ ศึกษาพิเศษ สำนักงานเขตพื้นที่การศึกษาประถมศึกษาชัยภูมิ เขต 1

² ผู้อำนวยการโรงเรียนบ้านซับชมภู สำนักงานเขตพื้นที่การศึกษาประถมศึกษาชัยภูมิ เขต 1

¹ Supervisor, Chaiyaphum Primary Educational Service Area Office 1

² Director of Ban Sab Chom Phu School Chaiyaphum Primary Educational Service Area Office 1

* Corresponding Author E-mail: jatupon@esdc.go.th

จากการเดา สะท้อนภาวะวิกฤตในระดับการคิดแบบการประเมินค่าและสมรรถนะการออกแบบและประเมินกระบวนการสืบเสาะหาความรู้ตามกรอบ PISA 2025 3) การวิเคราะห์ความต้องการการจำเป็นพบช่องว่างเชิงระบบ ใน 4 มิติ ได้แก่ การบริหารจัดการ บริบท การสอน และความรู้ นำไปสู่การออกแบบชุดนวัตกรรม "6-Stack Innovation" จำนวน 6 เล่ม ภายใต้รูปแบบ PCACK-4D เพื่อทำหน้าที่เป็นความทรงจำองค์กรสำหรับรักษาทักษะความรู้และส่งผลต่อแนวทางการจัดการเรียนรู้เชิงระบบอย่างยั่งยืน แม้มีการเปลี่ยนแปลงบุคลากร

คำสำคัญ การพัฒนาครูเชิงระบบ ความฉลาดรู้ด้านวิทยาศาสตร์ PISA 2025 รูปแบบ PCACK-4D

Abstract

This study aimed to: 1) diagnose critical learning management issues based on O-NET Science results for Grade 6 students over three academic years (2021–2023); 2) analyze critical test items using the Revised Bloom's Taxonomy and the PISA 2025 scientific literacy competency framework; and 3) conduct a needs assessment and propose guidelines for systemic teacher development using the PCACK-4D model. A mixed-methods design employing diagnostic research and needs assessment was utilized. The target group comprised 22 participants—11 school administrators and 11 Grade 6 science teachers—from 11 schools under a Quality Development Center, Chaiphaphum Primary Educational Service Area Office 1, selected through purposive sampling. Research instruments included the PROFILE program, a Revised Bloom's Taxonomy × PISA 2025 Matrix analysis form, and semi-structured interviews (IOC = 0.67–1.00). Data were analyzed using T-scores, percentages, content analysis, and data triangulation.

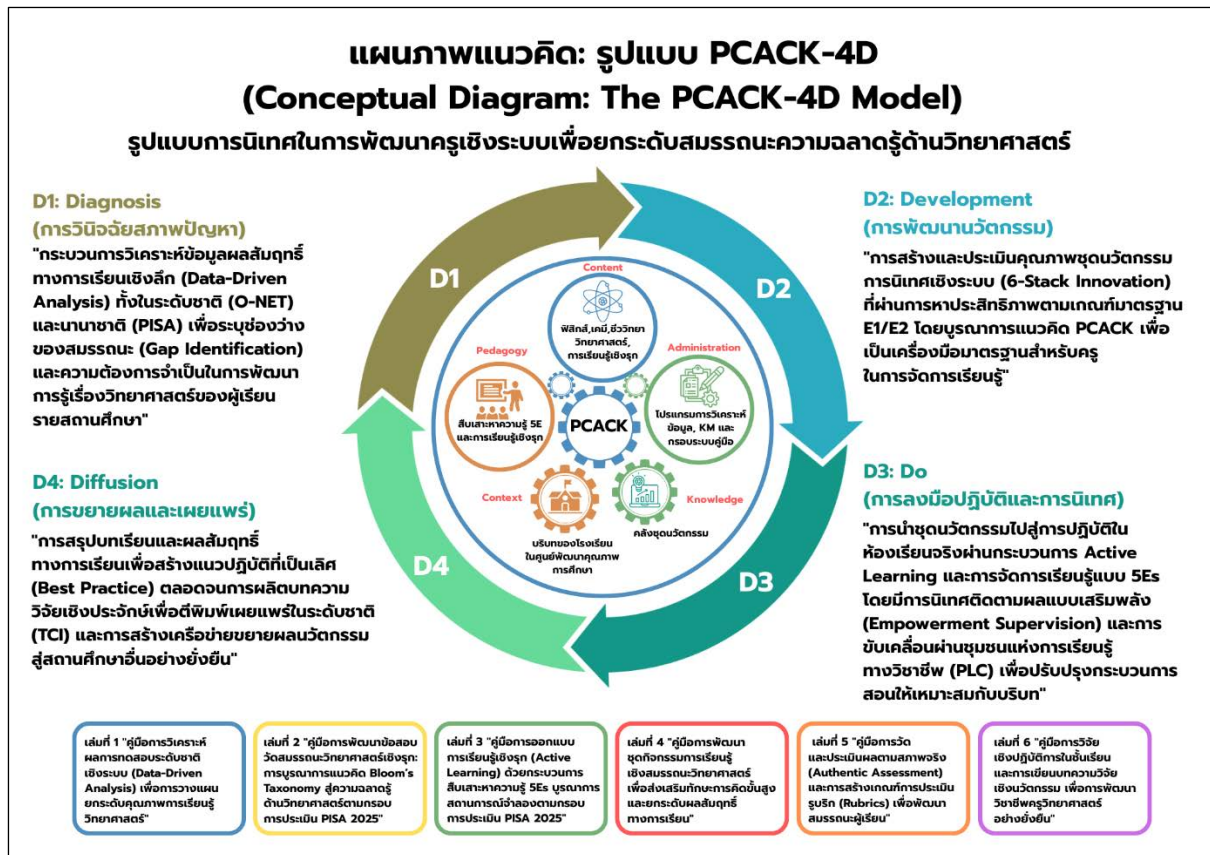
The findings revealed that: 1) All learning standards were classified at Level L1; the most critically recurring standards were SC 1.2, SC 2.1, and SC 2.3, each yielding the highest weighted score (10 points). 2) The most critical test item was Standard SC 2.1, Indicator P.4/1 (academic year 2021), with an average correct-response rate of only 6.33%—below chance-level performance—indicating severe deficiencies in the Evaluating cognitive level and the PISA 2025 competency of designing and evaluating scientific inquiry processes. 3) The qualitative needs assessment identified systemic gaps across four dimensions—administration, context, pedagogy, and knowledge—leading to the design of six Knowledge Stack innovations under the PCACK-4D model, serving as Organizational Memory to sustain knowledge transfer and systemic learning management despite personnel turnover.

Keywords: systemic teacher development, scientific literacy, PISA 2025, PCACK-4D model

บทนำ

การยกระดับสมรรถนะความฉลาดรู้ด้านวิทยาศาสตร์” (scientific literacy) ตามกรอบการประเมิน PISA 2025 เป็นเป้าหมายสำคัญของการศึกษาในศตวรรษที่ 21 ที่ประกอบด้วยสมรรถนะหลัก 3 ประการ ได้แก่ 1) การอธิบายปรากฏการณ์เชิงวิทยาศาสตร์ 2) การออกแบบ ประเมิน และแปลความหมายข้อมูลจาก กระบวนการสืบเสาะอย่างมีวิจารณญาณ และ 3) การศึกษา ประเมิน และใช้ข้อมูลทางวิทยาศาสตร์เพื่อการ ตัดสินใจ (Organisation for Economic Co-operation and Development (OECD), 2023) ซึ่งสอดคล้อง กับนโยบายของกระทรวงศึกษาธิการที่มุ่งขับเคลื่อนการจัดการเรียนรู้เชิงสมรรถนะเพื่อเตรียมความพร้อมสู่ มาตรฐานสากล (Office of the Private Education Commission, 2024) อย่างไรก็ตาม เมื่อพิจารณาบริบท การศึกษาไทยผ่านผลการทดสอบทางการศึกษาระดับชาตินัยพื้นฐาน (O-NET) วิชาวิทยาศาสตร์ ชั้น ประถมศึกษาปีที่ 6 ย้อนหลัง 3 ปีการศึกษา (พ.ศ. 2564-2566) ในโรงเรียนเครือข่ายศูนย์พัฒนาคุณภาพ การศึกษา สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษาชัยภูมิ เขต 1 พบวิกฤตการณ์เรียนรู้อย่างมีนัยสำคัญ ในมาตรฐาน ว 1.2, ว 2.1 และ ว 2.3 ซึ่งมีคะแนนเฉลี่ยต่ำกว่าระดับจังหวัดอย่างต่อเนื่อง โดยผลการวินิจฉัย ด้วยโปรแกรม PROFILE สะท้อนคุณภาพห้องเรียนในภาพรวมอยู่ในระดับ L1 ซึ่งหมายถึงภาวะสัมฤทธิ์ผลต่ำ ยกชั้นเรียน (Techakaew & Deeme, 2023) ซึ่งปรากฏการณ์ดังกล่าวชี้ชัดว่าผู้เรียนยังติดอยู่ใน “กับดัก ความจำ” (memorization trap) ทำให้ขาดทักษะการคิดขั้นสูง (Higher Order Thinking Skills: HOTS) ในระดับการวิเคราะห์และการประเมินค่าซึ่งเชื่อมโยงโดยตรงกับสมรรถนะความฉลาดรู้ด้านวิทยาศาสตร์ ตามกรอบการประเมิน PISA 2025 ที่ส่งผลให้การเรียนรู้แบบฐานเนื้อหา (content-based) ไม่สามารถตอบ โจทย์การพัฒนาสมรรถนะตามแนวคิด Revised Bloom’s Taxonomy ได้ โดยเฉพาะตัวชี้วัดที่มีความเป็น นามธรรมสูงและมีความซับซ้อนเกินกว่าทักษะการท่องจำ (Anderson & Krathwohl, 2001; Chiang Rai Primary Educational Service Area Office 1, 2024)

ผู้วิจัยในฐานะศึกษานิเทศก์และผู้บริหารสถานศึกษาได้วิเคราะห์ว่า การพัฒนาครูตามกรอบความรู้ ในเนื้อหาผนวกวิธีสอน (PCK) แบบดั้งเดิม (Shulman, 1986) อาจไม่เพียงพอต่อบริบทโรงเรียนขนาดเล็ก และ ขนาดกลางที่มีการโยกย้ายบุคลากรบ่อยครั้ง จนเกิดภาวะความไม่ต่อเนื่องทางวิชาการและ “ความรู้สูญหาย” (knowledge attrition) ผู้วิจัยจึงบูรณาการแนวคิดเชิงระบบ (general system theory) ของ Von Bertalanffy (1968) พัฒนาสู่ “รูปแบบการนิเทศ PCACK-4D Model” ซึ่งขยายขอบเขตไปสู่มิติด้านการบริหารจัดการ (Administration: A) บริบทเชิงพื้นที่ (Context: C) และการสร้างชุดความรู้ที่ยั่งยืน (Knowledge Stack: K) เพื่อสร้าง “ความทรงจำองค์กร” (organizational memory) ผ่านนวัตกรรมคู่มือการจัดการเรียนรู้ที่จะช่วย รักษาฐานสมรรถนะวิชาชีพครูให้คงอยู่กับสถานศึกษาอย่างยั่งยืน แม้มีการเปลี่ยนแปลงบุคลากร อันจะนำไปสู่ การยกระดับสมรรถนะผู้เรียนภายใต้บริบทที่มีข้อจำกัดสูงสุดได้อย่างแท้จริง



ภาพ 1 แผนภาพการทำงานตาม PCACK-4D Model

ความมุ่งหมายของการวิจัย

1. เพื่อวินิจฉัยปัญหาการจัดการเรียนรู้เชิงวิกฤตจากผลการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) วิชาวิทยาศาสตร์ ชั้นประถมศึกษาปีที่ 6 ย้อนหลัง 3 ปีการศึกษา (ปีการศึกษา 2564-2566)
2. เพื่อวิเคราะห์ข้อสอบจุดวิกฤตตามกรอบ Revised Bloom's Taxonomy และสมรรถนะความฉลาดรู้ด้านวิทยาศาสตร์ ตามกรอบการประเมิน PISA 2025
3. เพื่อวิเคราะห์ความต้องการจำเป็นและนำเสนอแนวทางการพัฒนาครูเชิงระบบด้วยรูปแบบ PCACK-4D

วิธีดำเนินการวิจัย

การวิจัยครั้งนี้ดำเนินการวิจัยเชิงวินิจฉัยและวิเคราะห์ความต้องการจำเป็น ซึ่งเป็นส่วนหนึ่งของโครงการวิจัยพหุระยะภายใต้รูปแบบ PCACK-4D โดยใช้รูปแบบการวิจัยแบบผสมผสานเชิงบูรณาการ (convergent mixed methods design) ผู้วิจัยดำเนินการเก็บและวิเคราะห์ข้อมูลเชิงปริมาณและเชิงคุณภาพในช่วงเวลาเดียวกัน (concurrent data collection) แล้วนำผลการวิเคราะห์ทั้งสองส่วนมาบูรณาการ (integration) เพื่ออธิบายปรากฏการณ์ร่วมกันอย่างเป็นระบบ โดยข้อมูลเชิงปริมาณจากผลการทดสอบ O-NET และการวิเคราะห์ด้วยโปรแกรม PROFILE ถูกใช้เพื่อระบุ "จุดวิกฤตเชิงโครงสร้าง" ของการจัดการเรียนรู้ ขณะที่ข้อมูลเชิงคุณภาพจากการสัมภาษณ์เชิงลึกถูกใช้เพื่ออธิบายสาเหตุเชิงบริบท กระบวนการจัดการ

เรียนรู้ และข้อจำกัดเชิงระบบที่อยู่เบื้องหลังผลสัมฤทธิ์ดังกล่าว จากนั้นผู้วิจัยจึงทำการเปรียบเทียบ ตรวจสอบความสอดคล้อง และสังเคราะห์ข้อค้นพบร่วมกัน (merging and interpretation) เพื่อระบุช่องว่างเชิงระบบ (systemic gap) ตามกรอบ PCACK และนำไปสู่การออกแบบแนวทางพัฒนาครูเชิงระบบ

1. กลุ่มเป้าหมาย

กลุ่มเป้าหมายที่ใช้ในการวิจัยครั้งนี้ คือ ผู้บริหารสถานศึกษาและครูผู้สอนวิชาวิทยาศาสตร์ระดับชั้นประถมศึกษาปีที่ 6 ในสถานศึกษาขนาดเล็กและขนาดกลางของศูนย์พัฒนาคุณภาพการศึกษาแห่งหนึ่ง สังกัดสำนักงานเขตพื้นที่การศึกษาประถมศึกษาชัยภูมิ เขต 1 จำนวน 11 แห่ง ซึ่งอยู่ในเขตพื้นที่รับผิดชอบการนิเทศของผู้วิจัย โดยผู้วิจัยใช้วิธีการเลือกแบบเจาะจง (purposive sampling) เพื่อให้ได้ข้อมูลที่สะท้อนสภาพการณ์ที่หลากหลาย โดยมีเกณฑ์การคัดเลือก ดังนี้ 1) เป็นสถานศึกษาที่มีบริบทใกล้เคียงกันในด้านขนาดโรงเรียน และจำนวนบุคลากร 2) มีความแตกต่างเชิงผลสัมฤทธิ์ โดยครอบคลุมทั้งโรงเรียนที่มีคะแนนเฉลี่ย O-NET วิชาวิทยาศาสตร์สูงกว่าและต่ำกว่าเกณฑ์เฉลี่ยระดับประเทศอย่างต่อเนื่อง 3 ปีการศึกษา เพื่อนำมาวิเคราะห์เปรียบเทียบหา "จุดวิกฤต" และ "ปัจจัยความสำเร็จ" และ 3) เป็นพื้นที่ที่ผู้วิจัยสามารถเข้าถึงข้อมูลเชิงลึกและนิเทศติดตามผลได้อย่างต่อเนื่องตามภาระหน้าที่และความสนใจในการพัฒนาเชิงระบบ โดยผู้ให้ข้อมูลสำคัญ ได้แก่ ผู้บริหารสถานศึกษา จำนวน 11 คน และครูผู้สอนวิชาวิทยาศาสตร์ชั้นประถมศึกษาปีที่ 6 จำนวน 11 คน รวมทั้งสิ้น 22 คน เพื่อร่วมกันวินิจฉัยปัญหาเชิงโครงสร้างตามกรอบ PCACK ในบริบทที่แตกต่างกัน

2. เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยในครั้งนี้ ประกอบด้วย 3 ชนิด ดังนี้ 1) โปรแกรม PROFILE สำหรับวิเคราะห์ผล O-NET โดยคำนวณค่าคะแนนมาตรฐานที่ (T-Score) เพื่อจัดระดับคุณภาพห้องเรียนเป็น 4 ระดับ (H1, H2, L1, L2) ตามกรอบของ Techakaew & Deeme (2023) 2) แบบวิเคราะห์ข้อสอบ (revised Bloom's Taxonomy × PISA 2025 Matrix) ซึ่งผ่านการตรวจสอบความเที่ยงตรงเชิงเนื้อหา (content validity) โดยผู้เชี่ยวชาญด้านการวัดผลและวิทยาศาสตร์ศึกษา 3 ท่าน พบว่ามีค่าดัชนีความสอดคล้อง (IOC) ระหว่าง 0.67–1.00 ผ่านเกณฑ์มาตรฐาน (IOC ≥ 0.50) ทุกรายการ และ 3) แบบสัมภาษณ์แบบกึ่งโครงสร้าง (semi-structured interview) สำหรับผู้บริหารและครู ซึ่งมีประเด็นสัมภาษณ์ 4 หมวดหลักตามกรอบ PCACK ได้แก่ 1) สภาพการบริหารจัดการข้อมูลผลสัมฤทธิ์ (administration) 2) บริบท และข้อจำกัดเชิงทรัพยากร (context) 3) วิธีการจัดการเรียนรู้และการวัดผล (pedagogy) และ 4) ความรู้เนื้อหาและมโนทัศน์ของครู (knowledge) แบบสัมภาษณ์ผ่านการตรวจสอบความเหมาะสมเชิงภาษาและโครงสร้างคำถามโดยผู้เชี่ยวชาญ 3 ท่าน (IOC = 0.67–1.00)

3. การเก็บรวบรวมข้อมูล

ผู้วิจัยดำเนินการเก็บข้อมูลจาก 2 แหล่ง ดังนี้ 1) ข้อมูลทุติยภูมิ โดยสกัดจากรายงานผล O-NET วิชาวิทยาศาสตร์ ชั้น ป.6 ย้อนหลัง 3 ปีการศึกษา (พ.ศ. 2564–2566) ฉบับที่ 2 (ค่าสถิติรายมาตรฐาน) และฉบับที่ 3 (ร้อยละการตอบถูกรายข้อ) จากระบบฐานข้อมูลของสถาบันทดสอบทางการศึกษาแห่งชาติ

(สทศ.) สำหรับโรงเรียนทั้ง 11 แห่ง และ 2) ข้อมูลปฐมภูมิ โดยดำเนินการสัมภาษณ์เชิงลึกกับกลุ่มเป้าหมาย โดยผู้วิจัยลงพื้นที่ด้วยตนเองระหว่างเดือนกรกฎาคม 2567 ใช้เวลาสัมภาษณ์คนละประมาณ 45-60 นาที บันทึกเสียงพร้อมจดบันทึก และถอดความภายใน 48 ชั่วโมงเพื่อความถูกต้องของข้อมูล

4. การวิเคราะห์ข้อมูล

ผู้วิจัยได้สรุปวิธีการวิเคราะห์ข้อมูลแยกตามความมุ่งหมายของการวิจัยและการตรวจสอบข้อมูลแบบสามเส้า เพื่อให้การวิเคราะห์ข้อมูลมีความเที่ยงตรงและครอบคลุม ดังตาราง 1

ตาราง 1 สรุปกระบวนการวิเคราะห์ข้อมูลแยกตามวัตถุประสงค์และการตรวจสอบข้อมูลแบบสามเส้า

ความมุ่งหมาย การวิจัย	เครื่องมือ/แหล่งข้อมูล	วิธีวิเคราะห์ข้อมูล	การตรวจสอบข้อมูล แบบสามเส้า
1. วินิจฉัยจุดวิกฤต การเรียนรู้จากผล O-NET วิทยาศาสตร์ ป.6 ย้อนหลัง 3 ปี (2564-2566)	(1) รายงาน O-NET ฉบับที่ 2 (2) โปรแกรม PROFILE (3) Excel	- Mean, SD, Percentage - Profile analysis (L1, L2, H2, H1) - frequency x weight analysis	Data Triangulation: เปรียบเทียบข้อมูล ต่างช่วงเวลา 3 ปี การศึกษา เพื่อยืนยัน จุดวิกฤตเชิงโครงสร้าง
2. วิเคราะห์ข้อสอบ จุดวิกฤตตามกรอบ Revised Bloom's Taxonomy และ PISA 2025	(1) รายงาน O-NET ฉบับที่ 3 (2) Excel (3) Revised Bloom's x PISA 2025 Matrix	- วิเคราะห์ร้อยละตอบถูก - วิเคราะห์ร้อยละ การตอบถูกเฉลี่ย ในแต่ละข้อของโรงเรียน ย้อนหลัง 3 ปีการศึกษา - content analysis จำแนกพฤติกรรมการคิด และสมรรถนะด้าน วิทยาศาสตร์	Investigator Triangulation: ผู้เชี่ยวชาญ 3 ท่าน ตรวจสอบ IOC และ ความสอดคล้องการ จำแนกข้อสอบ
3. วิเคราะห์ความ ต้องการจำเป็น ในการพัฒนาครู เชิงระบบด้วยรูปแบบ PCACK-4D	(1) แบบสัมภาษณ์ กึ่งโครงสร้างสำหรับครู (2) แบบสัมภาษณ์ กึ่งโครงสร้างสำหรับ ผู้บริหารสถานศึกษา	- content analysis - thematic analysis	Methodological Triangulation: เปรียบเทียบมุมมองครู และผู้บริหารเพื่อระบุ Systemic Gap ตาม กรอบ PCACK

การตรวจสอบข้อมูลแบบสามเส้าด้านข้อมูล (data triangulation) โดยการเปรียบเทียบผลการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) วิชาวิทยาศาสตร์ ระดับชั้นประถมศึกษาปีที่ 6 จากแหล่งเวลาที่แตกต่างกัน 3 ปีการศึกษา เพื่อ cross-check ให้เห็นความสอดคล้องของจุดวิกฤตทางการเรียนรู้ ยืนยันว่าปัญหาที่พบเป็นปัญหาเชิงโครงสร้างของระบบการจัดการเรียนรู้ที่ต้องได้รับการแก้ไขเร่งด่วน

การตรวจสอบข้อมูลแบบสามเส้าด้านผู้วิจัย (investigator triangulation) เพื่อลดความคลาดเคลื่อนจากการตีความข้อมูลของผู้วิจัยเพียงคนเดียว ผู้วิจัยได้นำผลการจำแนกพฤติกรรมในข้อสอบจุดวิกฤตตามกรอบ revised Bloom's x PISA 2025 Matrix ไปให้ผู้เชี่ยวชาญด้านการวัดผลและวิทยาศาสตร์ศึกษา จำนวน 3 ท่าน ตรวจสอบความเที่ยงตรงเชิงเนื้อหาและดัชนีความสอดคล้อง (IOC) เพื่อยืนยันความถูกต้องของการระบุจุดวิกฤตทางการเรียนรู้

การตรวจสอบข้อมูลแบบสามเส้าเชิงระเบียบวิธี (methodological triangulation) โดยเน้นไปที่การใช้เครื่องมือที่แตกต่างกัน (within-method) เพื่อมองปรากฏการณ์เดียวกันจากคนละมุมมอง ซึ่งตรวจสอบความสอดคล้องเชิงระบบโดยการเปรียบเทียบข้อมูลจากแบบสัมภาษณ์ผู้บริหารสถานศึกษา (เน้นมิตินโยบายและโครงสร้างการบริหาร) และแบบสัมภาษณ์ครูผู้สอน (เน้นมิติการปฏิบัติและมนทัศน์ทางวิทยาศาสตร์) เพื่อระบุช่องว่างเชิงระบบ (systemic gap) ที่ครอบคลุมทุกองค์ประกอบของ PCACK

ผลการวิจัย

1. ผลการวินิจฉัยปัญหาการจัดการเรียนรู้เชิงวิกฤตจากผลการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) วิชาวิทยาศาสตร์ ชั้นประถมศึกษาปีที่ 6 ย้อนหลัง 3 ปีการศึกษา (2564-2566)

จากการวิเคราะห์รายงานผลการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) วิชาวิทยาศาสตร์ ระดับชั้นประถมศึกษาปีที่ 6 ฉบับที่ 2 ค่าสถิติสำหรับโรงเรียนแยกตามมาตรฐานการเรียนรู้ย้อนหลัง 3 ปีการศึกษา (2564-2566) ด้วยโปรแกรม PROFILE ซึ่งเกิดจากการคิดคะแนน T-Score ของโรงเรียนเทียบกับของจังหวัดชัยภูมิ เพื่อให้มีความใกล้เคียงบริบทมากที่สุด โดยใช้สูตรในการคำนวณ ดังนี้

$$T = 50 + 10\left(\frac{X - \bar{X}}{S.D.}\right)$$

เมื่อ	X	คือ	ค่าเฉลี่ยของโรงเรียน
	$\bar{X}$	คือ	ค่าเฉลี่ยของจังหวัดชัยภูมิ
	$S.D.$	คือ	ส่วนเบี่ยงเบนมาตรฐานของจังหวัดชัยภูมิ

โดยแบ่งผลการวิเคราะห์ออกเป็น 4 กลุ่ม ดังนี้ 1) กลุ่มที่มีคะแนนเฉลี่ยมากกว่าจังหวัดชัยภูมิ จะแบ่งเป็น 2 ย่อย ได้แก่ H1 หมายถึง การจัดกิจกรรมการเรียนรู้ของครูมีประสิทธิภาพดีมากนักเรียนทุกกลุ่มสามารถเรียนรู้ได้ดี และ H2 หมายถึง การจัดกิจกรรมการเรียนรู้ของครูยังไม่มีประสิทธิภาพ นักเรียนบางกลุ่มสามารถเรียนรู้ได้ดี และ 2) กลุ่มที่มีคะแนนเฉลี่ยน้อยกว่าจังหวัดชัยภูมิ จะแบ่งเป็น 2 ย่อย ได้แก่ L1 หมายถึง เป็นมาตรฐานการเรียนรู้ที่ครูไม่ถนัด ทำให้นักเรียนเกิดความสับสนในเนื้อหา ครูควรปรับความคิดรวบยอด

เกี่ยวกับมาตรฐานนี้ และ L2 หมายถึง การจัดกิจกรรมการเรียนรู้ของครูยังไม่มีประสิทธิภาพนักเรียนทุกกลุ่มไม่สามารถเรียนรู้ได้ ผลการวิเคราะห์ พบว่า ทุกมาตรฐานการเรียนรู้ในระดับ L1 ผลตามตาราง 2

ตาราง 2 การวินิจฉัยมาตรฐานการเรียนรู้ที่เป็นปัญหาซ้ำซ้อนจากการทดสอบทางการศึกษาระดับชาตินิพื้นฐาน (O-NET) วิชาวิทยาศาสตร์ ชั้นประถมศึกษาปีที่ 6 ย้อนหลัง 3 ปีการศึกษา (2564-2566) ด้วยโปรแกรม PROFILE และโปรแกรม Excel

สาระที่	มาตรฐาน	รูปแบบจุดบกพร่องการเรียนรู้ ของปีการศึกษา			จำนวน (L1+L2)	ผลคูณ น้ำหนัก	สรุปผล การวินิจฉัย
		2564	2565	2566			
1 วิทยาศาสตร์ ชีวภาพ	ว 1.1	L1	H2	L1	2	6	ระดับ L1 (ปัญหาทั่วไป)
	ว 1.2	L1	L1	L2	3	10	ระดับ L1 (ปัญหาซ้ำซ้อน)
	ว 1.3	L1	L1	L1	3	9	ระดับ L1 (ปัญหาทั่วไป)
2 วิทยาศาสตร์ กายภาพ	ว 2.1	L2	L1	L1	3	10	ระดับ L1 (ปัญหาซ้ำซ้อน)
	ว 2.2	H2	L1	L1	2	6	ระดับ L1 (ปัญหาทั่วไป)
	ว 2.3	L2	L1	L1	3	10	ระดับ L1 (ปัญหาซ้ำซ้อน)
3 วิทยาศาสตร์ โลก และอวกาศ	ว 3.1	L1	L1	L1	3	9	ระดับ L1 (ปัญหาทั่วไป)
	ว 3.2	L1	L1	L1	3	9	ระดับ L1 (ปัญหาทั่วไป)
4 เทคโนโลยี	ว 4.2	L1	L1	H2	2	6	ระดับ L1 (ปัญหาทั่วไป)

หมายเหตุ คำนวณผลคูณน้ำหนักจากความถี่ของระดับคุณภาพ (L 1 = 3 , L 2 = 4) มาตรฐานที่มีผลคูณรวมสูงสุด (10 คะแนน) ถือเป็นปัญหาซ้ำซ้อนที่ต้องพัฒนาเร่งด่วน

2. ผลการวิเคราะห์ข้อสอบจุดวิกฤตตามกรอบ Revised Bloom's Taxonomy และสมรรถนะความฉลาดรู้ด้านวิทยาศาสตร์ ตามกรอบการประเมิน PISA 2025

ผู้วิจัยเจาะลึกสาระที่ 2 (ว 2.1 และ ว 2.3) ที่พบจุดวิกฤตหนาแน่น โดยวิเคราะห์รายละเอียดการตอบถูกรายข้อ (O-NET) ย้อนหลัง 3 ปี เพื่อระบุ จุดวิกฤตสูงสุด (highest critical point) พบในข้อสอบข้อที่ 20 (ว 2.1) มีอัตราการตอบถูกเฉลี่ยเพียง 6.33% (ปี 2564) และ 18.69% (ปี 2566) ซึ่งต่ำกว่าค่าความน่าจะเป็น

จากการเคา ซึ่งสะท้อนภาวะวิกฤตในการจัดการเรียนรู้และความยากลำบากของผู้เรียนต่อโจทย์ซับซ้อนประเภทเลือกตอบเชิงซ้อน (ใช่/ไม่ใช่) ทั้งนี้ กระบวนการจำแนกพฤติกรรมด้วย Revised Bloom x PISA Matrix ผ่านการตรวจสอบความถูกต้องโดยผู้เชี่ยวชาญ 3 ท่าน (investigator triangulation) มีค่า IOC ระหว่าง 0.67–1.00 ดังรายละเอียดในตาราง 3

ตาราง 3 การวิเคราะห์ข้อสอบจุดวิกฤตตามกรอบ Revised Bloom’s Taxonomy และสมรรถนะความฉลาดรู้ด้านวิทยาศาสตร์ ตามกรอบการประเมิน PISA 2025 (Revised Bloom’s Taxonomy x PISA 2025 Matrix)

ข้อที่	ปีการศึกษา	มาตรฐาน/ ตัวชี้วัด	ประเภท ตัวชี้วัด	ระดับการคิด (Revised Bloom’s Taxonomy)	สมรรถนะความ ฉลาดรู้ด้าน วิทยาศาสตร์ (PISA 2025)	ร้อยละ การ ตอบถูก เฉลี่ย
4	2564	ว 2.1 ป.4/1	ระหว่างทาง	การประยุกต์ใช้ (Applying)	สมรรถนะที่ 2 (Focus B)	48.34
	2565	ว 2.1 ป.4/1	ระหว่างทาง	การวิเคราะห์ (Analyzing)	สมรรถนะที่ 2 (Focus B)	51.99
	2566	ว 2.1 ป.4/3	ระหว่างทาง	การวิเคราะห์ (Analyzing)	สมรรถนะที่ 2 (Focus B)	40.71
5	2564	ว 2.1 ป.5/4	ปลายทาง	ความเข้าใจ (Understanding)	สมรรถนะที่ 1	14.09
	2565	ว 2.1 ป.6/1	ปลายทาง	การประยุกต์ใช้ (Applying)	สมรรถนะที่ 1	39.81
	2566	ว 2.1 ป.5/3	ระหว่างทาง	ความเข้าใจ (Understanding)	สมรรถนะที่ 1	41.85
6	2564	ว 2.1 ป.6/1	ปลายทาง	การประเมินค่า (Evaluating)	สมรรถนะที่ 1	61.96
	2565	ว 2.1 ป.5/2	ระหว่างทาง	ความเข้าใจ (Understanding)	สมรรถนะที่ 1	39.38
	2566	ว 2.1 ป.6/1	ปลายทาง	การประยุกต์ใช้ (Applying)	สมรรถนะที่ 1	61.84
20	2564	ว 2.1 ป.4/1	ระหว่างทาง	การประเมินค่า (Evaluating)	สมรรถนะที่ 2 (Focus A)	6.33
	2565	ว 2.1 ป.4/1	ระหว่างทาง	การวิเคราะห์ (Analyzing)	สมรรถนะที่ 2 (Focus B)	32.14
	2566	ว 2.1 ป.4/1	ระหว่างทาง	การวิเคราะห์ (Analyzing)	สมรรถนะที่ 2 (Focus B)	18.69

หมายเหตุ สมรรถนะความฉลาดรู้ด้านวิทยาศาสตร์ ตามกรอบการประเมิน PISA 2025 (Organisation for Economic Co-operation and Development (OECD), 2023) โดยผู้วิจัยจำแนกย่อยเพื่อการวินิจฉัย ดังนี้ สมรรถนะที่ 1 การอธิบายปรากฏการณ์ในเชิงวิทยาศาสตร์ สมรรถนะที่ 2 (Focus A) คือ การออกแบบและประเมินกระบวนการสืบเสาะหาความรู้ทางวิทยาศาสตร์ สมรรถนะที่ 2 (Focus B) คือ การแปลความหมายข้อมูลและใช้ประจักษ์พยานในเชิงวิทยาศาสตร์อย่างมีวิจารณญาณ และ สมรรถนะที่ 3 การศึกษาค้นคว้า ประเมิน และใช้ข้อมูลทางวิทยาศาสตร์เพื่อการตัดสินใจ โดยผลการวิเคราะห์เมทริกซ์ในตาราง 3 สะท้อนว่าข้อสอบมาตรฐาน ว 2.1 มีการยกระดับการวัดผลสู่ฐานสมรรถนะ โดยมุ่งเน้นระดับการคิดขั้นการวิเคราะห์ (analyzing) และการประเมินค่า (evaluating) ซึ่งสอดคล้องกับกรอบการประเมิน PISA 2025 ทั้งด้านการอธิบายปรากฏการณ์และการออกแบบการสืบเสาะ (ข้อที่ 6 และ 20) ซึ่งสอดคล้องเชิงโครงสร้างนี้ยืนยันมาตรฐานการวัดผลระดับสากลที่ท้าทายขีดความสามารถทางปัญญาของผู้เรียนในบริบทปัจจุบัน"

3. ผลการวิเคราะห์ความต้องการจำเป็นและนำเสนอแนวทางการพัฒนาครูเชิงระบบด้วยรูปแบบ PCACK-4D

ผู้วิจัยดำเนินการวิเคราะห์เนื้อหา (content analysis) และวิเคราะห์เชิงธีม (thematic analysis) จากการสัมภาษณ์เชิงลึกครูผู้สอนวิทยาศาสตร์ระดับชั้น ป.6 และผู้บริหารสถานศึกษา ผ่านกระบวนการถอดเทปแบบคำต่อคำ การกำหนดรหัสข้อมูล และการสังเคราะห์ประเด็นเชิงธีม เพื่อระบุช่องว่างเชิงระบบ (systemic gap) และนำไปสู่การออกแบบนวัตกรรม "6-stack innovation" ใน 4 มิติ ดังนี้

1) การบริหารจัดการข้อมูลและการตั้งเป้าหมาย (Administration: A) ซึ่งจากการศึกษาพบว่า สถานศึกษามีการใช้ข้อมูลผลการทดสอบทางการศึกษาระดับชาติด้านพื้นฐาน (O-NET) เป็นฐานในการกำหนดเป้าหมายเชิงปริมาณ แต่ยังขาดการวิเคราะห์ในเชิงลึกถึง "พฤติกรรมความคิดของผู้เรียน" (cognitive behaviors) ทำให้การนำข้อมูลไปใช้ยังไม่สามารถยกระดับสมรรถนะในระดับวิเคราะห์ได้จริง โดยเสียงสะท้อนจากครู: "เราเห็นคะแนนร้อยละเด็กตกสาระไหน แต่เรายังแยกแยะไม่ได้ว่าที่เขตก เพราะเขาจำเนื้อหาไม่ได้ หรือเขาตีความสถานการณ์ในโจทย์ไม่เป็น" (ครูผู้สอน, 9 กรกฎาคม 2567) และเสียงสะท้อนจากผู้บริหาร: "นโยบายคือต้องเพิ่มขึ้นร้อยละ 3 ทุกปี แต่วิธีการส่วนใหญ่ยังเป็นการติวเพื่อสอบ ไม่ใช่การปรับการเรียนเปลี่ยนการสอนที่เน้นสมรรถนะจริงๆ" (ผู้บริหารสถานศึกษา, 16 กรกฎาคม 2567)

2) บริบท และอุปสรรคเชิงทรัพยากร (Context: C) ที่สะท้อนปัญหาความเหลื่อมล้ำทางทรัพยากรส่งผลให้เกิดปรากฏการณ์ "ปฏิบัติการแห้ง" (dry lab) ซึ่งเป็นอุปสรรคสำคัญในการสร้างมนต์ทัศน์ที่ถูกต้องในวิชาวิทยาศาสตร์ โดยเฉพาะในมาตรฐาน ว 2.1 ที่เป็นนามธรรมสูง โดยเสียงสะท้อนจากครู: "อุปกรณ์การทดลองชำรุดและจำกัด ส่วนใหญ่จึงเป็นการอธิบายผ่านคลิปวิดีโอใน YouTube แทน เด็กไม่ได้สัมผัสของจริง ทำให้จินตนาการถึงปรากฏการณ์ได้ยาก" (ครูผู้สอน, 11 กรกฎาคม 2567) และเสียงสะท้อนจากผู้บริหาร: "อุปสรรคสำคัญคือโรงเรียนไม่มีห้องปฏิบัติการที่เป็นสัดส่วน ความต่อเนื่องของครูก็สำคัญ ครูผู้ช่วยพอเริ่มชำนาญก็ถึงเกณฑ์ย้าย ทำให้ความรู้ในการพัฒนาเด็กขาดช่วง" (ผู้บริหารสถานศึกษา, 17 กรกฎาคม 2567)

3) พฤติกรรมการจัดการเรียนรู้และกับดักความจำ (Pedagogy & Knowledge: P&K) ซึ่งพบภาวะ "กับดักความจำ" (memorization trap) อย่างเด่นชัดในชั้นเรียนวิทยาศาสตร์ แม้จะมีการรณรงค์เรื่อง active learning แต่ครูยังติดอยู่กับกระบวนการถ่ายทอดเนื้อหา (content-based) มากกว่าการสืบเสาะ (inquiry-based) โดยเสียงสะท้อนจากครู: "ความเข้าใจคลาดเคลื่อนเรื่องแรงและการละลายแก้ได้ยากมาก เพราะเด็กจำต่อๆ กันมา ครูเองก็กังวลว่าถ้ามันแต่ทำกิจกรรมสืบเสาะ เนื้อหาจะสอนไม่จบตามหลักสูตร" (ครูผู้สอน, 10 กรกฎาคม 2567) และเสียงสะท้อนจากผู้บริหาร: "ครูส่วนใหญ่ยังยึดเยียดคำตอบ ไม่ค่อยยืดหยุ่นรอให้เด็กสรุปผลเอง ในการออกข้อสอบก็ยังเน้นความจำมากกว่า เพราะมันง่ายและสะดวกต่อการตรวจ" (ผู้บริหารสถานศึกษา, 17 กรกฎาคม 2567)

4) ความท้าทายในการวัดและประเมินผลสมรรถนะระดับสูง โดยครูสะท้อนความต้องการจำเป็นอย่างเร่งด่วนในการพัฒนาทักษะการสร้างเครื่องมือวัดผลที่สอดคล้องกับมาตรฐานสากล (PISA) เพื่อก้าวข้ามการวัดผลแบบเดิม โดยเสียงสะท้อนจากครู: "ความท้าทายคือจะสร้างสถานการณ์อย่างไรให้เด็กได้วิเคราะห์ โดยไม่ยวบจนเด็กไม่อยากอ่าน และต้องไม่ไปคำตอบในตัวโจทย์เอง" (ครูผู้สอน, 10 กรกฎาคม 2567) และเสียงสะท้อนจากผู้บริหาร: "เราต้องการชุดกิจกรรมที่เป็นต้นแบบ หรือคู่มือที่ช่วยให้ครูเดินตามขั้นตอนได้ชัดเจน เพื่อให้เกิดมาตรฐานเดียวกันทั้งโรงเรียนและยั่งยืนแม้มีการย้ายบุคลากร" (ผู้บริหารสถานศึกษา, 16 กรกฎาคม 2567)

จากการวินิจฉัยข้างต้นแสดงให้เห็นว่า "ช่องว่างเชิงระบบ" (systemic gap) คือสาเหตุสำคัญของจุดวิกฤตในมาตรฐาน ว 2.1 ผู้วิจัยจึงได้นำเสนอแนวทางการพัฒนาครูเชิงระบบด้วยรูปแบบ PCACK-4D โดยการพัฒนาครูเชิงระบบ ในการวิจัยครั้งนี้ หมายถึง กระบวนการพัฒนาครูที่มองสถานศึกษาเป็นระบบที่มีองค์ประกอบเชื่อมโยงกัน ได้แก่ ครู ผู้บริหาร บริบทโรงเรียน และชุดความรู้ โดยมุ่งแก้ปัญหาเชิงโครงสร้างอย่างครอบคลุมและต่อเนื่อง มิใช่การอบรมแบบครั้งเดียวแต่เป็นการสร้างระบบนิเทศและความรู้ที่ยังคงอยู่กับสถานศึกษาแม้เมื่อบุคลากรเปลี่ยนแปลง (Von Bertalanffy, 1968) เพื่อเปลี่ยนผ่านจากการจัดการเรียนรู้ฐานเนื้อหาสู่ฐานสมรรถนะ โดยมุ่งสร้างความทรงจำขององค์กร (organizational memory) ที่จะช่วยพัฒนาที่ยั่งยืน แม้มีการเปลี่ยนแปลงบุคลากร ดังตาราง 4

ตาราง 4 ผลการวินิจฉัยสภาพปัญหาและวิเคราะห์ความต้องการจำเป็นสู่การพัฒนาชุดนวัตกรรมกรณีศึกษา เพื่อพัฒนาครูวิทยาศาสตร์ 6 เล่ม (6-Stack Innovation) ในระยะที่ 2 (Phase 2)

ข้อค้นพบจากการวิจัย (Phase 1 Findings)	องค์ประกอบ PCACK	นวัตกรรมที่สร้าง (6-Stack Innovation)	วัตถุประสงค์ ของคู่มือ
1. ปัญหาการขาดข้อมูลเชิง ประจักษ์: ข้อมูลผลสัมฤทธิ์ ยังไม่ถูกนำมาวิเคราะห์ราย สมรรถนะ เพื่อการวางแผน เชิงกลยุทธ์	A (Administration)	เล่มที่ 1 คู่มือการวิเคราะห์ผลการ ทดสอบระดับชาติเชิงระบบ (Data- Driven Analysis) เพื่อการวางแผน ยกระดับคุณภาพการเรียนรู้วิทยาศาสตร์	เพื่อบริหารจัดการข้อมูล และวินิจฉัยจุดวิกฤตของ ผู้เรียนเป็นรายบุคคล

ตาราง 4 (ต่อ)

ข้อค้นพบจากการวิจัย (Phase 1 Findings)	องค์ประกอบ PCACK	นวัตกรรมที่สร้าง (6-Stack Innovation)	วัตถุประสงค์ ของคู่มือ
2. ความไม่ชัดเจนใน สมรรถนะใหม่: ครูยังขาด ความเข้าใจนิยามและกรอบ การประเมิน PISA 2025 ในเชิงลึก	C (Content)	เล่มที่ 2 คู่มือการพัฒนาข้อสอบวัด สมรรถนะวิทยาศาสตร์เชิงรุก: การบูร ณาการแนวคิด Revised Bloom's Taxonomy สู่ความฉลาดรู้ด้าน วิทยาศาสตร์ ตามกรอบ PISA 2025	เพื่อสร้างความเข้าใจ ในเนื้อหาและสมรรถนะ ระดับสากลให้แก่ครู
3. กับดักความจำ (memorization trap): การสอนที่เน้นครูเป็น ศูนย์กลางและการบรรยาย เนื้อหา ทำให้นักเรียนไม่ได้ สรุปผลการสืบเสาะด้วย ตนเอง	P (Pedagogy)	เล่มที่ 3 คู่มือการออกแบบการเรียนรู้เชิง รุก (active learning) ด้วยกระบวนการ สืบเสาะ 5Es บูรณาการสถานการณ์ จำลองตามกรอบ PISA 2025	เพื่อเปลี่ยนบทบาทครู และกระตุ้นกระบวนการ สืบเสาะ (inquiry) เพื่อ ทำลายกับดักความจำ
4. จุดวิกฤตและมโนทัศน์ คลาดเคลื่อน (misconceptions): นักเรียนขาดทักษะการ ตีความข้อมูลในมาตรฐาน ว 2.1 และมีแนวคิดไม่ สอดคล้องกับหลัก วิทยาศาสตร์	K (Knowledge)	เล่มที่ 4 คู่มือการพัฒนาชุดกิจกรรมการ เรียนรู้เชิงสมรรถนะวิทยาศาสตร์ เพื่อ ส่งเสริมทักษะการคิดขั้นสูงและแก้ไข มโนทัศน์คลาดเคลื่อนในจุดวิกฤต	เพื่อแก้ปัญหาจุดวิกฤต ทางความรู้ด้วยชุดฝึก ปฏิบัติที่เน้นการ ปรับเปลี่ยนแนวคิด (conceptual change)
5. การประเมินผลไม่ตรงเป้า: เครื่องมือวัดผลยังไม่สะท้อน การคิดขั้นสูง (HOTS) และ ขาดเกณฑ์การประเมินที่ ชัดเจน	P (Pedagogy)	เล่มที่ 5 คู่มือการวัดและประเมินผลตาม สภาพจริง (authentic assessment) และการสร้างเกณฑ์การประเมินรูบริก (rubrics) เพื่อพัฒนาสมรรถนะผู้เรียน	เพื่อสร้างเครื่องมือ ประเมินผลที่สอดคล้อง กับรูปแบบข้อสอบ สมรรถนะขั้นสูง
6. ขาดความทรงจำองค์กร (organizational memory): ความรู้กระจกระบายและ สูญหายเมื่อมีการโยกย้าย บุคลากร ขาดความต่อเนื่อง เชิงระบบในสถานศึกษา	C (Context)	เล่มที่ 6 คู่มือการวิจัยเชิงปฏิบัติการใน ชั้นเรียนและการเขียนบทความวิจัยเชิง นวัตกรรม เพื่อการสร้างความยั่งยืนใน วิชาชีพครูวิทยาศาสตร์	เพื่อสร้างความยั่งยืนของ องค์ความรู้ (KM) และ รักษารูขี้อยู่ใน นวัตกรรมไว้ใน สถานศึกษา อย่างต่อเนื่อง

ข้อค้นพบจากการวิจัยถือเป็น "จุดเปลี่ยนเชิงยุทธศาสตร์" (strategic turning point) ที่นำไปสู่การพัฒนา นวัตกรรม Knowledge Stack 6 เล่ม ตามรูปแบบ PCACK-4D โดยจำแนกความเชื่อมโยงได้ 3 มิติสำคัญ ได้แก่ 1) มิติการบริหารและบริบท (A & C) สู่ความทรงจำองค์กร: เพื่อแก้ปัญหา ในภาวะความรู้สูญหาย (knowledge attrition) จากการโยกย้ายบุคลากร ผู้วิจัยจึงออกแบบคู่มือเล่มที่ 1 และ 6 ให้ทำหน้าที่เป็น

ความทรงจำองค์กร (organizational memory) เพื่อรักษาฐานสมรรถนะวิชาการให้คงอยู่กับสถานศึกษาอย่างต่อเนื่อง 2) มิติเนื้อหาและจุดวิกฤต (C) สู่การปิดช่องว่างทางปัญญา: จากข้อมูลเชิงประจักษ์ในมาตรฐาน ว 2.1 (ร้อยละการตอบถูก 6.33) ผู้วิจัยมุ่งออกแบบคู่มือเล่มที่ 2 และ 4 ให้เจาะลึกเนื้อหานามธรรมและการฝึกสมรรถนะระดับ “ประเมินค่า” (evaluating) เพื่ออุดช่องว่างทางปัญญา (learning gap) ตามกรอบการประเมิน PISA 2025 3) มิติวิธีสอน (P) สู่การเป็นผู้อำนวยการความสะดวกทางปัญญา: เพื่อทลายกับดักการสอนแบบบรรยาย โดยผู้วิจัยออกแบบคู่มือเล่มที่ 3 และ 5 เพื่อเปลี่ยนบทบาทครูสู่การเป็น “cognitive facilitator” ผ่านกระบวนการสืบเสาะ (5Es) ที่เน้นการสร้าง ความขัดแย้งทางปัญญาให้ผู้เรียนสร้างองค์ความรู้ด้วยตนเอง ดังนั้น การวินิจฉัยในระยะนี้จึงเป็นการเปลี่ยนวิกฤตการณ์เรียนรู้ให้เป็นยุทธศาสตร์นวัตกรรมที่มีฐานรองรับเชิงวิชาการที่เข้มแข็ง เพื่อเป็นหลักประกันในการยกระดับความฉลาดรู้ด้านวิทยาศาสตร์อย่างยั่งยืนในระยะการนำไปใช้ (do) และการขยายผล (diffusion) ต่อไป

อภิปรายผล

จากผลการวิจัยในครั้งนี้ ผู้วิจัยพบประเด็นสำคัญที่นำมาอภิปรายผลเชิงวิชาการเพื่อยกระดับสมรรถนะความฉลาดรู้ด้านวิทยาศาสตร์ตามแนวทาง PCACK-4D Model ดังนี้

1. วิกฤตการณ์สมรรถนะการคิดขั้นสูงในสาระวิทยาศาสตร์กายภาพ: จากกับดักความจำสู่ช่องว่างทางปัญญา

ผลการวินิจฉัยเชิงปริมาณเผยให้เห็นช่องว่างการเรียนรู้วิกฤตในมาตรฐาน ว 2.1 โดยเฉพาะทักษะการคิดขั้นสูง (HOTS) ระดับ “ประเมินค่า” (evaluating) ข้อมูลเชิงประจักษ์จากข้อสอบข้อที่ 20 ซึ่งมีอัตราการตอบถูกเพียงร้อยละ 6.33 สะท้อนถึงข้อจำกัดในการบูรณาการความรู้สู่สมรรถนะ “การประเมินและออกแบบการสืบเสาะหาความรู้ทางวิทยาศาสตร์” ตามกรอบการประเมิน PISA 2025 ซึ่งวิกฤตดังกล่าวชี้ชัดว่าผู้เรียนยังติดอยู่ในกับดักความจำ (memorization trap) ขาดทักษะการตรวจสอบ และวิพากษ์ซึ่งเป็นแกนกลางของความฉลาดรู้ด้านวิทยาศาสตร์ ปัญหารากเหง้านี้จึงถูกสังเคราะห์เป็นโจทย์ยุทธศาสตร์ในการพัฒนา “knowledge stack 6 เล่ม” ภายใต้รูปแบบ PCACK-4D เพื่อเปลี่ยนบทบาทครูสู่การเป็น “โค้ชทางปัญญา” (cognitive coach) ที่มุ่งทลายกับดักความจำและสร้างฐานสมรรถนะการคิดที่เท่าทันมาตรฐานสากลอย่างยั่งยืน

2. ช่องว่างระหว่าง “การปฏิบัติ” และ “การสรุปองค์ความรู้”: บทบาทที่ครอบงำของครู

จากการวิเคราะห์เชิงคุณภาพพบว่า ปัญหามีได้เกิดจากการขาดกิจกรรมทดลอง แต่เกิดจากความบกพร่องใน “กระบวนการสร้างความหมาย” (sense-making process) โดยครูมักเร่งรัดสรุปผลเพื่อประหยัดเวลา ทำให้ผู้เรียนขาดโอกาสในการฝึกสมรรถนะ “การแปลความหมายข้อมูล” (interpreting data) ตามกรอบ PISA 2025 (OECD, 2023) นอกจากนี้ การเร่งรัดกระบวนการดังกล่าวยังปิดกั้นโอกาสให้ผู้เรียนเผชิญกับความขัดแย้งทางปัญญา (cognitive conflict) ส่งผลให้เกิดโครงสร้างทางปัญญาที่ผิดพลาดในเนื้อหาที่เป็นนามธรรม เช่น การสับสนระหว่าง “การละลาย” และ “การหลอมเหลว” ในมาตรฐาน ว 2.1 ซึ่งสะท้อนว่าผู้เรียนยังไม่สามารถใช้ประจักษ์พยานเชิงวิทยาศาสตร์ (evidence-based) ในการสร้างองค์ความรู้ด้วยตนเองได้อย่างแท้จริง

3. บทสรุปเชิงข้อเสนอแนะจากการอภิปราย

ข้อค้นพบพบคะแนนความยากสูงสุดในข้อสอบข้อที่ 20 (ร้อยละ 6.33) ซึ่งวัดทักษะ "การประเมินค่า" (evaluating) และสมรรถนะการออกแบบการสืบเสาะหาความรู้ทางวิทยาศาสตร์ สะท้อนถึงวิกฤตความรู้เชิงกระบวนการ (procedural knowledge) ของผู้เรียนอย่างรุนแรง ผลการวิจัยชี้ชัดว่าสาเหตุหลักเกิดจากวัฒนธรรมการสอนแบบถ่ายทอดทางตรงที่ครูมักเป็นผู้สรุปผลการทดลองด้วยตนเอง ดังนั้นรูปแบบ PCACK-4D จึงมุ่งเน้นการปรับบทบาทครูสู่การเป็น "ผู้อำนวยความสะดวกทางปัญญา" (cognitive facilitator) ที่มีความสามารถในการ "รอคอยอย่างมีกลยุทธ์" (strategic waiting) เพื่อเปิดพื้นที่ให้ผู้เรียนได้เผชิญกับความขัดแย้งทางปัญญาและเชื่อมโยงประจักษ์พยานเชิงวิทยาศาสตร์ด้วยตนเอง นวัตกรรม knowledge stack 6 เล่ม จึงถูกออกแบบมาเพื่อเป็นเครื่องมือเชิงระบบในการเปลี่ยนผ่านครูจากการเป็นผู้ส่งต่อเนื้อหา สู่การเป็นผู้นำกระบวนการสืบเสาะระดับสูง เพื่อยกระดับสมรรถนะความฉลาดรู้ให้ยั่งยืนตามกรอบ PISA 2025

4. การวินิจฉัยผ่านเลนส์ PCACK: ความเปราะบางเชิงโครงสร้างและบริบทสถานศึกษา

การประยุกต์ใช้รูปแบบ PCACK-4D เผยให้เห็นว่ารากเหง้าของวิกฤตการเรียนรู้ก้าวข้ามเพียงมิติวิธีสอน (P) และเนื้อหา (K) แต่สัมพันธ์อย่างมีนัยสำคัญกับมิติการบริหาร (A) และบริบท (C) โดยเฉพาะปัญหา "ความไม่ต่อเนื่องทางวิชาการ" ในโรงเรียนขนาดเล็กและกลางที่เผชิญภาวะ "ความรู้สูญหาย" (Knowledge Attrition) จากการโยกย้ายบุคลากรบ่อยครั้ง โดยข้อค้นพบนี้ได้ขยายกรอบแนวคิด PCK เดิม (Shulman, 1986) สู่การเป็นระบบนิเวศแห่งความยั่งยืนที่ต้องบูรณาการปัจจัยเชิงโครงสร้างและการบริหารจัดการ (Woratadasawat, 2023) รูปแบบ PCACK-4D จึงมุ่งสร้าง "Knowledge Stack" เพื่อทำหน้าที่เป็น "ความทรงจำองค์กร" (Organizational Memory) ที่สามารถส่งต่อฐานสมรรถนะเชิงวิชาชีพสู่บุคลากรใหม่ได้ทันที การจัดการเชิงระบบในมิติ A และ C นี้ จึงเป็นยุทธศาสตร์ที่ทรงประสิทธิภาพกว่าการพัฒนาครูรายบุคคลแบบดั้งเดิม อันจะนำไปสู่การปิดช่องว่างทางวิชาการและยกระดับสมรรถนะผู้เรียนได้อย่างเป็นรูปธรรมและยั่งยืน

5. ทางออกเชิงระบบด้วยนวัตกรรมชุดความรู้ (Knowledge Stack)

จากการวิเคราะห์ความต้องการจำเป็นจากการสัมภาษณ์กลุ่มเป้าหมายด้วยแบบสัมภาษณ์กึ่งโครงสร้าง พบว่า ทั้งครูและผู้บริหารต้องการเครื่องมือเชิงระบบเพื่อแก้ปัญหาเชิงโครงสร้าง โดยเน้นเอกสารต้นแบบที่เชื่อมโยงมาตรฐานอย่างชัดเจนและชุดกิจกรรมที่เน้น active learning (5E) ควบคู่ไปกับการพัฒนาทักษะการวัดผลเชิงวิเคราะห์ การสร้างชุดความรู้ (Knowledge Stack: K) ในรูปแบบนวัตกรรมคู่มือ 6 เล่ม จึงไม่ใช่เพียงการผลิตสื่อ แต่เป็นการสร้าง "ความทรงจำขององค์กร (organizational memory)" ตามทฤษฎีระบบของ Von Bertalanffy (1968) ที่มองความสำเร็จจากการประสานองค์ประกอบย่อยอย่างสัมพันธ์กัน เพื่อให้เกิด "ทุนทางปัญญา" ที่ยั่งยืนแม้มีการเปลี่ยนแปลงบุคลากร

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

1.1 หน่วยงานต้นสังกัดควรนำ "นวัตกรรมชุดความรู้ 6 เล่ม (knowledge stack)" ไปใช้เป็นแนวทางในการนิเทศทางไกลสำหรับโรงเรียนขนาดเล็ก เพื่อสร้างมาตรฐานการสอนวิชาวิทยาศาสตร์ที่เป็นระบบเดียวกัน

1.2 การพัฒนาครูควรเลิกเน้นเพียงแค่เนื้อหา (content) แต่ต้องเน้นไปที่การปรับบทบาทครูให้เป็น "ผู้อำนวยความสะดวกทางปัญญา" (cognitive facilitator) ที่รู้วิธีการใช้คำถามเพื่อกระตุ้นให้นักเรียนประเมินค่าและสรุปผลการทดลองด้วยตนเองตามกรอบการประเมิน PISA 2025

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 ควรมีการศึกษาเชิงทดลอง (experimental research) ในระยะที่ 2 และระยะที่ 3 เพื่อประเมินประสิทธิภาพของรูปแบบ PCACK-4D และนวัตกรรมชุดความรู้ 6 เล่ม (knowledge stack) ที่มีต่อการพัฒนาสมรรถนะความฉลาดรู้ด้านวิทยาศาสตร์ของนักเรียนในเชิงประจักษ์

2.2 ควรมีการพัฒนาระบบนิเทศติดตามออนไลน์ (digital supervision platform) เพื่อสนับสนุนการใช้รูปแบบ PCACK ในโรงเรียนขนาดเล็กที่อยู่ในพื้นที่ห่างไกล เพื่อลดช่องว่างในการเข้าถึงความเชี่ยวชาญทางวิชาการและส่งเสริมการสร้างชุมชนแห่งการเรียนรู้ทางวิชาชีพ (PLC) อย่างต่อเนื่อง

References

- Anderson, L. W., & Krathwohl, D. R. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. Longman.
- Organisation for Economic Co-operation and Development. (2023). *PISA 2025 science framework* (Draft). https://pisa-framework.oecd.org/science-2025/assets/docs/PISA_2025_Science_Framework.pdf
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14. <https://doi.org/10.3102/0013189X015002004>
- Von Bertalanffy, L. (1968). *General system theory: Foundations, development, applications*. George Braziller.

Translate Thai References

- Chiang Rai Primary Educational Service Area Office 1. (2024). *O-NET result report: Academic year 2023*. Group of Supervision, Monitoring and Evaluation. (in Thai)
- Office of the Private Education Commission. (2024). *Guidelines for driving PISA 2025 score enhancement in educational institutions*. Ministry of Education. (in Thai)
- Techakaew, S., & Deemee, P. (2023). *Manual for using the Basic National Educational Test (O-NET) analysis program to determine the efficiency of teacher learning management in Phrae educational institutions, academic year 2023 (PROFILE O-NET 2023)*. Phrae Provincial Education Office. (in Thai)
- Woratadasawat, W. (2023). Supervision model for promoting teacher's learning management competency in small schools. *Journal of Educational Innovation and Research*, 7(2), 445-460. (in Thai)

การพัฒนาแบบวัดสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์

เรื่อง คลื่นเสียง สำหรับนักเรียนระดับมัธยมศึกษาตอนปลาย

Development of a Scientific Explanation Competency Test on Sound Waves
for Upper Secondary School Studentsกฤษฎา ทองประไพ^{1*}Krissada Tongprapai^{1*}

(Received: March 31, 2026; Revised: April 26, 2026; Accepted: April 27, 2026)

บทคัดย่อ

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแบบวัดสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ เรื่อง คลื่นเสียง สำหรับนักเรียนระดับชั้นมัธยมศึกษาตอนปลาย มีขั้นตอนในการพัฒนาแบบวัดดังนี้ 1) กำหนดนิยามเชิงปฏิบัติการของสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ เรื่อง คลื่นเสียง 2) วิเคราะห์เนื้อหา กำหนดโครงสร้างและจำนวนข้อของแบบวัด 3) สร้างข้อคำถามและเกณฑ์การให้คะแนนแบบจำเพาะ 4) ประเมินความเที่ยงตรงและความเหมาะสมต่อการนำไปใช้ 5) ปรับปรุงแบบวัดแล้วนำไปทดลองใช้และหาคุณภาพของแบบวัด ได้แก่ ความยากง่าย อำนาจจำแนก ความเชื่อมั่นทั้งฉบับ และคุณภาพของเกณฑ์การให้คะแนน

ผลการวิจัยพบว่า แบบวัดอัตนัยแบบความเรียงที่ไม่จำกัดคำตอบที่สร้างขึ้นมีความเที่ยงตรงเชิงเนื้อหาและมีความเหมาะสมต่อการนำไปใช้ เมื่อนำไปทดลองใช้กับกลุ่มตัวอย่าง ซึ่งเป็นนักเรียนระดับชั้นมัธยมศึกษา ปีที่ 5 จำนวน 40 คน จากโรงเรียนขนาดใหญ่แห่งหนึ่งในจังหวัดแพร่ที่เรียน เรื่อง คลื่นเสียง มาแล้ว แต่ยังไม่เคยได้รับการฝึกฝนด้านสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ พบว่า ความยากง่ายมีค่า 0.469 - 0.500 อำนาจจำแนกมีค่า 0.325 - 0.363 สัมประสิทธิ์แอลฟาของครอนบาคมีค่า 0.719 และค่าสัมประสิทธิ์สหสัมพันธ์ภายในชั้นมีค่า 0.766 - 0.913 ดังนั้นแบบวัดที่สร้างขึ้นสามารถนำไปใช้ประเมินสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ เรื่อง คลื่นเสียง ของนักเรียนระดับชั้นมัธยมศึกษาตอนปลายได้อย่างมีประสิทธิภาพ

คำสำคัญ สมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ การประเมิน ฟิสิกส์

Abstract

The objective of this research was to develop a scientific explanation competency test on the topic of sound waves for upper secondary school students. The development process comprised five stages: 1) defining the operational definitions of scientific explanation

¹ ตำแหน่งครู วิทยฐานะครูชำนาญการพิเศษ สาขาวิชาฟิสิกส์ โรงเรียนพิริยาลัยจังหวัดแพร่

¹ Senior Professional Level Teacher (Physics), Phiriyalai School Phrae.

competency regarding sound waves; 2) conducting content analysis and determining the test structure and number of items; 3) constructing test items and specific scoring rubrics; 4) evaluating the validity and suitability of the instrument; and 5) refining the test for pilot testing to determine its psychometric properties, including difficulty, discrimination, reliability, and the quality of the scoring rubrics.

The results indicated that the developed extended-response subjective test possessed content validity and was suitable for implementation. The instrument was pilot tested with a sample of 40 Grade 11 students from a large secondary school in Phrae Province, who had previously studied sound waves but had no formal training in constructing scientific explanations. The analysis revealed that the difficulty index ranged from 0.469 to 0.500, and the discrimination index ranged from 0.325 to 0.363. The test achieved a Cronbach's alpha coefficient of 0.719, while the Intraclass Correlation Coefficient for inter-rater reliability of the scoring rubrics ranged from 0.766 to 0.913. Consequently, the developed test is an effective tool for assessing the scientific explanation competency of upper secondary school students regarding sound waves.

Keywords: scientific explanation competency, assessment, physics

บทนำ

สมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ (scientific explanation competency) คือความสามารถองค์รวมของบุคคลในการประยุกต์ใช้ความรู้ทางวิทยาศาสตร์ (scientific knowledge) เพื่อสร้างข้อกล่าวอ้าง (claim) ที่แสดงให้เห็นถึงความสัมพันธ์เชิงสาเหตุ (causal links) ระหว่างปรากฏการณ์ธรรมชาติกับทฤษฎีหรือหลักการทางวิทยาศาสตร์ที่เกี่ยวข้อง โดยมีการใช้หลักฐานเชิงประจักษ์ (Evidence) มาสนับสนุนข้อกล่าวอ้างอย่างสมเหตุสมผล (Alameh et al., 2022; McNeill & Krajcik, 2008) ส่งเสริมให้นักเรียนใช้ความรู้วิทยาศาสตร์เชิงลึกและใช้ทักษะทางวิทยาศาสตร์ ทักษะการคิดวิเคราะห์ การใช้เหตุผลเชิงตรรกะ และการเชื่อมโยงความรู้เดิมกับหลักฐานใหม่ที่ได้จากการสำรวจตรวจสอบทำความเข้าใจปรากฏการณ์ทางธรรมชาติ คล้ายกับการทำงานของนักวิทยาศาสตร์ (McNeill & Krajcik, 2008; Ruiz-Primo et al., 2010) ทั้งนี้องค์การเพื่อความร่วมมือทางเศรษฐกิจและการพัฒนา (Organization for Economic Co-operation and Development หรือ OECD) จึงประเมินคุณภาพของระบบการศึกษาในการเตรียมความพร้อมให้ประชาชนมีศักยภาพหรือความสามารถพื้นฐานที่จำเป็นต่อการดำรงชีวิตในโลกที่มีการเปลี่ยนแปลง ได้มีการประเมินสมรรถนะนักเรียนมาตรฐานสากล (Programme for International Student Assessment; PISA) จากการศึกษางานวิจัยและเอกสารที่เกี่ยวข้องพบว่าผลการประเมิน PISA ของนักเรียนไทยในช่วงทศวรรษที่ผ่านมา (พ.ศ. 2555–2565) กลับอยู่ในสถานะถดถอยอย่างมีนัยสำคัญ โดยเฉพาะในปี พ.ศ. 2565 ที่คะแนนด้านวิทยาศาสตร์ลดลงถึง

17 คะแนน ต่ำกว่าค่าเฉลี่ยของ OECD อย่างต่อเนื่อง (Institute for the Promotion of Teaching Science and Technology, 2023) สะท้อนให้เห็นว่านักเรียนไทยส่วนใหญ่ยังขาดสมรรถนะในการเชื่อมโยงหลักการทางวิทยาศาสตร์เพื่ออธิบายปรากฏการณ์ทางธรรมชาติ สอดคล้องกับงานวิจัยของ Janto et al. (2020); Yohyong & Pitiporntapin (2019) พบว่านักเรียนไม่คุ้นชินกับการคิดวิเคราะห์และการตีความข้อมูลเชิงประจักษ์ ทำให้นักเรียนสับสนระหว่างหลักฐานกับข้อสรุป และใช้อารมณ์ความรู้สึกแทนเหตุผลเชิงวิทยาศาสตร์ซึ่งสอดคล้องกับประสบการณ์ตรงจากชั้นเรียนของผู้วิจัยเองพบว่านักเรียนมักระบุคำตอบเพียงสั้น ๆ ขาดการอ้างอิงหลักฐานสนับสนุนและไม่สามารถเชื่อมโยงเหตุและผลได้อย่างเป็นระบบ

หากพิจารณาผ่านมุมมองของธรรมชาติของวิชาฟิสิกส์ซึ่งมุ่งเน้นการทำความเข้าใจความสัมพันธ์เชิงกลไกของธรรมชาติตั้งแต่สิ่งที่มีความซับซ้อนน้อย เช่น การตกอย่างอิสระ การเคลื่อนที่แนวตรง ไปจนกระทั่งปรากฏการณ์ที่มีความซับซ้อนมาก เช่น แสงเชิงคลื่น คลื่นเสียง (Sari et al., 2021) ความซับซ้อนของปรากฏการณ์ทำให้นักเรียนมีสมรรถนะดังกล่าวอยู่ในระดับที่แตกต่างกัน ผู้วิจัยได้วิเคราะห์หลักสูตรฟิสิกส์ของระดับชั้นมัธยมศึกษาตอนปลาย พบว่าผลการเรียนรู้ส่วนใหญ่เน้นย้ำถึงการอธิบายเป็นสำคัญโดยเริ่มตั้งแต่อธิบายจากสิ่งที่สังเกตด้วยตาได้ไปจนกระทั่งไม่สามารถสังเกตได้ เช่น อธิบายการสั่นพ้องของเสียงในอากาศ อธิบายการเกิดเสียง หากนักเรียนมีสมรรถนะดังกล่าว นักเรียนต้องใช้ความรู้ที่สามารถอธิบายความสัมพันธ์เชิงสาเหตุได้สะท้อนให้เห็นถึงสมรรถนะ (competency) ซึ่งเป็นศักยภาพที่ยั่งยืนกว่าการจดจำเนื้อหา เพราะนักเรียนต้องแสดงความสามารถในการใช้ความเข้าใจเชิงมนทัศน์ อธิบายกลไกปรากฏการณ์ เพื่อพยากรณ์และให้เหตุผลต่อปรากฏการณ์ธรรมชาติซึ่งถือเป็นหัวใจสำคัญของการฉลาดรู้วิทยาศาสตร์ (Alameh et al., 2022; McNeill & Krajcik, 2008)

สมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ไม่ใช่เพียงการพรรณนาสิ่งที่มองเห็นแต่เป็นการใช้ความรู้ที่ได้จากการสืบเสาะเพื่อตอบคำถามว่าปรากฏการณ์หนึ่ง ๆ เกิดขึ้นได้อย่างไร (How) และทำไม (Why) ประกอบด้วยองค์ประกอบที่สำคัญ 3 ส่วน คือ ข้อกล่าวอ้าง (claim) หลักฐาน (evidence) ที่ได้จากข้อมูลเชิงประจักษ์ และการใช้เหตุผล (reasoning) ซึ่งเป็นส่วนสำคัญในการเชื่อมโยงหลักฐานเข้ากับทฤษฎี (McNeill & Krajcik, 2008) อย่างไรก็ตามผู้วิจัยพบว่าแบบวัดในประเทศไทยส่วนใหญ่เน้นองค์ประกอบของคำอธิบาย แต่ยังขาดการให้ความสำคัญกับกลไกการเกิดปรากฏการณ์ (mechanism) และความสัมพันธ์เชิงสาเหตุ (causal links) ที่สะท้อนถึงระดับความเข้าใจมนทัศน์หรือแนวคิดจากระดับพื้นฐานไปยังระดับที่ซับซ้อนมากขึ้น อีกทั้งระดับของการใช้ความรู้ทางวิทยาศาสตร์ที่ซับซ้อนจะช่วยให้เด็กนักเรียนมีสมรรถนะการอธิบายปรากฏการณ์อย่างเป็นวิทยาศาสตร์ได้ดีมากขึ้น (Andrade et al., 2019)

จากการศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องผู้วิจัยพบว่าแบบวัดสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ส่วนใหญ่วัดจากการเขียนซึ่งประกอบไปด้วยข้อคำถาม 4-5 ข้อ มีระยะเวลาในการทำประมาณ 60 นาที และเป็นแบบวัดแบบอัตนัยไม่จำกัดคำตอบเนื่องจากมีประสิทธิภาพสูงกว่าแบบจำกัดคำตอบ (Wannathai & Pruekpramool, 2021) แบบวัดส่วนใหญ่ที่พบเป็นเนื้อหาเกี่ยวกับวิชาเคมี เช่น แก๊ส (Pinchai & Seetee, 2025) เนื้อหาวิชาชีววิทยา เช่น ระบบหมุนเวียนเลือด (Suwannahong et al., 2025) ส่วนเนื้อหาวิชาฟิสิกส์ยังมีจำนวนจำกัดพบเฉพาะเนื้อหากลุ่มกลศาสตร์ เช่น แรงในชีวิตประจำวัน (Veeraphan et al.,

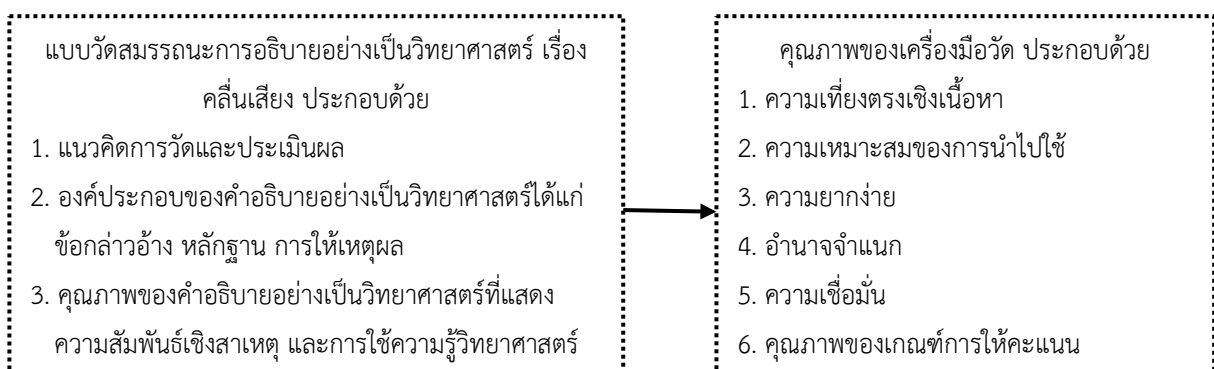
2025) แต่ยังไม่พบเนื้อหาในกลุ่มคลื่น อาจเป็นเพราะเนื้อหาในกลุ่มคลื่นนี้มีลักษณะเป็นนามธรรมสูง และเนื้อหา
มีแนวคิดที่ซับซ้อน ดังนั้นการอธิบายปรากฏการณ์ในกลุ่มคลื่นจึงมีกลไกที่ซับซ้อนเกินกว่าจะทำความเข้าใจ
ได้เพียงการท่องจำสูตร เช่น คลื่นเสียงที่นักเรียนรับรู้ได้แต่ไม่สามารถมองเห็นได้ด้วยตาเปล่า (Sari et al., 2021)
แม้จะมีการพัฒนาแบบวัดสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ในหลายเนื้อหา แต่ก็ยังไม่พบเครื่องมือ
ที่พัฒนามาบนพื้นฐานกรอบแนวคิดขององค์ประกอบการอธิบายอย่างเป็นวิทยาศาสตร์ ความสัมพันธ์เชิงสาเหตุ
(causal links) และความเชื่อมโยงมโนทัศน์ (conceptual connectivity) หากครูผู้สอนมีเครื่องมือในการ
วินิจฉัยสมรรถนะของนักเรียนในเนื้อหาเฉพาะก็จะช่วยให้ครูผู้สอนได้ข้อมูลนำมาออกแบบการจัดการเรียนรู้
ได้อย่างเหมาะสม ช่วยให้ครูผู้สอนสามารถประเมินระดับสมรรถนะของผู้เรียนในเนื้อหาวิชาเฉพาะได้ ผู้วิจัย
จึงเห็นความสำคัญที่จะพัฒนาแบบวัดสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ ที่เป็นประโยชน์ต่อครูผู้สอน
ในพัฒนาสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ เรื่อง คลื่นเสียง ของนักเรียนชั้นมัธยมศึกษาตอนปลายได้
ให้มีประสิทธิภาพและนำไปใช้ประโยชน์ได้

ความมุ่งหมายของการวิจัย

เพื่อพัฒนาและตรวจสอบคุณภาพแบบวัดสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ เรื่องคลื่นเสียง
สำหรับนักเรียนระดับชั้นมัธยมศึกษาตอนปลาย

กรอบแนวคิดการวิจัย

การวิจัยนี้มุ่งพัฒนาแบบวัดสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ เรื่อง คลื่นเสียง บูรณาการ
การกรอบแนวคิดของ McNeill & Krajcik (2008) ประกอบด้วย ข้อกล่าวอ้าง (claim) หลักฐาน (evidence)
และการให้เหตุผล (reasoning) และกรอบแนวคิดของ Andrade et al. (2019) ที่ให้ความสำคัญการสร้าง
ความสัมพันธ์เชิงสาเหตุ (causal links) และความเชื่อมโยงมโนทัศน์ (conceptual connectivity) เพื่อแสดง
ถึงกลไกของการเกิดปรากฏการณ์ (mechanism) การบูรณาการดังกล่าวช่วยให้เครื่องมือสามารถประเมินได้
ครอบคลุมทั้งองค์ประกอบและคุณภาพของคำอธิบายอย่างเป็นวิทยาศาสตร์ นำไปสู่การจำแนกสมรรถนะของ
ผู้เรียนตามลำดับขั้นพัฒนาการ โดยผ่านการตรวจสอบคุณภาพของเครื่องมือ ดังภาพ 1



ภาพ 1 กรอบแนวคิดการวิจัย

วิธีดำเนินการวิจัย

งานวิจัยนี้เป็นการพัฒนาเครื่องมือประเมินสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ของนักเรียนในระดับชั้นมัธยมศึกษาตอนปลายในเนื้อหาคลื่นเสียง ซึ่งมีรายละเอียดดังนี้

1. ประชากรและกลุ่มตัวอย่าง

ประชากร คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 5 แผนการเรียนวิทยาศาสตร์และคณิตศาสตร์ จำนวน 320 คน ของโรงเรียนมัธยมศึกษาขนาดใหญ่แห่งหนึ่งในจังหวัดแพร่

กลุ่มตัวอย่าง คือ นักเรียนระดับชั้นมัธยมศึกษาปีที่ 5 แผนการเรียนวิทยาศาสตร์และคณิตศาสตร์ จำนวน 40 คน จำนวน 1 ห้องเรียน ได้มาจากการการเลือกตามความสะดวก (convenience sampling) เนื่องจากผู้วิจัยรับผิดชอบในการจัดการเรียนรู้ และโรงเรียนอนุญาตให้เก็บข้อมูลในห้องเรียนปกติที่มีการละความสามารถของผู้เรียน อย่างไรก็ตามนักเรียนห้องนี้ได้เรียนเนื้อหา เรื่อง คลื่นเสียง เรียบร้อยแล้ว

2. เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ผู้วิจัยใช้ได้แก่แบบวัดสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ ซึ่งมีขั้นตอนในการสร้างเครื่องมือดังนี้

1. ศึกษาเอกสารและงานวิจัยที่เกี่ยวข้องเพื่อกำหนดนิยาม ความสำคัญ และองค์ประกอบของสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ สำหรับเป็นแนวทางในการพัฒนาเครื่องมือให้ครอบคลุมและลดข้อจำกัดในการวัดผล

2. วิเคราะห์ผลการเรียนรู้ รายวิชาเพิ่มเติมวิทยาศาสตร์ ฟิสิกส์ เล่ม 4 ตามผลการเรียนรู้กลุ่มสาระการเรียนรู้วิทยาศาสตร์ (ฉบับปรับปรุง พ.ศ. 2560) ตามหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 เพื่อกำหนดพฤติกรรมการเรียนรู้ที่ต้องการวัดให้สอดคล้องกับเนื้อหาเรื่องคลื่นเสียง

3. กำหนดนิยามของ สมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ เรื่อง คลื่นเสียง หมายถึง ความสามารถของผู้เรียนในการสร้าง ข้อกล่าวอ้าง (claim) ตอปรากฏการณ์ที่กำหนด โดยใช้หลักฐาน (evidence) มาเชื่อมโยงด้วยการให้เหตุผล (reasoning) เพื่ออธิบายกลไกการเกิดปรากฏการณ์ (mechanism) และความสัมพันธ์เชิงสาเหตุ (causal links) ได้อย่างถูกต้อง

4. กำหนดจำนวนข้อของแบบวัดให้ครอบคลุมเนื้อหาหลัก 5 หัวข้อของบทเรียน เรื่อง คลื่นเสียง สำหรับนักเรียนชั้นมัธยมศึกษาปีที่ 5 ประกอบด้วย ธรรมชาติของเสียง เครื่องดนตรี การได้ยินของมนุษย์ ปรากฏการณ์เกี่ยวกับเสียง และการประยุกต์เกี่ยวกับเสียง

5. ผู้วิจัยสร้างแบบวัดสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ เป็นข้อสอบอัตนัยแบบไม่จำกัดคำตอบจำนวน 5 ข้อ ใช้เวลาทำ 100 นาที โดยแต่ละข้อจะเริ่มต้นด้วยการนำเสนอสถานการณ์ใกล้ตัวพร้อมให้ข้อมูลประกอบ เช่น กราฟ ตาราง หรือภาพจำลอง และใช้คำถามปลายเปิดให้ผู้เรียนระบุข้อกล่าวอ้าง (claim) หลักฐาน (evidence) และเขียนการให้เหตุผล (reasoning) โดยเน้นการร้อยเรียงหลักฐานสนับสนุน ข้อกล่าวอ้างให้มีความสัมพันธ์เชิงสาเหตุและถูกต้องตามหลักฟิสิกส์แบบวัดนี้เป็นข้อสอบอัตนัยไม่จำกัดคำตอบ นักเรียนสามารถเขียนตอบได้อย่างอิสระตัวอย่าง ดังภาพ 2

สถานการณ์ที่ 5 : การประยุกต์ใช้ความรู้เสียง

5. บริษัทฟรุตเฟรช (FruitFresh Co., Ltd.) เป็นบริษัทส่งออกผลไม้ไทยไปต่างประเทศ โดยบริษัทต้องการคัดแยกผลไม้ตามขนาด เพื่อให้ได้มาตรฐานเดียวกันก่อนบรรจุลงกล่องส่งออก ฝ่ายวิศวกรรมของบริษัทเสนอแนวคิดการใช้ประโยชน์ของเสียง คือ คลื่นเสียงอัลตราโซนิก (Ultrasonic Waves) เพื่อตรวจวัดขนาดของผลไม้ที่กำลังไหลผ่านตามรางน้ำ โดยมีอัตราเร็วเสียงในน้ำเท่ากับ 1500 m/s ซึ่งจะแยกผลไม้ที่มีขนาดใหญ่กว่าและเล็กกว่าเส้นผ่านศูนย์กลาง ออกจากกันได้ตามมาตรฐาน เมื่อผลไม้แต่ละลูกเคลื่อนผ่านบริเวณนั้น ผิวของผลไม้จะส่งความถี่คลื่นไปยังเครื่องตรวจจับ (Detector Frequency) ดังตาราง

ชนิดของผลไม้	ขนาดเส้นผ่านศูนย์กลางเฉลี่ยที่ได้มาตรฐาน (cm)	ความถี่ของคลื่นสะท้อน (KHz)
ส้ม	6.00	25.00
มะม่วง	10.00	15.00
แอปเปิล	8.00	20.00
มะละกอ	12.00	12.50

นักเรียนคิดว่าผลไม้ชนิดใดที่ผลการประเมินไม่ได้มาตรฐาน ทราบได้อย่างไร เพราะเหตุใดจึงคิดเช่นนั้น จงอธิบายว่า (สามารถใช้หลักการทางวิทยาศาสตร์ หลักฐานที่เกี่ยวข้องมาอ้างอิงเพิ่มเติมได้)

ภาพ 2 ตัวอย่างสถานการณ์ ลักษณะของคำถามในรูปแบบเขียนตอบอย่างอิสระ

6. การสร้างเกณฑ์การให้คะแนน ผู้วิจัยพัฒนาเกณฑ์การประเมิน (rubrics) โดยสังเคราะห์ร่วมกันระหว่างแนวคิดของ McNeill & Krajcik (2008) และ Andrade et al. (2019) เพื่อใช้วัดสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ใน 2 มิติหลัก คือ มิติองค์ประกอบและมิติคุณภาพของคำอธิบายอย่างเป็นวิทยาศาสตร์ เพื่อแบ่งระดับสมรรถนะออกเป็น 5 ระดับ ตั้งแต่ระดับดีมากไปจนกระทั่งระดับควรปรับปรุง รายละเอียดดังตาราง 1

ตาราง 1 เกณฑ์การประเมินระดับสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์

ระดับสมรรถนะ		เกณฑ์การประเมินระดับสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์
(คะแนน)	องค์ประกอบ	คุณภาพของการอธิบายอย่างเป็นวิทยาศาสตร์
ระดับดีมาก (4 คะแนน)	นักเรียนสามารถแสดงองค์ประกอบได้ครบถ้วน ได้แก่ ข้อกล่าวอ้าง (Claim) หลักฐาน (Evidence) และการให้เหตุผล (Reasoning)	คำตอบของนักเรียนแสดงองค์ประกอบได้ครบถ้วน 3 องค์ประกอบ ได้อย่างถูกต้อง โดยมีข้อกล่าวอ้าง มีการวิเคราะห์ความน่าเชื่อถือของหลักฐาน มีการให้เหตุผลเชื่อมโยงกับหลักการทางวิทยาศาสตร์ มีการโดยระบุพารามิเตอร์ที่เกี่ยวข้องกับปรากฏการณ์และเชื่อมโยงความสัมพันธ์ด้วยแบบจำลอง หรือข้อจำกัด หรือเงื่อนไขการเกิดปรากฏการณ์ มีการอธิบายที่ซับซ้อน หรือมีการขยายแนวคิดทางวิทยาศาสตร์ได้อย่างถูกต้อง

ตาราง 1 (ต่อ)

ระดับสมรรถนะ		เกณฑ์การประเมินระดับสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์	
(คะแนน)		องค์ประกอบ	คุณภาพของการอธิบายอย่างเป็นวิทยาศาสตร์
ระดับดี (3 คะแนน)	นักเรียนสามารถแสดงองค์ประกอบได้ครบถ้วน ได้แก่ ข้อกล่าวอ้าง (Claim) หลักฐาน (Evidence) และการให้เหตุผล (Reasoning)		<u>ตัวอย่างคำตอบ</u> "แอปเปิลไม่เป็นไปตามมาตรฐาน (ข้อกล่าวอ้าง) เนื่องจากข้อมูลความเร็วเสียงในน้ำ 1500 m/s และความถี่ที่วัดได้ 20.0 KHz (หลักฐาน) เมื่อพิจารณาตามหลักการการสะท้อนของคลื่น ขนาดของวัตถุจะสัมพันธ์กับความยาวคลื่น ($\lambda = v/f$) คำนวณได้เป็น 0.075 เมตร หรือ 7.50 cm ซึ่งต่ำกว่ามาตรฐานที่กำหนดไว้ 8.00 cm (เหตุผล) โดยคำอธิบายนี้มีเงื่อนไขว่าอุณหภูมิอากาศต้องคงที่เพื่อให้ความเร็วเสียงไม่เปลี่ยนแปลง และเครื่องวัดต้องใช้หลักการการสะท้อนเพื่อระบุระยะหน้าคลื่นที่ตกกระทบผิวผลไม้"
			คำตอบของนักเรียนแสดงองค์ประกอบได้ครบถ้วน 3 องค์ประกอบ ได้อย่างถูกต้อง โดยมีข้อกล่าวอ้าง ให้หลักฐาน และให้เหตุผล เชื่อมโยงหลักฐานกับข้อกล่าวอ้างโดยมีการอธิบายกลไกด้วยแบบจำลองและหลักการทางวิทยาศาสตร์ที่ถูกต้องสมบูรณ์ในระดับพื้นฐาน เป็นการอธิบายอย่างง่ายหรือระบุพารามิเตอร์ที่เกี่ยวข้องกับปรากฏการณ์ได้ แต่ยังไม่สามารถระบุความสัมพันธ์ของพารามิเตอร์นั้นได้ถูกต้อง <u>ตัวอย่างคำตอบ</u> "แอปเปิล (ข้อกล่าวอ้าง) เพราะเมื่อคำนวณจากสูตร $v = f\lambda$ โดยใช้ความเร็วในน้ำ 1500 m/s และความถี่ 20.0 KHz (หลักฐาน) จะได้ขนาดของแอปเปิลเท่ากับ 7.50 cm ซึ่งเมื่อนำไปเทียบกับค่ามาตรฐานในตารางที่กำหนดไว้เพียง 8.0 cm พบว่าแอปเปิลนี้มีขนาดน้อยกว่ามาตรฐาน จึงไม่ผ่านเกณฑ์การส่งออก (เหตุผล)"
			คำตอบของนักเรียนแสดงองค์ประกอบได้ครบถ้วน 2 ใน 3 องค์ประกอบ แต่คุณภาพของคำอธิบายยังไม่ถูกต้องสมบูรณ์ กล่าวคือ <u>กรณีที่ 1</u> ข้อกล่าวอ้างไม่ถูกต้องและไม่สอดคล้องกับหลักฐานและเหตุผล <u>กรณีที่ 2</u> มีข้อกล่าวอ้าง และนำข้อมูลหรือหลักฐานเชิงประจักษ์ มาสนับสนุนข้อกล่าวอ้างได้ แต่การเชื่อมโยงหลักฐานเป็นเพียงความสัมพันธ์เชิงเหตุการณ์หรือเชิงประสบการณ์ หรือยังไม่สามารถอธิบายกลไกด้วยแบบจำลองหรือแนวคิดทางวิทยาศาสตร์ได้ถูกต้อง <u>กรณีที่ 3</u> มีข้อกล่าวอ้าง มีหลักฐาน แต่การให้เหตุผลไม่สอดคล้องกันระหว่างข้อกล่าวอ้างและหลักฐาน <u>ตัวอย่างคำตอบ</u> "แอปเปิล (ข้อกล่าวอ้าง) เพราะเครื่องวัดค่าความถี่ได้ 2,000 Hz ซึ่งเมื่อลองคำนวณดูแล้วได้ค่าออกมาน้อยกว่าค่ามาตรฐาน 8.00 cm (หลักฐาน) แสดงว่าความถี่ที่เปลี่ยนไปมีผลต่อการบอกขนาดที่ผิดพลาดจากที่กำหนดไว้"
ระดับปานกลาง (2 คะแนน)	นักเรียนสามารถแสดงองค์ประกอบได้ 2 ใน 3 องค์ประกอบทั้งหมด ได้แก่ ข้อกล่าวอ้าง (Claim) หลักฐาน (Evidence) และการให้เหตุผล (Reasoning)		

ตาราง 1 (ต่อ)

ระดับสมรรถนะ (คะแนน)	เกณฑ์การประเมินระดับสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์	
	องค์ประกอบ	คุณภาพของการอธิบายอย่างเป็นวิทยาศาสตร์
ระดับพอใช้ (1 คะแนน)	นักเรียนสามารถแสดงองค์ประกอบ ได้ 1 ใน 3 องค์ประกอบ ได้แก่ ข้อ กล่าวอ้าง (Claim) หลักฐาน (Evidence) และการให้เหตุผล (Reasoning)	คำตอบของนักเรียนแสดงองค์ประกอบของคำอธิบายได้เพียง 1 องค์ประกอบที่ถูกต้อง เช่น มีเฉพาะข้อกล่าวอ้าง แต่ยังไม่มีการนำ ข้อมูลหรือหลักฐานเชิงประจักษ์มาใช้สนับสนุน หรือยังไม่การใช้ แบบจำลองหรือสมการทางฟิสิกส์มาสนับสนุนหรือบรรยายลักษณะ ข้อมูลด้วยการพรรณนา <u>ตัวอย่างคำตอบ</u> แอปเปิล (ข้อกล่าวอ้าง) เพราะดูจากตัวเลขในตารางแล้ว ค่าความถี่ ดูเป็นตัวเลขล้วนๆ
ระดับ ควรปรับปรุง (0 คะแนน)	นักเรียนไม่สามารถแสดง องค์ประกอบได้ครบทุก องค์ประกอบ ได้แก่ ข้อกล่าวอ้าง (Claim) หลักฐาน (Evidence) และการให้เหตุผล (Reasoning)	คำตอบของนักเรียนแสดงถึงการอธิบายปรากฏการณ์โดยไม่มีการ ตั้งข้อกล่าวอ้าง ไม่อ้างอิงข้อมูลหรือหลักฐานเชิงประจักษ์ และไม่ แสดงการให้เหตุผล หรืออาจเป็นเพียงการตอบซ้ำคำถาม เดา คำตอบ หรือคำตอบนั้นไม่ถูกต้อง หรือไม่เกี่ยวกับปรากฏการณ์ <u>ตัวอย่างคำตอบ</u> "คำตอบอื่นที่ไม่ใช่แอปเปิล"

หมายเหตุ ระดับสมรรถนะในระดับดีมากและระดับดี ทั้ง 2 ระดับนี้ พิจารณาจากทั้ง 3 องค์ประกอบเหมือนกัน แต่คุณภาพของการอธิบายอย่างเป็นวิทยาศาสตร์มีความแตกต่างกัน

7. นำแบบวัดสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์และเกณฑ์การให้คะแนนรายข้อ
เสนอผู้เชี่ยวชาญ จำนวน 5 ท่าน ประกอบด้วย นักวิทยาศาสตร์ศึกษา 2 ท่าน นักฟิสิกส์ศึกษา 1 ท่าน และ
ครูผู้สอนที่มีประสบการณ์การสอนฟิสิกส์มากกว่า 10 ปี จำนวน 2 ท่าน พิจารณาความตรงเชิงเนื้อหา
(content validity) จากค่าดัชนีความสอดคล้องระหว่างข้อคำถามกับนิยามเชิงปฏิบัติการ (Index of Item
Objective Congruence: IOC) และปรับปรุงแบบวัดตามข้อเสนอแนะเพิ่มเติมของเสนอของผู้เชี่ยวชาญ

8. นำแบบวัดไปทดลองใช้กับกลุ่มตัวอย่างที่เรียนเนื้อหาเรื่องคลื่นเสียงมาเรียบร้อยแล้ว
เพื่อหาคุณภาพของแบบวัด โดยการตรวจสอบค่าความยาก (p) และค่าอำนาจจำแนก (r) และค่าความเชื่อมั่น
(reliability) ทั้งฉบับ รวมทั้งสอบถามนักเรียนเพิ่มเติมเกี่ยวกับความชัดเจนของภาษาและความเข้าใจของข้อมูล
ที่กำหนดในสถานการณ์ของข้อคำถามแต่ละข้อ จากนั้นปรับปรุงแบบวัดให้มีความสมบูรณ์

3. การเก็บรวบรวมข้อมูล

ผู้วิจัยนำแบบวัดที่ผ่านการตรวจสอบคุณภาพไปทดลองใช้กับกลุ่มตัวอย่างในภาคเรียนที่ 2 ปี
การศึกษา 2568 จำนวน 40 คน โดยนักเรียนได้เรียนเนื้อหา เรื่อง คลื่นเสียง เรียบร้อยแล้วก่อนเก็บข้อมูลผู้วิจัย
ชี้แจงให้นักเรียนทราบเกี่ยวกับการวิจัยที่ไม่มีผลต่อการตัดสินผลการเรียน ข้อมูลส่วนบุคคลเป็นความลับ
ตามจรรยาบรรณของนักวิจัย

4. การวิเคราะห์ข้อมูล

ผู้วิจัยได้ดำเนินการหาคุณภาพแบบวัด รายละเอียดดังต่อไปนี้

1. ความเที่ยงตรงเชิงเนื้อหา ผู้วิจัยนำความเห็นของผู้เชี่ยวชาญทั้ง 5 ท่านมาคำนวณหาค่าดัชนีความสอดคล้อง (IOC) ในการประเมินความสอดคล้องของคำตอบกับองค์ประกอบโดยพิจารณาจากค่าเฉลี่ย จากนั้นคัดเลือกข้อคำถามที่มีค่าดัชนีความสอดคล้องมากกว่า 0.5 (Srisa-ard, 2017)

2. ความเหมาะสมของเกณฑ์การให้คะแนนต่อการนำไปใช้ ผู้วิจัยนำความคิดเห็นผู้เชี่ยวชาญทั้ง 5 ท่าน มาคำนวณหาค่าเฉลี่ยของความเหมาะสมต่อการนำไปใช้และแปลความหมายของความเหมาะสมต่อการนำไปใช้ตามเกณฑ์ ได้กำหนดเกณฑ์ความเหมาะสมในการพิจารณา 5 ช่วง (rating scale) คือ 4.51 - 5.00 หมายถึง เหมาะสมมากที่สุด 3.51 - 4.50 หมายถึง เหมาะสมมาก 2.51 - 3.50 หมายถึง เหมาะสมปานกลาง 1.51 - 2.50 หมายถึง เหมาะสมน้อย 1.00 - 1.50 หมายถึง เหมาะสมน้อยที่สุด และปรับปรุงแก้ไขตามข้อเสนอแนะของผู้เชี่ยวชาญ (Srisa-ard, 2017)

3. ค่าความยากง่าย (difficulty) และอำนาจจำแนก (discrimination) ใช้เทคนิค 50% แบ่งคะแนนที่ได้จากการทำแบบวัดของแต่ละข้อ โดยเรียงคะแนนจากมากไปน้อย แบ่งเป็น 2 กลุ่ม ได้แก่ กลุ่มนักเรียนคะแนนสูง และกลุ่มนักเรียนคะแนนต่ำ คำนวณหาค่าความยากง่าย และค่าอำนาจจำแนก ทั้งนี้แต่ละข้อคำถามควรมีค่าความยากง่ายอยู่ระหว่าง 0.20 - 0.80 และมีค่าอำนาจจำแนก 0.20 ขึ้นไป (Srisa-ard, 2017)

4. ความเชื่อมั่น (reliability) ของแบบวัดทั้งฉบับ นำคะแนนที่ได้จากแบบวัดของนักเรียนรายบุคคลไปคำนวณหาค่าสัมประสิทธิ์แอลฟาครอนบาค (Cronbach's Alpha coefficient; α) แบบวัดที่มีความเชื่อมั่นสูง ควรมีค่าสัมประสิทธิ์ตั้งแต่ 0.70 ขึ้นไป (Srisa-ard, 2017)

5. ความน่าเชื่อถือของเกณฑ์การให้คะแนนเพื่อพิจารณาความเป็นปรนัยของเกณฑ์การให้คะแนนรายข้อ โดยผู้ประเมิน 3 ท่านประกอบด้วยผู้วิจัยและผู้ตรวจจำนวน 2 ท่าน โดยแยกตรวจอย่างอิสระ จากนั้นนำคะแนนของแต่ละคนที่ได้จากการตรวจแบบวัดเดียวกัน หาค่าสหสัมพันธ์ภายในชั้น (intraclass correlation ควรมีค่าตั้งแต่ 0.70 ขึ้นไป (Secolsky & Denison, 2017)

ผลการวิจัย

แบบวัดสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ เรื่อง คลื่นเสียง ที่ผู้วิจัยสร้างขึ้นมีค่าความตรงเชิงเนื้อหาและค่าความเหมาะสมต่อการนำไปใช้ดังตาราง 2

ตาราง 2 ค่าความเที่ยงตรงเชิงเนื้อหาและค่าเฉลี่ยความเหมาะสมของการนำไปใช้

ลำดับสถานการณ์	ความเที่ยงตรง เชิงเนื้อหา	การแปล ความหมาย	ค่าเฉลี่ยความเหมาะสม ของการนำไปใช้	การแปล ความหมาย
1. ธรรมชาติของเสียง	+1.00	สอดคล้อง	5.00	เหมาะสมมากที่สุด
2. เครื่องดนตรี	+1.00	สอดคล้อง	4.80	เหมาะสมมากที่สุด
3. การได้ยินของมนุษย์	+1.00	สอดคล้อง	5.00	เหมาะสมมากที่สุด
4. ปรากฏการณ์เกี่ยวกับเสียง	+1.00	สอดคล้อง	4.80	เหมาะสมมากที่สุด
5. การประยุกต์เกี่ยวกับเสียง	+1.00	สอดคล้อง	5.00	เหมาะสมมากที่สุด

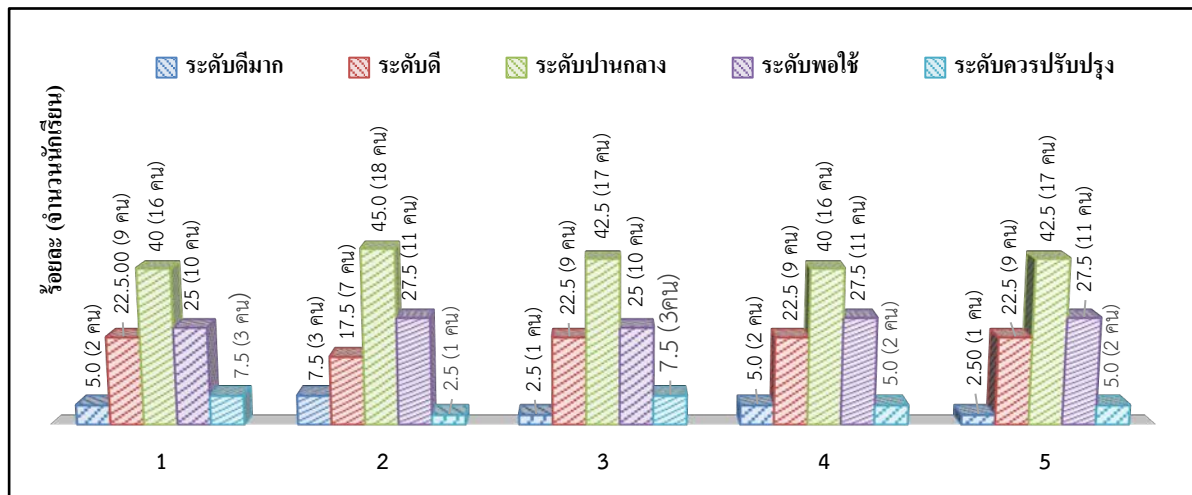
จากตาราง 2 พบว่า ข้อคำถามทุกข้อมีค่าความตรงเชิงเนื้อหา เท่ากับ +1.00 ซึ่งมากกว่า 0.50 แสดงว่าข้อคำถามทุกข้อสอดคล้องกับนิยามเชิงปฏิบัติการ และทุกข้อมีความเหมาะสมต่อการนำไปใช้เฉลี่ยอยู่ในช่วง 4.80 - 5.00 ซึ่งอยู่ในระดับเหมาะสมมากที่สุด โดยผู้เชี่ยวชาญมีความคิดเห็นว่าสถานการณ์ที่ใช้เป็นสิ่งใกล้ตัวนักเรียน และมีข้อมูลให้นักเรียนได้ตีความหมายของข้อมูลเหมาะสมกับนักเรียนระดับชั้นมัธยมศึกษาตอนปลาย โดยมีข้อเสนอแนะจากผู้เชี่ยวชาญ คือ ควรตรวจสอบองค์ประกอบของคำอธิบายของนักเรียนแต่ละคนได้ดีขึ้น ควรใช้คำถามเฉพาะเจาะจง เช่น นักเรียนทราบได้อย่างไร เพราะเหตุใดคิดเช่นนั้นและควรให้นักเรียนใช้ข้อมูลเพิ่มเติมได้เพื่อตรวจสอบความเข้าใจเชิงลึกของนักเรียน ผู้วิจัยจึงได้ปรับข้อคำถามตามข้อเสนอแนะ ตัวอย่าง ดังภาพที่ 2 ส่วนเกณฑ์การให้คะแนนรายข้อ ผู้เชี่ยวชาญเสนอแนะให้ใช้ระดับที่แสดงถึงพัฒนาการของนักเรียน จึงได้ปรับเป็นระดับสมรรถนะเป็น 5 ระดับ ดังตาราง 2

เมื่อผู้วิจัยตรวจให้คะแนนตามเกณฑ์การให้คะแนนรายข้อ มีจำนวน 5 สถานการณ์ จำนวนทั้งหมด 5 ข้อซึ่งในแต่ละข้อมีคะแนนเต็ม 4 คะแนน นำคะแนนมาหาค่าเฉลี่ย ($\bar{X}$) และส่วนเบี่ยงเบนมาตรฐาน (SD) ดังตาราง 3

ตาราง 3 คะแนนเฉลี่ยและส่วนเบี่ยงเบนมาตรฐาน

สถานการณ์	คะแนนเฉลี่ย (คะแนนเต็ม 4 คะแนน)	ส่วนเบี่ยงเบนมาตรฐาน
1. ธรรมชาติของเสียง	1.925	0.997
2. เครื่องดนตรี	2.000	0.934
3. การได้ยินของมนุษย์	1.875	0.939
4. ปรากฏการณ์เกี่ยวกับเสียง	1.950	0.959
5. การประยุกต์เกี่ยวกับเสียง	1.900	0.900

จากตาราง 3 จะเห็นได้ว่าสถานการณ์ที่ 2 มีคะแนนเฉลี่ยมากที่สุดเท่ากับ 2.00 คะแนน และสถานการณ์ที่ 3 มีคะแนนเฉลี่ยน้อยที่สุดเท่ากับ 1.875 คะแนน โดยสถานการณ์ที่ 1 มีส่วนเบี่ยงเบนมาตรฐานมากที่สุด เท่ากับ 0.934 และสถานการณ์ที่ 5 มีส่วนเบี่ยงเบนมาตรฐานและน้อยที่สุด เท่ากับ 0.900 ตามลำดับ เมื่อพิจารณาคะแนนของกลุ่มตัวอย่างรายบุคคลในแต่ละข้อคำถาม จำนวน 5 ข้อ โดยใช้เกณฑ์ที่ผู้วิจัยสร้างขึ้นตามตัวอย่างตาราง 1 พบว่านักเรียนส่วนใหญ่มีสมรรถนะอยู่ในระดับปานกลางมีจำนวนมากที่สุด กล่าวคือนักเรียนสามารถแสดงองค์ประกอบของคำอธิบายทางวิทยาศาสตร์ได้ 2 ใน 3 องค์ประกอบทั้งหมด และระดับดีมีจำนวนน้อยที่สุด ดังภาพ 3



ภาพ 3 ร้อยละของนักเรียนที่แสดงระดับสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ในแต่ละข้อคำถาม

ผู้วิจัยนำคะแนนของนักเรียนแต่ละคนมาเรียงลำดับจากสูงที่สุดไปยังต่ำสุด และใช้เทคนิคร้อยละ 50 แบ่งนักเรียนออกเป็น 2 กลุ่ม ได้แก่ กลุ่มนักเรียนคะแนนสูง และกลุ่มนักเรียนคะแนนต่ำ จากนั้น คำนวณหาค่าความยากง่ายและค่าอำนาจจำแนกของคำถามแต่ละข้อ และค่าสหสัมพันธ์ภายในชั้นของแบบวัด แสดงดังตาราง 4

ตาราง 4 ค่าความยากง่าย ค่าอำนาจจำแนก ค่าสหสัมพันธ์ภายในชั้นของแบบวัด

สถานการณ์	ค่าความยากง่าย (p)	แปลความหมาย	ค่าอำนาจจำแนก (r)	แปลความหมาย	ค่าสหสัมพันธ์ภายในชั้น (ICC)	แปลความหมาย	ผลการพิจารณา
1. ธรรมชาติของเสียง	0.481	ปานกลาง	0.363	พอใช้	0.842	น่าเชื่อถือ	ใช้ได้
2. เครื่องดนตรี	0.500	ปานกลาง	0.325	พอใช้	0.898	น่าเชื่อถือ	ใช้ได้
3. การได้ยินของมนุษย์	0.469	ปานกลาง	0.338	พอใช้	0.913	น่าเชื่อถือ	ใช้ได้
4. ปรากฏการณ์เกี่ยวกับเสียง	0.488	ปานกลาง	0.350	พอใช้	0.852	น่าเชื่อถือ	ใช้ได้
5. การประยุกต์เกี่ยวกับเสียง	0.475	ปานกลาง	0.325	พอใช้	0.766	น่าเชื่อถือ	ใช้ได้

จากตาราง 4 ผู้วิจัยพบว่าข้อคำถามมีค่าความยากง่ายอยู่ระหว่าง 0.469 - 0.500 แสดงว่าข้อคำถามแต่ละข้อมีความยากง่ายในระดับปานกลาง ค่าอำนาจจำแนกอยู่ระหว่าง 0.325 - 0.363 แสดงว่ามีอำนาจจำแนกในระดับพอใช้ ด้านความน่าเชื่อถือของเกณฑ์การให้คะแนน พิจารณาความเป็นปรนัยของเกณฑ์การให้คะแนนรายข้อ โดยผู้ประเมิน 3 ท่าน นำคะแนนไปคำนวณหาค่าสหสัมพันธ์ภายในชั้น (ICC) มีค่าอยู่ระหว่าง 0.766 - 0.913 แสดงว่าเกณฑ์การให้คะแนนมีความน่าเชื่อถือและชัดเจน ผู้ตรวจสามารถตรวจให้คะแนนได้

ไม่แตกต่างกัน นอกจากนี้ผู้วิจัยได้นำคะแนนรวมของนักเรียนทุกข้อที่ได้จากแบบวัดฉบับนี้ไปคำนวณหาค่าสัมประสิทธิ์แอลฟาครอนบาค มีค่าสัมประสิทธิ์ 0.719 แสดงว่าแบบวัดที่มีความเชื่อมั่นสูง

อภิปรายผล

การวิจัยนี้มีวัตถุประสงค์เพื่อพัฒนาแบบวัดสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ เรื่องคลื่นเสียง สำหรับนักเรียนระดับชั้นมัธยมศึกษาตอนปลายให้มีคุณภาพและสามารถนำไปประเมินสมรรถนะของนักเรียนได้ โดยแบบวัดมีความเที่ยงตรงเชิงเนื้อหา ความเหมาะสมของการนำไปใช้อยู่ในระดับเหมาะสมมากจนถึงเหมาะสมมากที่สุด เนื่องมาจากประการแรกผู้วิจัยใช้สถานการณ์ที่สอดคล้องกับชีวิตจริงของนักเรียน เช่น การเล่นกีตาร์ การได้ยินเสียงเชียร์กีฬา และการคัดเลือกผลไม้โดยใช้การสะท้อนของเสียง สถานการณ์ในลักษณะนี้จะช่วยส่งเสริมให้นักเรียนพิจารณาข้อมูลและใช้ความรู้ทางวิทยาศาสตร์ การคิดวิเคราะห์และทักษะอื่น ๆ เช่น การตีความหมายของข้อมูล เพื่ออธิบายข้อคำถามนั้น ๆ สะท้อนถึงสมรรถนะของนักเรียนได้อย่างแท้จริง ซึ่งสอดคล้องกับแนวทางการพัฒนาของ Pinchai & Seetee (2025) กล่าวว่าการใช้สถานการณ์ที่ใกล้เคียงกับบริบทของนักเรียนจะสามารถวัดสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ได้ดีกว่าสถานการณ์ที่นักเรียนไม่คุ้นชิน ประการที่สองผู้วิจัยเลือกใช้รูปแบบคำถาม เพราะเหตุใด นักเรียนทราบได้อย่างไร เพราะเหตุใดจึงคิดเช่นนั้น เป็นคำถามปลายเปิดที่ให้นักเรียนตอบได้อย่างอิสระเพื่อเปิดโอกาสให้นักเรียนแสดงหลักฐานและเหตุผลของนักเรียนได้อย่างเต็มความสามารถ และสามารถนำหลักฐานอื่นมาอ้างอิงเพิ่มเติมได้ ทำให้นักเรียนมีอิสระในการเรียบเรียงความสัมพันธ์เชิงสาเหตุและสะท้อนการใช้ความรู้ทางวิทยาศาสตร์ที่ซับซ้อนได้ สามารถประเมินสมรรถนะได้ดีกว่าข้อสอบแบบจำกัดคำตอบ ซึ่งประเด็นนี้สอดคล้องกับผลการศึกษาของ Wannathai & Pruekpramool (2021) ที่พบว่าข้อสอบอัตนัยไม่จำกัดคำตอบมีประสิทธิภาพสูงกว่าแบบจำกัดคำตอบ

แบบวัดนี้มีความเหมาะสมต่อการนำไปใช้มีค่า 4.40 - 5.00 เนื่องมาจากแบบวัดฉบับนี้ไม่ได้วัดความรู้ความจำของนักเรียนเพียงอย่างเดียว แต่วัดสมรรถนะของผู้เรียนที่ต้องใช้ทั้งความรู้ ทักษะการตีความจากข้อมูล เช่น ตาราง กราฟ หรือข้อความสถานการณ์ที่กำหนดให้มาใช้ในการอธิบายปรากฏการณ์ (Tang, 2023) ค่าความยากง่ายของแบบวัดมีค่าอยู่ระหว่าง 0.469 - 0.500 แสดงถึงข้อสอบแต่ละข้อมีความยากง่ายในระดับปานกลาง ค่าอำนาจจำแนกอยู่ระหว่าง 0.325 - 0.363 มีอำนาจจำแนกในระดับพอใช้ สะท้อนให้เห็นว่าแบบวัดสามารถจำแนกกลุ่มนักเรียนได้แต่เนื่องจากสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ต้องใช้ความรู้วิทยาศาสตร์อธิบายความสัมพันธ์เชิงสาเหตุ และต้องระบุข้อกล่าวอ้าง หลักฐาน และให้เหตุผลให้ครบตามองค์ประกอบจึงจะได้คะแนนสูง แต่เนื่องจากนักเรียนในกลุ่มนี้ไม่เคยได้รับการฝึกฝนสมรรถนะดังกล่าวมาก่อน ทำให้การกระจายของคะแนนในกลุ่มเก่งและกลุ่มอ่อนยังไม่ห่างกันมาก สอดคล้องกับงานวิจัยของ Janto et al. (2020); Yohyong & Pitipornatapin (2019) พบว่าการอธิบายอย่างเป็นวิทยาศาสตร์ของนักเรียนอยู่ในระดับต่ำ เนื่องจากนักเรียนขาดประสบการณ์ในการฝึกฝน และใช้ความรู้สีกส่วนตัวในการอธิบายปรากฏการณ์ อีกทั้งนักเรียนขาดการวิเคราะห์และเชื่อมโยงหลักฐาน รวมทั้งมีความเข้าใจเนื้อหาที่เรียนไม่ดีเท่าที่ควร (Lertdechapat & Promrat, 2018)

เกณฑ์การให้คะแนนของแบบวัดมีลักษณะเป็นเกณฑ์การให้คะแนนรายข้อ ให้คะแนนตามองค์ประกอบ และคุณภาพของคำอธิบาย พร้อมยกตัวอย่างคำตอบที่ถูกต้องในแต่ละระดับคะแนนทำให้มีเกณฑ์ที่ใช้ตรวจได้ง่ายและเหมาะสมต่อการนำไปใช้ ทั้งนี้เป็นเพราะระดับสมรรถนะแต่ละระดับมีการระบุพฤติกรรมและแสดงตัวอย่างที่สะท้อนสมรรถนะของนักเรียนได้ทุกระดับอย่างชัดเจน รวมทั้งผ่านผู้เชี่ยวชาญและการตรวจสอบความเป็นปรนัยโดยวิเคราะห์ค่าความเชื่อมั่นระหว่างผู้ประเมิน 3 ท่าน โดยใช้ค่าสหสัมพันธ์ภายในชั้น (ICC) ที่ความเชื่อมั่น 95% พบว่าในภาพรวมมีค่าระหว่าง 0.766-0.913 ซึ่งถือว่าผู้ประเมินแต่ละท่านมีการประเมินที่สอดคล้องกันสูง (Bajpai et al., 2015) สะท้อนให้เห็นถึงความเป็นปรนัยของเกณฑ์การให้คะแนนที่พัฒนาขึ้น สอดคล้องกับ Murphy et al. (2019) กล่าวว่าการวิเคราะห์ค่าความเชื่อมั่นระหว่างผู้ประเมินโดยใช้ค่าสหสัมพันธ์ภายในชั้น (ICC) มีค่าสูง จะสะท้อนถึงความเป็นปรนัยสูง อย่างไรก็ตามผู้วิจัยพบว่า สถานการณ์ที่ 3 เรื่อง การได้ยินของมนุษย์ มีค่าความยากง่ายต่ำสุด ($p = 0.469$) แต่มีค่าความเชื่อมั่นระหว่างผู้ประเมิน (ICC) สูงสุดเท่ากับ 0.913 สะท้อนให้เห็นว่าเกณฑ์การให้คะแนน (rubrics) ในหัวข้อนี้ มีความชัดเจนและเป็นปรนัยสูง ทำให้ผู้ประเมินสามารถตัดสินคุณภาพคำตอบที่ซับซ้อนของนักเรียนได้สอดคล้องกัน ส่วนสถานการณ์ที่ 1 เรื่อง การประยุกต์เกี่ยวกับเสียง มีค่าความเชื่อมั่นระหว่างผู้ประเมิน (ICC) ต่ำสุดเท่ากับ 0.766 เนื่องจากลักษณะของข้อคำถามที่เปิดกว้างให้นักเรียนตอบได้หลากหลายรูปแบบ ทำให้ผู้ประเมินมีการตีความความเชิงลึกแตกต่างกันในบางจุด ทั้งนี้ผู้วิจัยมีแนวทางในการปรับปรุงเกณฑ์การให้คะแนนโดยการเพิ่มนิยามเชิงพฤติกรรมให้เฉพาะเจาะจงยิ่งขึ้นและมีตัวอย่างคำตอบที่ชัดเจนเพื่อเป็นตัวอย่างให้ผู้ประเมินก่อนทำการตรวจให้คะแนนจริง เพิ่มความเป็นปรนัยและความเชื่อมั่นของการประเมิน นอกจากนี้แบบวัดมีค่าความเชื่อมั่นสูงโดยมีค่าสัมประสิทธิ์แอลฟาของครอนบาคเท่ากับ 0.719 สะท้อนให้เห็นถึงความสม่ำเสมอของคะแนนที่ได้จากการวัด เป็นเครื่องมือที่มีคุณภาพและเหมาะสมต่อการนำไปใช้ (Ningsi et al., 2019; Sener & Tas, 2017) จะเห็นได้ว่าแบบวัดที่ผู้วิจัยสร้างขึ้นสามารถนำไปใช้ประเมินสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ เรื่อง คลื่นเสียงของนักเรียนระดับชั้นมัธยมศึกษาตอนปลายได้อย่างมีประสิทธิภาพ

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

1.1 แบบวัดฉบับนี้เป็นแบบวัดสมรรถนะการอธิบายอย่างเป็นวิทยาศาสตร์ที่มีความจำเพาะต่อเนื้อหา เรื่อง คลื่นเสียง รายวิชาฟิสิกส์สำหรับนักเรียนระดับชั้นมัธยมศึกษาตอนปลาย ครูผู้สอนสามารถนำไปใช้ในการตรวจสอบสมรรถนะของนักเรียนเรื่อง คลื่นเสียงได้ เนื่องจากสถานการณ์ที่ใช้มีความสอดคล้องกับประสบการณ์ของนักเรียน

1.2 จากผลการวิจัยพบว่า แบบวัดมีค่าอำนาจจำแนกผ่านเกณฑ์แต่อยู่ในระดับพอใช้ หากนำไปใช้ควรคำนึงถึงระดับสมรรถนะของผู้เรียน และควรใช้หลังการจัดการเรียนรู้เพื่อให้นักเรียนมีความคุ้นชินกับการเขียนคำอธิบายอย่างเป็นวิทยาศาสตร์โดยใช้ความรู้ทางวิทยาศาสตร์อธิบายความสัมพันธ์เชิงสาเหตุ สร้างข้อกล่าวอ้าง หลักฐาน และการให้เหตุผล เพื่อตรวจสอบสมรรถนะของนักเรียนได้อย่างแท้จริง

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 ผลการวิจัยครั้งนี้ได้มาจากกลุ่มตัวอย่างที่เป็นนักเรียนจากโรงเรียนมัธยมศึกษาขนาดใหญ่พิเศษในจังหวัดแพร่เพียงโรงเรียนเดียว ซึ่งอาจจะมีบริบทที่แตกต่างจากโรงเรียนขนาดเล็กในต่างจังหวัด หรือโรงเรียนขนาดกลาง ในการวิจัยครั้งต่อไปจึงควรนำแบบวัดฉบับเดียวกันนี้ไปทดลองใช้กับนักเรียนในโรงเรียนอื่น ๆ ที่มีบริบทแตกต่างไปจากเดิม

2.2 การวิจัยครั้งนี้ใช้ทฤษฎีการทดสอบแบบดั้งเดิม (Classical Test Theory: CTT) ในการวิเคราะห์คุณภาพของเครื่องมือซึ่งมีข้อจำกัดที่ค่าพารามิเตอร์อาจเปลี่ยนแปลงตามกลุ่มผู้สอบ ในการวิจัยครั้งถัดไปควรขยายกลุ่มตัวอย่างให้มีความหลากหลายและจำนวนมากขึ้น และใช้ทฤษฎีการตอบสนอง (Item Response Theory: IRT) ในการวิเคราะห์คุณภาพเครื่องมือเพื่อสร้างคลังข้อสอบ (item bank) ต่อไป

References

- Alameh, S., Abd-El-Khalick, F., & Brown, D. (2022). The nature of scientific explanation: Examining the perceptions of the nature, quality, and “goodness” of explanation among college students, science teachers, and scientists. *Journal of Research in Science Teaching*, 59(10), 1833–1873.
- Andrade, V., Freire, S., & Baptista, M. (2019). Constructing scientific explanations: A system of analysis for students’ explanations. *Research in Science Education*, 49, 1323–1349. <https://doi.org/10.1007/s11165-017-9648-9>
- Bajpai, N., Bajpai, R., & Chaturvedi, K. (2015). Role of scientific explanation in science education. *International Journal of Science and Research (IJSR)*, 4(11), 1431–1434.
- McNeill, K. L., & Krajcik, J. S. (2008). Scientific explanations: Characterizing and evaluating the effects of explicit instructional support on student learning. *Science Education*, 92(1), 53–87.
- Murphy, P. K., Greene, J. A., Firetto, C. M., Hendrick, B. D., Li, M., Montalbano, C., & Wei, L. (2018). Quality Talk: Developing students’ discourse to promote high-level comprehension. *American Educational Research Journal*, 55(5), 1113–1160. <https://doi.org/10.3102/0002831218771303>
- Ningsi, S., Sappaile, B. I., & Darsikin, D. (2019). The effect of discovery learning model on the ability of scientific explanation of students in class XI SMA Negeri 1 Marawola. *Journal of Physics: Conference Series*, 1157(3), Article 032013. <https://doi.org/10.1088/1742-6596/1157/3/032013>

- Ruiz-Primo, M. A., Li, M., Tsai, S. P., & Schneider, J. (2010). Testing one premise of scientific inquiry in science classrooms: Examining students' scientific explanations and student learning. *Journal of Research in Science Teaching*, 47(5), 583–608.
- Sari, B. K., Sahyar, & Sani, R. A. (2021). The development of physics objective tests based on PISA for sound wave topics in high school. *Journal of Physics: Conference Series*, 1811(1), Article 012027. <https://doi.org/10.1088/1742-6596/1811/1/012027>
- Secolsky, C., & Denison, D. B. (2017). *Handbook on measurement, assessment, and evaluation in higher education*. Routledge.
- Sener, N., & Tas, E. (2017). The analysis of the scientific explanation levels of the primary school teacher candidates. *Universal Journal of Educational Research*, 5(8), 1311–1319.
- Tang, K. S. (2023). The characteristics of diagrams in scientific explanations: Multimodal integration of written and visual modes of representation in junior high school textbooks. *Science Education*, 107(3), 741–772.

Translate Thai References

- Institute for the Promotion of Teaching Science and Technology. (2023). *Results of PISA 2022: Reading, mathematics, and science*. Institute for the Promotion of Teaching Science and Technology. (in Thai)
- Janto, N., Kitkeunggul, S., & Chapu, S. (2020). Development of argumentative learning management using forensic science issues to promote scientific explanation construction abilities of 10th grade students on human homeostasis. *Journal of Education Naresuan University*, 24(1), 176–187. (in Thai)
- Lertdechapat, K., & Limpanont Promrat, P. (2018). The effects of collaborative inquiry on lower secondary school students' scientific explanation certificates. *Journal of Education Studies*, 46(2), 174–191. (in Thai)
- Pinchai, J., & Seetee, N. (2025). Development of a Scientific Explanation Test on the topic of Gases for Upper Secondary School Students. *Journal of Educational Measurement Mahasarakham University*, 31(1), 121–137. (in Thai)
- Srisa-ard, B. (2017). *Introduction to research*. Suwiriyan. (in Thai)
- Suwannahong, P., Seetee, N., & Khumwong, P. (2025). High school students' construction of scientific explanations on the circulatory system. *Journal of Education Studies, Chulalongkorn University*, 53(3), 1-18. (in Thai)

- Veeraphan, D., Wuttisela, K., Khampu, C., Thanaheuang, C., & Wuttirom, S. (2025). Inquiry-based physics learning on force in everyday life to enhance scientific argumentation skills. *Journal of Research Unit on Science, Technology and Environment for Learning*, 16(2), 143-155. (in Thai)
- Wannathai, P., & Pruekpramool, C. (2021). Comparing the efficiency between ability in constructing scientific explanation restricted- and extended-response subjective tests for Mathayom Suksa 4 students. *Journal of Graduate Research*, 12(2), 118–132. (in Thai)
- Yohyong, S., & Pitipornatapin, S. (2019). Development of grade 10 students' argumentation skills through science, technology, engineering, and mathematics (STEM) education by the 6E learning cycle in the topic of environmental conservation mission. *Journal of Research Unit on Science, Technology and Environment for Learning*, 10(2), 183–200. (in Thai)

การพัฒนาเครื่องมือวัดความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK ของครูมัธยมศึกษาในจังหวัดพิษณุโลก

Development of an Instrument to Measure Artificial Intelligence Technological Pedagogical Content Knowledge (AI-TPACK) Competencies of Secondary School Teachers in Phitsanulok Province

จิราลักษณ์ จันท์ดี^{1*} และ ณัฐกานต์ ประจันบาน²

Tiralak Jandee^{1*} and Nattakan Prachanban²

(Received: April 29, 2026; Revised: May 29, 2026; Accepted: May 29, 2026)

บทคัดย่อ

การวิจัยครั้งนี้มุ่งนำเสนอการพัฒนาเครื่องมือวัดความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK สำหรับครูมัธยมศึกษาในจังหวัดพิษณุโลก โดยประยุกต์ใช้รูปแบบแบบวัดเชิงสถานการณ์ มุ่งประเมินวิจารณ์ญาณเชิงวิชาชีพในการจัดการปัญหาในสถานการณ์ปฏิบัติงานจริง กลุ่มตัวอย่างครูมัธยมศึกษา จำนวน 405 คน ด้วยวิธีการสุ่มแบบหลายขั้นตอน เครื่องมือที่พัฒนาขึ้นประกอบด้วยองค์ประกอบ 4 ด้าน 16 ตัวบ่งชี้ รวม 45 ข้อคำถาม โดยใช้เกณฑ์การให้คะแนนแบบรูปริก 4 ระดับ ผลการตรวจสอบคุณภาพพบว่า แบบวัดมีค่าอำนาจจำแนกรายข้อสูงกว่า 0.20 ทุกข้อ และมีค่าความเที่ยงแบบความสอดคล้องภายในทั้งฉบับเท่ากับ 0.917 ผลการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง ยืนยันว่าโมเดลการวัดมีความตรงเชิงโครงสร้างและสอดคล้องกับข้อมูลเชิงประจักษ์ (Chi-square = 105.743, df = 86, p-value = 0.073, CFI = 0.992, TLI = 0.989, RMSEA = 0.024, SRMR = 0.028) เกณฑ์ปกติระดับท้องถิ่นที่สร้างขึ้นด้วยคะแนนมาตรฐานที่ปกติ (Normalized T-score) สามารถจำแนกความสามารถออกเป็น 5 ระดับ ผลการประเมินพบว่า ในภาพรวมครูมัธยมศึกษาส่วนใหญ่มีความสามารถในการบูรณาการปัญญาประดิษฐ์อยู่ในระดับปานกลาง (ร้อยละ 40.25) รองลงมาคือระดับสูง (ร้อยละ 25.68) และระดับต่ำ (ร้อยละ 23.70)

คำสำคัญ การพัฒนาเครื่องมือ ปัญญาประดิษฐ์ ครูมัธยมศึกษา AI-TPACK

¹ นิสิตระดับปริญญาโท สาขาวิจัยและประเมินทางการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยนเรศวร

² ผู้ช่วยศาสตราจารย์ ภาควิชาบริหาร วิจัย และพัฒนานวัตกรรมทางการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยนเรศวร

¹ Master's degree student, Educational Research and Evaluation, Faculty of Education, Naresuan University

² Assistant Professor, Educational Research and Evaluation, Faculty of Education, Naresuan University

* Corresponding Author E-mail: tjiraluck@gmail.com

Abstract

This research presents the development of an instrument to measure the Artificial Intelligence Technological Pedagogical Content Knowledge (AI-TPACK) competencies of secondary school teachers in Phitsanulok province. A Situational Judgment Test (SJT) format was applied to overcome the limitations of traditional self-report measures and assess professional judgment in real-world scenarios. Data were collected from a sample of 405 secondary school teachers selected through multi-stage random sampling. The developed instrument consisted of 4 main components and 16 behavioral indicators, totaling 45 items utilizing a 4-level rubric scoring system. The quality examination revealed that all items exhibited discrimination power greater than 0.20, and the overall internal consistency reliability was high at 0.917. Furthermore, the Second-Order Confirmatory Factor Analysis (CFA) confirmed strong construct validity and an excellent fit with the empirical data (Chi-square = 105.743, df = 86, p-value = 0.073, CFI = 0.992, TLI = 0.989, RMSEA = 0.024, SRMR = 0.028). Local norms were established using normalized T-scores, categorizing competencies into five levels. The assessment showed that the majority of secondary school teachers possessed a moderate level of AI-TPACK competency (40.25%), followed by high (25.68%) and low (23.70%) levels. This research provides a standardized measurement tool and essential information for planning teachers' professional development in the artificial intelligence era.

Keywords: instrument development, artificial intelligence, secondary school teachers, AI-TPACK

บทนำ

กรอบแนวคิด TPACK (Technological Pedagogical Content Knowledge) ของ Mishra & Koehler (2006) เป็นรากฐานสำคัญที่ใช้อธิบายและวัดสมรรถนะการบูรณาการเทคโนโลยีเข้ากับการจัดการเรียนการสอนของครู โดยเน้นการผสมผสานความรู้ด้านเทคโนโลยี เนื้อหา และวิธีการสอนเข้าด้วยกัน เพื่อให้การจัดการเรียนรู้มีประสิทธิภาพ เมื่อการศึกษาเข้าสู่ยุคดิจิทัลที่ปัญญาประดิษฐ์ (Artificial Intelligence: AI) เข้ามามีบทบาทสำคัญ กรอบแนวคิด TPACK เดิมจึงเริ่มมีข้อจำกัด เนื่องจาก AI ไม่ใช่เพียงเครื่องมือทั่วไปที่รอรับคำสั่ง แต่มีสถานะเสมือนตัวแทนที่มีความเป็นอิสระ สามารถเรียนรู้และตัดสินใจได้ในระดับหนึ่ง องค์การเพื่อการศึกษา วิทยาศาสตร์ และวัฒนธรรมแห่งสหประชาชาติ รวมถึงองค์การเพื่อความร่วมมือทางเศรษฐกิจและการพัฒนา (OECD, 2017) จึงได้เน้นย้ำถึงความจำเป็นเร่งด่วนที่ระบบการศึกษาต้องพัฒนาสมรรถนะของครูในการบูรณาการเทคโนโลยีปัญญาประดิษฐ์ การเปลี่ยนแปลงนี้ทำให้กรอบแนวคิด TPACK ดั้งเดิมต้องขยายขอบเขตสู่กรอบแนวคิด AI-TPACK ซึ่งเป็นการบูรณาการความรู้ด้านเทคโนโลยีปัญญาประดิษฐ์เข้ากับวิธีสอน เนื้อหาวิชา และเพิ่มมิติด้านจริยธรรมปัญญาประดิษฐ์ (AI Ethics) เข้าไปอย่าง

ชัดเจน (Celik, 2023; Ning et al., 2024) การประเมินและทำความเข้าใจความสามารถตามกรอบ AI-TPACK ของครูมัธยมศึกษาอย่างถูกต้องแม่นยำจึงเป็นสิ่งสำคัญ เพื่อให้ครูสามารถนำปัญญาประดิษฐ์ไปประยุกต์ใช้ในการจัดการเรียนการสอนได้อย่างมีประสิทธิภาพและมีความรับผิดชอบทางจริยธรรม (Chiu et al., 2025)

การประเมินความสามารถด้านเทคโนโลยีของครูมักอาศัยมาตรฐานค่าแบบประเมินตนเอง ซึ่งมีข้อจำกัดด้านความตรงเชิงโครงสร้างจากอคติในการเข้าข้างตนเอง (Kanjanawasee, 2013) ทำให้ผู้ตอบประเมินตนเองสูงกว่าจริงหรือเลือกตอบค่ากลาง ผลการประเมินจึงสะท้อนเพียงความรู้สึก ไม่อาจวัดความสามารถและวิจรรณญาณที่แท้จริงในการแก้ปัญหาหน้างานได้

เพื่อแก้ไขข้อจำกัดดังกล่าว การวิจัยนี้จึงนำแบบวัดเชิงสถานการณ์ (Situational Judgment Tests: SJTs) มาใช้พัฒนาเครื่องมือวัด (Wichitwanna, 2022) โดยจำลองปัญหาความขัดแย้งในยุคดิจิทัลเพื่อกระตุ้นให้ผู้ตอบใช้วิจรรณญาณทางวิชาชีพตัดสินใจแก้ปัญหา ครอบคลุม 4 องค์ประกอบ 16 ตัวบ่งชี้ ได้แก่ 1) ความรู้ด้านเทคโนโลยีปัญญาประดิษฐ์ (AI-TK) 2) ความรู้ด้านการสอนที่บูรณาการปัญญาประดิษฐ์ (AI-TPK) 3) ความรู้ด้านเนื้อหาที่บูรณาการปัญญาประดิษฐ์ (AI-TCK) และ 4) ความรู้ด้านจริยธรรมปัญญาประดิษฐ์ (AI Ethics) ให้คะแนนแบบรูปกรวยหลายค่าตามกรอบสมรรถนะของยูเนสโก (UNESCO, 2024) พร้อมตรวจสอบคุณภาพอย่างเป็นระบบ ทั้งความตรงเชิงเนื้อหา อำนาจจำแนก ความเที่ยง และความตรงเชิงโครงสร้างด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง (second-order CFA) รวมถึงสร้างเกณฑ์ปกติระดับท้องถิ่น (local norms) ด้วยคะแนนมาตรฐานที่ปกติ เพื่อประเมินความสามารถที่แท้จริงภายใต้บริบทของพื้นที่จังหวัดพิษณุโลก เนื่องจากจังหวัดพิษณุโลกเป็นศูนย์กลางทางการศึกษาของภาคเหนือตอนล่างที่มีสถานศึกษาระดับมัธยมศึกษาครอบคลุมทุกขนาด ได้แก่ ขนาดเล็ก ขนาดกลาง ขนาดใหญ่ และขนาดใหญ่พิเศษ จึงเป็นพื้นที่ที่มีความหลากหลายทางโครงสร้างพื้นฐานที่เหมาะสมอย่างยิ่งสำหรับการวิจัย เพื่อให้การสร้างเกณฑ์ปกติระดับท้องถิ่นมีความแม่นยำ และสามารถสะท้อนศักยภาพของพื้นที่ได้อย่างแท้จริง

การวิจัยครั้งนี้มีความมุ่งหมายเพื่อพัฒนาตัวบ่งชี้และเครื่องมือวัดความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK ของครูมัธยมศึกษา พร้อมทั้งตรวจสอบคุณภาพของเครื่องมือวัดความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK ตลอดจนเพื่อสร้างเกณฑ์ปกติและประเมินระดับความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK ของครูมัธยมศึกษาในจังหวัดพิษณุโลก ทั้งนี้เพื่อให้ได้เครื่องมือวัดเชิงสถานการณ์ที่ได้ตามหลักวิชาการ ปราศจากความคลาดเคลื่อนจากการประเมินตนเอง เนื่องจากการประเมินสมรรถนะ (competency-based assessment) ซึ่งมุ่งเน้นการประยุกต์ใช้ความรู้และทักษะในการปฏิบัติงานจริงนั้น แบบวัดเชิงสถานการณ์ (SJTs) มีความเหมาะสมและแตกต่างจากข้อสอบแบบเลือกตอบ (multiple choice) ทั่วไปอย่างมีนัยสำคัญ โดยข้อสอบแบบเลือกตอบมักมุ่งวัดเพียงระดับความรู้ความจำว่าถูกหรือผิด แต่ไม่สามารถรับประกันได้ว่าผู้ตอบจะสามารถนำความรู้นั้นไปประยุกต์ใช้แก้ปัญหาหน้างานได้จริง ในขณะที่แบบวัดเชิงสถานการณ์ถูกออกแบบมาเพื่อกระตุ้นให้ผู้ตอบต้องใช้ "วิจรรณญาณทางวิชาชีพ" (professional judgment) ในการตัดสินใจ ซึ่งจะทำให้ได้ข้อมูลสารสนเทศเชิงประจักษ์สำหรับให้หน่วยงานทางการศึกษาใช้ในการวางแผนพัฒนาวิชาชีพของครูให้ก้าวทันความเปลี่ยนแปลงในยุคปัญญาประดิษฐ์ได้อย่างมีประสิทธิภาพและยั่งยืน

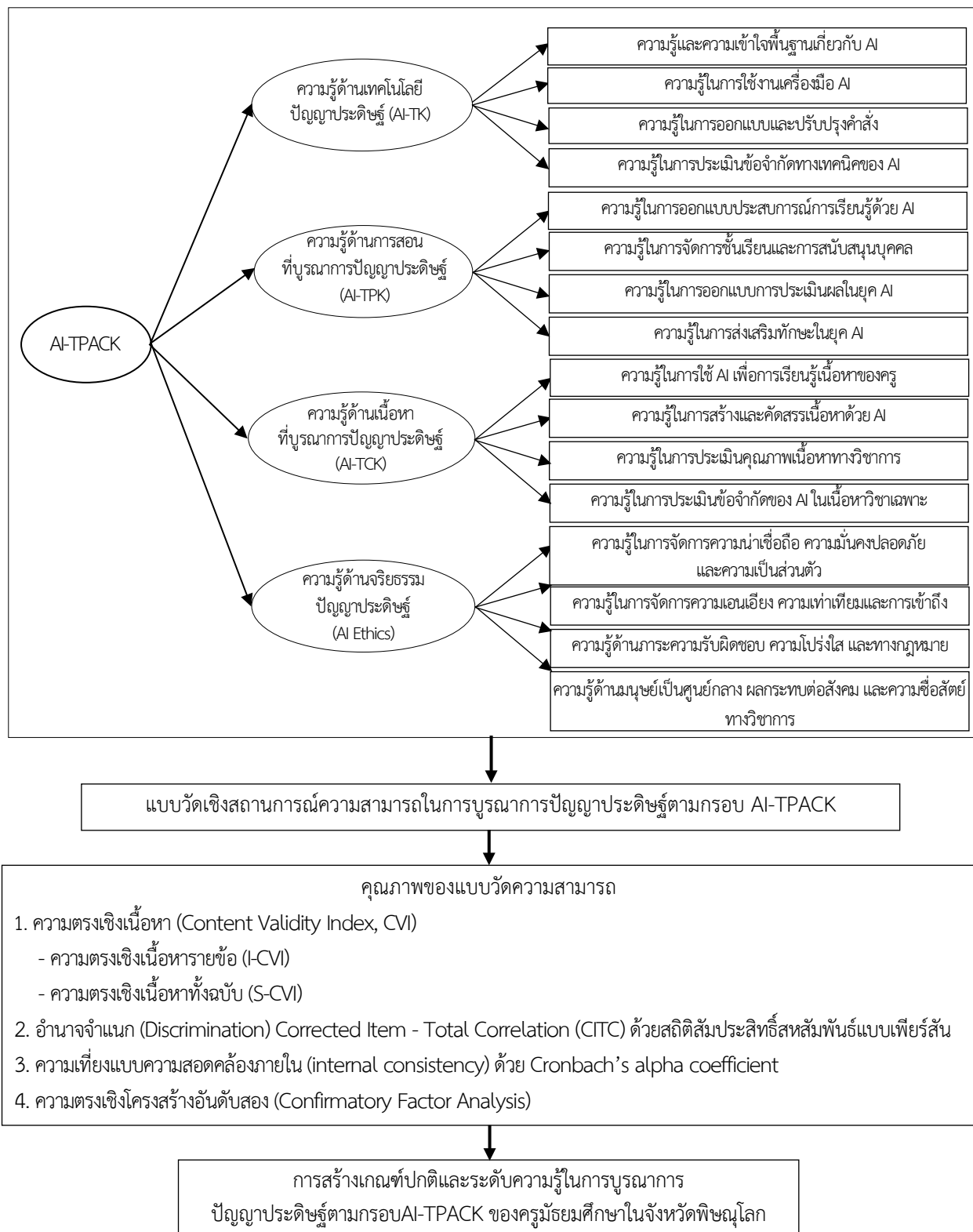
ความมุ่งหมายของการวิจัย

1. พัฒนาตัวบ่งชี้และเครื่องมือวัดความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK ของครูมัธยมศึกษา
2. ตรวจสอบคุณภาพของเครื่องมือวัดความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK ดังกล่าว
3. สร้างเกณฑ์ปกติและประเมินระดับความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK ของครูมัธยมศึกษาในจังหวัดพิษณุโลก

กรอบแนวคิดการวิจัย

การวิจัยครั้งนี้ ผู้วิจัยได้สังเคราะห์แนวคิดและทฤษฎีที่เกี่ยวข้อง โดยประยุกต์ใช้กรอบแนวคิด Intelligent-TPACK ของ Celik (2023) และโครงสร้างความรู้ AI-TPACK ของ Ning et al. (2024) ผสมผสานร่วมกับกรอบสมรรถนะปัญญาประดิษฐ์สำหรับครูขององค์การยูเนสโก (UNESCO, 2024) เพื่อกำหนดเป็นโครงสร้างความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK ของครูมัธยมศึกษา ซึ่งประกอบด้วย 4 องค์ประกอบหลัก 16 ตัวบ่งชี้ ได้แก่ 1) ความรู้ด้านเทคโนโลยีปัญญาประดิษฐ์ (AI-TK) 2) ความรู้ด้านการสอนที่บูรณาการปัญญาประดิษฐ์ (AI-TPK) 3) ความรู้ด้านเนื้อหาที่บูรณาการปัญญาประดิษฐ์ (AI-TCK) และ 4) ความรู้ด้านจริยธรรมปัญญาประดิษฐ์ (AI Ethics)

จากโครงสร้างองค์ประกอบดังกล่าว ผู้วิจัยได้นำมาพัฒนาเป็นเครื่องมือวัดรูปแบบแบบวัดเชิงสถานการณ์ (Situational Judgment Tests: SJTs) ชนิด 4 ตัวเลือก จำนวน 45 ข้อคำถาม ที่มุ่งวัดความรู้ในการจัดการปัญหาในสถานการณ์ปฏิบัติงานจริง โดยใช้เกณฑ์การให้คะแนนแบบหลายค่า (polytomous scoring) ตามเกณฑ์รูปทรี 4 ระดับ คือ 3, 2, 1 และ 0 คะแนน ซึ่งสอดคล้องกับระดับพัฒนาการตามกรอบของยูเนสโก ได้แก่ ระดับเชี่ยวชาญ ระดับชำนาญ ระดับกำลังพัฒนา และระดับเริ่มต้น จากนั้นจึงกำหนดกระบวนการตรวจสอบคุณภาพของแบบวัดอย่างเป็นระบบ ทั้งการตรวจสอบความตรงเชิงเนื้อหา ค่าอำนาจจำแนก ความเที่ยงแบบความสอดคล้องภายใน และความตรงเชิงโครงสร้างด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง (second-order CFA) ตลอดจนการนำผลลัพธ์ไปสร้างเกณฑ์ปกติระดับท้องถิ่นด้วยคะแนนมาตรฐานที่ปกติ เพื่อจัดแบ่งและประเมินระดับความสามารถในการบูรณาการปัญญาประดิษฐ์ของครูมัธยมศึกษาในจังหวัดพิษณุโลกออกเป็น 5 ระดับ ซึ่งสามารถสรุปเป็นกรอบแนวคิดการวิจัยได้ดังภาพ 1



ภาพ 1 กรอบแนวคิดการวิจัย

วิธีดำเนินการวิจัย

1. ประชากรและกลุ่มตัวอย่าง

การวิจัยครั้งนี้มีการกำหนดแหล่งข้อมูล ประชากร และกลุ่มตัวอย่าง โดยแบ่งการดำเนินการออกเป็น 3 ขั้นตอน ดังนี้

ขั้นตอนที่ 1 การพัฒนาตัวบ่งชี้ แหล่งข้อมูล ได้แก่ ผู้เชี่ยวชาญ จำนวน 5 คน ที่ได้จากการเลือกแบบเจาะจง (purposive selection) โดยมีคุณสมบัติเป็นอาจารย์ในสถาบันอุดมศึกษาที่มีคุณวุฒิ การศึกษาระดับปริญญาเอกหรือดำรงตำแหน่งทางวิชาการระดับผู้ช่วยศาสตราจารย์ขึ้นไป หรือเป็นศึกษานิเทศก์ที่ดำรงตำแหน่งวิทยฐานะชำนาญการพิเศษขึ้นไป ที่เชี่ยวชาญด้านเทคโนโลยีการศึกษาและคอมพิวเตอร์ศึกษา จำนวน 4 คน และอาจารย์ในสถาบันอุดมศึกษาที่มีความเชี่ยวชาญด้านการวัดและประเมินผลทางการศึกษา จำนวน 1 คน

ขั้นตอนที่ 2 การพัฒนาและตรวจสอบคุณภาพเครื่องมือวัด แบ่งกลุ่มตัวอย่างเป็น 2 กลุ่ม ได้แก่

1) กลุ่มตัวอย่างสำหรับการตรวจสอบคุณภาพด้านอำนาจจำแนกและความเที่ยง ประชากร ได้แก่ ครูมัธยมศึกษาที่ปฏิบัติหน้าที่ในสถานศึกษาจังหวัดอุดรดิตถ์ จำนวน 663 คน กลุ่มตัวอย่าง ได้แก่ ครูมัธยมศึกษาในจังหวัดอุดรดิตถ์ จำนวน 56 คน ได้มาโดยการสุ่มแบบหลายขั้นตอน โดยขั้นตอนที่ 1 ใช้วิธีการสุ่มแบบแบ่งชั้นภูมิ ตามขนาดสถานศึกษา 4 ขนาด ได้แก่ ขนาดเล็ก ขนาดกลาง ขนาดใหญ่ และขนาดใหญ่พิเศษ โดยคำนวณสัดส่วนให้เป็นไปตามสัดส่วนจริงของประชากร และขั้นตอนที่ 2 ใช้วิธีการสุ่มอย่างง่าย ด้วยการจับสลากรายชื่อโรงเรียน

2) กลุ่มตัวอย่างสำหรับการวิเคราะห์องค์ประกอบเชิงยืนยัน ประชากร ได้แก่ ครูมัธยมศึกษาที่ปฏิบัติหน้าที่ในสถานศึกษาจังหวัดพิษณุโลก ปีการศึกษา 2568 จำนวน 1,429 คน กลุ่มตัวอย่าง ได้แก่ ครูมัธยมศึกษาในจังหวัดพิษณุโลก จำนวน 405 คน ซึ่งได้มาโดยการสุ่มแบบหลายขั้นตอน โดยขั้นตอนที่ 1 ใช้วิธีการสุ่มแบบแบ่งชั้นภูมิ ตามขนาดสถานศึกษา 4 ขนาด และขั้นตอนที่ 2 ใช้วิธีการสุ่มอย่างง่าย ด้วยการจับสลากรายชื่อโรงเรียน ได้สถานศึกษาเป้าหมายจำนวน 21 โรงเรียน การกำหนดขนาดตัวอย่าง โดยใช้เกณฑ์ของ Hair et al. (2010) โดยกำหนดอัตราส่วนจำนวนหน่วยตัวอย่างต่อจำนวนพารามิเตอร์ที่ต้องประมาณค่าที่ 10:1 ซึ่งโมเดลการวิจัยนี้มีพารามิเตอร์ที่ต้องประมาณค่าจำนวน 36 พารามิเตอร์ จึงต้องการกลุ่มตัวอย่างขั้นต่ำ 360 คน ทั้งนี้ ผู้วิจัยได้กำหนดขนาดกลุ่มตัวอย่างเพิ่มเป็น 405 คน เพื่อป้องกันความคลาดเคลื่อนและให้เพียงพอต่อการวิเคราะห์องค์ประกอบ

ขั้นตอนที่ 3 การสร้างเกณฑ์ปกติ กลุ่มตัวอย่างที่ใช้เป็นกลุ่มตัวอย่างชุดเดียวกันกับกลุ่มตรวจสอบความตรงเชิงโครงสร้างในขั้นตอนที่ 2 ได้แก่ ครูมัธยมศึกษาในจังหวัดพิษณุโลก จำนวน 405 คน

2. เครื่องมือที่ใช้ในการวิจัย

เครื่องมือที่ใช้ในการวิจัยครั้งนี้ คือ แบบวัดเชิงสถานการณ์วัดความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK ของครูมัธยมศึกษา จำนวน 45 ข้อ ครอบคลุม 4 องค์ประกอบ 16 ตัวบ่งชี้ ได้แก่ 1) ความรู้ด้านเทคโนโลยีปัญญาประดิษฐ์ (AI-TK) จำนวน 12 ข้อ 2) ความรู้ด้านการสอน

ที่บูรณาการปัญญาประดิษฐ์ (AI-TPK) จำนวน 11 ข้อ 3) ความรู้ด้านเนื้อหาที่บูรณาการปัญญาประดิษฐ์ (AI-TCK) จำนวน 10 ข้อ และ 4) ความรู้ด้านจริยธรรมปัญญาประดิษฐ์ (AI ethics) จำนวน 12 ข้อ ลักษณะของข้อคำถามเป็นสถานการณ์จำลองที่มีตัวเลือกเชิงพฤติกรรมการตัดสินใจ 4 ตัวเลือก โดยใช้เกณฑ์การให้คะแนนแบบหลายค่า (polytomous scoring) แบบรูบริก (rubric scoring) คือ 3, 2, 1, และ 0 คะแนนลดหลั่นตามคุณภาพในการตัดสินใจแก้ปัญหาของวิจารณ์ญาณเชิงวิชาชีพครู โดยอิงระดับพัฒนาการตามกรอบสมรรถนะปัญญาประดิษฐ์ของยูเนสโก (UNESCO, 2024) ซึ่งสร้างและพัฒนาขึ้นใน ขั้นตอนที่ 2 ผู้วิจัยมีขั้นตอนการสร้างและตรวจสอบคุณภาพเครื่องมือ ดังนี้

2.1 การศึกษาและสังเคราะห์ตัวบ่งชี้ ผู้วิจัยศึกษาแนวคิด ทฤษฎี และงานวิจัยที่เกี่ยวข้องโดยใช้กรอบแนวคิดของ Celik (2023) และ Ning et al. (2024) ผสานกับกรอบสมรรถนะปัญญาประดิษฐ์สำหรับครูของยูเนสโก (2024) เพื่อนำมาสังเคราะห์เป็นร่างองค์ประกอบ ตัวบ่งชี้ และนิยามเชิงปฏิบัติการ AI-TPACK หมายถึง ความรู้ของครูในการตัดสินใจประยุกต์ใช้เทคโนโลยีปัญญาประดิษฐ์เข้ากับการจัดการเรียนการสอน ถูกต้องตามหลักวิชาการ และมีความรับผิดชอบทางจริยธรรม โดยโครงสร้างความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK ในการวิจัยครั้งนี้ ประกอบด้วยองค์ประกอบ 4 ด้าน ซึ่งมีนิยามเชิงปฏิบัติการและพฤติกรรมบ่งชี้ ดังตาราง 1

ตาราง 1 แสดงนิยามเชิงปฏิบัติการและพฤติกรรมบ่งชี้ของความสามารถ AI-TPACK

องค์ประกอบ	นิยามเชิงปฏิบัติการ	พฤติกรรมบ่งชี้
1. ความรู้ด้านเทคโนโลยีปัญญาประดิษฐ์ (AI-TK)	ความรู้ของครูที่เกี่ยวข้องกับเทคโนโลยีปัญญาประดิษฐ์เพื่อการสอน โดยเน้นความเข้าใจในหลักการทำงาน แนวคิดพื้นฐาน และการใช้งานเครื่องมือ AI ในระดับเทคนิค	1) มีความรู้และเข้าใจกลไกทำงาน ศักยภาพ และข้อจำกัดของ AI 2) สามารถสร้างบัญชี ใช้งาน และจัดการผลลัพธ์จากเครื่องมือ AI ได้ 3) สามารถออกแบบและปรับปรุงคำสั่ง (Prompt) ขั้นสูงได้ 4) ประเมินความเสี่ยงทางเทคนิค เช่น ข้อมูลเท็จและอคติของอัลกอริทึมได้
2. ความรู้ด้านการสอนที่บูรณาการปัญญาประดิษฐ์ (AI-Technological Pedagogical Knowledge: AI-TPK)	ความรู้ของครูว่า AI สามารถเปลี่ยนแปลงและส่งเสริมกระบวนการทางการสอนและการเรียนรู้ได้อย่างไร โดยไม่ขึ้นกับวิชาใดวิชาหนึ่งโดยเฉพาะ รวมถึงความรู้ในการเลือกใช้ AI เพื่อสนับสนุนกลยุทธ์การสอนที่แตกต่างกัน	1) ออกแบบประสบการณ์การเรียนรู้ที่รักษาสมดุลระหว่างการใช้และไม่ใช้ AI 2) ใช้ AI จัดการชั้นเรียนและสนับสนุนความต้องการผู้เรียนรายบุคคล 3) ออกแบบเกณฑ์การประเมินผลและการให้ข้อมูลป้อนกลับในยุค AI 4) ส่งเสริมให้ผู้เรียนถามคำถามและประเมินผลลัพธ์จาก AI อย่างมีประสิทธิภาพ

ตาราง 1 (ต่อ)

องค์ประกอบ	นิยามเชิงปฏิบัติการ	พฤติกรรมบ่งชี้
3. ความรู้ด้านเนื้อหาที่บูรณาการปัญญาประดิษฐ์ (AI-Technological Content Knowledge: AI-TCK)	ความรู้ของครูระหว่างเทคโนโลยี AI กับเนื้อหาวิชาเฉพาะทาง ครอบคลุมความรู้ในการพิจารณาว่า AI สามารถนำเสนอ จำลอง หรือสร้างเนื้อหาในรูปแบบใหม่ ๆ ได้ รวมถึงความรู้ในการประเมินความถูกต้องและความน่าเชื่อถือของเนื้อหาที่ AI สร้างขึ้นตามหลักการของสาขาวิชานั้นๆ	1) ใช้ AI สืบหาข้อมูลใหม่ ๆ เพื่อยกระดับความเชี่ยวชาญในวิชาของตน 2) ใช้ AI สร้างและคัดสรรสื่อ หรือแบบฝึกหัดที่เฉพาะเจาะจงในสาขาวิชา 3) ตรวจสอบความถูกต้องและระบุข้อผิดพลาดทางวิชาการของ AI 4) รู้ข้อจำกัดของ AI ในวิชาเฉพาะและรู้ว่าเมื่อใดควรใช้แหล่งข้อมูลอ้างอิงอื่น
4. ความรู้ด้านจริยธรรมปัญญาประดิษฐ์ (AI Ethics)	ความรู้ของครูในการวิเคราะห์ ประเมิน และตัดสินใจเลือกใช้หรือออกแบบการใช้ปัญญาประดิษฐ์ในบริบททางการศึกษา โดยยึดหลักการทางจริยธรรม กฎหมาย และความรับผิดชอบต่อสังคม เพื่อป้องกันผลกระทบเชิงลบที่อาจเกิดขึ้นกับผู้เรียนและผู้มีส่วนได้ส่วนเสีย ตลอดจนสามารถถ่ายทอดและปลูกฝังค่านิยมด้านความเท่าเทียม ความโปร่งใส และความน่าเชื่อถือของเทคโนโลยี AI ให้แก่ผู้เรียน	1) จัดการความน่าเชื่อถือ ความมั่นคงปลอดภัย และความความเป็นส่วนตัว (PDPA) 2) ลดความเหลื่อมล้ำและจัดการกับความเอนเอียงของอัลกอริทึม 3) มีความโปร่งใส เปิดเผยการใช้ AI และตระหนักถึงภาระความรับผิดชอบ 4) ให้นักเรียนเป็นศูนย์กลางในการตัดสินใจ และรักษาความซื่อสัตย์ทางวิชาการ

2.2 การตรวจสอบความเหมาะสมของตัวบ่งชี้ นำร่างตัวบ่งชี้ไปสร้างเป็นแบบประเมินชนิดมาตรประมาณค่า 5 ระดับ และเสนอต่อผู้เชี่ยวชาญจำนวน 5 คน เพื่อพิจารณาความเหมาะสม แม้การประเมินนี้จะมีผู้เชี่ยวชาญจำนวนน้อย แต่เนื่องจากข้อมูลลักษณะมาตรประมาณค่าเป็นตัวแปรระดับอันดับ ผู้วิจัยจึงเลือกใช้สถิตินอนพาราเมตริก ซึ่งมีความเหมาะสมต่อข้อตกลงเบื้องต้นมากกว่าการใช้สถิติพาราเมตริก (Kanjana-wasee, 2013) รวมทั้งความไว (sensitivity) ของสถิติดังกล่าวยังช่วยเพิ่มความเข้มงวดในการหา

ฉันทามติ ร่วมกันของกลุ่มผู้เชี่ยวชาญ ผู้วิจัยจึงกำหนดเกณฑ์คัดเลือกตัวบ่งชี้ โดยต้องมีความมัธยฐาน (median) ตั้งแต่ 3.50 ขึ้นไป และมีความเห็นสอดคล้องกันโดยมีค่าพิสัยระหว่างควอไทล์ (IQR) ไม่เกิน 1.50

2.3 การสร้างแบบวัด นำตัวบ่งชี้และนิยามศัพท์เชิงปฏิบัติการที่ผ่านการประเมินมาสร้างเป็นแบบวัดเชิงสถานการณ์ (SJTs) จำนวน 45 ข้อ โดยจำลองเหตุการณ์หรือความขัดแย้งทางการศึกษาในยุคดิจิทัล ที่ครูต้องเผชิญในการปฏิบัติงานจริง พร้อมสร้างตัวเลือกและเกณฑ์การให้คะแนนแบบรูปกริก 4 ระดับ ตามคุณภาพของการตัดสินใจ

2.4 การตรวจสอบความตรงเชิงเนื้อหา (content validity) นำแบบทดสอบที่สร้างขึ้นเสนอต่อผู้เชี่ยวชาญชุดเดิมจำนวน 5 คน เพื่อพิจารณาความสอดคล้องระหว่างสถานการณ์กับนิยามตัวบ่งชี้ รวมถึงความถูกต้องของเกณฑ์การให้คะแนน โดยวิเคราะห์ค่าดัชนีความตรงเชิงเนื้อหารายข้อ (I-CVI) และดัชนีความตรงทั้งฉบับ (S-CVI/Ave)

2.5 การหาค่าอำนาจจำแนกและความเที่ยง นำแบบทดสอบไปจัดทำในรูปแบบอิเล็กทรอนิกส์ (Google form) และประสานงานส่งหนังสือขอความอนุเคราะห์เก็บข้อมูลกับกลุ่มตัวอย่างครูมัธยมศึกษา ในจังหวัดอุดรดิตถ์ จำนวน 56 คน โดยให้ผู้ตอบใช้เวลาในการทำแบบทดสอบประมาณ 20 - 30 นาที และใช้ระยะเวลาในการเก็บรวบรวมข้อมูลรวม 1 สัปดาห์ จากนั้นนำชุดข้อมูลมาวิเคราะห์ค่าอำนาจจำแนกรายข้อ (item discrimination) ด้วยการหาค่าสัมประสิทธิ์สหสัมพันธ์ระหว่างคะแนนรายข้อกับคะแนนรวมแบบปรับแก้ (Corrected Item-Total Correlation: CITC) และวิเคราะห์หาความเที่ยงแบบความสอดคล้องภายใน (internal consistency reliability) ด้วยสูตรสัมประสิทธิ์แอลฟาของครอนบาค (Cronbach's Alpha coefficient)

2.6 การตรวจสอบความตรงเชิงโครงสร้าง (construct validity) นำแบบทดสอบฉบับสมบูรณ์ไปเก็บข้อมูลกับกลุ่มตัวอย่าง ได้แก่ ครูมัธยมศึกษาในจังหวัดพิษณุโลก จำนวน 405 คน จากนั้นนำข้อมูลเชิงประจักษ์มาวิเคราะห์ความสอดคล้องกลมกลืนของโมเดลการวัดด้วยสถิติการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสอง (second-order Confirmatory Factor Analysis: second-order CFA)

3. การเก็บรวบรวมข้อมูล

ผู้วิจัยได้ดำเนินการเก็บรวบรวมข้อมูลตามลำดับขั้นตอน ดังนี้

ขั้นตอนที่ 1 การเก็บรวบรวมข้อมูลเพื่อตรวจสอบความเหมาะสมของตัวบ่งชี้จากผู้เชี่ยวชาญ 5 คน โดยส่งแบบประเมินทางไปรษณีย์และด้วยตนเอง ใช้ระยะเวลา 5 สัปดาห์

ขั้นตอนที่ 2 การเก็บรวบรวมข้อมูลเพื่อตรวจสอบความตรงเชิงเนื้อหาจากผู้เชี่ยวชาญชุดเดิม จากนั้นประสานขอความอนุเคราะห์นำแบบทดสอบอิเล็กทรอนิกส์ (Google form) ไปเก็บข้อมูลกับครูมัธยมศึกษาในจังหวัดอุดรดิตถ์ จำนวน 56 คน เพื่อวิเคราะห์หาค่าอำนาจจำแนกและความเที่ยง ใช้ระยะเวลา 1 สัปดาห์

ขั้นตอนที่ 3 การเก็บรวบรวมข้อมูลเพื่อวิเคราะห์ความตรงเชิงโครงสร้างและสร้างเกณฑ์ปกติ ผู้วิจัยประสานขอความอนุเคราะห์เก็บข้อมูลกับครูมัธยมศึกษาในจังหวัดพิษณุโลก โดยบริหารการทดสอบอย่างเป็นมาตรฐานผ่านแบบวัดอิเล็กทรอนิกส์ (Google form) ที่ระบุคำชี้แจงชัดเจนและตั้งค่าบังคับตอบครบ

ทุกข้อเพื่อป้องกันข้อมูลสูญหาย โดยให้ผู้ตอบใช้เวลาประมาณ 20-30 นาทีต่อคน ใช้ระยะเวลาเก็บข้อมูลรวม 4 สัปดาห์ ได้รับข้อมูลตอบกลับอย่างสมบูรณ์จำนวนทั้งสิ้น 405 คน

4. การวิเคราะห์ข้อมูล

ขั้นตอนที่ 1 วิเคราะห์ความเหมาะสมของตัวบ่งชี้ด้วยค่ามัธยฐาน (Median) และค่าพิสัยระหว่างควอไทล์ (IQR)

ขั้นตอนที่ 2 วิเคราะห์ความตรงเชิงเนื้อหาด้วยค่าดัชนีความตรงเชิงเนื้อหารายข้อ (I-CVI) และดัชนีความตรงเชิงเนื้อหาทั้งฉบับแบบเฉลี่ย (S-CVI/Ave) จากนั้นวิเคราะห์ค่าอำนาจจำแนกรายข้อ (Corrected Item-Total Correlation: CITC) และหาค่าความเที่ยงแบบความสอดคล้องภายในด้วยสูตรสัมประสิทธิ์แอลฟาของครอนบาค ต่อมาจึงวิเคราะห์ข้อมูลพื้นฐาน ตรวจสอบข้อตกลงเบื้องต้นด้วย Bartlett's Test และดัชนี KMO และตรวจสอบความตรงเชิงโครงสร้าง ด้วยสถิติการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองด้วยโปรแกรมสำเร็จรูป Mplus

ขั้นตอนที่ 3 สร้างเกณฑ์ปกติระดับท้องถิ่นด้วยการแปลงคะแนนดิบเป็นคะแนนมาตรฐานที่ปกติ และหาค่าตำแหน่งเปอร์เซ็นต์ไทล์เพื่อจัดแบ่งระดับความสามารถออกเป็น 5 ระดับ

ผลการวิจัย

1. ผลการพัฒนาตัวบ่งชี้ความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK ของครูมัธยมศึกษา ผลการสังเคราะห์และประเมินความเหมาะสมโดยผู้เชี่ยวชาญจำนวน 5 คน พบว่าโครงสร้างความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK ประกอบด้วยองค์ประกอบ 4 ด้าน 16 ตัวบ่งชี้ ผลการประเมินพบว่าตัวบ่งชี้ทั้งหมดมีความเหมาะสมอยู่ในระดับมากที่สุด โดยมีค่ามัธยฐาน (median) อยู่ระหว่าง 4.50 – 5.00 และมีความสอดคล้องกันสูง โดยมีค่าพิสัยระหว่างควอไทล์ (IQR) อยู่ระหว่าง 0.00 – 1.00 ซึ่งผ่านเกณฑ์ที่กำหนดทุกตัวบ่งชี้

2. ผลการพัฒนาและตรวจสอบคุณภาพของเครื่องมือวัดความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK แบบวัดสถานการณ์ (SJTs) ชนิดเลือกตอบ 4 ตัวเลือก ที่พัฒนาขึ้นจำนวน 45 ข้อ มีเกณฑ์การให้คะแนนแบบรูปrik 4 ระดับ (3, 2, 1, 0) ตัวอย่างข้อคำถาม สถานการณ์ ครูท่านหนึ่งแสดงความกังวลว่า AI จะมาแทนที่ครู เพราะมีความรู้สึกลึกซึ้งและเข้าใจโลกได้ดีกว่ามนุษย์ คำถาม ท่านจะอธิบายหลักการทำงานของ AI อย่างไรให้ถูกต้องที่สุด ตัวเลือก 1. อธิบายว่า AI ประมวลผลทางสถิติเพื่อคาดเดาผลลัพธ์ ไม่ได้มีความรู้สึกลึกซึ้ง (3 คะแนน) 2. อธิบายว่า AI ฉลาดจากการจดจำข้อมูล (2 คะแนน) 3. อธิบายว่า AI คิดเหมือนมนุษย์แต่ไม่มีอารมณ์ (1 คะแนน) 4. สนับสนุนความเชื่อว่า AI มีความรู้สึกลึกซึ้งจริงและครูควรเตรียมตัวทำงาน (0 คะแนน) ผลการตรวจสอบคุณภาพพบว่า 1) ความตรงเชิงเนื้อหา มีค่าดัชนีความตรงรายข้อ (I-CVI) อยู่ระหว่าง 0.80 – 1.00 และดัชนีความตรงทั้งฉบับ (S-CVI/Ave) เท่ากับ 0.973 2) ค่าอำนาจจำแนกรายข้อ (CITC) อยู่ระหว่าง 0.301 – 0.746 3) ค่าความเที่ยงแบบความสอดคล้องภายในทั้งฉบับ (Cronbach's alpha) ในระดับสูงเท่ากับ 0.917 เมื่อพิจารณาค่าความเที่ยงรายองค์ประกอบพบว่า ด้านความรู้ด้านเทคโนโลยีปัญญาประดิษฐ์ (AI-TK) เท่ากับ 0.754 ด้านความรู้

ด้านการสอนที่บูรณาการปัญญาประดิษฐ์ (AI-TPK) เท่ากับ 0.727 ด้านความรู้ด้านเนื้อหาที่บูรณาการปัญญาประดิษฐ์ (AI-TCK) เท่ากับ 0.775 และด้านความรู้ด้านจริยธรรมปัญญาประดิษฐ์ (AI Ethics) เท่ากับ 0.841 และ 4) ความตรงเชิงโครงสร้างด้วยการวิเคราะห์องค์ประกอบเชิงยืนยันอันดับสองกับกลุ่มตัวอย่าง 405 คน พบว่าโมเดลมีความสอดคล้องกับข้อมูลเชิงประจักษ์เป็นอย่างดี ค่าน้ำหนักองค์ประกอบทุกตัวมีนัยสำคัญทางสถิติ ดังตาราง 2

ตาราง 2 แสดงค่าน้ำหนักองค์ประกอบ ค่าความคลาดเคลื่อนมาตรฐานของน้ำหนักองค์ประกอบ ค่าสถิติทดสอบ t และค่าสัมประสิทธิ์การพยากรณ์ขององค์ประกอบความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK (n=405) จากการวิเคราะห์องค์ประกอบเชิงยืนยันของโมเดลการวัดความสามารถในการบูรณาการปัญญาประดิษฐ์

องค์ประกอบและตัวบ่งชี้	น้ำหนักองค์ประกอบ			
	Factor loading	SE	t	R ²
การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่สอง (Second-Order CFA)				
1. ความรู้ด้านเทคโนโลยีปัญญาประดิษฐ์ (AI-TK)	0.839	0.030	27.792**	0.704
2. ความรู้ด้านการสอนที่บูรณาการ AI (AI-TPK)	0.890	0.023	38.910**	0.792
3. ความรู้ด้านเนื้อหาที่บูรณาการ AI (AI-TCK)	0.929	0.029	31.756**	0.863
4. ความรู้ด้านจริยธรรมปัญญาประดิษฐ์ (AI Ethics)	0.947	0.023	41.974**	0.897
การวิเคราะห์องค์ประกอบเชิงยืนยันอันดับที่หนึ่ง (First-Order CFA)				
TK1- ความรู้และความเข้าใจพื้นฐานเกี่ยวกับ AI	0.548	0.054	10.097**	0.300
TK2- ความรู้ในการใช้งานเครื่องมือ AI	0.678	0.034	20.011**	0.460
TK3- ความรู้ในการออกแบบและปรับปรุงคำสั่ง	0.701	0.033	21.453**	0.491
TK4- ความรู้ในการประเมินข้อจำกัดทางเทคนิคของ AI	0.756	0.030	25.521**	0.572
TPK1- ความรู้ในการออกแบบประสบการณ์การเรียนรู้ด้วย AI	0.541	0.040	13.644**	0.292
TPK2- ความรู้ในการจัดการชั้นเรียนและการสนับสนุนบุคคล	0.806	0.024	33.285**	0.649
TPK3- ความรู้ในการออกแบบการประเมินผลในยุค AI	0.629	0.035	18.078**	0.396
TPK4- ความรู้ในการส่งเสริมทักษะในยุค AI	0.756	0.027	27.861**	0.572
TCK1- ความรู้ในการใช้ AI เพื่อการเรียนรู้เนื้อหาของครู	0.607	0.038	16.093**	0.368
TCK2- ความรู้ในการสร้างและคัดสรรเนื้อหาด้วย AI	0.629	0.039	16.108**	0.396
TCK3- ความรู้ในการประเมินคุณภาพเนื้อหาทางวิชาการ	0.536	0.043	12.496**	0.287
TCK4- ความรู้ในการประเมินข้อจำกัดของ AI ในเนื้อหาวิชาเฉพาะ	0.614	0.039	15.643**	0.377
ETH1- ความรู้ในการจัดการความน่าเชื่อถือ ความมั่นคงปลอดภัย และความเป็นส่วนตัว	0.597	0.038	15.593**	0.357
ETH2- ความรู้ในการจัดการความเอนเอียง ความเท่าเทียมและการเข้าถึง	0.748	0.027	27.422**	0.560
ETH3- ความรู้ด้านการระมัดระวังความรับผิดชอบ ความโปร่งใส และทางกฎหมาย	0.616	0.036	17.104**	0.379
ETH4- ความรู้ด้านมนุษยเป็นศูนย์กลาง ผลกระทบต่อสังคม และความซื่อสัตย์ทางวิชาการ	0.782	0.027	29.360**	0.612
X ² = 105.743, df = 86, p-value = 0.073, RMSEA = 0.024, CFI = 0.992, TLI = 0.989, SRMR = 0.028				

3. ผลการสร้างเกณฑ์ปกติของเครื่องมือวัดความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK ผู้วิจัยสร้างเกณฑ์ปกติระดับท้องถิ่น (Local Norms) ด้วยคะแนนมาตรฐานที่ปกติ (Normalized T-score) จากกลุ่มตัวอย่างครูมัธยมศึกษาจำนวน 405 คน เพื่อจำแนกระดับความสามารถออกเป็น 5 ระดับ ผลการประเมินพบว่า ในภาพรวมครูมัธยมศึกษาส่วนใหญ่มีความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK อยู่ในระดับปานกลาง (ร้อยละ 40.25) รองลงมาคือระดับสูง (ร้อยละ 25.68) และระดับต่ำ (ร้อยละ 23.70) ตามลำดับ ดังตาราง 3

ตาราง 3 เกณฑ์ปกติและการแปลความหมายระดับความสามารถ AI-TPACK ในภาพรวมและรายองค์ประกอบ

AI-TPACK	คะแนนมาตรฐานที่ปกติ	คะแนนดิบ	การแปลความหมาย	จำนวน	ร้อยละ
เกณฑ์ปกติของความสามารถ AI-TPACK (ภาพรวม)					
	ตั้งแต่ 65 ขึ้นไป	133 – 135	ระดับสูงมาก	15	3.70
	55 – 64	127 – 132	ระดับสูง	104	25.68
	45 – 54	114 – 126	ระดับปานกลาง	163	40.25
	35 – 44	85 – 113	ระดับต่ำ	96	23.70
	ตั้งแต่ 34 ลงมา	57 – 84	ระดับต่ำมาก	27	6.670
รวม				405	100.00
เกณฑ์ปกติด้านความรู้ด้านเทคโนโลยีปัญญาประดิษฐ์ (AI-TK)					
	ตั้งแต่ 65 ขึ้นไป	-	ระดับสูงมาก	0	0
	55 – 64	35 – 36	ระดับสูง	105	25.93
	45 – 54	30 – 34	ระดับปานกลาง	182	44.94
	35 – 44	21 – 29	ระดับต่ำ	92	22.72
	ตั้งแต่ 34 ลงมา	16 – 20	ระดับต่ำมาก	26	6.41
รวม				405	100.00
เกณฑ์ปกติด้านความรู้ด้านการสอนที่บูรณาการปัญญาประดิษฐ์ (AI-TPK)					
	ตั้งแต่ 65 ขึ้นไป	-	ระดับสูงมาก	0	0
	55 – 64	32 – 33	ระดับสูง	116	28.64
	45 – 54	28 – 31	ระดับปานกลาง	161	39.75
	35 – 44	20 – 27	ระดับต่ำ	101	24.94
	ตั้งแต่ 34 ลงมา	10 – 19	ระดับต่ำมาก	27	6.67
รวม				405	100.00
เกณฑ์ปกติด้านความรู้ด้านเนื้อหาที่บูรณาการปัญญาประดิษฐ์ (AI-TCK)					
	ตั้งแต่ 65 ขึ้นไป	30 – 30	ระดับสูงมาก	28	6.91
	55 – 64	29 – 29	ระดับสูง	72	17.78
	45 – 54	26 – 28	ระดับปานกลาง	174	42.97
	35 – 44	17 – 25	ระดับต่ำ	106	26.17

ตาราง 3 (ต่อ)

AI-TPACK	คะแนนมาตรฐานที่ปกติ	คะแนนดิบ	การแปลความหมาย	จำนวน	ร้อยละ
	ตั้งแต่ 34 ลงมา	7 – 16	ระดับต่ำมาก	25	6.17
			รวม	405	100.00
เกณฑ์ปกติด้านความรู้ด้านจริยธรรมปัญญาประดิษฐ์ (AI Ethics)					
	ตั้งแต่ 65.00 ขึ้นไป	36 – 36	ระดับสูงมาก	51	12.59
	55.00 – 64.99	35 – 35	ระดับสูง	88	21.73
	45.00 – 54.99	32 – 34	ระดับปานกลาง	141	34.81
	35.00 – 44.99	19 – 31	ระดับต่ำ	97	23.95
	ตั้งแต่ 34.99 ลงมา	8 – 18	ระดับต่ำมาก	28	6.92
			รวม	405	100.00

อภิปรายผล

การวิจัยครั้งนี้สามารถอภิปรายประเด็นสำคัญจากการพัฒนาและประเมินความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK ได้ดังนี้

ความสอดคล้องของโครงสร้างเครื่องมือวัดกับแนวคิดสากล การที่โมเดลการวัดความสามารถ AI-TPACK มีความตรงเชิงโครงสร้างและสอดคล้องกับข้อมูลเชิงประจักษ์ เป็นผลมาจากการสังเคราะห์กรอบแนวคิดที่บูรณาการมิติด้านจริยธรรม (AI ethics) เข้าไปเป็นแกนหลักร่วมกับเทคโนโลยี วิธีสอน และเนื้อหา ซึ่งสะท้อนให้เห็นว่าการวัดสมรรถนะครูในยุคปัญญาประดิษฐ์ไม่อาจใช้กรอบ TPACK แบบดั้งเดิมได้ โครงสร้างที่พัฒนาขึ้นนี้มีความสอดคล้องกับแนวทางการสร้างเครื่องมือวัด Intelligent-TPACK ของ Celik (2023) และมาตรวัดของ Chiu et al. (2025) อย่างไรก็ตาม เครื่องมือวัดที่พัฒนาขึ้นในการวิจัยครั้งนี้มีความแตกต่างและอุดช่องว่างทางวิชาการของเครื่องมือดังกล่าว โดยเปลี่ยนจากการใช้มาตราประมาณค่าแบบประเมินตนเอง (Self-report rating scale) มาเป็นการวัดเชิงสมรรถนะผ่านแบบวัดเชิงสถานการณ์ ซึ่งช่วยให้สามารถประเมินวิจารณ์ญาณที่แท้จริงได้มากกว่าการสอบถามเพียงระดับความรู้สึหรือความมั่นใจ

การพัฒนาเครื่องมือวัดเชิงสถานการณ์ (SJTs) คุณภาพของเครื่องมือวัดที่อยู่ในเกณฑ์สูง ทั้งด้านความตรงเชิงเนื้อหา ค่าอำนาจจำแนก และความเที่ยง เป็นผลโดยตรงจากกระบวนการพัฒนาเครื่องมือที่ออกแบบข้อคำถามในลักษณะความขัดแย้งทางการศึกษาที่ครูต้องพบเจอจริงในชั้นเรียน (Wichitwannna, 2022) ผสานกับการใช้เกณฑ์การให้คะแนนแบบรูปริก 4 ระดับ ที่อิงตามกรอบพัฒนาการสมรรถนะปัญญาประดิษฐ์ของยูเนสโก (UNESCO, 2024) การกำหนดให้มีตัวเลือกที่สะท้อนถึงการหลีกเลี่ยงนวัตกรรมหรือสุ่มเสี่ยงต่อการละเมิดจริยธรรมเป็นเกณฑ์ 0 คะแนน ช่วยลดความคลาดเคลื่อนจากความลำเอียงและการเข้าข้างตนเอง (Kanjana-wasee, 2013) ทำให้เครื่องมือฉบับนี้มีอำนาจจำแนกสูงและสามารถวัดความสามารถของครูผู้สอนได้อย่างเที่ยงตรงปราศจากการบิดเบือนคำตอบ

จากการประเมินที่พบว่าครุมัธยมศึกษาส่วนใหญ่มีความสามารถในภาพรวมอยู่ในระดับปานกลาง เมื่อพิจารณาในรายองค์ประกอบพบว่า องค์ประกอบด้านความรู้ด้านการสอนที่บูรณาการปัญญาประดิษฐ์

(AI-TPK) เป็นด้านที่ครูมีฐานคะแนนต่ำที่สุด ข้อค้นพบเชิงประจักษ์จากงานวิจัยนี้เป็นตัวอธิบายและขยายความให้เห็นว่า แม้ครูจะมีความสามารถในการใช้งานเทคโนโลยีปัญญาประดิษฐ์พื้นฐาน (AI-TK) แต่การจะนำ AI มาประยุกต์ใช้เพื่อออกแบบกิจกรรมการเรียนรู้หรือประเมินผลผู้เรียนนั้นยังคงเป็นความท้าทาย ซึ่งผลลัพธ์นี้สอดคล้องกับงานวิจัยของ Hava & Babayigit (2025) และ Xie & Luo (2025) ที่ระบุว่า ความเชี่ยวชาญทางดิจิทัลทั่วไปเพียงอย่างเดียวไม่เพียงพอต่อการยกระดับสมรรถนะ AI-TPACK ครูจึงจำเป็นต้องได้รับการฝึกอบรมเชิงปฏิบัติการที่เจาะจงไปที่กลยุทธ์การสอน (Pedagogical strategies) เพื่อให้เกิดการเปลี่ยนผ่านจากการเป็นเพียงผู้ใช้เทคโนโลยีไปสู่ผู้บูรณาการเทคโนโลยีเพื่อการจัดการเรียนรู้ได้อย่างแท้จริง

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลวิจัยไปใช้

1.1 ข้อพึงระวังในการแปลความหมายคะแนนจากแบบวัดเชิงสถานการณ์ เนื่องจากการได้ 0 คะแนนจากเกณฑ์รูบริกไม่ได้สะท้อนการตอบผิดแบบข้อสอบความรู้ทั่วไป แต่บ่งชี้ถึงการหลีกเลี่ยงนวัตกรรมหรือสับสนเสี่ยงต่อการละเมิดจริยธรรม ผู้นำผลประเมินไปใช้จึงควรทำความเข้าใจบริบทนี้เพื่อป้องกันการแปลผลที่คลาดเคลื่อน

1.2 จากข้อค้นพบที่ครูในจังหวัดพิษณุโลกมีคะแนนด้านการสอน (AI-TPK) ต่ำที่สุด หน่วยงานควรปรับหลักสูตรพัฒนาบุคลากร โดยเปลี่ยนจากการอบรมทักษะ AI พื้นฐาน (AI-TK) ไปสู่การอบรมเชิงปฏิบัติการที่เน้นกลยุทธ์การสอน การออกแบบกิจกรรม และการประเมินผลที่บูรณาการ AI เข้าสู่ชั้นเรียนจริง

1.3 เกณฑ์การแปลความหมายระดับความสามารถในแบบวัดฉบับนี้ เนื่องจากเกณฑ์แปลผลนี้สร้างจากเกณฑ์ปกติระดับท้องถิ่น (Local Norms) เฉพาะจังหวัดพิษณุโลก หากพื้นที่อื่นต้องการนำไปใช้ ควรจัดทำเกณฑ์ปกติใหม่ด้วยการแปลงคะแนนเป็นมาตรฐานที่ปกติ เพื่อให้การแปลผลแม่นยำ เป็นธรรม และสอดคล้องกับบริบทที่แท้จริงของจังหวัดนั้น

2. ข้อเสนอแนะในการวิจัยครั้งต่อไป

2.1 ควรมีการตรวจสอบคุณภาพเครื่องมือด้วยสถิติขั้นสูงเพิ่มเติมในมิติอื่น เช่น การหาความตรงตามสภาพ (concurrent validity) หรือความตรงเชิงพยากรณ์ (predictive validity) เพื่อยืนยันว่าคะแนนจากแบบวัดเชิงสถานการณ์ (SJTs) สามารถพยากรณ์พฤติกรรมการสอนจริงของครูได้อย่างแม่นยำ

2.2 ควรนำเครื่องมือแบบวัดเชิงสถานการณ์ (SJTs) ไปใช้ประเมินกับกลุ่มประชากรครูในภูมิภาคอื่นหรือเปรียบเทียบบริบทเชิงสังคมที่มีความแตกต่างกัน เช่น โรงเรียนขนาดเล็กในชนบทกับโรงเรียนขนาดใหญ่พิเศษในเมือง หรือครูระดับประถมศึกษา เนื่องจากสภาพแวดล้อม โครงสร้างพื้นฐานทางเทคโนโลยี และวัฒนธรรมองค์กรที่ต่างกัน อาจส่งผลต่อความสามารถในการบูรณาการปัญญาประดิษฐ์ตามกรอบ AI-TPACK ที่แตกต่างกัน

References

- Celik, I. (2023). Towards Intelligent-TPACK: An empirical study on teachers' professional knowledge to ethically integrate artificial intelligence (AI)-based tools into education. *Computers in Human Behavior*, 138, Article 107468. <https://doi.org/10.1016/j.chb.2022.107468>
- Chiu, T. K. F., Ahmad, Z., & Çoban, M. (2025). Development and validation of teacher artificial intelligence (AI) competence self-efficacy (TAICS) scale. *Education and Information Technologies*, 30(5), 6667–6685. <https://doi.org/10.1007/s10639-024-13094-z>
- Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed.). Pearson Prentice Hall.
- Hava, K., & Babayiğit, Ö. (2025). Exploring the relationship between teachers' competencies in AI-TPACK and digital proficiency. *Education and Information Technologies*, 30(3), 3491–3508. <https://doi.org/10.1007/s10639-024-12939-x>
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054. <https://doi.org/10.1111/j.1467-9620.2006.00684.x>
- Ning, Y., Zhang, C., Xu, B., Zhou, Y., & Wijaya, T. T. (2024). Teachers' AI-TPACK: Exploring the relationship between knowledge elements. *Sustainability*, 16(3), 978. <https://doi.org/10.3390/su16030978>
- OECD. (2017). *Skills for a High Performing Civil Service*. OECD Publishing. <https://doi.org/10.1787/9789264280724-en>
- UNESCO. (2024). *AI competency framework for teachers*. UNESCO. <https://unesdoc.unesco.org/ark:/48223/pf0000391104>
- Xie, M., & Luo, L. (2025). *The status quo and future of AI-TPACK for mathematics teacher education students: A case study in Chinese universities* [Preprint]. arXiv. <https://doi.org/10.48550/arXiv.2503.13533>

Translate Thai References

- Kanjanawasee, S. (2013). *Classical test theory* (7th ed.). Chulalongkorn University Press. (in Thai)
- Wichitwanna, S. (2022). Situational judgment tests: Application for assessing character traits. *Academic Journal*, 3(1), 11–19. (in Thai)

การประเมินโครงการคณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชน กรณีศึกษาโรงเรียนบ้านหนองเล

Evaluation of Mathematics for Life and Community through the Sufficiency Economy Project Case Study of Ban Nongle School

วริศรา วงศ์หล่มแก้ว^{1*} และ สายฝน วิบูลรังสรรค์²

Warisara Wonglomkaew ^{1*} and Saifon Vibulrangson²

(Received: May 2, 2026; Revised: May 23, 2026; Accepted: May 25, 2026)

บทคัดย่อ

การประเมินโครงการครั้งนี้ ใช้รูปแบบการประเมินตามแนวคิดของไทเลอร์ (Tyler, 1949) มีวัตถุประสงค์เพื่อ 1) ประเมินวัตถุประสงค์ของโครงการ 2) ประเมินแผนการดำเนินโครงการ 3) ประเมินแนวทางในการพัฒนาโครงการ 4) ประเมินการนำโครงการไปปฏิบัติ 5) ประเมินผลลัพธ์ของโครงการ และ 6) ติดตามและประเมินผลกระทบของโครงการ แหล่งข้อมูล ประกอบด้วย ผู้บริหาร จำนวน 1 คน คณะครู จำนวน 6 คน นักการภารโรง จำนวน 1 คน นักเรียนชั้น ป. 1 – 6 จำนวน 43 คน ผู้ปกครองนักเรียน จำนวน 40 คน และเอกสารแสดงผลการเรียนรู้ระดับชั้น ป. 1 – 6 ปีการศึกษา 2567 และปีการศึกษา 2568 เครื่องมือที่ใช้ในการประเมิน ได้แก่ แบบประเมิน แบบสัมภาษณ์ แบบสังเกตพฤติกรรม แบบสอบถาม และแบบวัดเจตคติ มีผลการประเมิน ดังนี้ 1) ผลการประเมินวัตถุประสงค์ของโครงการ พบว่า วัตถุประสงค์ของโครงการ มีความสอดคล้องกับปัญหาและความต้องการในการพัฒนานักเรียน และความต้องการของผู้ปกครองในการพัฒนาทักษะทางคณิตศาสตร์ของนักเรียน และมีความเหมาะสมกับเนื้อหาในรายวิชาคณิตศาสตร์ บูรณาการร่วมกับเศรษฐกิจพอเพียง 2) ผลการประเมินแผนการดำเนินโครงการ พบว่า รูปแบบกิจกรรมในโครงการมีความเหมาะสม 3) ผลการประเมินแนวทางในการพัฒนาโครงการ พบว่า สื่อ อุปกรณ์ คณะครู อาคารและสถานที่ มีความพร้อม 4) ผลการประเมินการนำโครงการไปปฏิบัติ พบว่า ทั้ง 8 กิจกรรมมีความเหมาะสม เนื่องจากใช้กระบวนการเรียนรู้ที่ยึดโยงกับบริบทจริง แต่อย่างไรก็ตามมีข้อจำกัดด้านเวลาในบางกิจกรรมที่ซับซ้อน การนำเสนอผลงาน และความรอบคอบในการคิดเลขรวมถึงการใช้เวลานานในการคิดเลข 5) ผลการประเมินผลลัพธ์ของโครงการ พบว่า นักเรียนมีทักษะและกระบวนการทางคณิตศาสตร์ และทักษะการทำงานกลุ่ม ในภาพรวม

¹ นิสิตระดับปริญญาโท สาขาวิจัยและประเมินทางการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยนเรศวร

² ผู้ช่วยศาสตราจารย์ สาขาวิจัยและประเมินทางการศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยนเรศวร

¹ Master's degree student, Educational Research and Evaluation Program, Faculty of Education, Naresuan University

² Assistant Professor, Educational Research and Evaluation Program, Faculty of Education, Naresuan University

* Corresponding Author E-mail: warisarawo67@nu.ac.th

อยู่ในระดับดี นักเรียนมีความพึงพอใจต่อโครงการ และเจตคติในรายวิชาคณิตศาสตร์ ในภาพรวมอยู่ในระดับมาก และ 6) ผลการติดตามและประเมินผลกระทบของโครงการ พบว่า นักเรียนในระดับชั้น ป. 1 – 4 มีผลการเรียนรายวิชาคณิตศาสตร์เพิ่มขึ้น นอกจากนี้ นักเรียนที่เข้าร่วมโครงการส่วนใหญ่มีความตระหนักในการออมเงิน สามารถนำความรู้เรื่องการคำนวณและสถิติไปปรับใช้กับชีวิตประจำวัน

คำสำคัญ การประเมินโครงการ การเรียนรู้ตามสภาพจริง การเรียนรู้จากชีวิตจริงและชุมชน
คณิตศาสตร์ระดับประถมศึกษา

Abstract

This project evaluation was conducted based on Tyler's Evaluation Model (1949). The objectives of this evaluation were to assess: 1) project objectives, 2) implementation plans, 3) developmental guidelines, 4) project execution, 5) project outcomes, and 6) project impacts. The data sources comprised one administrator, six teachers, one janitor, 43 students (Grades 1–6), 40 parents, and academic achievement records from the 2024 and 2025 academic years. The evaluation instruments comprised assessment forms, interview schedules, behavioral observation scales, questionnaires, and attitude scales. The evaluation findings are as follows 1) The objectives were highly congruent with student development needs and parental expectations regarding mathematical skills, demonstrating a strong alignment with the mathematics curriculum integrated with Sufficiency Economy principles. 2) The activity formats were found to be appropriate and well-structured. 3) The project was well-supported by the readiness of instructional media, equipment, personnel, and facilities. 4) All eight activities were considered effective as they utilized learning processes rooted in authentic contexts. However, some limitations were noted, particularly time constraints for complex activities, student presentations, and the need for greater calculation speed and accuracy. 5) Students demonstrated a "Good" level of mathematical process skills and collaborative abilities. Overall, student satisfaction and attitudes toward mathematics were at a "High" level. And 6) The impact assessment indicated an improvement in mathematics academic performance among students in Grades 1–4. Most participants exhibited a heightened awareness of financial saving and were able to effectively apply calculation and statistical knowledge to their daily lives.

Keywords: program evaluation, authentic learning, life- and community-based learning, elementary mathematics

บทนำ

ระบบการศึกษาในศตวรรษที่ 21 ไม่ต้องการผู้เรียนที่เก่งกาจท่องจำ แต่คาดหวังให้เกิดผู้เรียนที่มีความสุขในการเรียน และสามารถสร้างทักษะการเรียนรู้ (learning skills) ได้ด้วยตนเอง เพื่อพัฒนาความรู้ของตนเองในการศึกษาหรือการประกอบอาชีพในอนาคต (Kongsilp, 2018) โดยเฉพาะทักษะทางคณิตศาสตร์ซึ่งถือเป็นเครื่องมือสำคัญที่ช่วยให้ผู้เรียนสามารถใช้เหตุผลทางตรรกะ เข้าใจความสัมพันธ์เชิงตัวเลข และนำไปประยุกต์ใช้ได้จริงในชีวิตประจำวัน โดยหลักสูตรแกนกลางการศึกษาขั้นพื้นฐาน พุทธศักราช 2551 (ฉบับปรับปรุง พ.ศ.2560) ได้มีการปรับสาระการเรียนรู้คณิตศาสตร์ เพื่อให้เหมาะสมกับช่วงวัยของผู้เรียน เน้นการเชื่อมโยงเนื้อหาคณิตศาสตร์สู่ชีวิตจริงมากขึ้น

อย่างไรก็ตามปัญหาที่พบทั่วไปในสถานศึกษา โดยเฉพาะโรงเรียนขนาดเล็กในชนบท คือ ผู้เรียนส่วนใหญ่ยังขาดความเชื่อมโยงระหว่างเนื้อหาทางคณิตศาสตร์กับสถานการณ์ในชีวิตจริง ทำให้การเรียนรู้คณิตศาสตร์เป็นเพียงการท่องจำสูตรหรือวิธีคิด โดยไม่ได้เกิดการพัฒนาสมรรถนะเชิงประยุกต์เท่าที่ควร ซึ่งโรงเรียนบ้านหนองเลห์เป็นโรงเรียนขนาดเล็กในพื้นที่ชนบท ก็ได้พบปัญหาเช่นเดียวกัน จึงได้เล็งเห็นถึงความสำคัญในการพัฒนากระบวนการเรียนรู้ทางคณิตศาสตร์ให้สอดคล้องกับแนวคิด “เรียนรู้จริงจากชีวิตและชุมชน” จึงได้ออกแบบโครงการ “คณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชน” ขึ้น มีลักษณะเป็นโครงการที่เชื่อมโยงคณิตศาสตร์เข้ากับวิถีชุมชนและปรัชญาของเศรษฐกิจพอเพียง ตามแนวพระราชดำริของพระบาทสมเด็จพระเจ้าอยู่หัวรัชกาลที่ 9 เพื่อพัฒนานักเรียนให้มีทักษะและกระบวนการทางคณิตศาสตร์ที่สามารถใช้ในชีวิตจริง เน้นการลงมือทำ ในบริบทที่แท้จริงของนักเรียน จัดขึ้นในภาคเรียนที่ 2 ปีการศึกษา 2568 ที่ดำเนินการจัดกิจกรรมในโครงการเป็นระยะเวลา 2 เดือน รวมทั้งสิ้น 8 กิจกรรม ได้แก่ 1) กิจกรรมสำรวจตัวตน เชื่อมชุมชนด้วยคณิต 2) กิจกรรมจิตอาสา ทำความดีตามรอยพ่อหลวง 3) กิจกรรมเดินตามรอยพ่อ ปลูกพืชพอเพียง 4) สำรวจร้านค้าชุมชน สร้างกราฟ 5) กิจกรรมออมเงินวันละบาท คำนวณเป้าหมาย 6) กิจกรรมผ่านมดย้อม ลายคณิตคิดสร้างสรรค์ 7) กิจกรรมคณิตคิดเมนูพอเพียง และ 8) กิจกรรมตลาดนัดจำลอง ฝึกซื้อขาย เพื่อเป็นเวทีให้ผู้เรียนได้ฝึกฝนทักษะคณิตศาสตร์เชิงประยุกต์จากกิจกรรมหลากหลายรูปแบบ ให้สนองกลยุทธ์สำนักงานเขตพื้นที่การศึกษาประถมศึกษา เพชรบูรณ์ เขต 2 ข้อที่ 2 ในการเพิ่มโอกาสและความเสมอภาคทางการศึกษาลดความเหลื่อมล้ำทางการศึกษา และกลยุทธ์ของโรงเรียนข้อที่ 1 พัฒนาคุณธรรมนำความรู้น้อมนำปรัชญาเศรษฐกิจพอเพียงสู่การเรียนการสอน (Ban Nongle School, 2026) ซึ่งการดำเนินโครงการใด ๆ การประเมินถือเป็นสิ่งสำคัญ เพราะโครงการมีรูปแบบของกิจกรรมที่ต้องการสารสนเทศในการปรับปรุง และพัฒนาในแต่ละขั้นของการดำเนินโครงการ

ทางผู้ประเมินต้องการดำเนินการตรวจสอบทุกกระบวนการของการดำเนินโครงการ ตั้งแต่ก่อนเริ่มดำเนินโครงการ ดำเนินโครงการ สิ้นสุดโครงการ จนถึงการทำกับติดตามโครงการอย่างต่อเนื่อง จากการศึกษาจึงได้เลือกการประเมินตามแนวคิดของไทเลอร์ (Tyler, 1949) ที่มีการพิจารณาถึงกระบวนการระหว่างทาง โดยเน้นให้การประเมินมีความเป็นระบบ สามารถนำไปใช้ตรวจสอบ ปรับปรุงโครงการได้อย่างมีประสิทธิภาพทุกขั้นตอน และนำผลการประเมินที่ได้ไปปรับปรุงโครงการให้ดียิ่งขึ้น โดย Pitiyanuwat (2013) และ Promchui (2009) กล่าวถึงกรอบความคิดของการประเมินโครงการใหม่ของไทเลอร์ (1986) ที่มีการประเมิน

ครอบคลุม 6 ประเด็น ดังนี้ 1) การประเมินวัตถุประสงค์ของโครงการ 2) การประเมินแผนการดำเนินโครงการ 3) การประเมินแนวทางในการพัฒนาโครงการ 4) การประเมินการนำโครงการไปปฏิบัติ 5) การประเมินผลลัพธ์ของโครงการ และ 6) การติดตามและประเมินผลกระทบของโครงการอย่างต่อเนื่อง

จากความสำคัญของโครงการและความสำคัญของการประเมินโครงการ ผู้ประเมินในฐานะผู้มีส่วนร่วมในการดำเนินโครงการ จึงได้นำรูปแบบการประเมินของไทเลอร์ (Tyler, 1949) มาเป็นแนวทางในการประเมินโครงการ ทั้งนี้เพื่อให้ได้สารสนเทศที่สำคัญในการปรับปรุงและพัฒนาโครงการในด้านต่าง ๆ อย่างมีประสิทธิภาพ ช่วยให้การดำเนินงานในอนาคตมีประสิทธิภาพมากขึ้น รวมถึงช่วยยกระดับคุณภาพการจัดกิจกรรมและสร้างประโยชน์สูงสุดต่อผู้เรียน ตลอดจนสนับสนุนการตัดสินใจในการพัฒนาโครงการที่ตอบสนองต่อเป้าหมายและบริบทของโรงเรียนได้อย่างยั่งยืน

ความมุ่งหมายของการวิจัย

เพื่อประเมินโครงการคณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชน กรณีศึกษาโรงเรียนบ้านหนองเล โดยการใช้รูปแบบการประเมินตามแนวคิดของไทเลอร์ (Tyler, 1949) โดยมีวัตถุประสงค์ย่อย ดังนี้

1. เพื่อประเมินวัตถุประสงค์ของโครงการคณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชน
2. เพื่อประเมินแผนการดำเนินของโครงการคณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชน
3. เพื่อประเมินแนวทางในการพัฒนาโครงการคณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชน
4. เพื่อประเมินการนำโครงการคณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชน ไปปฏิบัติ
5. เพื่อประเมินผลลัพธ์ของโครงการคณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชน
6. เพื่อติดตามและประเมินผลกระทบของโครงการคณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชน

วิธีดำเนินการวิจัย

1. แหล่งข้อมูล

1.1 ผู้บริหารโรงเรียน	จำนวน	1	คน
1.2 คณะครูที่เข้าร่วมโครงการ	จำนวน	6	คน
1.3 นักการภารโรง	จำนวน	1	คน
1.4 นักเรียนชั้นประถมศึกษาปีที่ 1 – 6	จำนวน	43	คน
1.5 ผู้ปกครอง	จำนวน	40	คน
1.6 เอกสารแสดงผลการเรียนรู้ระดับชั้น ป. 1 – 6 ปีการศึกษา 2567 และปีการศึกษา 2568			

2. รูปแบบการประเมิน

การประเมินครั้งนี้ ใช้รูปแบบการประเมินของไทเลอร์ (Tyler, 1949) โดยมุ่งประเมิน 6 ประเด็น สามารถเขียนเป็นกรอบการประเมิน ดังตาราง 1

ตาราง 1 แสดงกรอบการประเมินโครงการคณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชนตามแนวคิดของ ไทเลอร์ (Tyler, 1949)

วัตถุประสงค์/ตัวบ่งชี้การประเมิน	แหล่งข้อมูล	เครื่องมือ/วิธีการเก็บข้อมูล	วิเคราะห์ข้อมูล	เกณฑ์การตัดสิน
1. เพื่อประเมินวัตถุประสงค์ของโครงการคณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชน				
1.1 ความสอดคล้องของ วัตถุประสงค์ของโครงการกับ ปัญหาและความต้องการใน การพัฒนานักเรียน	- เอกสาร แสดงผลการ เรียน ปี 2567 - โครงร่าง โครงการ	แบบสรุปผลการเรียน ของนักเรียนชั้น ป. 1- 6 ปีการศึกษา 2567	- วิเคราะห์ แบบอุปนัย	มีหลักฐานเชิงประจักษ์ ที่สะท้อนให้เห็นถึง ปัญหาและความ ต้องการในการพัฒนา นักเรียน
1.2 ความสอดคล้องของ วัตถุประสงค์โครงการกับความ ต้องการของผู้ปกครองในการ พัฒนาทักษะทางคณิตศาสตร์	ผู้ปกครอง นักเรียน	แบบประเมิน วัตถุประสงค์ - ด้านความสอดคล้อง กับความต้องการของ	- ความถี่ - ร้อยละ	ผู้ประเมินเห็นว่ามีความ สอดคล้อง ร้อยละ 80 ขึ้นไป
1.3 ความเหมาะสมของ วัตถุประสงค์โครงการกับ เนื้อหาวิชาคณิตศาสตร์ บูรณา การร่วมกับเศรษฐกิจพอเพียง	ผู้บริหารและ คณะครูที่เข้า ร่วมโครงการ	ผู้ปกครองในการพัฒนา ทักษะของนักเรียน - ด้านความเหมาะสม กับเนื้อหา	<i>M, SD</i>	วัตถุประสงค์ของ โครงการมีความ เหมาะสมกับเนื้อหา ตั้งแต่ระดับมากขึ้นไป
2. เพื่อประเมินแผนการดำเนินของโครงการคณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชน				
ความเหมาะสมข้อกิจกรรม วัตถุประสงค์ ระยะเวลาใน แต่ละกิจกรรม วิธีดำเนินการ ผู้รับผิดชอบ และวัสดุอุปกรณ์	คณะครูที่เข้า ร่วมโครงการ	แบบสัมภาษณ์แผนการ ดำเนินของโครงการ	- วิเคราะห์ แบบอุปนัย	- มีหลักฐานเชิง ประจักษ์ที่สะท้อนให้ เห็นถึงความเหมาะสม ของแผน
3. เพื่อประเมินแนวทางในการพัฒนาโครงการคณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชน				
3.1 ความพร้อมด้านสื่อและ อุปกรณ์	คณะครูที่เข้า ร่วมโครงการ	แบบสัมภาษณ์แนวทาง ในการพัฒนาโครงการ	- วิเคราะห์ แบบอุปนัย	- มีหลักฐานเชิง ประจักษ์ที่สะท้อนให้ เห็นถึงความพร้อมด้าน สื่อและอุปกรณ์ คณะครู และอาคารและสถานที่
3.2 ความพร้อมของคณะครู คณะครูที่เข้าร่วมโครงการ	คณะครูที่เข้า ร่วมโครงการ	ของความพร้อมด้าน ต่าง ๆ		
3.3 ความพร้อมด้านอาคารและ สถานที่	- ครูสถานที่ - ภารโรง			
4. เพื่อประเมินการนำโครงการคณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชน ไปปฏิบัติ				
4.1 ความเหมาะสมของ กิจกรรมทั้ง 8 กิจกรรม	คณะครูที่เข้า ร่วมโครงการ	แบบสัมภาษณ์ การดำเนินงาน ของโครงการ	- วิเคราะห์ แบบอุปนัย	- มีหลักฐานเชิง ประจักษ์ที่สะท้อนให้ เห็นถึงความเหมาะสม ของกิจกรรม
4.2 ปัญหาและอุปสรรค				

ตาราง 1 (ต่อ)

วัตถุประสงค์/ตัวบ่งชี้การประเมิน	แหล่งข้อมูล	เครื่องมือ/ วิธีการเก็บข้อมูล	วิเคราะห์ข้อมูล	เกณฑ์การตัดสิน
5. เพื่อประเมินผลลัพธ์ของโครงการคณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชน				
5.1 ทักษะและกระบวนการทางคณิตศาสตร์ - ทักษะการแก้ปัญหา - ทักษะการให้เหตุผล - ทักษะการสื่อสาร - ทักษะการเชื่อมโยงความรู้	- นักเรียนชั้น ป.1 – 6	แบบประเมินพฤติกรรม ด้านทักษะและ กระบวนการทาง คณิตศาสตร์ของ นักเรียน	<i>M, SD</i>	- นักเรียนมีค่าเฉลี่ย ตั้งแต่ระดับดี ขึ้นไป
5.2 ทักษะการทำงานกลุ่ม - ด้านความสามัคคี - ด้านสัมพันธภาพในกลุ่ม - ด้านความรับผิดชอบ	- นักเรียนชั้น ป.1 – 6	แบบสังเกตการทำงาน กลุ่ม	<i>M, SD</i>	- นักเรียนมีค่าเฉลี่ย ตั้งแต่ระดับดี ขึ้นไป
5.3 ความพึงพอใจของ นักเรียนที่มีโครงการ และ กิจกรรมในโครงการ	- นักเรียนชั้น ป.1 – 6	แบบสอบถามความพึง พอใจต่อโครงการ แบบสะท้อนความรู้สึก ความพึงพอใจ	<i>M, SD.</i> <i>M, SD</i>	มีคะแนนตั้งแต่ระดับ มากขึ้นไป มีคะแนนตั้งแต่ระดับ ปานกลางขึ้นไป
5.4 เจตคติของนักเรียนใน รายวิชาคณิตศาสตร์	- นักเรียนชั้น ป.1 – 6	แบบวัดเจตคติ	<i>M, SD</i>	มีเจตคติ อยู่ในระดับ มากขึ้นไป
6. เพื่อติดตามและประเมินผลกระทบของโครงการคณิตคิดพอเพียง เรียนรู้จริงจากชีวิตและชุมชน อย่างต่อเนื่อง				
6.1 ผลสัมฤทธิ์ทางการเรียนใน รายวิชาคณิตศาสตร์	- นักเรียนชั้น ป.1 – 6	เอกสารแสดงผลการ เรียน ปีการศึกษา 2568	<i>M</i>	- ผลสัมฤทธิ์วิชา คณิตศาสตร์ เฉลี่ยสูงขึ้น เมื่อเทียบกับปีก่อนหน้า
6.2 พฤติกรรมการเรียนของ นักเรียนหลังเสร็จสิ้นโครงการ	คณะครูที่เข้า ร่วมโครงการ	แบบสัมภาษณ์เกี่ยวกับ พฤติกรรมของนักเรียน ที่เปลี่ยนแปลงไป	- วิเคราะห์ แบบอุปนัย	- มีหลักฐานเชิง ประจักษ์ที่สะท้อนให้ เห็นถึงพฤติกรรมที่ เปลี่ยนแปลงไป

3. เครื่องมือที่ใช้ในการประเมิน แบ่งเป็น 6 ประเภท ประกอบด้วย

3.1 แบบประเมิน จำนวน 3 ฉบับ ได้แก่

1) แบบประเมินวัตถุประสงค์ของโครงการ ด้านความสอดคล้องของวัตถุประสงค์กับความต้องการของผู้ปกครองในการพัฒนาทักษะของนักเรียน สำหรับผู้ปกครองนักเรียน มีลักษณะเป็นแบบตรวจสอบรายการ โดยมีให้เลือก 3 คำตอบ คือ สอดคล้อง ไม่แน่ใจ และไม่สอดคล้อง จำนวน 4 ข้อ

2) แบบประเมินวัตถุประสงค์ของโครงการ ด้านความเหมาะสมของวัตถุประสงค์กับเนื้อหาในรายวิชาคณิตศาสตร์ บูรณาการร่วมกับเศรษฐกิจพอเพียง สำหรับผู้บริหารและคณะครู ที่เข้าร่วมโครงการ เป็นแบบมาตราประมาณค่า 5 ระดับ โดยมีเกณฑ์การวัดระดับความเหมาะสม ดังนี้ ระดับ 5 หมายถึง เหมาะสมมากที่สุด ระดับ 4 หมายถึง เหมาะสมมาก ระดับ 3 หมายถึง เหมาะสมปานกลาง ระดับ 2 หมายถึง เหมาะสมน้อย ระดับ 1 หมายถึง เหมาะสมน้อยที่สุด จำนวน 11 ข้อ

3) แบบประเมินพฤติกรรมด้านทักษะและกระบวนการทางคณิตศาสตร์ของนักเรียน เกณฑ์การให้คะแนนแบบรูบิก 4 ระดับ

3.2 แบบสัมภาษณ์ จำนวน 6 ฉบับ ได้แก่

1) แบบสัมภาษณ์แผนการดำเนินของโครงการ ด้านความเหมาะสมของรูปแบบกิจกรรม สำหรับคณะครูที่เข้าร่วมโครงการ ลักษณะแบบมีโครงสร้าง จำนวน 12 ข้อ

2) แบบสัมภาษณ์แนวทางในการพัฒนาโครงการ ด้านความพร้อมของสื่อและอุปกรณ์ที่ใช้ในการดำเนินกิจกรรมในโครงการ สำหรับคณะครูที่เข้าร่วมโครงการ ลักษณะแบบมีโครงสร้าง จำนวน 3 ข้อ

3) แบบสัมภาษณ์แนวทางในการพัฒนาโครงการ ด้านความพร้อมของคณะครูที่เข้าร่วมโครงการ สำหรับคณะครูที่เข้าร่วมโครงการ ลักษณะแบบมีโครงสร้าง จำนวน 3 ข้อ

4) แบบสัมภาษณ์แนวทางในการพัฒนาโครงการ ด้านความพร้อมของอาคารและสถานที่ในการจัดกิจกรรมในโครงการ จำนวน 2 ฉบับ ได้แก่ ฉบับที่ 1 สำหรับครูหัวหน้าฝ่ายสถานที่ และภารโรง ลักษณะแบบมีโครงสร้าง จำนวน 5 ข้อ และฉบับที่ 2 สำหรับคณะครูที่เข้าร่วมโครงการ ลักษณะแบบมีโครงสร้าง จำนวน 4 ข้อ

5) แบบสัมภาษณ์การดำเนินงานของโครงการ ด้านความเหมาะสมของกิจกรรมในโครงการ และปัญหาและอุปสรรคในระหว่างการดำเนินการ สำหรับคณะครูที่เข้าร่วมโครงการ ลักษณะแบบมีโครงสร้าง จำนวน 4 ข้อ

6) แบบสัมภาษณ์เกี่ยวกับพฤติกรรมของนักเรียนที่เปลี่ยนแปลงไป หลังเสร็จสิ้นโครงการ สำหรับคณะครูที่เข้าร่วมโครงการ ลักษณะแบบมีโครงสร้าง จำนวน 5 ข้อ

3.3 แบบสอบถามความพึงพอใจ ของนักเรียนที่มีต่อโครงการ ได้แก่ ด้านครูผู้สอน ด้านเนื้อหา ด้านการจัดกิจกรรม และด้านการวัดและประเมินผล เป็นแบบมาตราประมาณค่า 4 ระดับ โดยมีเกณฑ์การวัด ดังนี้ 4 หมายถึง ระดับมากที่สุด 3 หมายถึง ระดับมาก 2 หมายถึง ระดับน้อย และ 1 หมายถึง ระดับน้อยที่สุด จำนวน 17 ข้อ

3.4 แบบสังเกตพฤติกรรม จากการทำกิจกรรม ด้านความสามัคคี ด้านสัมพันธภาพในกลุ่ม และด้านความรับผิดชอบ เกณฑ์การให้คะแนนแบบรูบิก 4 ระดับ

3.5 แบบวัดเจตคติ ต่อยาวิชาคณิตศาสตร์ เป็นแบบมาตราประมาณค่า 4 ระดับ โดยมีเกณฑ์การวัด ดังนี้ ระดับ 4 หมายถึง เห็นด้วยมากที่สุด ระดับ 3 หมายถึง เห็นด้วยมาก ระดับ 2 หมายถึง เห็นด้วยน้อย และ ระดับ 1 หมายถึง เห็นด้วยน้อยที่สุด จำนวน 8 ข้อ

3.6 แบบสะท้อนความรู้สึกความพึงพอใจ ของนักเรียนที่มีต่อกิจกรรมแต่ละสัปดาห์ โดยใช้โพสอิท ทั้งหมด 8 กิจกรรม มีลักษณะเป็นการให้นักเรียนเลือกระดับความพึงพอใจของกิจกรรมในโครงการ โดยใช้สติ๊กเกอร์ที่ครูผู้ควบคุมกิจกรรมเป็นผู้แจกให้ โดยมีระดับความพึงพอใจ ซึ่งจะใช้สัญลักษณ์แทนความรู้สึก เพื่อให้นักเรียนเห็นภาพชัดเจนขึ้น ดังนี้ 3 หมายถึง ระดับมาก 2 หมายถึง ระดับปานกลาง และ 1 หมายถึง มีความพึงพอใจในระดับน้อย

4. การสร้างเครื่องมือและตรวจสอบคุณภาพเครื่องมือประเมิน

4.1 ศึกษาเอกสาร แนวคิด ทฤษฎีเกี่ยวกับการสร้างเครื่องมือ และกำหนดกรอบนิยามตามประเด็นการประเมิน และสร้างเครื่องมือตามนิยามของประเด็นการประเมิน ตามลักษณะของเครื่องมือที่กำหนด

4.2 นำเครื่องมือทั้งหมดไปหาความเที่ยงตรงเชิงเนื้อหา โดยให้ผู้เชี่ยวชาญ จำนวน 3 ท่าน พิจารณาสอดคล้องของข้อความกับนิยามศัพท์ของประเด็นการประเมินครั้งนี้ โดยการหาค่าดัชนีความสอดคล้อง พบว่า แบบประเมิน แบบสอบถาม แบบสังเกต แบบวัดทัศนคติ และแบบสัมภาษณ์ มีค่าดัชนีความสอดคล้อง เท่ากับ 0.67 - 1.00 จึงถือว่ามีความเที่ยงตรงเชิงเนื้อหา สามารถนำไปสอบถามได้ และปรับปรุงข้อความคำถามตามคำแนะนำของผู้เชี่ยวชาญ มีการปรับแก้ในส่วนข้อความที่ยังไม่ครอบคลุม และภาษาที่เข้าใจยากสำหรับผู้ตอบในกลุ่มของนักเรียน ให้มีความเข้าใจง่าย และชัดเจนมากยิ่งขึ้น

5. การเก็บรวบรวมข้อมูล

5.1 จัดทำหนังสือจากคณะศึกษาศาสตร์ มหาวิทยาลัยนเรศวร ขอความร่วมมือในการเก็บข้อมูลของนักเรียนโรงเรียนบ้านหนองเล

5.2 การเก็บรวบรวมข้อมูล มีขั้นตอนในการดำเนินงาน ตามตาราง 2 ดังนี้

ตาราง 2 แสดงขั้นตอนในการดำเนินงานและการเลือกใช้เครื่องมือในแต่ละขั้นตอน

ช่วงเวลา	ขั้นตอน	เครื่องมือที่ใช้
10 ต.ค. 2568 - 31 ต.ค. 2568	ประเมินวัตถุประสงค์ของโครงการ	1) แบบสรุปผลการเรียนนักเรียน ปีการศึกษา 2567 2) แบบประเมินวัตถุประสงค์ของโครงการ
11 พ.ย. 2568	ประเมินแผนการดำเนินของโครงการ	แบบสัมภาษณ์แผนการดำเนินของโครงการ
18 พ.ย. 2568	ประเมินแนวทางในการพัฒนาโครงการ	แบบสัมภาษณ์แนวทางในการพัฒนาโครงการ
25 พ.ย. 2568 - 28 ก.พ. 2569	ประเมินการนำโครงการไปปฏิบัติ	1) แบบสัมภาษณ์การดำเนินงานของโครงการ 2) แบบสอบถามสะท้อนความรู้สึกความพึงพอใจ
28 ก.พ. 2569	ประเมินผลลัพธ์ของโครงการ	1) แบบประเมินพฤติกรรม ด้านทักษะและกระบวนการทางคณิตศาสตร์ของนักเรียน 2) แบบสังเกตการทำงานกลุ่ม 3) แบบสอบถามความพึงพอใจของนักเรียน 4) แบบวัดเจตคติของนักเรียน
31 มี.ค. 2569	ติดตามและประเมินผลกระทบของโครงการ	1) แบบสรุปผลการเรียนนักเรียน ปีการศึกษา 2568 2) แบบสัมภาษณ์เกี่ยวกับพฤติกรรมที่เปลี่ยนแปลงไป

6. การวิเคราะห์ข้อมูล

6.1 การวิเคราะห์ข้อมูลเชิงปริมาณ

1) วิเคราะห์ข้อมูลความสอดคล้องของวัตถุประสงค์โครงการกับความต้องการของผู้ปกครอง ในการพัฒนาศักยภาพของนักเรียน โดยใช้ความถี่ ร้อยละ

2) วิเคราะห์ข้อมูลแบบประเมินวัตถุประสงค์ของโครงการ ความพึงพอใจของนักเรียนที่มีต่อการในโครงการ ความพึงพอใจของนักเรียนที่มีต่อกิจกรรมในโครงการทั้ง 8 กิจกรรม พฤติกรรมด้านทักษะและกระบวนการทางคณิตศาสตร์ของนักเรียน ด้านทักษะการทำงานกลุ่มของนักเรียน และเจตคติของนักเรียนหลังเสร็จสิ้นโครงการ โดยใช้ค่าเฉลี่ย และส่วนเบี่ยงเบนมาตรฐาน

6.2 การวิเคราะห์ข้อมูลเชิงคุณภาพ

แบบสัมภาษณ์ใช้การวิเคราะห์แบบอุปนัย ร่วมกับหลักฐานเชิงประจักษ์

ผลการประเมิน

ผู้ประเมินได้สรุปผลการประเมิน ตามวัตถุประสงค์โดยแบ่งออกเป็น 6 ตอน ดังนี้

ตอนที่ 1 ผลการประเมินวัตถุประสงค์ของโครงการ พบว่า วัตถุประสงค์มีความสอดคล้องกับปัญหาและความต้องการในการพัฒนานักเรียน พิจารณาจากผลการเรียนเฉลี่ยในรายวิชาคณิตศาสตร์มีเกรดเฉลี่ยต่ำที่สุดคือ 3.22 2) ผู้ปกครองร้อยละ 93.75 เห็นว่าวัตถุประสงค์โครงการมีความสอดคล้องกับความต้องการในการพัฒนาทักษะทางคณิตศาสตร์ของนักเรียนในการนำไปใช้ในชีวิตประจำวัน และ 3) วัตถุประสงค์มีความเหมาะสมกับเนื้อหาในรายวิชาคณิตศาสตร์ บูรณาการร่วมกับเศรษฐกิจพอเพียง ในภาพรวมอยู่ในระดับมาก

ตอนที่ 2 ผลการประเมินแผนการดำเนินของโครงการ พบว่า รูปแบบกิจกรรมในโครงการมีความเหมาะสม ในด้านชื่อกิจกรรมมีความชัดเจน น่าสนใจ และสะท้อนเป้าหมายด้านเศรษฐกิจพอเพียงได้ดี ระยะเวลาที่กำหนดมีความเหมาะสม ไม่กระทบต่อการจัดการเรียนการสอนปกติ มีการเรียงลำดับกิจกรรมจากง่ายไปยาก วิธีดำเนินการสอดคล้องกับบริบทของโรงเรียนขนาดเล็กที่มีทรัพยากรจำกัด โดยเน้นการใช้ทรัพยากรที่มีอยู่ให้เกิดประโยชน์สูงสุด คณะครูผู้รับผิดชอบมีความรู้ความสามารถตรงตามลักษณะกิจกรรม ส่วนวัสดุอุปกรณ์มีความเพียงพอต่อนักเรียน

ตอนที่ 3 ผลการประเมินแนวทางในการพัฒนาโครงการ พบว่า 1) สื่อและอุปกรณ์ มีความพร้อมเนื่องจากมีความหลากหลายทั้งรูปแบบสื่อทำมือและสื่อเทคโนโลยี อย่างไรก็ตาม ยังมีข้อจำกัดเรื่องความเพียงพอของจำนวนอุปกรณ์ และอุปกรณ์ที่ไม่ได้ใช้งานเป็นเวลานานอาจชำรุด 2) คณะครูที่เข้าร่วมโครงการ มีความสามารถในการใช้สื่อและอุปกรณ์ แต่มีข้อจำกัดด้านภาระงานอื่นนอกจากการสอน อีกทั้งครูผู้สอนส่วนใหญ่ในโรงเรียนไม่ได้จบวิชาเอกคณิตศาสตร์โดยตรงทำให้ขาดความเข้าใจในบางเนื้อหาที่มีความซับซ้อน และ 3) อาคารและสถานที่มีความเหมาะสม สะอาด และปลอดภัย อย่างไรก็ตามมีข้อจำกัดเรื่องระบบน้ำที่ไม่เสถียรซึ่งอาจจะส่งผลกระทบต่อกิจกรรมปลูกผัก

ตอนที่ 4 ผลการประเมินการนำโครงการไปปฏิบัติ จากการสัมภาษณ์ พบว่า ทั้ง 8 กิจกรรมมีความเหมาะสม โดยมีกระบวนการเรียนรู้ที่ยึดโยงกับบริบทจริง ซึ่งนักเรียนได้มีการลงมือปฏิบัติจริงทั้งจากการทำงานกลุ่ม

และงานเดี่ยว อีกทั้งยังได้รับความรู้ในรายวิชาคณิตศาสตร์บูรณาการร่วมกับเศรษฐกิจพอเพียง นักเรียนจึงไม่เพียงเกิดความสนุกสนานเท่านั้นแต่ยังรู้สึกมีความสุขจากการได้เข้าร่วมโครงการ แต่อย่างไรก็ตามในบางกิจกรรมที่มีความซับซ้อนจะใช้เวลานานมากกว่าที่ตั้งไว้ รวมถึงนักเรียนบางคนที่มีความเขินอายหรือขาดความมั่นใจในการนำเสนอผลงาน และขาดความรอบคอบในการคิดเลขรวมถึงใช้เวลานานในการคิดเลข

ตอนที่ 5 ผลการประเมินผลลัพธ์ของโครงการ พบว่า 1) ทักษะและกระบวนการทางคณิตศาสตร์ของนักเรียน ในภาพรวมอยู่ในระดับดี โดยทักษะการเชื่อมโยงความรู้คณิตศาสตร์ มีค่าเฉลี่ยสูงสุด อยู่ในระดับดี ($M = 3.19, SD = 0.66$) 2) ทักษะการทำงานกลุ่ม ในภาพรวมอยู่ในระดับดี ($M = 3.33, SD = 0.66$) ซึ่งด้านความสามัคคี มีค่าเฉลี่ยสูงสุด อยู่ในระดับดีมาก และด้านสัมพันธภาพในกลุ่ม มีค่าเฉลี่ยต่ำที่สุด อยู่ในระดับดี 3) ความพึงพอใจของนักเรียนที่มีต่อโครงการ ในภาพรวมอยู่ในระดับมาก ($M = 3.35, SD = 0.62$) เมื่อเรียงลำดับด้านที่มีความพึงพอใจจากมากไปน้อย คือ ด้านการจัดกิจกรรม ด้านเนื้อหา ด้านครูผู้สอน และด้านการวัดและประเมินผล ตามลำดับ โดยเฉพาะกิจกรรมตลาดนัดจำลอง ฝึกซื้อขาย นักเรียนมีความพึงพอใจมากที่สุด และ 4) เจตคติของนักเรียนในรายวิชาคณิตศาสตร์ ในภาพรวมอยู่ในระดับมาก มาก ($M = 2.51, SD = 0.57$) โดยนักเรียนเห็นว่าคณิตศาสตร์มีประโยชน์ต่อชีวิตประจำวันและการใช้ในชุมชน มีค่าเฉลี่ยสูงสุด

ตอนที่ 6 ผลการติดตามและประเมินผลกระทบของโครงการ พบว่า 1) นักเรียนในระดับชั้นประถมศึกษาปีที่ 1 – 4 มีผลการเรียนเฉลี่ยในรายวิชาคณิตศาสตร์เพิ่มขึ้น แต่ในระดับชั้นประถมศึกษาปีที่ 5 - 6 มีผลการเรียนลดลงจากเดิม และ 2) นักเรียนส่วนใหญ่มีความตระหนักในการออมเงินตามหลักเศรษฐกิจพอเพียง สามารถนำความรู้เรื่องการคำนวณและสถิติไปปรับใช้เพื่อสร้างรายได้ให้ครอบครัว และมีทักษะการทำงานกลุ่มที่เข้มแข็งขึ้นผ่านการสื่อสารกันระหว่างการทำกิจกรรมกลุ่ม

อภิปรายผล

1) ผลการประเมินวัตถุประสงค์ของโครงการ ในด้านความสอดคล้องกับปัญหาและความต้องการพัฒนานักเรียน พบว่า วัตถุประสงค์สามารถตอบโจทย์การแก้ปัญหาผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ที่มีเกรดเฉลี่ยต่ำที่สุดได้ โดยเปลี่ยนจากการเรียนแบบท่องจำสูตร มาเป็นการสร้างความเข้าใจผ่านเรื่องราวใกล้ตัวในชุมชน ซึ่งการกำหนดวัตถุประสงค์ที่มุ่งเน้นทักษะการเงินและอาชีพ ทำให้นักเรียนไม่ได้มองว่าคณิตศาสตร์เป็นเพียงตัวเลขในสมุด แต่เป็นเครื่องมือที่สามารถใช้แก้ปัญหาชีวิตจริงและเป็นเรื่องที่พบเห็นได้โดยทั่วไป สอดคล้องกับแนวคิดของ Freudenthal (1991) ผู้เสนอแนวคิด Realistic Mathematics Education (RME) ที่มองว่าการเรียนคณิตศาสตร์ควรเริ่มจากสถานการณ์จริงใกล้ตัวผู้เรียน เพื่อให้ผู้เรียนสร้างความเข้าใจด้วยตนเองและเห็นคุณค่าของคณิตศาสตร์ในการแก้ปัญหาชีวิตประจำวัน ในด้านความต้องการของผู้ปกครองในการพัฒนาทักษะทางคณิตศาสตร์ พบว่า วัตถุประสงค์ของโครงการทุกข้อได้รับความเห็นชอบ โดยเฉพาะการส่งเสริมทักษะคณิตศาสตร์ที่สามารถประยุกต์ใช้ในชีวิตจริงและส่งเสริมคุณลักษณะนิสัยรักการทำงาน รักการเรียนรู้ และการมีส่วนร่วมกับผู้อื่น สอดคล้องกับบริบทของโรงเรียนที่มีความใกล้ชิดกับชุมชน ที่ผู้ปกครองมุ่งหวังให้นักเรียนมีทักษะชีวิตและคุณธรรมควบคู่ไปกับวิชาการ และด้านความเหมาะสมของวัตถุประสงค์กับเนื้อหาในรายวิชาคณิตศาสตร์บูรณาการร่วมกับเศรษฐกิจพอเพียง พบว่า มีความเหมาะสมในระดับมาก โดยเฉพาะในการนำเนื้อหาคณิตศาสตร์

มาเชื่อมโยงกับการใช้จริงในชีวิตประจำวันที่มีความเหมาะสมมากที่สุด แต่ความสอดคล้องกับมาตรฐานการเรียนรู้ และตัวชี้วัดกลับอยู่ในระดับปานกลาง ซึ่งสะท้อนให้เห็นถึงข้อจำกัดของครูในโรงเรียนที่ต้องสอนควบทุกวิชา ทำให้ การตีความตัวชี้วัดให้สอดคล้องกับกิจกรรมที่มีการบูรณาการหลายเรื่องเป็นเรื่องที่ยากและท้าทาย

2) ผลการประเมินแผนการดำเนินโครงการ พบว่า มีความเหมาะสมทุกองค์ประกอบ ได้แก่ ชื่อกิจกรรม วัตถุประสงค์ของกิจกรรม ระยะเวลาดำเนินการ วิธีดำเนินการ ผู้รับผิดชอบ และวัสดุอุปกรณ์ ซึ่งทั้งหมด ถูกออกแบบมาอย่างเป็นระบบ สอดคล้องกับหลักการบริหารโครงการของ Kerzner (2017) ที่เน้นการวางแผนอย่างเป็น ขั้นตอน การกำหนดบทบาทหน้าที่และการจัดสรรทรัพยากรให้สอดคล้องกับเป้าหมาย ช่วยให้คณะครูสามารถทำงาน ประสานกันและมองเห็นภาพความสำเร็จของโครงการได้อย่างชัดเจน ในด้านระยะเวลาและวิธีดำเนินการ พบว่า มีความเหมาะสมและสอดคล้องกับบริบทของโรงเรียนที่มีทรัพยากรจำกัด มีการเรียงลำดับกิจกรรมจากง่ายไปหายาก โดยการออกแบบวิธีดำเนินการที่เน้นการใช้ทรัพยากรที่มีอยู่ให้เกิดประโยชน์สูงสุด สอดคล้องกับแนวคิดของ Pitiyanuwat (2013) ที่กล่าวว่าแผนการดำเนินงานที่มีประสิทธิภาพในบริบทสถานศึกษาไทย คือแผนที่สามารถ ปรับใช้ทรัพยากรในท้องถิ่นที่มีอยู่เดิมมาสร้างประสบการณ์การเรียนรู้ที่คุ้มค่า ด้านบุคลากรและวัสดุอุปกรณ์ พบว่า คณะครูมีความสามารถในการถ่ายทอด เข้าใจในลักษณะกิจกรรมและมีเจตคติที่ดีต่อการพัฒนา ช่วยให้การบริหารจัดการวัสดุอุปกรณ์ที่ส่วนใหญ่เป็นสื่อทำมือมีประสิทธิภาพสูงสุด สอดคล้องกับงานวิจัยของ Darling-Hammond (2017) ที่ระบุว่า คุณภาพของครูและความเชี่ยวชาญในการจัดการเรียนรู้เป็นปัจจัยสำคัญที่ส่งผลต่อประสิทธิภาพ ของการใช้ทรัพยากรทางการศึกษา

3) ผลการประเมินแนวทางในการพัฒนาโครงการ ในด้านสื่อและอุปกรณ์ที่ใช้ในการดำเนินกิจกรรม พบว่า มีความพร้อม โดยเฉพาะความหลากหลายของสื่อที่มีทั้งรูปแบบสื่อทำมือและสื่อเทคโนโลยี สามารถ ใช้งานได้ง่าย ช่วยกระตุ้นความสนใจ และส่งเสริมการเรียนรู้เชิงรุกได้อย่างมีประสิทธิภาพ แม้จะมีข้อจำกัดเรื่อง จำนวนอุปกรณ์ที่ไม่เพียงพอต่อความต้องการและร่อยรอยการชำรุด แต่โรงเรียนยังสามารถบริหารจัดการได้ ในด้านความพร้อมของคณะครู พบว่า ครูมีใจเปิดรับและมีทักษะการถ่ายทอดที่ดี ช่วยให้นักเรียนเข้าใจเนื้อหา ได้ง่ายขึ้น อย่างไรก็ตามโรงเรียนบ้านหนองเลเป็นโรงเรียนที่มีครูวิชาเอกคณิตศาสตร์เพียงท่านเดียว ซึ่งการที่ครู มีวุฒิไม่ตรงกับวิชาเอกคณิตศาสตร์ต้องมาทำความเข้าใจเนื้อหาคณิตศาสตร์ที่ซับซ้อนถือเป็นงานที่ทำหายและ ต้องใช้เวลาศึกษามากกว่าปกติ ประกอบกับครูแต่ละท่านยังมีภาระงานนอกเหนือจากการสอน ทำให้เมื่อถึง เวลาจัดกิจกรรมบูรณาการ ครูบางส่วนจึงไม่สามารถเข้าร่วมได้อย่างเต็มที่หรือขาดช่วงไป ส่งผลให้การดูแล นักเรียนที่มีความแตกต่างรายบุคคลทำได้ไม่ทั่วถึงเท่าที่ควร ในด้านอาคารและสถานที่ พบว่า โรงเรียนมีความ โดดเด่นเรื่องความสะอาดและปลอดภัย ซึ่งช่วยสร้างบรรยากาศที่น่าเรียนรู้ให้แก่นักเรียน อย่างไรก็ตาม มีข้อจำกัดคือปัญหาระบบน้ำที่ไม่เสถียร ซึ่งส่งผลกระทบโดยตรงต่อกิจกรรมภาคปฏิบัติ อย่างเช่นการปลูกผักที่ ต้องใช้น้ำสม่ำเสมอ สอดคล้องกับแนวคิดของ Dewey (1938) ที่ชี้ให้เห็นว่าสภาพแวดล้อมทางการเรียนรู้มี บทบาทสำคัญต่อประสบการณ์และคุณภาพการเรียนรู้ของผู้เรียน โดยเฉพาะการจัดสภาพแวดล้อมที่เอื้อต่อการ ลงมือปฏิบัติจริง หากสภาพแวดล้อมหรือทรัพยากรพื้นฐานยังไม่พร้อม ย่อมส่งผลกระทบต่อประสิทธิภาพของการ เรียนรู้เชิงประสบการณ์ได้ ซึ่งโรงเรียนได้รับความร่วมมือจากนักการภารโรง ในการซ่อมแซมและดูแลสถานที่

อยู่เสมอ ถือเป็นการแก้ปัญหาตามบริบทของโรงเรียน ช่วยประคับประคองให้โครงการเดินหน้าต่อไปได้ ตามหลักปรัชญาของเศรษฐกิจพอเพียง คือการรู้จักใช้สิ่งที่มียูให้เกิดประโยชน์สูงสุดภายใต้ข้อจำกัดที่มีอยู่จริง

4) ผลการประเมินการนำโครงการไปปฏิบัติ พบว่า กิจกรรมทั้ง 8 กิจกรรม มีความเหมาะสมเนื่องจาก ใช้กระบวนการเรียนรู้ที่เชื่อมโยงความรู้ในห้องเรียนกับบริบทจริง ซึ่งนักเรียนได้มีการลงมือปฏิบัติจริงจากการทำงานกลุ่มและงานเดี่ยว นักเรียนจึงไม่เพียงเกิดความรู้เท่านั้นแต่ยังรู้สึกมีความสุขจากการได้เข้าร่วมโครงการ ซึ่งกิจกรรมคณิตศาสตร์ที่น่าบริบทชุมชนและตลาดจำลองมาใช้ ช่วยให้ผู้เรียนได้ฝึกการประมาณค่า และการวางแผนใช้ทรัพยากร เช่น การคำนวณวัสดุต่าง ๆ ต้นทุน กำไร หรือการตัดสินใจเลือกซื้อ-ขายอย่างเหมาะสม สิ่งเหล่านี้สะท้อนหลักสำคัญของเศรษฐกิจพอเพียง คือ ความพอประมาณและความมีเหตุผล เพราะผู้เรียนต้องคิดก่อนใช้ทรัพยากร ไม่ใช่เงินจำเป็น และเลือกวิธีที่เหมาะสมกับสถานการณ์ ในขณะเดียวกัน กิจกรรมตลาดจำลองยังช่วยปลูกฝังเจตคติด้านความรู้และคุณธรรม โดยผู้เรียนต้องใช้ทักษะคณิตศาสตร์ควบคู่กับความซื่อสัตย์ในการซื้อขาย ไม่โกง ไม่เอาเปรียบ และรับผิดชอบต่อการตัดสินใจของตนเอง ทำให้คณิตศาสตร์ไม่ได้เป็นเพียงวิชาที่ต้องท่องจำสูตร แต่กลายเป็นเครื่องมือที่ใช้ในชีวิตจริง เพื่อช่วยให้ผู้เรียนคิดอย่างรอบคอบ มีภูมิคุ้มกัน และดำรงชีวิตได้อย่างเหมาะสม สอดคล้องกับแนวคิดของ Khammanee (2018) ที่ชี้ว่าการเรียนรู้ที่ผู้เรียนได้ลงมือปฏิบัติแก้ปัญหาและเชื่อมโยงกับชีวิตจริงจะช่วยให้เกิดการเรียนรู้ที่มีความหมายและคงทน ซึ่งเป็นการตอบสนองต่อเป้าหมายของโรงเรียนในการน้อมนำปรัชญาของเศรษฐกิจพอเพียงมาใช้ในการลงมือทำอย่างเป็นรูปธรรม

อย่างไรก็ตามในข้อจำกัดด้านการใช้เวลาในกิจกรรมที่ซับซ้อน สภาพอากาศหรือเสียงรบกวนภายนอกที่ทำให้ความสนใจของนักเรียนไม่ต่อเนื่อง ความรอบคอบและการใช้เวลาในการคิดเลข สะท้อนให้เห็นความจำเป็นในการออกแบบกิจกรรมให้ยืดหยุ่น เพื่อให้ผู้เรียนทุกคนสามารถมีส่วนร่วมได้อย่างเต็มศักยภาพ ทั้งยังเป็นโอกาสสำคัญในการพัฒนารูปแบบกิจกรรมให้เหมาะสมกับบริบทโรงเรียน สอดคล้องกับทฤษฎีการเรียนรู้ของ Vygotsky (1978) ที่มองว่าบรรยากาศทางสังคมและเครื่องมือช่วยคิดรอบตัวเป็นปัจจัยสำคัญต่อการพัฒนาสมอง นอกจากนี้ นักเรียนยังขาดความมั่นใจในตนเองเมื่อต้องนำเสนอ ซึ่งสะท้อนว่าโรงเรียนต้องเพิ่มพื้นที่ให้นักเรียนได้ฝึกทักษะการสื่อสาร ควบคู่ไปกับวิชาการ เพื่อสร้างความมั่นใจให้นักเรียนกล้าสื่อสารในโลกแห่งความเป็นจริงได้ และเกิดความมั่นใจในตนเอง

5) ผลการประเมินผลลัพธ์ของโครงการ ในด้านทักษะและกระบวนการทางคณิตศาสตร์และการทำงานกลุ่ม พบว่า ในภาพรวมอยู่ในระดับดี โดยเฉพาะทักษะการเชื่อมโยงความรู้คณิตศาสตร์กับชีวิตประจำวันที่ได้ค่าเฉลี่ยสูงที่สุด เช่นเดียวกับความสามัคคีในการทำงานกลุ่ม ที่อยู่ในระดับดีมาก สะท้อนให้เห็นว่าบริบทของโรงเรียนบ้านหนองเลซึ่งเป็นโรงเรียนขนาดเล็ก ถือเป็นปัจจัยสำคัญที่เอื้อต่อความสำเร็จของการดำเนินโครงการอย่างชัดเจน เนื่องจากจำนวนนักเรียนต่อชั้นมีไม่มาก ครูจึงสามารถดูแล ให้คำแนะนำ และติดตามพัฒนาการของผู้เรียนได้อย่างใกล้ชิดและทั่วถึง ส่งผลให้การจัดกิจกรรมบูรณาการสามารถปรับให้เหมาะสมกับความแตกต่างระหว่างบุคคลได้อย่างมีประสิทธิภาพ อีกทั้งบริบทดังกล่าวยังเอื้อให้เกิดการจัดการเรียนรู้แบบข้ามระดับชั้น โดยนักเรียนรุ่นพี่สามารถช่วยแนะนำและสนับสนุนนักเรียนรุ่นน้อง เกิดการเรียนรู้ร่วมกันตามธรรมชาติ สอดคล้องกับแนวคิดการcaffolding (Scaffolding) ของ Vygotsky (1978) ที่เน้นบทบาทของผู้ที่มีความสามารถมากกว่าในการช่วยให้ผู้เรียนพัฒนาศักยภาพได้เต็มที่ ประเด็นนี้จึงชี้ให้เห็นว่าขนาดของ

สถานศึกษาเป็นเงื่อนไขสำคัญที่ส่งเสริมการจัดกิจกรรมลักษณะนี้ และสามารถเป็นแนวทางที่เป็นประโยชน์ต่อโรงเรียนขนาดเล็กอื่น ๆ ในการนำไปประยุกต์ใช้ต่อไป

อย่างไรก็ตามการที่ทักษะการแก้ปัญหา แม้จะผ่านเกณฑ์แต่กลับมีค่าเฉลี่ยต่ำที่สุดเมื่อเทียบกับด้านอื่น ๆ เป็นเพราะนักเรียนส่วนใหญ่คุ้นเคยกับการเรียนคณิตศาสตร์แบบทำตามตัวอย่างในสมุด พอต้องมาเจอโจทย์ปัญหาในสถานการณ์จริงที่มีความซับซ้อนและต้องใช้การตัดสินใจ นักเรียนจึงยังปรับตัวไม่ทัน ในด้านความพึงพอใจของนักเรียนที่มีต่อโครงการ พบว่า ในภาพรวมอยู่ในระดับมาก โดยเฉพาะในด้านเนื้อหาและกิจกรรมที่ตรงกับความสนใจ ซึ่งกิจกรรมที่ได้รับความนิยมสูงสุดคือ “ตลาดนัดจำลอง-ฝึกซื้อขาย” และ “ผ้าผัดย้อม ลายคณิตคิดสร้างสรรค์” สะท้อนให้เห็นว่า ผู้เรียนมีความสนใจต่อกิจกรรมที่ได้ลงมือปฏิบัติจริงและสร้างผลงานด้วยตนเอง ซึ่งสอดคล้องกับ Kolb (1984) ที่กล่าวถึงแนวคิดการเรียนรู้จากประสบการณ์ (experiential learning) ที่เชื่อว่าผู้เรียนจะเกิดการเรียนรู้อย่างลึกซึ้งเมื่อได้ลงมือปฏิบัติ ทดลอง และสะท้อนผลจากประสบการณ์ตรง กิจกรรมลักษณะนี้จึงช่วยเพิ่มแรงจูงใจ ความสนุกในการเรียน และทำให้ผู้เรียนเชื่อมโยงความรู้กับชีวิตจริงได้ดีกว่าการเรียนรู้แบบบรรยายเพียงอย่างเดียว สำหรับด้านเจตคติต่อวิชาคณิตศาสตร์ พบว่า ในภาพรวมอยู่ในระดับมาก โดยเฉพาะประเด็นที่ว่าคณิตศาสตร์มีประโยชน์ต่อชีวิตประจำวันและการใช้ในชุมชน มีค่าเฉลี่ยสูงสุด สะท้อนให้เห็นว่าโครงการประสบความสำเร็จในการเปลี่ยนภาพจำในการเรียนวิชาคณิตศาสตร์จากการสอนในห้องเรียนที่นักเรียนมองว่ายากและน่าเบื่อ เป็นการดึงเอาบริบทของชุมชนและปรัชญาของเศรษฐกิจพอเพียงมาเป็นบทเรียน ทำให้คณิตศาสตร์กลายเป็นวิชาที่จับต้องได้จริง แม้ความมั่นใจในการทำแบบฝึกหัดจะมีค่าเฉลี่ยต่ำที่สุดแต่ก็ยังอยู่ในระดับมาก ซึ่งถือเป็นสัญญาณที่ดีในการพัฒนาความสามารถทางการเรียนในระยะยาว เนื่องจากเมื่อนักเรียนมีเจตคติที่ดีต่อสิ่งที่เรียนจะทำให้นักเรียนมีใจจดจ่อ และมีความมุ่งมั่นมากขึ้น ซึ่งจะส่งผลให้ผลสัมฤทธิ์ทางการเรียนมีแนวโน้มสูงขึ้นตามไปด้วย สอดคล้องกับแนวคิดของ Dechawatanapaisal (2022) ที่กล่าวว่า เจตคติที่ดีจะนำไปสู่ความคิดที่ดี ความคิดที่ดีก็จะนำไปสู่การกระทำหรือพฤติกรรมในการทำงานที่ดี และเมื่อการกระทำดี ผลของการกระทำที่ดีก็มักจะดีไปด้วยเช่นกัน

6) ผลการติดตามและประเมินผลกระทบของโครงการ ในด้านผลสัมฤทธิ์ทางการเรียนวิชาคณิตศาสตร์ พบว่า นักเรียนชั้นประถมศึกษาปีที่ 1 - 4 มีผลการเรียนเพิ่มขึ้นหลังเข้าร่วมโครงการ แสดงว่าโครงการส่งผลดีต่อผู้เรียนในระดับประถมต้น สอดคล้องกับทฤษฎีของ Piaget (1964) ที่กล่าวถึงพัฒนาการทางสติปัญญาในวัยที่ต้องอาศัยสิ่งรูปธรรม (concrete operational stage) ซึ่งเด็กในช่วงวัยนี้สามารถคิดอย่างมีเหตุผลมากขึ้น การที่นักเรียนได้เรียนรู้ผ่านการขายสินค้าหรือการปลูกผักช่วยให้โมโนทัศน์ทางคณิตศาสตร์มีความหมายและจับต้องได้จริง ส่งผลให้เกิดความกระตือรือร้นและมีระดับคะแนนที่สูงขึ้น อย่างไรก็ตามผลการเรียนของนักเรียนชั้นประถมศึกษาปีที่ 5 - 6 กลับมีแนวโน้มลดลง แม้ว่าจะระดับความพึงพอใจและเจตคติที่มีต่อวิชาคณิตศาสตร์จะอยู่ในระดับมากก็ตาม ซึ่งเมื่อพิจารณาโครงสร้างเนื้อหาคณิตศาสตร์ของนักเรียนระดับชั้น ป.5 - 6 พบว่ามีลักษณะเชิงนามธรรมสูงกว่านักเรียนระดับชั้นต้น ดังนั้นแม้การเชื่อมโยงบริบทชุมชนจะช่วยเพิ่มความหมายในการเรียนรู้ แต่หากกิจกรรมถูกออกแบบให้มีความซับซ้อนมากเกินไป อาจก่อให้เกิดภาระทางปัญญาที่สูง (cognitive overload) ส่งผลให้ประสิทธิภาพในการเรียนรู้ลดลง ซึ่งสอดคล้องกับแนวคิดของ Sweller (1988) ที่ชี้ว่างานการเรียนรู้ที่มีความยากหรือซับซ้อนเกินระดับความสามารถของผู้เรียนจะลดทอนประสิทธิภาพของการเรียนรู้ อีกทั้งในบริบทของโรงเรียน

บ้านหนองเล่ นักเรียนระดับประถมปลายบางคนมีพื้นฐานทางคณิตศาสตร์ที่ยังไม่ดี ซึ่งไม่สามารถใช้เพียงทักษะพื้นฐานจากชีวิตประจำวันมาแก้โจทย์ปัญหาที่มีการบูรณาการหลายบทเรียนเข้าด้วยกันได้ สอดคล้องกับแนวคิดของ Khammanee (2018) ที่ระบุว่าการจัดการเรียนรู้บูรณาการต้องระวังกว้างความสมดุลระหว่างกิจกรรมและเนื้อหาเชิงลึกในด้านพฤติกรรมและการเปลี่ยนแปลงหลังเสร็จสิ้นโครงการ พบว่า นักเรียนมีความตระหนักในการออมเงินตามหลักเศรษฐกิจพอเพียง และสามารถนำความรู้เรื่องการเงินไปใช้ในการสร้างรายได้ให้ครอบครัวได้ ซึ่งโครงการนี้ประสบความสำเร็จในการสร้างภูมิคุ้มกัน และทักษะชีวิตให้แก่ผู้เรียน สอดคล้องกับหลักปรัชญาเศรษฐกิจพอเพียงในเรื่องการมีภูมิคุ้มกันในตัวที่ดี คือการเตรียมความพร้อมรับผลกระทบและการเปลี่ยนแปลงด้านต่าง ๆ ที่เกิดขึ้น โดยคำนึงถึงความเป็นไปได้ของสถานการณ์ต่าง ๆ ที่คาดว่าจะเกิดขึ้นในอนาคต ถือเป็นจุดแข็งสำคัญของโรงเรียน (National Moral Promotion Committee, 2018)

ข้อเสนอแนะ

1. ข้อเสนอแนะในการนำผลการประเมินไปใช้

1.1 จากการประเมินพบว่า นักเรียนยังมีจุดอ่อนสำคัญคือความประมาทในการสื่อสารและความล่าช้าในการคิดคำนวณโจทย์ที่ซับซ้อน ดังนั้น ควรใช้ระบบข้อมูลสะท้อนกลับ จากผลการประเมินแจ้งให้นักเรียนทราบจุดที่ต้องพัฒนาของตนเอง เพื่อให้นักเรียนนำผลการประเมินไปปรับปรุงทักษะต่าง ๆ ให้ดียิ่งขึ้น

1.2 จากการประเมินพบข้อจำกัดด้านพื้นที่จัดกิจกรรมที่มีความแออัด และระบบน้ำที่ยังไม่เสถียร รวมถึงภาระงานนอกเหนือการสอนของครู ดังนั้น ควรนำสารสนเทศเหล่านี้มาประกอบการตัดสินใจเพื่อปรับปรุงแหล่งเรียนรู้ในโรงเรียน นอกจากนี้ควรสนับสนุนการลดภาระงานที่ไม่จำเป็นของครู และสร้างเครือข่ายความร่วมมือกับชาวบ้าน เพื่อดึงประสบการณ์จริงมาลดช่องว่างระหว่างทฤษฎีในตำรากับวิถีชุมชน

2. ข้อเสนอแนะในการประเมินครั้งต่อไป

2.1 ในการประเมินครั้งนี้ ได้ยึดกรอบแนวคิดการประเมินของไทเลอร์ (Tyler, 1949) ที่ประกอบด้วย 6 ขั้นตอน ซึ่งครอบคลุมทั้งระบบ อย่างไรก็ตาม สารสนเทศบางประการอาจยังไม่ครอบคลุมหรือไม่สามารถตอบสนองต่อความต้องการของผู้มีส่วนได้ส่วนเสียได้ทุกกลุ่ม เพื่อให้การประเมินสามารถนำไปใช้ประโยชน์ได้อย่างแท้จริง อาจเลือกใช้แนวทางการประเมินที่เน้นการมีส่วนร่วมของผู้มีส่วนได้ส่วนเสียมุ่งเน้นการใช้ ประโยชน์ เช่น Utilization-focused Evaluation, Responsive Evaluation เป็นต้น

2.2 โครงการที่ดำเนินการในครั้งนี้เป็นเพียงส่วนหนึ่งของกระบวนการจัดการเรียนรู้ทั้งหมดของโรงเรียน จึงอาจไม่สามารถสรุปได้อย่างชัดเจนว่าผลสัมฤทธิ์ทางการเรียนที่เพิ่มขึ้นหรือลดลงของนักเรียนเป็นผลมาจากโครงการเพียงปัจจัยเดียว ทั้งนี้ ผลการเรียนรู้ของนักเรียนอาจได้รับอิทธิพลจากปัจจัยอื่นร่วมด้วย ดังนั้นในการดำเนินงานครั้งต่อไป ควรมีการศึกษาปัจจัยแวดล้อมต่าง ๆ ที่อาจส่งผลต่อผลสัมฤทธิ์ทางการเรียนของนักเรียน เช่น ความแตกต่างระหว่างบุคคล พื้นฐานความรู้เดิม บริบทครอบครัว และแรงจูงใจในการเรียน เพื่อนำข้อมูลมาใช้ประกอบการปรับปรุงและพัฒนาโครงการให้เหมาะสมยิ่งขึ้น นอกจากนี้ ควรติดตามพัฒนาการของนักเรียนกลุ่มเดิมอย่างต่อเนื่องในช่วงหลายปีการศึกษา เพื่อเปรียบเทียบพัฒนาการของนักเรียนอย่างแท้จริง และช่วยให้เกิดประโยชน์สูงสุดต่อการพัฒนาผู้เรียนในระยะยาว

References

- Darling-Hammond, L. (2017). *Empowered educators: How high-performing systems shape teaching quality around the world*. Jossey-Bass.
- Dewey, J. (1938). *Experience and education*. Macmillan.
- Freudenthal, H. (1991). *Revisiting mathematics education: China lectures*. Kluwer Academic Publishers.
- Kerzner, H. (2017). *Project management: A systems approach to planning, scheduling, and controlling* (12th ed.). Wiley.
- Kolb, D. A. (1984). *Experiential learning: Experience as the source of learning and development*. Prentice Hall.
- Piaget, J. (1964). Cognitive development in children: Piaget development and learning. *Journal of Research in Science Teaching*, 2(3), 176 – 186.
<https://doi.org/10.1002/tea.3660020306>
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
- Tyler, R. W. (1949). *Basic principles of curriculum and instruction*. University of Chicago Press.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

Translate Thai References

- Ban Nongle School. (2026). *Action plan for fiscal year 2026*. Unpublished internal document. (in Thai)
- Dechawatanapaisal, D. (2022). *Organizational Behavior Theories and Practices*. Chulalongkorn University Press. (in Thai)
- Khammanee, T. (2018). *Pedagogy: Knowledge for effective learning process management* (22nd ed.). Chulalongkorn University Press. (in Thai)
- Kongsilp, N. (2018). *Teaching Mathematics in the 21st Century* (1st ed.). Pimphijit Co., Ltd. (in Thai)
- National Moral Promotion Committee. (2018). *Moral promotion: Sufficiency, discipline, integrity, and public mind — Building good people for society* (2nd ed.). Pimdee Co., Ltd. (in Thai)
- Pitiyanuwat, S. (2013). *Evaluation methodology: Science of value* (6th ed.). Chulalongkorn University Press. (in Thai)
- Promchui, S. (2009). *Project evaluation techniques* (6th ed.). Jatuporn Design. (in Thai)



ภาคผนวก

- รายชื่อผู้ทรงคุณวุฒิพิจารณาบทความ วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม
ปีที่ 32 ฉบับที่ 1: มกราคม - มิถุนายน 2569
- หลักเกณฑ์และคำแนะนำสำหรับผู้เขียน



รายชื่อผู้ทรงคุณวุฒิพิจารณาบทความวารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม
ปีที่ 32 ฉบับที่ 1: มกราคม - มิถุนายน 2569

- | | |
|------------------------------------|---|
| 1. รศ.ดร.วราพร เอรารวรรณ์ | มหาวิทยาลัยมหาสารคาม |
| 2. รศ.ดร.กัญญารัตน์ โคจร | มหาวิทยาลัยมหาสารคาม |
| 3. รศ.ดร.รัชนีวรรณ ตั้งภักดี | มหาวิทยาลัยมหาสารคาม |
| 4. รศ.ดร.อพนันตรี พูลพุทธา | มหาวิทยาลัยมหาสารคาม |
| 5. ผศ.ดร.กันยารัตน์ สอนสุภาพ | มหาวิทยาลัยมหาสารคาม |
| 6. ผศ.ดร.ฉัตรเกล้า เจริญผล | มหาวิทยาลัยมหาสารคาม |
| 7. พันตรีหญิง ดร.ดุขฎิ เล็บขาว | มหาวิทยาลัยมหาสารคาม |
| 8. อ.ดร.อุษา แก้วกำก | มหาวิทยาลัยมหาสารคาม |
| 9. อ.ดร.บวรพจน์ ชมภูนุช | มหาวิทยาลัยมหาสารคาม |
| 10. ผศ.ดร.สมทรง สิทธิ | มหาวิทยาลัยมหาสารคาม (เกษียณอายุ) |
| 11. รศ.ดร.พัชร จันทรเพ็ง | มหาวิทยาลัยขอนแก่น |
| 12. รศ.ดร.ชัยวิชิต เขียรชนะ | มหาวิทยาลัยเทคโนโลยีพระจอมเกล้าพระนครเหนือ |
| 13. รศ.ดร.เมษา นวลศรี | มหาวิทยาลัยราชภัฏวไลยอลงกรณ์ ในพระบรมราชูปถัมภ์ |
| 14. รศ.ดร.กมลทิพย์ ศรีหาเศษ | มหาวิทยาลัยรามคำแหง |
| 15. รศ.ดร.สุนทรพจน์ ดำรงค์พานิช | มหาวิทยาลัยเชียงใหม่ |
| 16. รศ.ดร.พงศ์วัชร พองันทา | มหาวิทยาลัยราชภัฏลำปาง |
| 17. รศ.ดร.มณฑิรา ดวงสาพล | มหาวิทยาลัยธรรมศาสตร์ |
| 18. ผศ.ดร.สาธิตา สกุรัตนกุลชัย | มหาวิทยาลัยธรรมศาสตร์ |
| 19. ผศ.ดร.มัยดี แวดราแม | มหาวิทยาลัยสงขลานครินทร์ |
| 20. ผศ.ดร.พินดา วราสุนันท์ | มหาวิทยาลัยเกษตรศาสตร์ |
| 21. ผศ.ดร.เรืองเดช ศิริกิจ | มหาวิทยาลัยศรีนครินทรวิโรฒ |
| 22. อ.ดร.พนัส จันทรเปล่ง | มหาวิทยาลัยศรีนครินทรวิโรฒ |
| 23. ผศ.ดร.ศจี จิระโร | มหาวิทยาลัยสุโขทัยธรรมมาธิราช |
| 24. ผศ.ดร.สรัญญา จันทรชูสกุล | มหาวิทยาลัยศิลปากร |
| 25. อ.ดร.พรรณภัทร ปลั่งศรีเจริญสุข | มหาวิทยาลัยศิลปากร |
| 26. ผศ.ดร.สิรินธร สิ้นจินดาวงศ์ | มหาวิทยาลัยศรีปทุม |
| 27. ผศ.ดร.สิวะโชติ ศรีสุทธิยากร | จุฬาลงกรณ์มหาวิทยาลัย |
| 28. อ.ดร.ณรงค์ฤทธิ์ อินทนาม | มหาวิทยาลัยราชภัฏอุบลราชธานี |
| 29. อ.ดร.ณรัตน์ ศิริภาพ | โรงเรียนโพธิสารศึกษา |
| 30. ดร.ชลี ภัทรพิชญธรรม | Southern Methodist University, USA |

หลักเกณฑ์และคำแนะนำสำหรับผู้เขียน (Principles and Guideline for Author)

ประเภทผลงานวิชาการ

วารสารการวัดผลการศึกษามหาวิทยาลัยมหาสารคาม (Journal of Educational Measurement Mahasarakham University : JEM-MSU) จะรับพิจารณาผลงานวิชาการใน 3 ลักษณะ ได้แก่ บทความวิชาการ บทความวิจัย/วิทยานิพนธ์ และบทความทบทวนวรรณกรรม โดยผลงานทางวิชาการดังกล่าวจะต้องไม่เคยถูกนำไปตีพิมพ์เผยแพร่ในวารสารหรือเอกสารอื่นใดมาก่อน ซึ่งผู้เขียนต้องเป็นผู้รับผิดชอบในการตรวจสอบประเด็นดังกล่าวอย่างเคร่งครัด

การเตรียมต้นฉบับ

1. ภาษา สามารถนำเสนอได้ทั้งภาษาไทยหรือภาษาอังกฤษ กรณีเขียนบทความเป็นภาษาอังกฤษให้ใช้ภาษาอังกฤษอย่างเดียวได้ โดยไม่ต้องมีภาษาไทย

2. หัวข้อในการเขียนบทความ

2.1 บทความวิจัย มีหัวข้อดังนี้

- ชื่อเรื่อง มีทั้งภาษาไทยและภาษาอังกฤษ
- ชื่อ-สกุล/สถานภาพผู้เขียน มีทั้งภาษาไทยและภาษาอังกฤษ
- บทคัดย่อ มีทั้งภาษาไทยและภาษาอังกฤษ
- คำสำคัญ ภาษาไทยและภาษาอังกฤษ
- บทนำ
- ความมุ่งหมายของการวิจัย
- วิธีดำเนินการวิจัย
- ผลการวิจัย
- อภิปรายผล
- ข้อเสนอแนะ
- เอกสารอ้างอิง (อ้างอิงภาษาไทยต้องแปลเป็นภาษาอังกฤษทั้งหมด ตามรูปแบบ APA)

2.2 บทความวิชาการ มีหัวข้อดังนี้

- ชื่อเรื่อง มีทั้งภาษาไทยและภาษาอังกฤษ
- ชื่อ-สกุล/สถานภาพผู้เขียน มีทั้งภาษาไทยและภาษาอังกฤษ
- บทคัดย่อ มีทั้งภาษาไทยและภาษาอังกฤษ
- คำสำคัญ ภาษาไทยและภาษาอังกฤษ
- บทนำ
- เนื้อหา
- สรุป
- เอกสารอ้างอิง (อ้างอิงภาษาไทยต้องแปลเป็นภาษาอังกฤษทั้งหมด ตามรูปแบบ APA)

2.3 บทความวิจารณ์หนังสือ มีหัวข้อดังนี้

- ชื่อเรื่อง (ชื่อหนังสือ) มีทั้งภาษาไทยและภาษาอังกฤษ
- ชื่อ-สกุลผู้เขียนหนังสือ ภาษาไทยหรือภาษาอังกฤษ ขึ้นกับภาษาที่ใช้ในหนังสือ
- รายละเอียดเกี่ยวกับหนังสือที่นำมาวิจารณ์ เช่น สำนักพิมพ์ ภาษาที่ใช้ จำนวนหน้า ภาพปก
- ชื่อ-สกุล/สถานภาพผู้วิจารณ์ มีทั้งภาษาไทยและภาษาอังกฤษ
- เนื้อหาการวิจารณ์
- เอกสารอ้างอิง (อ้างอิงภาษาไทยต้องแปลเป็นภาษาอังกฤษทั้งหมด ตามรูปแบบ APA)

3. การพิมพ์

- 3.1 ให้จัดพิมพ์ด้วยโปรแกรม Microsoft Word โดยกำหนดขนาด A4 ตั้งค่าหน้ากระดาษสำหรับการพิมพ์ห่างจากขอบกระดาษทุกด้าน ๆ ละ 1 นิ้ว (2.54 เซนติเมตร) และจัดรูปแบบเป็นหนึ่งคอลัมน์เท่านั้น
- 3.2 ตัวอักษร TH SarabunPSK ขนาด 18 pt สำหรับชื่อเรื่อง จัดกลางหน้ากระดาษ
- 3.3 ตัวอักษร TH SarabunPSK ขนาด 16 pt สำหรับเนื้อหา จัดชิดขอบ
- 3.4 ตัวอักษร TH SarabunPSK ขนาด 15 pt สำหรับ FootNote จัดชิดซ้ายหน้ากระดาษ
- 3.5 ช่องว่างระหว่างคำให้ตั้งค่าเป็น Normal ไม่บีบหรือขยายตัวอักษร ยกเว้นในตารางหรือแผนภาพ
- 3.6 ช่องว่างระหว่างบรรทัดให้ตั้งค่าเป็น single space
- 3.7 การเว้นวรรค 1 เคาะเท่านั้น
- 3.8 ตาราง และ ภาพ ใช้คำว่า ตาราง 1 และ ภาพ 1
4. จำนวนหน้าของผลงาน ไม่ควรเกิน 15 หน้า (รวมเอกสารอ้างอิง หลังตัดอ้างอิงภาษาไทยออกแล้ว)
5. เอกสารอ้างอิง มีรายการอ้างอิงครบถ้วน สมบูรณ์ ตามหลักการเขียนการอ้างอิงเอกสารแบบ APA

การเขียนเอกสารอ้างอิง

ด้วยวารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม มีความพยายามที่จะพัฒนาคุณภาพวารสารสู่มาตรฐานสากล ดังนั้น กองบรรณาธิการวารสารฯ จึงกำหนดรูปแบบการเขียนเอกสารดังนี้

- กำหนดให้ใช้รูปแบบการอ้างอิงแบบ APA (7th edition)
- เอกสารอ้างอิงภาษาไทยต้องแปลเป็นภาษาอังกฤษทุกรายการและเติมคำว่า (in Thai) ต่อท้าย
- กรณีส่งบทความให้วารสารพิจารณา ยังคงเอกสารอ้างอิงภาษาไทยเดิมไว้ด้วย เขียนจัดเรียงคู่กัน เพื่อให้วารสารฯ สามารถตรวจสอบความถูกต้องของการเขียนอ้างอิงในรูปแบบ Translate Thai Reference และอ้างอิงภาษาไทยจะถูกตัดออกในขั้นตอนของการตีพิมพ์เผยแพร่

- การเรียงลำดับอ้างอิง กรณีเอกสารอ้างอิงที่แปลจากภาษาไทยเป็นภาษาอังกฤษให้ยึดตัวอักษรภาษาอังกฤษในการเรียงลำดับ

*การเขียนรายการอ้างอิงภาษาไทยเป็นภาษาอังกฤษ มีวิธีการเขียนที่คล้ายคลึงกับการเขียนรายการอ้างอิงตามรูปแบบของภาษาอังกฤษในแต่ละประเภท มีรายละเอียดดังนี้

- 1) ชื่อผู้เขียน ขึ้นต้นด้วยนามสกุลของผู้เขียนและต่อด้วยชื่อย่อ
- 2) ชื่อสำนักงานหรือกระทรวง เขียนเป็นชื่อภาษาอังกฤษ
- 3) ปีที่พิมพ์ เปลี่ยนจาก พ.ศ. เป็น ค.ศ.
- 4) ชื่อเรื่อง ชื่อวิทยานิพนธ์ ชื่อบทความ ชื่อวารสาร และชื่องานประชุมวิชาการ แปลเป็นภาษาอังกฤษ

Ary, D., Jacobs, L. C., Irvine, A. K. S., & Walker, D. A. (2018). *Introduction to research in education*. Wadsworth.

* กรณีไม่ปรากฏปีที่พิมพ์ให้ใส่ (ม.ป.ป.) หรือ (n.d.)

** กรณีพิมพ์ครั้งที่ 1 ไม่ต้องระบุครั้งที่พิมพ์

*** กรณีไม่ปรากฏสำนักพิมพ์ให้ใส่ (ม.ป.ท.) หรือ (n.p.)

2. การอ้างอิงบทความจากวารสารแบบรูปเล่ม/อิเล็กทรอนิกส์ที่ไม่มี DOI

ไทย ชื่อ/สกุล/(ปีที่พิมพ์)/ชื่อบทความ./ชื่อวารสาร,/ปีที่(ฉบับที่),/เลขหน้า.

อังกฤษ สกุล,/อักษรย่อชื่อ/(ปีที่พิมพ์)/ชื่อบทความ./ชื่อวารสาร,/ปีที่(ฉบับที่),/เลขหน้า.

สุนันท์ สipay และ ไพฑูรย์ สีนลารัตน์. (2561). เปลี่ยนผ่านการศึกษาไทยสู่ การศึกษา 4.0. *วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม*, 24(2), 13-27.

Wang, W., Song, L., Chen, P., Meng, Y., & Ding, S. (2015). Attribute level and pattern level classification consistency and accuracy indices for cognitive diagnostic assessment. *Journal of Educational Measurement*, 52(4), 457-476.

3. การอ้างอิงบทความจากวารสารอิเล็กทรอนิกส์ (มี DOI)

ไทย ชื่อ/สกุล/(ปีที่พิมพ์)/ชื่อบทความ./ชื่อวารสาร,/เลขของปีที่(เลขของฉบับที่), /
/////เลขหน้า./https://doi.org/เลขdoi

อังกฤษ สกุล,/อักษรย่อชื่อ/(ปีที่พิมพ์)/ชื่อบทความ./ชื่อวารสาร,/เลขของปีที่(เลขของฉบับที่),/
/////เลขหน้า./https://doi.org/doi

Rubright, J. D. (2018). Impact of both local item dependencies and cut-point locations on examinee classifications. *Educational Measurement: Issues and Practice*, 37(3), 40-45.
<https://doi.org/10.1111/emip.12183>

4. การอ้างอิงจากวิทยานิพนธ์ (รูปเล่ม)

ไทย ชื่อ/สกุล/(ปีที่เผยแพร่)/ชื่อวิทยานิพนธ์/[วิทยานิพนธ์ปริญญาดุษฎีบัณฑิต หรือ
/////วิทยานิพนธ์ปริญญาโท/ดุษฎีบัณฑิต]./ชื่อมหาวิทยาลัย.

อังกฤษ สกุล,/อักษรย่อชื่อ/(ปีที่เผยแพร่)/ชื่อวิทยานิพนธ์/[Doctoral dissertation or
/////Master's thesis]./ชื่อมหาวิทยาลัย.

สุจิตรา โชฬิมา. (2559). การพัฒนาตัวบ่งชี้สมรรถนะเชิงวัฒนธรรมของนักเรียนระดับชั้นมัธยมศึกษาตอนปลาย
สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 20 (อุดรธานี) [วิทยานิพนธ์ปริญญาโท].
มหาวิทยาลัยมหาสารคาม.

Ata, A. (2015). *Factor effecting teacher-child communication skill & self- efficacy beliefs: An investigation on preschool teachers* [Master's thesis]. Middle East Technical University.

5. การอ้างอิงวิทยานิพนธ์จากเว็บไซต์

ไทย	ชื่อ/สกุล./(ปีที่เผยแพร่)./ชื่อวิทยานิพนธ์/[วิทยานิพนธ์ปริญญาโทหรือปริญญาเอก] //ชื่อเว็บไซต์./URL
อังกฤษ	สกุล./อักษรย่อชื่อ./(ปีที่เผยแพร่)./ชื่อวิทยานิพนธ์/[Doctoral dissertation or Master's thesis]./ชื่อเว็บไซต์./URL

สุจิตรา โชฬิมา. (2559). การพัฒนาตัวบ่งชี้สมรรถนะเชิงวัฒนธรรมของนักเรียนระดับชั้นมัธยมศึกษาตอนปลาย
สังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 20 (อุดรธานี) [วิทยานิพนธ์ปริญญาโท].
Academic Resource Center Mahasarakham University. [http://khon.msu.ac.th/_dir/
fulltext/fulltextman/full4/sujitta11402/titlepage.pdf](http://khon.msu.ac.th/_dir/fulltext/fulltextman/full4/sujitta11402/titlepage.pdf)

Ata, A. (2015). *Factor effecting teacher-child communication skill & self- efficacy beliefs: An investigation on preschool teachers* [Master's thesis]. Middle East Technical University
Library. <http://etd.lib.metu.edu.tr/upload/12619101/index.pdf>

6. การอ้างอิงจากเว็บไซต์

ไทย	ชื่อ/สกุลผู้เขียน./(ปี./วัน/เดือนที่เผยแพร่)./ชื่อบทความ./ชื่อเว็บไซต์./URL
อังกฤษ	สกุล./อักษรย่อชื่อผู้เขียน./(ปี, เดือน/วันที่เผยแพร่)./ชื่อบทความ./ชื่อเว็บไซต์./URL
* กรณีที่ไม่มีวันที่เผยแพร่ปรากฏ ให้ใส่ (ม.ป.ป.) หรือ (n.d.)	
* กรณีที่มีปรากฏเฉพาะ พ.ศ. หรือ ค.ศ. ให้ใส่แค่ (ปี) เท่านั้น	
* กรณีชื่อผู้เขียนและชื่อเว็บไซต์เป็นชื่อเดียวกัน ให้ตัดชื่อเว็บไซต์ออก	

เกรียงศักดิ์ เจริญวงศ์ศักดิ์. (2552). *บทเรียนสหรัฐอเมริกา เปิดหลักสูตรสายอาชีพตามความต้องการ
ตลาดแรงงาน*. <http://www.kriengsak.com/>

การอ้างอิงในเนื้อหา

การอ้างอิงในเนื้อหาให้ใช้ระบบนาม-ปี โดยต้องแปลเป็นภาษาอังกฤษทั้งหมด โดยมีรูปแบบการเขียนดังนี้

1. กรณีผู้แต่ง 1 คน

สกุล/(ปี) หรือ (สกุล, ปี)

Parker (2019) พบว่า.....

.....การใช้สื่อสังคมออนไลน์และความเครียดของวัยรุ่นมีความสัมพันธ์กัน (Parker, 2019)

2. กรณีผู้แต่ง 2 คน ให้เชื่อมด้วย “and” หรือ “&” ดังนี้

สกุล/and/สกุล/(ปี) หรือ (สกุล/&/สกุล, ปี)

Sawangboon and Erawan (2020) กล่าวว่า

....รวมถึงยกระดับระบบการประเมินคุณภาพให้ได้มาตรฐาน (Sawangboon & Erawan, 2020)

3. กรณีผู้แต่ง 2 คน ขึ้นไป ให้เขียนนามสกุลเฉพาะคนแรกแล้วตามด้วย et al

สกุล/et al./ (ปี) หรือ (สกุล/et al., ปี)

Pearson et al. (1984) กล่าวว่า.....

.....(Pearson et al, 1984)

4. กรณีการอ้างอิงมากกว่าสองแหล่งข้อมูล ให้เขียนเรียงตามลำดับอักษร และค้นแต่ละแหล่งด้วยเครื่องหมายอัฒภาค (;)

Embretson & Reise (2000); Marvelde et al. (2006) ได้ข้อค้นพบว่าสอดคล้องกันว่า.....

..... (Embretson & Reise, 2000; Marvelde et al., 2006)

5. กรณีการอ้างอิงที่ไม่ได้อ้างจากต้นฉบับแต่เป็นการอ้างอิงต่อให้ใช้คำว่า “as cited in”

งานของผู้เขียน (1) ถูกอ้างอิงในงานของผู้เขียน (2)

สกุล⁽¹⁾/(ปี,/as cited in/สกุล⁽²⁾;/ปี).....หรือ(สกุล⁽¹⁾;/ปี/as cited in/สกุล⁽²⁾;/ปี)

Arnett (2000, as cited in Claiborne & Drewery, 2010) ให้ข้อเสนอแนะว่า.....

..... (Arnett, 2000 as cited in Claiborne & Drewery, 2010)

การเขียนรายการอ้างอิงภาษาไทยเป็นภาษาอังกฤษ (Translate Thai Reference)

การเขียนรายการอ้างอิงภาษาไทยเป็นภาษาอังกฤษ มีวิธีการเขียนที่คล้ายคลึงกับการเขียนรายการอ้างอิงตามรูปแบบของภาษาอังกฤษในแต่ละประเภทของสารสนเทศ มีรายละเอียดดังนี้

1. ชื่อผู้เขียน ขึ้นต้นด้วยนามสกุลของผู้เขียนและต่อด้วยชื่อย่อ
2. ชื่อสำนักงานหรือกระทรวง เขียนเป็นชื่อภาษาอังกฤษ
3. ปีที่พิมพ์ เปลี่ยนจาก พ.ศ. เป็น ค.ศ.
4. ชื่อเรื่อง ชื่อวิทยานิพนธ์ ชื่อบทความ ชื่อวารสาร และชื่องานประชุมวิชาการ แปลเป็นภาษาอังกฤษ

ตัวอย่างการอ้างอิงในเนื้อหา

ทัศนศิริินทร์ สว่างบุญ (2563) ได้ศึกษาวิจัยเรื่อง.....

Sawangboon (2020) ได้ศึกษาวิจัยเรื่อง.....

....รวมถึงยกระดับระบบการประเมินคุณภาพให้ได้มาตรฐาน (ทัศนศิริินทร์ สว่างบุญ, 2563)

....รวมถึงยกระดับระบบการประเมินคุณภาพให้ได้มาตรฐาน (Sawangboon, 2020)

รายการอ้างอิง (Reference)

ประวิต เอราวรรณ์. (2562). *การออกแบบโครงการทางการศึกษาและการประเมินแบบเร่งด่วน*. สำนักพิมพ์แห่งจุฬาลงกรณ์มหาวิทยาลัย.

Erawan, P. (2019). *Designing education program and rapid appraisal*. Chulalongkorn University Press. (in Thai)

พิสิษฐ์ พจนจารุวิทย์. (2559). การสร้างแบรนด์องค์กรของมหาวิทยาลัยราชภัฏในประเทศไทย. *วารสารวิทยาลัยพาณิชยศาสตร์บูรพาปริทัศน์*, 11(1), 17 – 49.

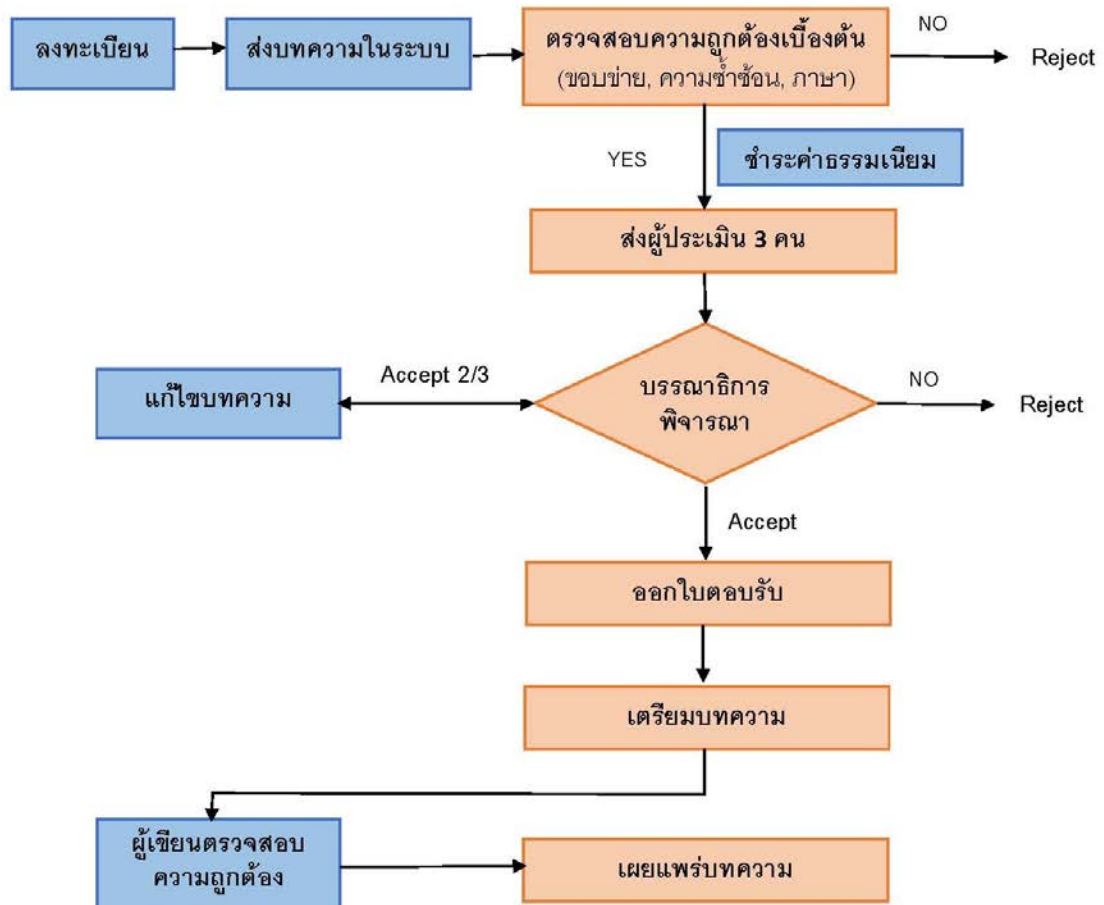
Potjanajruwit, P. (2016). The Corporate Branding Strategy of Rajabhat University in Thailand. *Journal of Graduate School of Commerce Burapha Review*, 11(1), 17 – 49. (in Thai)

สุจิตรา โชติพินาย. (2559). *การพัฒนาตัวบ่งชี้สมรรถนะเชิงวัฒนธรรมของนักเรียนระดับชั้นมัธยมศึกษาตอนปลายสังกัดสำนักงานเขตพื้นที่การศึกษามัธยมศึกษา เขต 20 (อุดรธานี) [วิทยานิพนธ์ปริญญาโท]*. มหาวิทยาลัยมหาสารคาม.

Sopimai. S. (2016). *Developing Indicators of Cultural Competence for High School Students under The Secondary Education Service Area 20 (Udon Thani)* [Master's thesis]. Mahasarakham University. (in Thai)

ขั้นตอนการยื่นบทความ

ขั้นตอนการยื่นบทความให้ผู้พิมพ์ (Author) สมัครในระบบ TCI-Thaijo โดยมีกระบวนการดังต่อไปนี้ (ศึกษารายละเอียดเพิ่มเติมจากเว็บไซต์)



อัตราค่าพิจารณาบทความ (Rate)

1. บทความภาษาไทย 3,000 บาท/1 บทความ
2. บทความภาษาอังกฤษ 5,000 บาท/1 บทความ

*ชำระค่าธรรมเนียมเมื่อผ่านการตรวจสอบคุณภาพเบื้องต้น ก่อนนำเสนอต่อผู้ทรงคุณวุฒิ

การชำระค่าธรรมเนียม

โปรดชำระค่าธรรมเนียมผ่านทาง

ธนาคารทหารไทยธนชาต สาขาเสริมไทยคอมเพล็กซ์ มหาสารคาม

ชื่อบัญชี วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม (JEM MSU)

ประเภท ออมทรัพย์ บัญชีเลขที่ 438-2-70054-2



วารสารการวัดผลการศึกษา มหาวิทยาลัยมหาสารคาม
Journal of Educational Measurement
Mahasarakham University



Website: <https://so02.tci-thaijo.org/index.php/jemmsu/index>

ภาควิชาวิจัยและพัฒนาศึกษา คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม

สำนักงาน: ชั้น 8 อาคารวิทยพัฒนา คณะศึกษาศาสตร์ มหาวิทยาลัยมหาสารคาม
ต.ตลาด อ.เมือง จ.มหาสารคาม 44000

โทรศัพท์: 043-754321-40 ต่อ 6272

E-mail : jem2msu@gmail.com