10th NETWORK OF INTER-ASIAN CHEMISTRY EDUCATORS [10NICE] CONFERENCE 2025

Proceedings

YAMAGATA 26-28 JULY 2025

Venue: YAMAGATA KOKUSAI HOTEL

Editor-in-Chief : YASUNAO KURIYAMA (Yamagata University)

Chemistry for STEAM, SDGs, Inquiry Based Learning

CONTENTS

A Study on Promoting Elementary Students’ Construction of Research Questions in Chemistry Science Fairs through Generative AI and Cognitive Apprenticeship-Based Instructional Design

Shih-Min Chang [1*], Chin-Cheng Chou [2]

[1,2] *Department of Science Education, National Taipei University of Education*

[1]*nice1040802@gmail.com*

Abstract

This study explores how elementary school teachers can integrate generative AI with cognitive apprenticeship to guide students in gradually constructing researchable questions during a chemistry science fair project. The participants were four upper-grade elementary students, each provided with a mobile device pre-configured with specific AI prompts. The instructional design comprised three phases: topic generation, question-formulation strategy guidance, and abstract analysis, integrating teacher scaffolding and AI assistance throughout the process. Data sources included students’ learning portfolios and their dialogue logs with the AI system. Analysis was conducted using the six dimensions of cognitive apprenticeship. The results indicate that teachers played the roles of cognitive modeling and strategic coaching, while the generative AI served as both a reasoning trigger and a real-time scaffolding tool. This dual-support system facilitated students' transition from vague, ill-defined interests to the formulation of operational research questions with awareness of variable control. The study suggests that teachers can effectively leverage generative AI as a support tool to enhance students’ quality of questioning and scientific inquiry thinking throughout the science fair process.

Keywords: Generative AI, cognitive apprenticeship, science fair, research question construction, chemistry inquiry

Introduction

In Taiwan’s science fairs, students must meet three fundamental requirements: propose researchable questions, effectively control variables, and design complete and feasible experimental plans. These requirements not only test students’ scientific knowledge and skills but also reflect their understanding and application of scientific inquiry methods. Through these requirements, students can experience the scientific research process in practice and cultivate a rigorous scientific attitude.

However, at the elementary school level, students’ research questions often tend to be interesting and entertaining but lack operability and verifiability. For example, some questions, while intriguing, cannot be tested through experiments or with controlled conditions. This suggests that younger students, when designing questions, often overlook the feasibility of scientific inquiry and the importance of variable control.

Award-winning science fair projects often share common features: rigorous and thorough research design, creative thinking with a clear problem focus, strong examples of how to ask questions and adjust variables, and effective development of scientific thinking and communication skills throughout the process. These projects stand out not only in results but also in their ability to inspire audiences and other participants, becoming exemplary learning models in science fairs.

To address the gap between students’ natural curiosity and the scientific rigor required in research, instructional support frameworks such as the Cognitive Apprenticeship model have been widely advocated (Collins, Brown, & Newman, 2018). This model consists of six important stages: “Modeling”—experts demonstrate and reveal their thought processes; “Coaching”—providing guidance and real-time feedback; “Scaffolding”—offering necessary support and gradually reducing dependence; “Articulation”—guiding learners to articulate their thinking and strategies; “Reflection”—comparing different methods and outcomes to promote improvement; and “Exploration”—encouraging independent application and extended learning. Through this model, teachers can effectively guide students from imitation to independent thinking, ultimately enabling them to engage in self-directed inquiry.

Moreover, students’ ability to improve their questioning skills involves not only external scaffolding but also the development of self-regulation skills, including self-monitoring, self-evaluation, and self-reflection (Zimmerman, 2000). Research also indicates that metacognitive prompts—whether provided by teachers or embedded in learning tools—can enhance both students’ understanding of content and their grasp of the nature of science (Peters & Kitsantas, 2010). These findings suggest that strategically designed teacher and AI supports could foster both cognitive and metacognitive growth in student question-asking.

In the task design of this study, Task 1 was entirely student-directed, with no teacher support and AI serving only as a “Responder”. This stage consisted of four student-led sessions aimed at observing whether students could generate high-quality chemistry research questions when relying solely on AI.

Task 2 included one teacher-led session and four student-directed activities. The teacher selected science fair project summaries that matched students’ abilities and provided AI prompts, with AI serving as a ”Coach” to help students learn how to interact strategically with AI and focus on chemistry research variables.

Task 3 further increased teacher guidance, with a total of seven sessions, including two teacher-led classes. The teacher guided students in analyzing science fair project summaries, provided the summary content and AI prompts, and AI acted as a “Co-constructor” to collaborate with students in generating research questions. This design aimed to observe whether students could demonstrate higher-level question-posing abilities with comprehensive teacher and AI collaborative support.

The three core research questions of this study were:

(1) Can students propose high-quality chemistry research questions when using only AI without teacher support?
(2) Does teacher-supported AI interaction help students focus on chemistry research variables?
(3) Can students’ chemistry question-posing abilities be enhanced through the six stages of the Cognitive Apprenticeship model?

Methodology

This study involved four fifth-grade students who demonstrated a high level of interest in science fair topics and had prior experience using generative AI. The research design was divided into three task stages, each focusing on a different chemistry concept. Task 1, titled “Making Bouncy Balls,” explored the principles of polymer elasticity. Task 2, “Natural pH Indicators,” utilized red cabbage and butterfly pea flowers as research materials. Task 3 investigated the relationship between cyanotype printing and pH, specifically examining how acidity affects the color changes in cyanotype images.

In terms of data collection, three main types of data were gathered. First, complete conversation transcripts between the students and the generative AI were collected to analyze the AI’s role and influence in the questioning process. Second, students' worksheets and the research questions they generated in each task were collected to examine the types and levels of questions posed. Third, students' reflective notes after each task were collected to gain insight into their shifts in thinking and learning experiences throughout the process.

For the analytical framework, students' questions were first categorized into five types based on their nature: Factual Questions, Procedural Questions, Reasoning Questions, Explanation Questions, and Hypothetical Questions. Next, according to Bloom's Taxonomy, questions were classified into six cognitive levels: Remember, Understand, Apply, Analyze, Evaluate, and Create. Through this dual classification framework, the study was able to examine students' questioning performance and changes in ability from both the perspective of question type and cognitive level, under varying tasks and instructional support conditions.

Results

Based on the conversation data between students and the generative AI across the three tasks (see Table 1), in Task 1, S1 had the highest number of interactions with the AI (10 times), indicating the greatest reliance on AI when no teacher support was provided. In Task 2, with the teacher providing summaries and prompt phrases, all students significantly increased their interactions with the AI (a total of 79 times), with S2 reaching as many as 44 interactions, reflecting that the support strategy effectively enhanced student–AI communication. By Task 3, although the total number of interactions decreased to 23, the quality of the research questions generated by the students improved noticeably.

In terms of research question quality (see Table 2), in Task 1, none of the students were able to propose acceptable research questions. In Task 2, some students were able to pose 1–2 qualifying research questions. By Task 3, S1, S2, and S3 each proposed four high-quality research questions, and S4 proposed one, indicating that through three consecutive tasks combined with teacher support, students' research question quality improved significantly.

Table 1. Number of Student–AI Interactions Across Three Tasks

Task	S1	S2	S3	S4	Total number
Task1	10	2	2	3	17
Task2	15	44	12	8	79
Task3	4	7	6	6	23

Table 2. Number of High-Quality Research Questions Proposed by Each Student Across Three Tasks

Task	S1	S2	S3	S4
Task1	No good question	No good question	No good question	No good question
Task2	2 good question	1 good question	No good question	1 good question
Task3	4 good question	4 good question	4 good question	1 good question

Using S1 as an example for qualitative analysis, the types of questions posed showed a clear shift over the course of the tasks. In terms of question levels, the proportion of factual questions (Q1) decreased, while the proportions of explanation questions (Q4) and hypothetical questions (Q5) increased (see Figure 1). This indicates a transition from merely seeking factual information to engaging in deeper reasoning and hypothesis construction. According to Bloom's taxonomy, the proportion of questions at the remembering level (B1) declined, while the proportions at the evaluating (B5) and creating (B6) levels increased (see Figure 2), reflecting a gradual development of S1's questioning ability toward higher-order thinking.

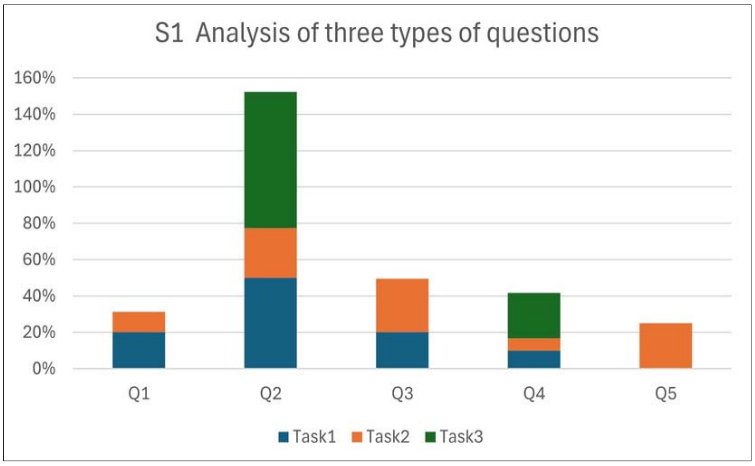


Figure 1. S1 Distribution of Question Types Across Three Tasks

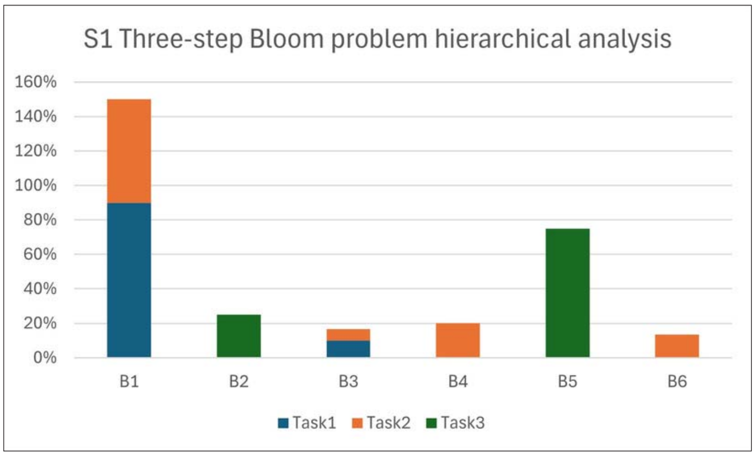


Figure 2. S1 Three-Step Bloom's Taxonomy Analysis Across Three Tasks

Based on the qualitative analysis of S2, the questions posed showed a clear transformation across the three tasks. In terms of question level distribution, the proportions of factual questions (Q1) and procedural questions (Q2) gradually decreased, while the proportion of hypothetical questions (Q5) increased (see Figure 3),. This indicates that after multiple tasks, S2's questioning shifted from focusing on basic knowledge and procedural steps to more exploratory and speculative thinking.

From the perspective of Bloom's taxonomy, the proportion of questions at the remembering level (B1) decreased significantly, indicating a reduced reliance on simple knowledge recall. At the same time, the proportions of questions at the analyzing level (B4) and creating level (B6) increased(see Figure 4), reflecting that S2, through interactions with AI, gradually demonstrated higher-order thinking skills, such as analyzing variable relationships, comparing different possibilities, and creatively generating new ideas or hypotheses. This shift suggests that, with teacher guidance and AI support, the student's questioning ability improved qualitatively from lower-order to higher-order levels.

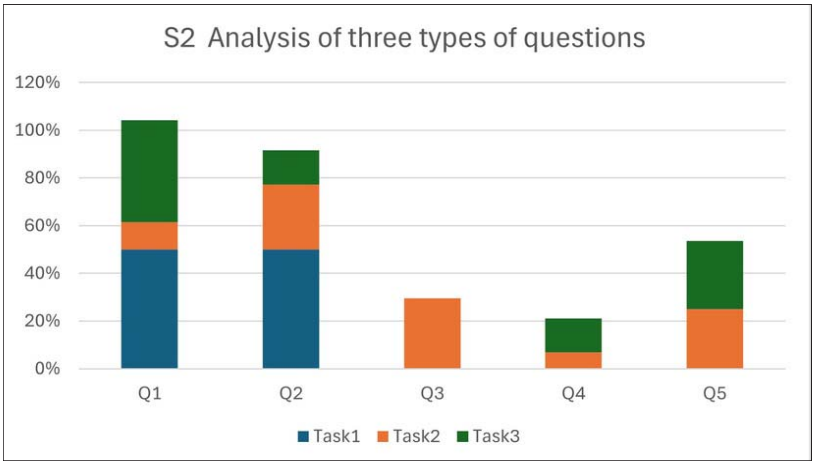


Figure 3. S2 Distribution of Question Types Across Three Tasks

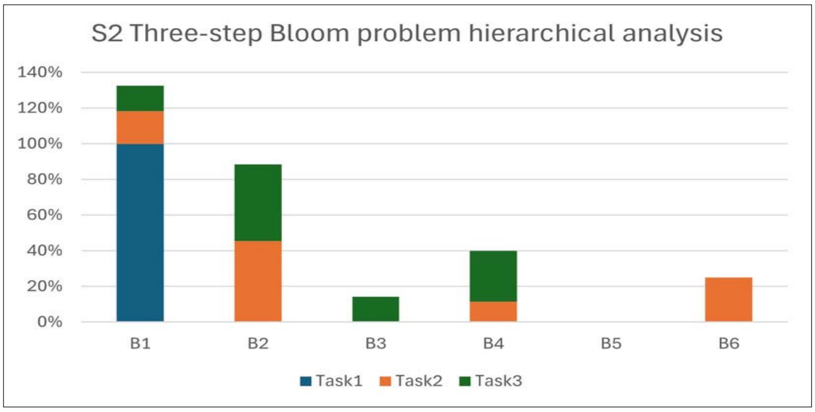


Figure 4. S2 Three-Step Bloom's Taxonomy Analysis Across Three Tasks

From the qualitative analysis of S3, it is evident that the types of questions posed gradually shifted toward higher-level thinking across the three tasks. In terms of question levels, the proportions of factual questions (Q1) and procedural questions (Q2) decreased, while the proportions of reasoning questions (Q3) and explanation questions (Q4) increased(see Figure 5). This indicates that the student gradually reduced questions focused solely on factual information and procedural steps, and began attempting reasoning and explanation.

According to Bloom's taxonomy, the proportion of questions at the remembering level (B1) showed a downward trend, while the proportion at the understanding level (B2) increased significantly(see Figure 6). This reflects that, over the course of the tasks, S3 transitioned from merely recalling facts to better understanding and interpreting scientific concepts. Although there was no clear performance yet in higher-order levels such as analyzing, evaluating, and creating, the cognitive process already showed a progression from lower-order to mid-level thinking.

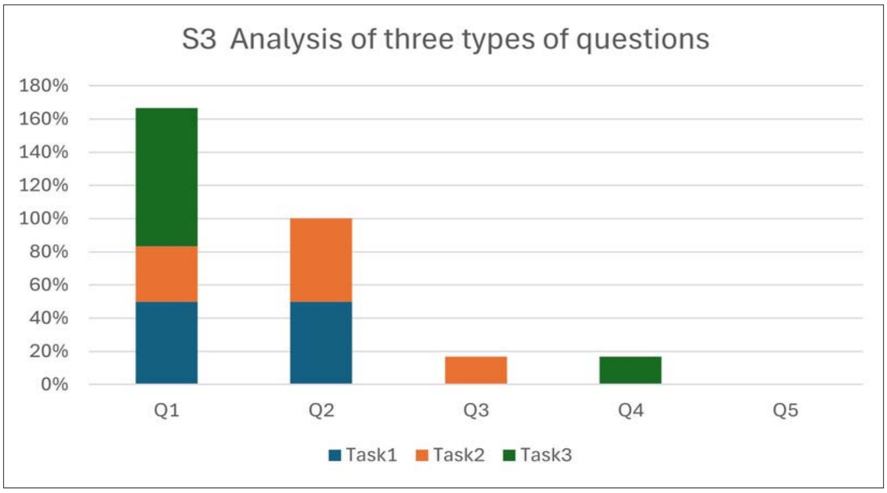


Figure 5. S3 Distribution of Question Types Across Three Tasks

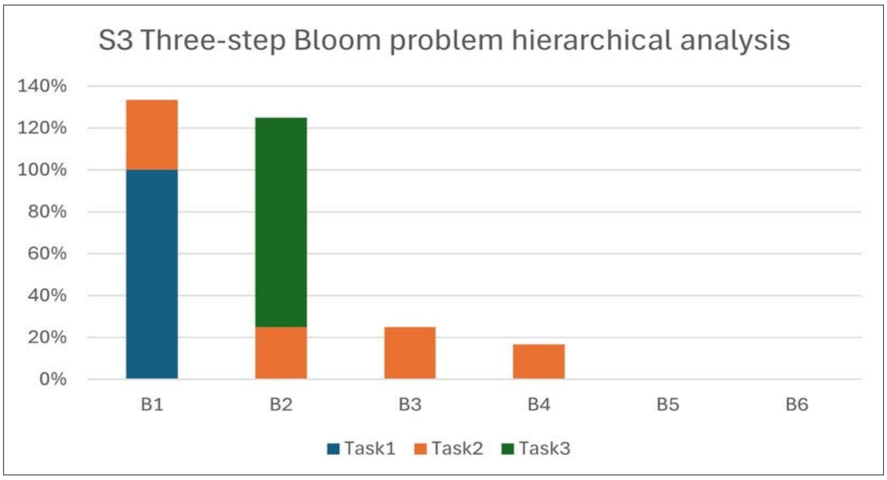


Figure 6. S3 Three-Step Bloom's Taxonomy Analysis Across Three Tasks

From the qualitative analysis of S4, it can be seen that the student's questioning showed a trend of shifting from lower-order to higher-order thinking across the three tasks. In terms of question levels, the proportions of factual questions (Q1) and procedural questions (Q2) gradually decreased, while the proportions of explanation questions (Q4) and hypothetical questions (Q5) increased(see Figure 7). This indicates that S4's questions gradually moved from a focus on factual knowledge and procedural aspects toward more reasoning-based and predictive inquiry.

In terms of Bloom's taxonomy, the proportion of questions at the remembering level (B1) decreased significantly, while the proportions at the evaluating level (B5) and creating level (B6) increased(see Figure 8). This reflects that, during interactions with AI, S4 was increasingly able to engage in critical judgment and creative thinking. Such a shift demonstrates that, in a context where teacher support and AI collaboration are provided, a student's questioning ability can progress from basic knowledge recall to the production of higher-order thinking.

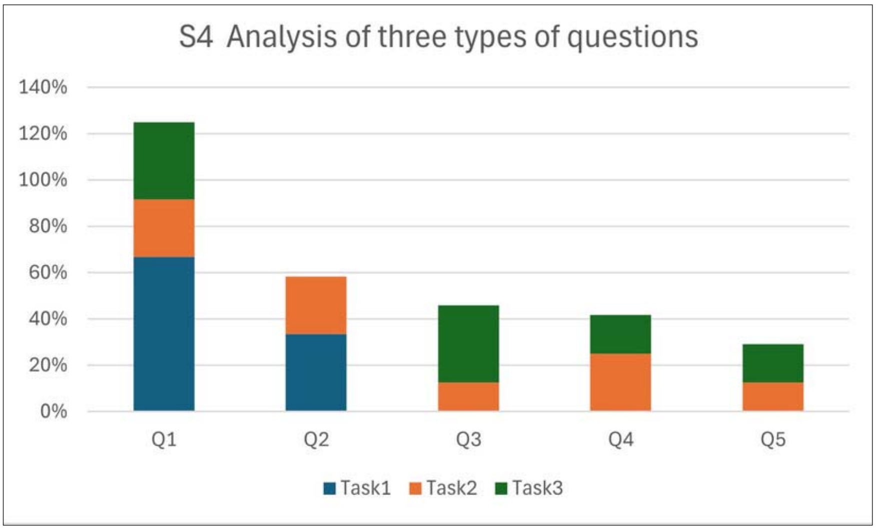


Figure 7. S4 Distribution of Question Types Across Three Tasks

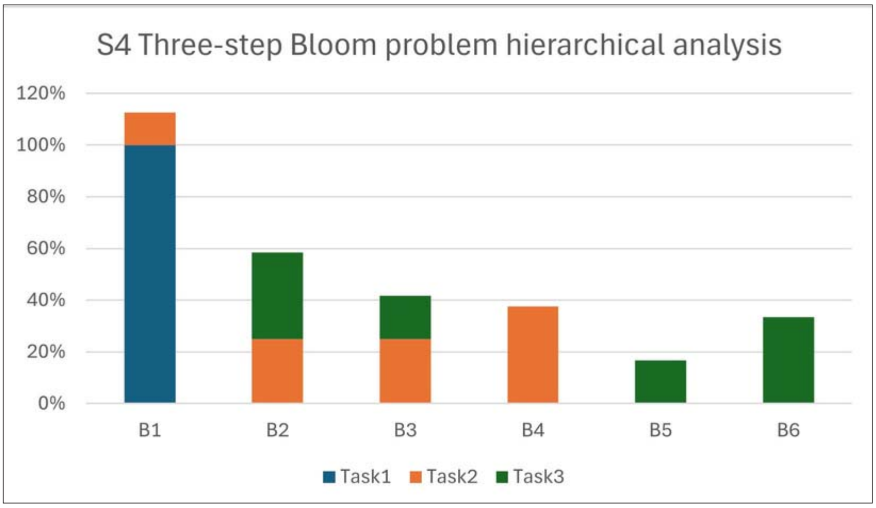


Figure 8. S4 Three-Step Bloom's Taxonomy Analysis Across Three Tasks

For the first research question—"Can students propose high-quality chemistry research questions relying solely on AI?"—in Task 1, students interacted independently with the AI, but most of the questions they posed were simple or vague and failed to focus on scientific variables. For example, S1 stated, "*I didn't know where to start,*" and S3 mentioned, "*The bouncy ball turned into slime.*" These responses indicate that, in the absence of guidance, students had difficulty identifying the key points of inquiry, reflecting that at the "Modeling" stage of the Cognitive Apprenticeship model, they still required more support.

For the second research question—"Does teacher-supported AI help students focus on chemistry variables?"—in Task 2, students were provided with science fair project summaries and key prompts, enabling them to focus their questions on variable control and pose more testable research questions. S4 commented, "*The summary made it easier for me to ask questions,*" while S2 said, "*The answers helped me come up with better questions.*" These findings suggest that, with teacher guidance and AI assistance, students were able to conduct more effective inquiry design. This stage corresponds to the "Coaching" and "Scaffolding" stages of the Cognitive Apprenticeship model.

For the third research question—"Can students' chemistry question-posing skills be enhanced through the six stages of the Cognitive Apprenticeship model?"—in Task 3, students were able to independently use prompts to interact with the AI and demonstrated greater reflection and deeper thinking, generating questions that were more specific and targeted. S3 remarked, "*AI also asked me harder questions,*" and S4 stated, "*It helped me think deeper.*" These statements reflect that students had reached the "Articulation", "Reflection", and "Exploration" stages of the Cognitive Apprenticeship model, showing significant progress in both questioning ability and cognitive level.

Conclusions

The results of this study show that when students engaged in chemistry inquiry without any support, their questions were mostly simple and factual in nature, lacking the characteristics of in-depth investigation and experimental validation. However, when teacher guidance was combined with generative AI support, the quality of students' questions improved significantly—not only becoming more creative, but also meeting the criteria for experimental testability and focusing more precisely on the control and examination of scientific variables. In addition, students' questioning abilities exhibited a gradual progression through the six stages of the Cognitive Apprenticeship model, moving from basic knowledge recall in the early stages to higher-order thinking characterized by critical and creative abilities. These findings indicate that the collaborative support of teachers and AI has a significant effect on fostering students' scientific questioning skills.

Reference

[1] Collins, A., Brown, J. S., & Newman, S. E. (2018). Cognitive apprenticeship: Teaching the crafts of reading, writing, and mathematics. In *Knowing, learning, and instruction* (pp. 453–494). Routledge. https://doi.org/10.4324/9780429492255-15

[2] 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). San Diego, CA: Academic Press.

[3] Peters, E., & Kitsantas, A. (2010). The effect of nature of science metacognitive prompts on science students' content and nature of science understanding. *The Journal of Educational Research, 103* (5), 295–307. https://doi.org/10.1111/j.1949-8594.2010.00050.x

Creating Art Installations That Combine Nature and Culture in Chemistry Class

KAJIYA Daisuke[1*], SUNAGA Hiroaki[1], SATO Shugo[1]

[1]Liberal Arts Education Center, Ashikaga University, Tochigi, Japan

** kajiya.daisuke@g.ashikaga.ac.jp*

Abstract

With the growth of the field of chemistry and the remarkable development of artificial intelligence technology, understanding how to live a meaningful and fulfilling life of well-being has become an important agenda in countries where material products are abundant and vast amounts of information is instantly accessible. In this context, a key societal concern is the interdisciplinary integration of humanities and natural sciences, particularly the incorporation of an aesthetic sensibility into natural science education. At this conference, we aim to share our demonstration experiment in chemistry to the educational community and foster discussions on the future of education that contributes to creating a sustainable society that feels natural and good to individuals and promotes their harmony. This study is characterized by its use of a widely known chemistry experiment—the acid–base color change of plant pigments—recontextualized within the regional and cultural background of Ashikaga City, renowned for its textile and dyeing industries and wisteria flowers, ultimately transforming it into an art installation.

Keywords: Botanical Dyeing; Spiral Evolution; Color; General

Introduction

One factor contributing to the difficulty students face in understanding chemistry compared to other scientific disciplines is its focus on invisible atoms and molecules. Increasing the number of experiments that allow students to visually observe changes could enhance their ability to vividly imagine the world of atoms and molecules with a sense of realism, potentially leading to more beneficial learning experiences. Indeed, as the old adage goes, "seeing is believing". This paper presents our efforts to visualize chemistry experiments.

At the Faculty of Engineering, Ashikaga University, first-year students are introduced to chemistry through general education courses. These include "Introduction to Chemistry" (matter structure) and "Chemistry I" (chemical reactions/equilibrium) in their first year, and "Chemistry II" (environmental chemistry) in their second year. Practical experience is gained through "Chemistry Experiment." Additionally, "Seminar: The Science of Color" integrates science and art for STEAM education, while "Chemistry for Life Design", an inquiry-based subject, and "Summer School Chemistry" offer further study.

Within this group of general education chemistry courses, we have introduced hands-on, experiential chemistry experiments that appeal to students' sensibilities, along with demonstration content that allows interaction with real-world materials. For example, we have reported on colorful experiments that involve the fun of rainbow colors [1], purple [2], shimmering effects [3], gray [4], teal [5], yellow [6], green [7], brown [8], blue [9], and magenta [10]. Figure 1 shows a schematic illustration of color changing experiment using purple juice to produce various colors [10], focusing on STEAM education, a key topic at this conference.

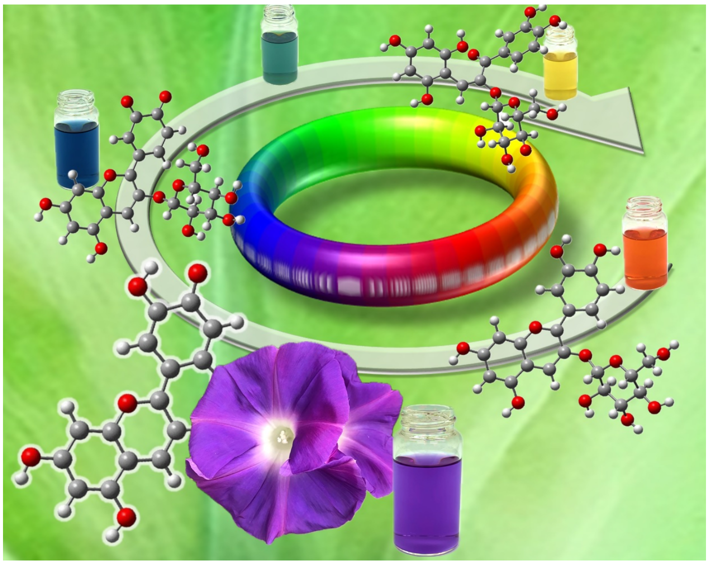


Figure 1. Students can experience a variety of color changes along the color wheel using plant-derived pigments including anthocyanin. Reprinted with permission from Kajiya, D, J. Chem. Educ. 2025, 102, 3671. Copyright 2025 American Chemical Society.

Collaboration between Local Characteristics and Chemistry

Educational content that offers genuine experiences, coupled with storytelling that allows individuals to find those experiences meaningful, is anticipated to be a source of significant value in the coming era. Ashikaga City, where Ashikaga University is located, possesses several distinctive characteristics: (i) Historically, the textile and dyeing industries flourished there. (ii) Asian cultures are deeply rooted, and the city is home to numerous temples. It is a regional city intimately familiar with the cyclical nature of changing seasons and cyclical thinking. This aligns well with a circular, sustainable development mindset, especially in an era transitioning from the linear consumption society prevalent since the Industrial Revolution to a circular society [11]. (iii) It boasts abundant flowers and nature, with strong winds and sunlight, indicating rich natural energy sources. The symbolic flower of Ashikaga City is wisteria.

Art Installations

One day, the wisteria flowers were carefully collected by students. An extract was prepared by adding water and gently rubbing the petals by hand. When acids or bases such as citric acid or baking soda were added to this extract, a range of color changes was observed. By utilizing these color variations, multi-colored gradient patterns were created on filter paper, resembling watercolor paintings. Figure 2 shows an artwork titled KAZAGURUMA (“Pinwheel”), a rotating piece made using wisteria extract, water, food-grade acids and bases, paper, wooden sticks (toothpicks), and paper straws. The pinwheel was crafted using origami techniques, reflecting an aspect of Japanese culture.

Figure 2. Model artwork made using wisteria.

This experiment follows the same procedure as popular and well-known chemical experiments worldwide that use red cabbage extract or butterfly pea solution to demonstrate color changes with acids and bases. Figure 3 shows an artwork created by first-year undergraduate students using butterfly pea tea, food grade acid-base, paper and scissors. A lighting artwork was designed by sticking the paper work onto a beaker and shining a flashlight on it.

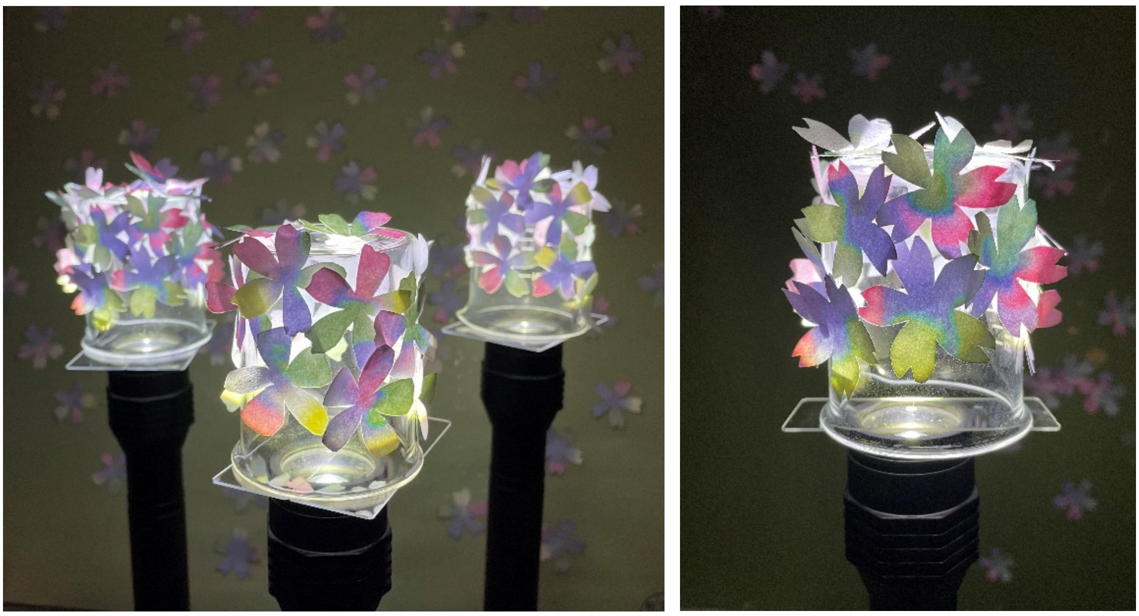


Figure 3. Model artwork made using butterfly pea [12].

By adding alum, which contains aluminum ions, and dyeing fabric fibers, students can also incorporate experiments demonstrating deep color formation through complexation [2,7]. Similar experiments can be conducted using purple azalea flowers, which are commonly planted in many regions of Japan. In Japan, elementary school students cultivate morning glories as part of their science classes, and analogous color-change experiments can be performed using morning glory petals.

Depending on the combination of these experiments, students can learn foundational chemical concepts such as atoms, ions, molecules, acids/bases, complexation, and redox reactions in an interdisciplinary manner through hands-on experiments, as if they were enjoying art. These activities provide an opportunity for non-chemistry major students to engage with chemistry in an accessible and creative way. This content can also be implemented as an outreach program for high school [13], junior high, and elementary school students, as well as video demonstrations shared with overseas universities [14] and activity for high school teachers [15]. Exciting and visually appealing experiments are conducted using abundant, readily available substances. Some of these raw materials are plants grown on campus from seeds, with only the necessary amounts harvested for use in classes. The campus chemical laboratory is surrounded by a flower garden.

Furthermore, student feedback indicated that this practice offered a meaningful opportunity to connect artistic expression with chemical learning. Several students commented that they were excited to see how the colors changed when acids or bases were added, that it was enjoyable to apply what they had learned in lectures to real experiments, and that it was interesting to think about structural changes and chemical mechanisms visually and logically. Test results also showed that incorporating this research into the class enhanced students' understanding of key concepts such as structural changes involved in acid–base reactions [10] and the formation of metal complexes [7].

Summary

This paper reports on a chemical experiment that utilized natural dye materials cultivated in the botanical garden to create colorful circular gradient patterns. This work contributes to the development of a lush campus environment while integrating a well-known chemistry experiment—the acid–base color change of plant pigments—with the regional and cultural identity of Ashikaga City, characterized by its long-standing textile and dyeing traditions and its symbolic wisteria flowers, thereby elevating the experiment into an educational art installation that connects science with local culture.

ASSOCIATED CONTENT

Hazards of Plant-Derived Materials in Experiments

When conducting experiments with materials extracted from natural plants, it is crucial to recognize the potential for toxicity. Operators must implement adequate safety precautions, including the use of protective eyewear and gloves, to mitigate risks.

Ethics

Regarding the artwork prepared by students, all students provided written consent for the photograph to be published. For all other content, the study falls under a general institutional protocol for classroom activities.

Acknowledgements

The experimental educational materials utilized in this research were developed with the invaluable cooperation of many individuals in the Ashikaga region and its surrounding areas. In particular, we extend our sincere gratitude to the adults involved in natural dyeing, textile dyeing, textile arts, children's science workshops, and scientific instrument analysis institutions. We also deeply appreciate the students who participated in experiments using these materials and provided valuable feedback, which significantly contributed to enhancing the learning experience. Finally, we express our heartfelt thanks to the organizers, secretariat, and all domestic and international participants of the 10th NICE conference for their friendly contributions and engaging discussions.

References

[1] Kajiya, D. Rainbow Colors Generated by Viewing Transparent Polymers through Polarizers, *J. Chem. Educ.* 97, 154–158 (2020).

[2] Kajiya, D. Demonstrating Purple Color Development to Students by Showing the Highly Visual Effects of Aluminum Ions and pH on Aqueous Anthocyanin Solutions, *J. Chem. Educ.* 97, 4084–4090 (2020).

[3] Kajiya, D. Formation of a Water Ball in a Water Bottle to Learn the Chemistry of Surfactants, *J. Chem. Educ.* 98, 1712–1717 (2021).

[4] Kajiya, D. Using Sodium Hydrogen Carbonate to Teach Chemical Concepts of Thermodynamics, *J. Chem. Educ.* 98, 3968–3974 (2021).

[5] Kajiya, D. Using RGBY Color Experiments to Demonstrate Redox Reactions in Terms of Hydrogen, *J. Chem. Educ.* 99, 3346–3351 (2022).

[6] Kajiya, D. Utilizing Yellow Compounds to Introduce π-Conjugated Molecular Structures: A Harmony with Blue Compounds, *J. Chem. Educ.* 100, 4147–4154 (2023).

[7] Kajiya, D. Dyeing Cloth Green Using Red Onion Skin: A pH-Responsive Dye and Aluminum-Ion Complex, *J. Chem. Educ.* 101, 1241–1247 (2024).

[8] Kajiya, D. Producing Ink That Emits Blue Light: Introduction of Dehydration Reaction to Combine Biomolecules Using Maillard Reaction, *J. Chem. Educ.* 101, 5134–5138 (2024).

[9] Kajiya, D. Alizarin Can Dye Nylon Blue: Curiosity-Driven Learning of a Basic Paradigm of Intermolecular Force, *J. Chem. Educ.* 102, 1340–1346 (2025).

[10] Kajiya, D. Creating Purple to Learn the Color Wheel Using a Colorful Anthocyanin Experiment, *J. Chem. Educ.* 102, 3671–3678 (2025).

[11] Ashikaga City “Japan Heritage Ashikaga Gakko: The Oldest School in Japan/The Intellectual Heritage”. https://www.youtube.com/watch?v=CjjrfrkIfUs (accessed Jul 2025).

[12] Ashikaga University News “Life Design Course Students Create Art Installation”. https://ashikaga.ac.jp/archives/2025/8785/ (accessed Jul 2025).

[13] Ashikaga University News “Let's All Make Art: A One-Day University Event”. https://ashikaga.ac.jp/archives/2025/9377/ (accessed Jul 2025).

[14] Ashikaga University News “Environmental Chemistry Class Video Content Shared with Overseas University”. https://ashikaga.ac.jp/archives/2025/8834/ (accessed Jul 2025).

[15] Ashikaga University News “A practical session for high school science teacher”. https://ashikaga.ac.jp/archives/2025/10354/ (accessed Oct 2025).

pH Changes of River Water in STEAM:
Practices of High School Chemistry and Science Club

HIRAI Toshio

Osaka Prefectural Nagao High School, Osaka 573-0102, Japan

t-hiraitos@e.osakamanabi.jp

Abstract

The objective of this study is to share information about the practices of high school science club activities and chemistry classes. The author announced that the water pH of rivers in Japan sometimes indicated alkaline at the 9th NICE Conference in 2023. Why is the river water alkaline? Twenty years ago, his science club students and he started to investigate the water quality of rivers near their school. In STEAM education at his high school, he treats his teaching material “pH changes of the river water” as follows: Science includes chemistry, biology, geology, scientific method and simulation, etc., Technology and Engineering includes the way to use pH meters, the calibration of them, and wastewater treatment etc., Art includes environmental issues and its regulation, etc., and Mathematics includes calculations of exponents and logarithms, and so on. When he taught about the pH of chemistry using these materials above, the students were surprised at the alkaline river water with the experiment of the water quality test, and amazed at the pH change mechanism by photosynthesis.

Keywords: pH changes of river water, alkaline, STEAM education, teaching material, high school chemistry, science club

Introduction

STS education in the 1990s

STS education is an education that relates to Science, Technology and Society shown in Fig. 1. Many stakeholders, including scientists, exist in every domain of STS. Science, Technology and Society interact with one another. So, they are not value free.

In the author’s educational experience, the environmental education was popular in Japanese high school in the 1990s. Shiokawa et al. and the author developed STS teaching materials about Minamata Disease [1] - [4], which included a training series of decision making and role plays.

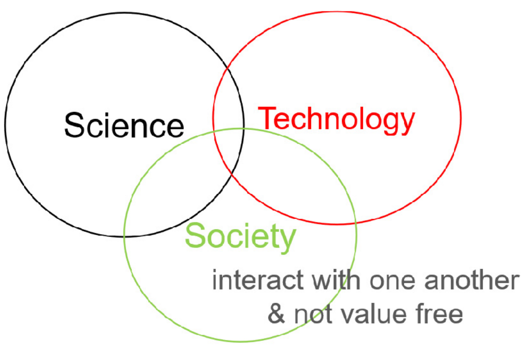


Fig. 1. STS education in the 1990s.

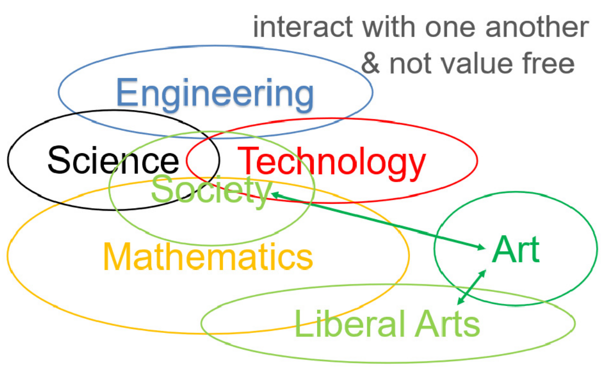


Fig. 2. STEAM education.

STEAM education

Here, STEM education that doesn't cover liberal arts and social science is skipped. Liberal arts and social science relate a lot to Society.

STEAM education is an education that relates to Science, Technology, Engineering, Art and Mathematics shown in Fig. 2. STEAM like STS has many stakeholders including scientists in every domain. Science, Technology, Engineering, Art and Mathematics interact with one another. So, they are not value free, too. Interpreting "Art" broadly, "Art" and "Society" are synonymous. STEAM education is, in a sense, almost Science, Technology, Engineering, Society and Mathematics education.

The purpose of this study is to share information about the practice of high school chemistry classes and science club, and provide ideas that may lead to your awareness of STEAM education.

Practices

The author introduces his concepts of pH changes and river water in STEAM education and then shares information on his high school science club activities and his high school teaching practices.

Concepts of pH changes and river water in STEAM education

Fig. 3. shows a part of concepts of pH changes and river water in STEAM education. Science includes chemistry [ionization, solubility and dissolution equilibrium, pH], biology [photosynthesis and respiration], geology, scientific method [to collect the sample water, analyze it, discuss data, make a report and present the findings], model experiment and simulation, etc., Technology and Engineering includes the way to use pH meters, wastewater treatment etc., Art [Society] includes environmental issues and its regulation, etc., and Mathematics includes calculations of exponents and logarithms, and so on.

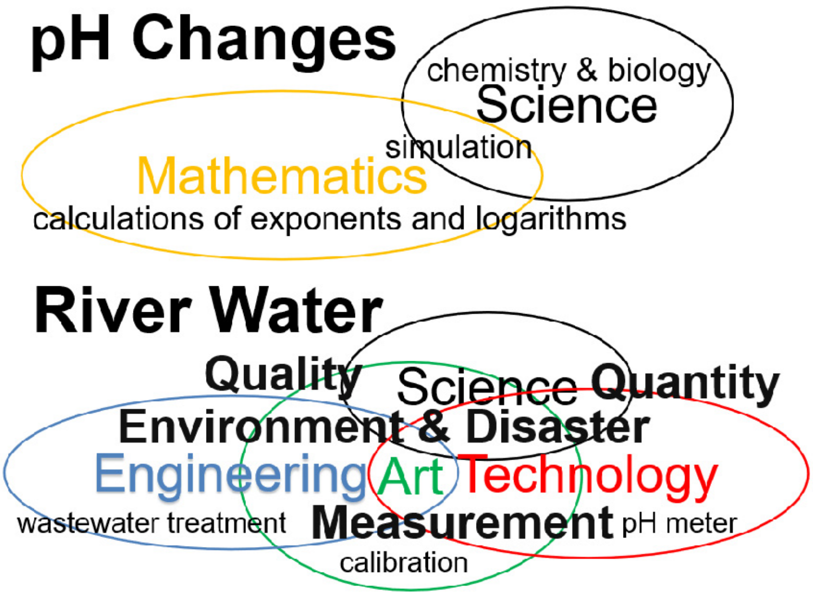


Fig. 3. Concepts of pH changes and river water in STEAM education.

High school science club activities

Survey about river water quality Fig. 4. shows the pH of the water versus the water temperature of the Funahashi River from 2014 to 2018. The line graph indicates the pH and the bar graph indicates the water temperature. In summer, the river water is almost always alkaline.

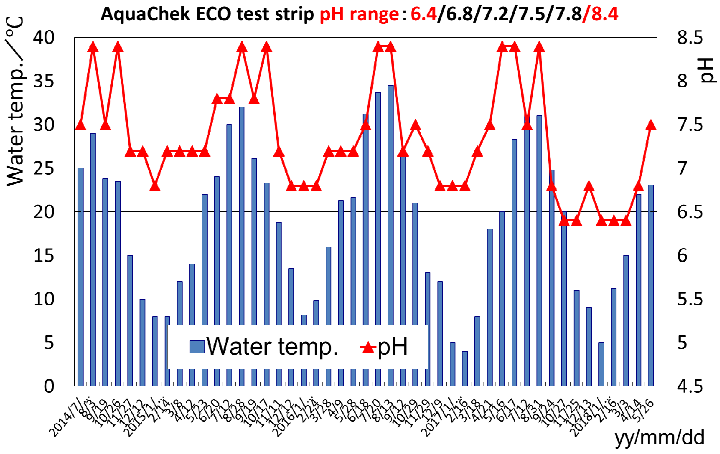


Fig. 4. pH and water temperature of the Funahashi River.

Some volunteers from the students to whom the author taught chemistry wanted to study the reasons why the river water became alkaline after school. A few of them made a science club of which the author has been the mentor for seven years. They took photosynthesis of both aquatic plants and phytoplankton in the river into consideration. The graph shows that the pH increased as the water temperature rose. An increase of the water temperature caused more active photosynthesis of the aquatic plants and phytoplankton in the river. They consumed more CO_2 dissolved in the water.

Chemical Equilibrium $CO_2 + H_2O \rightleftarrows H^+ + HCO_3^-$

This equation indicates that CO_2 reacts reversibly with water to form a solution of the weak acid, H_2CO_3 (carbonic acid). A decrease in concentration of CO_2 shifts this equilibrium to the left as this decrease effects a decrease in concentration of H^+, that is, an increase of the pH.

From these results, they concluded that the stronger photosynthesis of the plants and phytoplankton in the river caused the increase of pH in the summer.

$CO_2 + H_2O \rightleftarrows H^+ + HCO_3^-$

According to the Iwanami Rikagakujiten, pK1 = 6.35. ∴ $pK_1 = -\log K_1$, then $6.35 = -\log K_1$. Thus, $K_1 = 10^{-6.35} = 4.5 \times 10^{-7} ≒ 5 \times 10^{-7}$

If pH=5.7, then $[H^+] = 2 \times 10^{-6}$ mol/L
and $[HCO_3^-] = 2 \times 10^{-6}$ mol/L, too.

$K_1 = [H^+][HCO_3^-] / [H_2CO_3] = 5 \times 10^{-7}$
$[H_2CO_3] = 2 \times 10^{-6} \times 2 \times 10^{-6} / 5 \times 10^{-7}$
Thus, $[H_2CO_3] = 8 \times 10^{-6}$ mol/L.

If $[H_2CO_3]$ decreases to 2×10^{-10} mol/L,
$[H^+][HCO_3^-] / 2 \times 10^{-10} = 5 \times 10^{-7}$
then $[H^+]$ and $[HCO_3^-]$ equals 1×10^{-8} mol/L.
Thus, the pH =8.

As a result, the pH increases to 8.0 from 5.7.

Schema 1. Simulation of pH change to 8.0 from 5.7.

Simulation They simulated the pH changes shown in Schemata 1.-2. and confirmed the pH changes. They presented the result of this study by poster at NICEST (Nippon International Chemistry Expo for Students and Teachers) 2017 Tokyo (Fig. 5.).

The team simulated the pH change of the water from the viewpoint of the equilibrium equation shown below.

$H_2CO_3 \rightleftarrows H^+ + HCO_3^-$

They calculated the dissociation constant K_1 mentioned below from data in the Iwanami Rikagakujitenn (1998).

$K_1 = [H^+][HCO_3^-] / [H_2CO_3] = 5 \times 10^{-7}$

The initial conditions are as follows:
If pH=6.7 at equilibrium, then $[H^+] = 2 \times 10^{-7}$ mol/L.
Thus, $[H_2CO_3] = 8 \times 10^{-8}$ mol/L.

Next, the plants consume CO_2 dissolved in the water.
If $[H_2CO_3]$ decreases to 2×10^{-11} mol/L, then $[H^+]$ and $[HCO_3^-]$ equal $1 \times 10^{-8.5}$ mol/L. Thus, the pH = 8.5.

As a result, the pH increases to 8.5 from 6.7.

Schema 2. Simulation of pH change to 8.5 from 6.7.

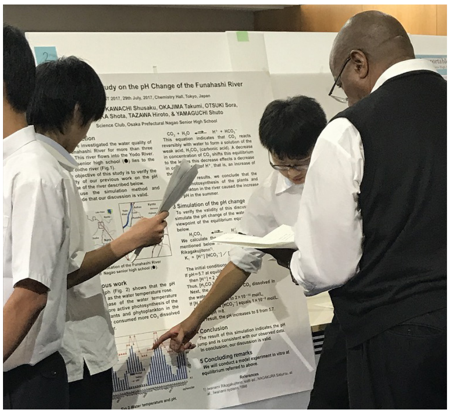


Fig. 5. Poster presentation in English at Chemistry Hall.

Fig. 6. A model experiment of the pH changes by photosynthesis.

Model experiment Some model experiments were conducted (one of them shown in Fig. 6.). Science club members measured the pH and dissolved oxygen (DO) of the water in both the beaker (water only) and the desiccator (water with aquatic plants) under the sunshine in the spring morning. The result of this experiment is shown in Fig. 7.

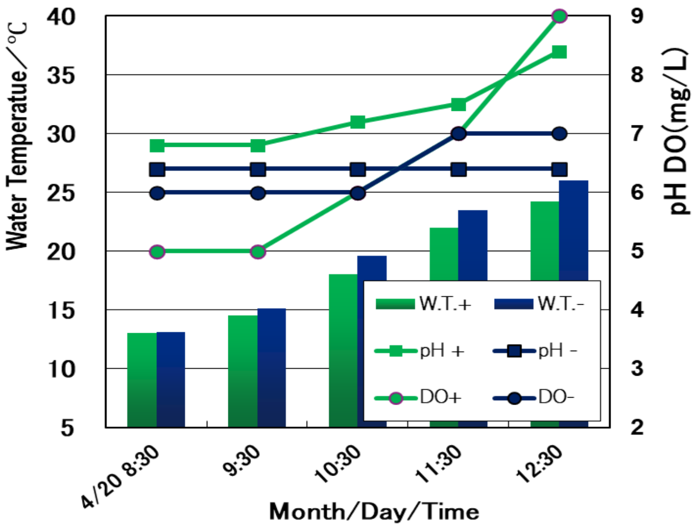


Fig. 7. The changes of water temperature, DO and pH of the water with plants (+) and water only (-).

As the water temperature rose, the pH+ (with plants) increased to 8.4 from 6.8 and also the DO+ increased to 9 from 5 remarkably, but the pH- (water only) did not change and the DO- increased a little to 7 from 6. These indicate that more active photosynthesis of the plants by higher water temperature decreased CO2 and increased O2 in the water. That raised both the pH+ and DO+. Science club members presented the findings by poster in Japanese at Kyoto University Academic Day 2019 (http://hdl.handle.net/2433/244419, in Japanese).

pH and RpH of the river water Tab. shows pH and RpH of the Yamashina River (Fig. 8.) two or three years ago. Reserved pH (RpH) [5] is defined as the pH at which the CO2 concentration of the water is in equilibrium with that of the atmosphere [6]. Science club members stirred [aerated] the river water in the container for more than 15 minutes with a SK-632PH pH Meter (measurement accuracy: +-0.4 pH, automatic temperature compensation function) shown in Fig. 9., and then measured the water pH (RpH).

Tab. pH and RpH of the Yamashina River.

Year	2022		2023		
Month/Day (Time)	8/16 (14:25)	11/22(14:25)	2/25(11:30)	3/28(10:45)	9/5 (8:15)
pH	9.55	8.75	8.48	8.81	8.78
RpH	8.01	8.01	8.05	8.22	8.25

Fig. 8. The Yamashina River in Kyoto Prefecture.

Fig. 9. A pH meter in a yogurt container attached to some cord.

The red shape shown in Tab. indicates that discrepancy between the pH and RpH values that exceeds about +-0.8 pH. In this case, the pH and RpH values are not the same. Because the pH meter has measurement accuracy: +-0.4 pH. There is a discrepancy between the pH and RpH values, shown above.

Biological fluctuations, such as the change of balance between photosynthesis and respiration in the river, cause the discrepancy [5] [6]. Besides, the water whose value of pH is higher than about 9 is weak alkaline. It may become an environmental risk not only for aquatic life but also for the human race. Science club members presented the new findings about pH and RpH of the Yodo River system by poster in English at the 10th NICE Conference 2025 Yamagata. They could improve their self-esteem and competences.

High school teaching practices

Water quality test from the "Acid and Base" unit of Grade 2's Basic Chemistry course Fig. 10. shows an experiment of high school chemistry in the 2000s. The students checked river water qualities by AquaChek ECO test strips. They were surprised at the alkaline river water and said, "Why is the river water alkaline?"

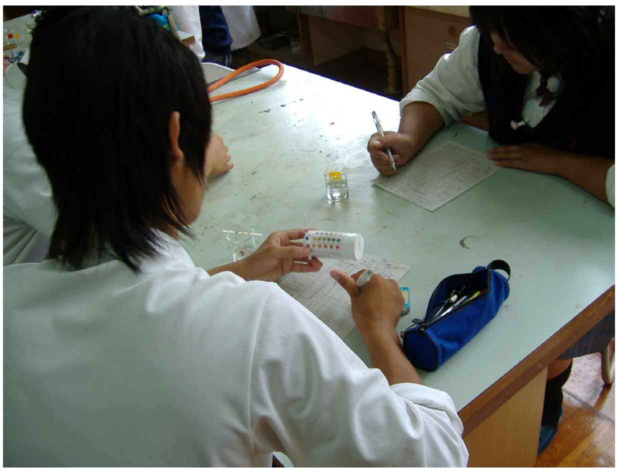


Fig. 10. A water quality test from the "Acid and Base" unit of Grade 2's Basic Chemistry course.

Ionization equilibrium of weak acid in Grade 3's Chemistry course In the unit of ionization equilibrium of weak acid and base of chemistry class, the author always showed the Fig. 4. and the students were surprised at the fact that the river water was sometimes alkaline. He asked, “Why is the river water alkaline?” They started to discuss with each other.

A few minutes later, he showed the Figs. 6.-7. and described the Schemata 1.-2. of the science club study mentioned before. The students were amazed at the pH change mechanism including ionization equilibrium of weak acid by photosynthesis. Most of them understood the mechanism with wonder and satisfaction.

Conclusion

In the past, the author mentioned “the water quality of rivers” as a teaching material [9] - [11]. This time, he has reviewed his practices relating to “the water quality of rivers”.

Considering the results of his practices described above, he concludes the following:
Firstly, the teaching material “pH changes of the river water” is valuable for STEAM education, because it contains many concepts in every STEAM domain.

Secondly, the material arouses high school students’ interest. This is because it is a curiosity for not only the students, but also the masses. People believe that river waters are neutral and they are sure that the waters are not alkaline or acidic.

Thirdly, the mechanism of the pH change including ionization equilibrium of weak acid by photosynthesis is amazing for the students. To learn “chemical equilibrium” in chemistry class, it is easy to understand the mechanism above. He saw many students understand the mechanism with wonder and satisfaction.

Acknowledgements The author wishes to thank Mr. A. Kirkham for his kind advice.

References

[1] Shiokawa, T.; et al., *Proceedings of 17th Japan Society for Science Education Annual Meeting*, 135-136 (1993, in Japanese).

[2] Shiokawa, T.; et al., *Proceedings of 18th Japan Society for Science Education Annual Meeting*, 81-82 (1994, in Japanese).

[3] Hirai, T.; et al., *ibid*., 83-84 (in Japanese).

[4] Hirai, T., *Kagaku to Kyoiku*, 43(1), 23-26 (1995, in Japanese).

[5] Yasutomi, R.; et al. *Uo to Mizu*, 49 (1), 13-22 (2012, in Japanese).

[6] Yamanouchi, T.; et al., *Hozenseitaigaku Kenkyu* (*Japanese J. Conserv. Ecol.*), 16, 169-179 (2011, in Japanese).

[7] Neal, C.; et al., 210/211, Sci. *Total. Environ.*, 205-217 (1998).

[8] Neal, C.; et al., 282-283, ibid., 205-231 (2002).

[9] Hirai, T., Chem. Edu. J., 17, No. 17-205, 1-9 (2017).

[10] Hirai, T., Proceedings of 8th NICE Conference, 28-31 (2019).

[11] Hirai, T., ibid, 176-178 (2019).

Investigating Sixth Graders' Mental Models of the Particle Model of Matter via a Modeling Curriculum on the Three States of Water

Chen-Yu Chen[1,2], Jing-Wen Lin[1*]

[1]*Department of Science Education, National Taipei University of Education, Taiwan*

[2]*Min-An Elementary School, New Taipei City, Taiwan*

jwlin@mail.ntue.edu.tw

Abstract

The particle model of matter (PMM), now included in Taiwan's elementary curriculum, is globally emphasized. However, research on student model development and conceptual difficulties remains limited. This study investigates how students develop and revise their PMM mental models and identifies where conceptual difficulties emerge during a structured curriculum on the three states of water. This design-based research with 36 sixth-graders implemented a curriculum structured into macroscopic, microscopic, and model transfer stages, each with observation and analogy substages. Expert science educators conducted interviews and collected student responses, which were analyzed using an eight-model framework by Chen & Lin [15]. Results showed that students primarily used Continuous, Descriptive, Mixed, and Basic Particle Models. In the macroscopic stage, 89% used Continuous Models, but this proportion dropped to 47% when analogies were introduced. In the microscopic stage, about half of the students demonstrated Mixed Models, with only a few reaching Basic Particle Models. In the transfer stage, 42% reverted to macroscopic reasoning, while 50% maintained Mixed Models and 17% demonstrated more advanced particle reasoning. Some students continued to show difficulties consolidating microscopic concepts, especially when applying models to novel contexts. Overall, analogy-based modeling tasks proved more effective than purely observational tasks. Findings highlight the importance of structured scaffolding and iterative practice in modeling curricula to deepen PMM understanding and to inform chemistry teachers on strategies that support scientifically accurate model development in the classroom.

Keywords: Analogy-based modeling, Particle model of matter, Design-based research, Chemistry education

1. Introduction

The Particle Model of Matter (PMM) is a key concept in chemistry education, as it explains the states of matter, phase changes, and material properties [1]. Learning PMM is not merely about memorizing facts—it also requires students to engage in scientific modeling to develop a deeper understanding of particle theory [2].

In recent years, many countries have introduced PMM into the elementary curriculum at earlier stages, recognizing its importance for science learning [3,4]. However, this shift also creates challenges. Research shows that students often form misconceptions. For example, even if they acquire particle-related terms, they may still perceive matter as a continuous whole rather than adopting a genuine particle perspective [5]. If particle ideas are introduced "too early" without sufficient support, students may become confused, lose confidence, or disengage from science learning [6].

In Taiwan's new curriculum guidelines, the content is explicitly stated: "Matter is composed of tiny particles, and these particles are in constant motion." Yet, many teachers still struggle to connect students' everyday reasoning with the progression toward abstract particle concepts [7]. This underscores the need for instructional scaffolds that align with students' developmental stages and systematically guide them toward more scientific models.

With this in mind, our study investigates how elementary students progress from macroscopic to microscopic particle concepts. We focus on how their mental models evolve during the process of model construction and how modeling-based instruction can support this learning. Our research questions are as follows:

1. What mental model of PMM do elementary school students hold regarding the three states of water?
2. How do elementary students' mental models of PMM change across six stages?
3. How does an analogical modeling curriculum influence students' model development, and what implications does this have for chemistry teaching practice?

2. Literature Review

2.1 Scope of PMM in Elementary School

The Particle Model of Matter (PMM) is central to modern science, and some even call it a "threshold concept" [2]. Many researchers have argued that PMM should be one of the core learning goals in K–12 science, which is why there has been a push to introduce it earlier in the curriculum [3].

Still, students frequently encounter misconceptions and difficulties when learning this concept. For example, Novick and colleagues [8] identified four basic particle properties that became the foundation for later studies. From a science education perspective, de Vos and colleagues [9] expanded this framework into eight key particle properties, directly connecting particle theory to phase changes, chemical reactions, and the structure of matter. In Taiwan's new elementary curriculum guidelines, the concept of atoms is deliberately omitted at this stage. To align with this approach, Lin [10] proposed eight simplified propositions for teaching PMM at the elementary level, providing a practical framework for teachers to design lessons on particle-related ideas.

2.2 How Students Learn PMM

Merritt and Krajcik [2] observed that while teachers often use analogies as tools for teaching PMM, they typically emphasize only the content of the analogy. What tends to be overlooked are the functions of models and the crucial role that modeling plays in helping students understand chemical behavior. For example, when students role-play atoms or molecules in different states, many mistakenly believe that liquid particles are spread out like in a gas rather than packed almost as closely as in a solid.

To address this, teachers need to recognize students' initial models and potential learning pathways. Instruction can then be designed using Learning Progression (LP) variables, so that students gradually move from descriptive models, through hybrid and basic models, toward a more complete particle model. This progression increases both the complexity and accuracy of students' scientific ideas, while also clarifying common points of confusion that teachers must address.

Other studies also suggest that instruction should be grounded in conceptual change theory [11] and supported by stronger teacher content knowledge of microscopic particle concepts [12]. In this context, analogy-based modeling provides valuable insight into how students actually learn PMM. Justi and Gilbert [12] and Mozzer and Justi [13] emphasized that analogies are not merely explanatory tools—they can also serve as creative entry points for modeling. Through analogies, students can engage in the full modeling cycle of generating, representing, testing, and revising models. This process helps them move from making surface-level comparisons of macroscopic events to constructing deeper connections that integrate both macroscopic and microscopic levels. They also learn to apply their models in new contexts while recognizing their limitations.

For students to learn PMM effectively, they require scaffolding across representational levels, explicit model-based language, and interactive tools or activities that encourage them to create and reflect on their own analogies. These strategies make abstract ideas more accessible and help students apply models more flexibly to novel problem situations.

However, most existing studies have only described general teaching strategies and outcomes. Few have examined the specific particle ideas that students generate in analogy-based modeling, or how well these align with what teachers and researchers expect. To address this gap, the present study designs a PMM teaching program centered on analogy-based modeling, focusing on how students' mental models and understanding develop across different learning stages, with the aim of providing practical and research-based insights for chemistry teachers.

3. Research Methods

3.1 Research Design and Participants

This study was part of a larger research project that initially examined teachers' instructional analyses [15]. While teacher-level data were collected, the present paper focuses exclusively on students' model development.

The participants were thirty-six sixth graders from an elementary school in Taipei City. The activities were conducted in small groups of six students with one teacher, and a total of six teachers participated. To ensure consistency across instruction, we organized preparatory interviews and demonstration lessons before the study began. Student selection included a mix of high-, medium-, and low-achievers, but ability-level differences were not the focus of this analysis.

3.2 Instruction Design

Before formal instruction, a short warm-up activity (Figure 1) was designed to introduce the idea of analogy. For example, students imagined ink diffusing in water as "fairies running around." Sample analogical mappings included:

- water particles → blindfolded fairies
- ink particles → fairies in ink-colored clothes
- temperature → a joyful atmosphere

This playful task offered students an accessible entry point into analogy, while also laying a foundation for subsequent lessons on the particle model.

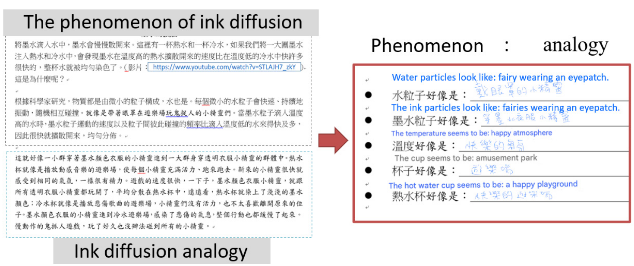


Figure 1. Preliminary activity design

Our formal lesson design followed Lin's analogy-based modeling framework [10], which emphasizes observation and analogy activities for the Particle Model of Matter (PMM). The curriculum focused on phase changes of water and was organized into three levels: macroscopic (Ma), microscopic (Mi), and transfer (Tr). Each level included two parts: phenomenon explanation (P) and analogy construction (A). The specific tasks and example items for each stage are summarized in Table 1.

Table 1. Framework of the analogy-based modeling curriculum on PMM.

Stage (Code)		Description	Example Item
Macroscopic models (Ma)	Phenomenon (P) Explaining	Explaining the Phenomenon of water's three states.	Why does ice gradually melt into water and then gradually evaporate into steam when heated by an alcohol lamp?
	Analogical modeling (A)	Construct analogical models to explain the three-state change of water during macroscopic observations.	Please design an analogy to explain the three-state changes of water.
Microscopic models (Mi)	Phenomenon Explaining (P)	Construct analogical models to explain the three-state change of water during microscopic observations by PhET.	Please use this powerful magnifier to observe and compare the smallest 'components and structures' of ice, water, and water vapor in PhET (https://reurl.cc/7joALN). Use <, >, = to compare the distance between particles: Particles (ice state) □Particles (water state) □ Particles (water vapor state)
	Analogical modeling.	Construct microscopic analogical models of water's three states.	Please design an analogy to explain how the speed of movement, the distance, and the chaotic degree of arrangement among particles (ice), particles (water), and particles (water vapor), as observed by a powerful magnifier, can result in the change of the three states of water.
Transferring models (Tr)	Phenomenon Explaining (P)	Explaining that phenomenon in the video water vapor in a conical bottle becomes water vapor after heating, and that the air in the bottle expands when heated, causing the balloon to grow larger.	After heating to 200 degrees for a while, use the powerful magnifier again to observe inside the balloon. Predict the observed phenomena: Will the size of the balloon change? Will the water level change? Predict and draw on the right diagram. After making your predictions, observe PhET (https://reurl.cc/pZxZvl), does it match your prediction?
	Analogical modeling (A)	Use self-generated analogical model to explain phenomenon of thermal expansion.	Using the analogy you designed for the "three-state change of water" in Question 4, try to explain why the size of the balloon changed.

3.3 Data Collection

We also carried out semi-structured interviews with the participating teachers. During the interviews, teachers used paper worksheets to write down key points or sketch their ideas. Each interview lasted between about 45 minutes and an hour and a half, and we recorded both video and audio. We also took field notes at the same time.

Before analyzing the data, we transcribed all recordings word-for-word. For coding, we used a simple labeling system: each teacher (TS1–TS3, TE1–TE3), student (SL1&2, SM1&2,SH1&2), the context (macroscopic = Ma, microscopic = Mi, transfer = Tr), and the task type (phenomenon explanation = P, analogy modeling = A). For example, "TS1–SL1-MaP" refers to Teacher TS1's student SL1 response when explaining a macroscopic phenomenon.

3.4 Data Analysis

After we transcribed the interviews, we coded the teachers' responses using the categories listed in Table 2. To make sure the coding was reliable, two researchers first coded the data separately. Their results matched almost completely—about 96% agreement. For the few cases where there were differences, we discussed them together with the corresponding author until we reached the same conclusion. The rest of the data was then coded by the first author. To double-check, the same researcher reviewed the coding again one month later and found it fully consistent, with 100% agreement.

3.5 Coding framework

For our analysis, we adapted the coding framework from Merritt and Krajcik [2] and Chen and Lin [15]. We conceptualized students' models along a macroscopic–microscopic continuum. To match the analytical needs of the Results section, we employed an eight-category scheme that explicitly distinguishes the four subtypes— M-ic, M-c, B-d, B-m —in addition to C, D, qS, and S (see Table 2). This granularity preserves instructionally meaningful differences (e.g., inconsistency vs. coherence within mixed reasoning; spacing- vs. motion-dominant basic models) and supports actionable implications for chemistry teaching.

Coders prioritized the dominant explanatory resource (e.g., spacing/density vs. motion/disorder) and judged self-consistency across states within a single explanation or drawing (for M-ic vs. M-c). Ambiguities were resolved by reviewing the full response set for the task and discussing until consensus.

Table 2. Categories of PMM as described by Merritt and Krajcik [2].

Model Type (Code)	Concise Description	Typical indicators / coding cues (student utterances or drawings)
Continuous (C)	Matter as a continuous whole; no particles or empty space.	"Ice melts and becomes steam when heated." *(no mention of particles, spacing, motion, or empty space)*
Descriptive (D)	Mentions parts but stays macroscopic; each part keeps original properties.	"It breaks into pieces but each piece is still the same thing." / "It just gets smaller or disappears."

Model Type (Code)	Concise Description	Typical indicators / coding cues (student utterances or drawings)
Mixed–inconsistent (M-ic)	Combines macroscopic and microscopic ideas, but they contradict or shift across states.	"Water vapor has the smallest particles and ice has the largest." / "With heat, particles get denser but also farther apart."
Mixed–consistent (M-c)	Macroscopic→microscopic mapping is self-consistent but partial.	"Solid is like a team in formation; in steam the team becomes loose and disordered." / "Heating makes particles move faster and spread further." *(forces/energy often missing)*
Basic Particle—distance-focused (B-d)	Recognizes particles & empty space; emphasizes spacing/density.	"After heating, the average distance increases and the density decreases." / "Different states have different particle spacing."
Basic Particle—motion-focused (B-m)	Recognizes particles & empty space; emphasizes motion/disorder.	"With heating, particles move faster and become more chaotic." / "Steam has the most vigorous motion."
Quasi-scientific (qS)	Multiple scientific features present but imprecise (e.g., forces, vacuum).	"There's attraction between particles, but I'm not sure how distance changes it." / "There is vacuum, but I can't explain it clearly."
Scientific (S)	Integrates spacing, motion, energy, inter-particle interactions with macro properties; distinguishes substance-specific behaviors.	"When temperature rises, kinetic energy increases; effective attractive influence decreases relative to motion; average spacing increases—hence volume expands. Different substances differ due to structure/composition."

4. Results

4.1 Common Mental Models of PMM among Students

In students' responses, we identified several recurring models (see Table 2 for definitions). Typical examples include:

- **C (Continuum):** “When something hot touches ice, it melts; when water is heated, it turns into steam.” TE1-SL1-MaP
- **D (Descriptive):** “It’s one whole thing, but it gradually gets smaller and disappears—like something old that turns moldy.” TE1-SL1-MaA
- **M-ic (Mixed–inconsistent):** “Water vapor has the smallest particles, ice has the largest.” TS2-SH1-MiP
- **M-c (Mixed–consistent):** “Ice is like an organized team that stays in place. When it becomes steam, the team becomes messy, so shape and volume can’t stay fixed.” TS3-SM2-MiA
- **B-d (Basic–distance):** “After heating, the particles are farther apart, so density decreases.” TS2-SH2-TrP
- **B-m (Basic–motion):** “Particles move faster, spread further apart, and become more disordered.” TS1-SM2-TrP

Overall, most students’ explanations stayed at the macroscopic level, while only a few managed to consistently use microscopic particle ideas. Notably, no student responses reached the **quasi-scientific (qS)** or **scientific (S)** levels, indicating that their conceptual understanding had not yet progressed to higher-level mental models. This highlights the challenge of helping elementary students construct more advanced and coherent particle-based explanations.

4.2 Changes in PMM Mental Models across Six Stages

The results are summarized in Figure 2, which combines the classification of mental models from Merritt and Krajcik [2] with our six-stage teaching sequence. The models are arranged hierarchically: darker shading indicates closer alignment with the scientific particle model, while lighter shading shows more basic reasoning.

Across the sequence, several clear patterns emerged. In the macroscopic observation stage (Ma-P), most students relied on the Continuum Model (C). With analogy tasks (Ma-A), some began shifting toward Mixed models, showing the first attempts to bring in microscopic ideas. In the microscopic stages (Mi-P and Mi-A), the Mixed-consistent Model (M-c) became dominant, and a few students advanced to Basic models, beginning to describe particle motion and spacing. In the transfer stages (Tr-P and Tr-A), some students regressed to C models, but many maintained M-c, with only a small number extending to Basic models.

Overall, the developmental trend shows a stepwise movement from macroscopic (C) to microscopic reasoning (M-c), with M-c acting as a key transition point. Only a few students moved further toward particle-level explanations of density or motion, and almost none reached more advanced models involving inter-particle forces. This suggests that stronger scaffolding is needed to help students consolidate M-c and extend into higher-level particle models.

	Macroscopic Models (Ma)		Microscopic Models (Mi)		Transferring Models (Tr)	
	P	A	P	A	P	A
TS1-SL1	C	D	C	M-ch	D	M-ch
TS1-SL2	C	M-ich	B-m	M-ch	C	M-ch
TS1-SM1	C	M-ch	M-ch	M-ch	M-ch	M-ch
TS1-SM2	C	M-ich	B-m	M-ich	M-ch	C
TS1-SH1	C	C	M-ch	B-d	B-d	B-d
TS1-SH2	C	M-ich	B-m	C	M-ich	B-m
TS2-SL1	C	D	M-ch	M-ch	M-ch	B-m
TS2-SL2	C	C	M-ich	M-ich	M-ch	M-ch
TS2-SM1	C	C	M-ich	M-ich	M-ich	M-ich
TS2-SM2	C	M-ich	M-ich	M-ich	M-ich	M-ich
TS2-SH1	M-ich	C	M-ich	M-ch	M-ich	M-ch
TS2-SH2	C	C	B-d	B-d	B-d	B-d
TS3-SL1	C	M-ch	M-ch	M-ch	C	M-ch
TS3-SL2	C	X	M-ch	M-ch	M-ch	M-ch
TS3-SM1	C	C	M-ch	M-ch	M-ch	M-ch
TS3-SM2	C	C	M-ch	M-ch	M-ch	M-ch
TS3-SH1	C	C	M-ch	M-ch	M-ch	B-m
TS3-SH2	C	M-ch	M-ch	M-ch	C	M-ch
TE1-SL1	C	D	C	D	C	C
TE1-SL2	C	X	C	X	C	M-ch
TE1-SM1	C	X	X	X	C	X
TE1-SM2	C	C	D	D	C	C
TE1-SH1	C	C	M-ch	M-ch	M-ich	M-ch
TE1-SH2	C	X	C	M-ch	M-ich	M-ch
TE2-SL1	C	C	B-m	M-ch	M-ch	M-ch
TE2-SL2	C	C	X	C	X	X
TE2-SM1	D	C	C	X	D	X
TE2-SM2	C	C	B-m	M-ch	C	X
TE2-SH1	C	C	C	M-ch	C	M-ch
TE2-SH2	C	X	C	X	C	X
TE3-SL1	C	X	D	X	C	X
TE3-SL2	C	X	C	X	C	X
TE3-SM1	C	C	M-ch	M-ch	C	M-ch
TE3-SM2	C	C	D	M-ich	C	M-ich
TE3-SH1	M-ch	B-d	B-m	B-d	M-ch	B-m
TE3-SH2	M-ich	X	M-ch	M-ch	C	M-ch

Figure 2. Changes in students' PMM across six instructional stages.

Note: Each row represents one student's responses across the six stages of instruction (Ma-P, Ma-A, Mi-P, Mi-A, Tr-P, Tr-A). Codes indicate model types (C = Continuum, D = Descriptive, M-ic = Mixed-inconsistent, M-c = Mixed-consistent, B-d = Basic–density, B-m = Basic–motion). Shading intensity reflects proximity to the scientific particle model: darker shading indicates more advanced particle reasoning; lighter shading indicates more basic macroscopic reasoning.

4.3 The Impact of Analogy-Based Modeling Instruction on Students' Model Development

Looking at the six-stage teaching sequence (Table 3), students' explanations began primarily in the Continuum Model (C), focusing only on macroscopic descriptions. With the introduction of analogy-based modeling, however, many gradually shifted toward the Mixed-consistent Model (M-c), where particle ideas were more logically integrated. Still, students' reasoning was not yet fully stable, showing that particle concepts remained "under construction." Even so, analogy structures encouraged them to connect visible phenomena with particle-level explanations.

In the transfer stages, some students reverted to macroscopic reasoning, reminding us that without cross-context practice and stronger scaffolding, they may easily return to familiar ways of thinking when facing new problems.

Table 3. Distribution of mental model development.

	Macroscopic models(Ma)		**Microscopic models(Ma)**		**Transferring models**	
	P	**A**	**P**	**A**	**P**	**A**
X	0%	22%	6%	17%	3%	19%
C	89%	47%	22%	6%	42%	8%
D	3%	8%	8%	6%	6%	0%
M-ic	6%	11%	11%	14%	17%	8%
M-c	3%	8%	33%	50%	28%	47%
B-d	0%	0%	17%	0%	0%	11%
B-m	0%	2%	3%	8%	6%	6%

In terms of the eight core propositions (see Table 3), students showed progress in several areas:

- Proposition 1 (matter is composed of particles): Most students recognized that matter is made of invisible, discrete units.
- Propositions 2.1 and 2.2 (random motion and kinetic energy): Many began to describe particles as moving randomly and faster when heated.
- Propositions 3.1 and 3.2 (spacing and vacuum): Some progress was observed, as students noted that particles could be closer or farther apart, though the idea of empty space remained fragile.
- Propositions 4.1–4.3 (interactions and temperature effects): A small number of students connected temperature changes with changes in particle motion and interactions.

However, deeper concepts such as inter-particle forces and emergent properties remained out of reach. This is partly due to their abstract nature, and partly because our curriculum design placed limited emphasis on this aspect. Future teaching should therefore provide stronger scaffolds—helping students not only build particle-level explanations but also understand how particle interactions and structures shape macroscopic properties.

5. Conclusions and Implications

Our findings show that the analogy-based modeling curriculum supported elementary students in moving from the macroscopic Continuum Model (C) toward the Mixed-consistent Model (M-c), where microscopic ideas were more logically integrated. This shift was likely driven by the alternation between observation and analogy, combined with scaffolding from microscopic simulations and teacher guidance. The M-c model became relatively stable across lessons, indicating its role as a key transitional stage.

That said, only a small number of students advanced to more sophisticated particle models—such as explaining inter-particle forces, attraction–distance relationships, or kinetic energy changes with temperature. These abstract concepts remain highly challenging for elementary learners. In addition, during transfer stages some students regressed to macroscopic reasoning (C model), suggesting that microscopic ideas had not yet been fully internalized as flexible knowledge. Their understanding of spatial structures such as vacuum and particle spacing was also fragile.

As Chen and Lin [15] noted, many teachers themselves possess only a basic grasp of particle modeling. Our results confirm that limited teacher knowledge and modeling language constrain how far students' models can develop. This highlights the central role of teachers in advancing students' conceptual growth.

Implications for practice are clear: teacher professional development should include targeted preparation on deeper aspects of the PMM (e.g., inter-particle interactions, spatial structures, emergent properties), along with concrete strategies for scaffolding and analogy use. Integrating such training with analogy-based modeling curricula can help students not only consolidate microscopic reasoning in class but also apply particle models flexibly in new contexts—laying the groundwork for long-term development in scientific thinking.

References

[1] Ayas, A.; Özmen, H.; Çalik, M. Students' Conceptions of the Particulate Nature of Matter at Secondary and Tertiary Level. Int. J. Sci. Math. Educ. 2009, 8, 165–184. https://doi.org/10.1007/s10763-009-9167-x.

[2] Merritt, J.; Krajcik, J. Learning Progression Developed to Support Students in Building a Particle Model of Matter. In Concepts of Matter in Science Education; Tsaparlis, G., Sevian, H., Eds.; Springer: Dordrecht, 2013; pp 11–45. https://doi.org/10.1007/978-94-007-5914-5_2.

[3] National Research Council. Next Generation Science Standards: For States, by States; National Academies Press: Washington, DC, 2013.

[4] Sim, B.; Yoon, H. International Comparison of National Elementary Science Curriculum and Science Textbook on Introduction of Particulate Concept. J. Korean Elem. Sci. Educ. 2018, 37 (2), 147–160.

[5] Novick, S.; Nussbaum, J. High School Students' Understanding of the Particulate Nature of Matter: An Interview Study. Sci. Educ. 1981, 65 (2), 187–196. https://doi.org/10.1002/sce.3730650209.

[6] Lin, J.-W. Analogy-Based Modeling of the Particle Model of Matter as the Core of Science and Bilingual Science Teaching and Learning (Project No. NSTC 112-2410-H-152-011-MY3); National Science and Technology Council: Taipei, 2024.

[7] Wu, W.-L.; Huang, M.-Z. A Study on Inquiry-Based Experiments and Analogy-Based Modeling of the Particle Model of Matter in Junior High School Biology Classes; National Science and Technology Council: Taipei, 2024.

[8] Novick, S.; Nussbaum, J. Junior High School Pupils' Understanding of the Particulate Nature of Matter: An Interview Study. Sci. Educ. 1978, 62 (3), 273–281.

[9] de Vos, W.; Verdonk, A. H. The Particulate Nature of Matter in Science Education and in Science. J. Res. Sci. Teach. 1996, 33 (6), 657–664. https://doi.org/10.1002/(SICI)1098-2736(199608)33:6<657::AID-TEA4>3.0.CO;2-N.

[10] Lin, J. W. Science and Bilingual Science Teaching and Learning Centered on Analogical Modeling of Particle Models of Matter (Report No. NSTC 112-2410-H-152-011-MY3); National Science and Technology Council: Taipei, Taiwan, 2023.

[11] Karataş, F. Ö.; Ünal, S.; Durland, G.; Bodner, G. What Do We Know about Students' Beliefs? Changes in Students' Conceptions of the Particulate Nature of Matter from Pre-Instruction to College. In Concepts of Matter in Science Education; Tsaparlis, G., Sevian, H., Eds.; Springer: Dordrecht, 2013; pp 231–247.

[12] Justi, R.; Gilbert, J. K. Science Teachers' Knowledge about and Attitudes towards the Use of Models and Modelling in Learning Science. Int. J. Sci. Educ. 2002, 24 (12), 1273–1292.

[13] Mozzer, N. B.; Justi, R. Science Teachers' Analogical Reasoning. Res. Sci. Educ. 2013, 43, 1689–1713. https://doi.org/10.1007/s11165-012-9328-8.

[14] Özmen, H.; Ayas, A.; Coştu, B. Determination of the Science Student Teachers' Understanding Level and Misunderstandings about the Particulate Nature of the Matter. Educ. Sci. Theory Pract. 2002, 2, 506–529.

[15] Chen, C. Y.; Lin, J. W. Unveiling Elementary School Teachers' Mental Models: Utilizing the Particulate Nature of Matter to Explain Water's Three States and Constructing Analogical Models for Their Students; Oral presentation at 27th IUPAC International Conference on Chemistry Education, Pattaya, Thailand, July 15–19, 2024.

Exploring the Feasibility of AI-Based Analysis of Elementary Chemistry Science Fair Reports in Taiwan

Chao-Min Hu [1*], Chin-Cheng Chou [1]

[1] *Department of Science Education, National Taipei University of Education, Taipei 106, Taiwan*

jimmyhu19908047@gmail.com

Abstract

Science fairs in Taiwan serve as crucial platforms for fostering students' scientific literacy, inquiry competence, and creativity. However, teachers often struggle to guide elementary students in developing feasible yet innovative research projects due to limited time and ambiguous evaluation criteria. This study explores the feasibility of using ChatGPT-4o—a large language model—to analyze and predict award-winning reports in the elementary chemistry division of Taiwan's National Science Fair. A total of 38 reports from 2021 to 2024 were analyzed, with 29 reports (2021–2023) used for model training and 9 reports (2024) for prediction testing. The model successfully identified all three top-award reports from 2024, achieving 100% prediction accuracy. Repeated trials indicated high test–retest consistency under identical prompts. The results suggest that constructing a feature-based evaluation model enhances the reliability and interpretability of AI-assisted assessments. This stu dy highlights the potential of AI tools to complement human judgment in science fair evaluation, offering new directions for integrating AI into inquiry-based science education and assessment practices.

Keywords: AI-assisted assessment, ChatGPT, elementary chemistry reports, inquiry-based learning, science fair.

Introduction

1.1 Background of Science Fairs in Taiwan

Since their establishment in 1955, Taiwan's science fairs have become essential mechanisms for promoting students' inquiry and problem-solving abilities [1,2]. The fairs follow a three-tier progression system—school, municipal, and national levels—creating a structured pathway for project advancement [3]. The elementary division currently covers eight disciplinary categories [4], with top-performing projects often linked to academic recognition and secondary school admissions. These competitions therefore serve both educational and selection functions.

Despite their long history and institutional support, disparities in mentoring quality and resource allocation persist. Teachers often face difficulties providing consistent guidance, especially when assisting students with limited disciplinary knowledge but strong creative potential. Consequently, establishing more objective and transparent evaluation mechanisms has become essential to ensure fairness and improve instructional alignment.

1.2 Challenges in Guiding Elementary Student Research

Guiding elementary students through independent research projects is both pedagogically demanding and time-intensive [5,6]. Students at this level often exhibit high imagination but limited understanding of experimental control, data interpretation, or logical structure. Teachers must therefore strike a delicate balance between offering necessary scaffolding and maintaining student ownership of the work—a dilemma reinforced by national regulations requiring students to complete and defend their projects independently [5].

Drawing from Vygotsky's Zone of Proximal Development [7,8], appropriate scaffolding can help students transcend their current competencies. However, interdisciplinary projects—such as those combining chemistry, physics, and environmental science—often exceed a single teacher's expertise, highlighting the need for collaborative or technology-assisted approaches [6].

1.3 AI as a Collaborative Assistant in Science Education

Artificial intelligence (AI) has recently emerged as a potential "second mentor" in education [9,10]. Large language models (LLMs) can support teachers by accelerating literature reviews, suggesting research ideas, and simulating rubric-aligned feedback [11]. Prior studies have shown that AI can identify learning gaps, provide formative feedback, and enhance scoring consistency [9,10]. Within science education, AI-assisted evaluation could reduce teachers' workload while maintaining fairness and transparency in assessment.

Recent evidence also supports the concept of human–AI collaboration. According to Shneiderman's Human-Centered AI (HCAI) framework [12], AI should be designed to augment rather than replace human capacities, emphasizing safety, explain ability, and empowerment. In both educational and industrial contexts, AI-augmented teams have demonstrated improved performance, creativity, and collaboration quality [13,14].

The integration of AI into science fairs presents an opportunity to bridge the gap between student creativity and systematic evaluation. By analyzing past award-winning projects, AI can help identify latent patterns and features associated with successful reports—insights that teachers can use to design instructional scaffolds and guide students' project development [9,11].

1.4 Research Purpose and Significance

This study investigates whether an AI model, specifically ChatGPT-4o, can accurately identify and predict award-winning reports in Taiwan's National Elementary Science Fair (Chemistry Division). The research aims to:

A. Develop a feature-based AI model derived from historical science fair reports.

B. Evaluate the model's predictive accuracy and stability.

C. Explore the implications of AI-assisted assessment for supporting teacher decision-making in science education.

By examining the feasibility of AI-based report evaluation, this study contributes to emerging discussions on AI-supported inquiry assessment, providing empirical evidence for how AI can function as a formative evaluation assistant in science learning environments.

2. Methodology

This study used OpenAI's GPT-4o, released on 13 May 2024, which operates at approximately twice the speed of GPT-4 at roughly 50% lower cost, and remains accessible to free-tier users with limited quotas [15,16]. According to official OpenAI documentation, free users are limited to a small number of messages within a rolling five-hour window [16], Plus subscribers can send approximately 80 messages per three-hour window [17], and Team/Pro workspaces have higher caps [14]. Selecting GPT-4o therefore enables secondary school teachers to replicate the analytic pipeline without subscription fees, aligning with the practitioner-oriented objectives of this study.

To assess output stability, each prompt was re-run three times between June and July 2025 under identical parameter settings (e.g., temperature, random seed), and the results were compared for consistency. This procedure follows the test–retest reliability framework proposed by Mondal et al., who reported a Pearson correlation coefficient of $r = .71$ for ChatGPT-3.5 in statistical-test recommendation tasks, and extends this approach to GPT-4o [18]. Upon completion of all report processing, GPT-4o re-examined and summarized the main writing features of award-winning reports based on the entire dataset.

While this study primarily focuses on developing and analyzing an AI-assisted evaluation model, the model itself was intentionally designed to be applicable across different subject domains rather than restricted to a specific discipline. By training the model on datasets containing student work from multiple subjects, we aimed to identify common evaluative features that transcend disciplinary boundaries. Interestingly, the extracted features showed strong correlations with domain-specific patterns, and models trained on diverse data often yielded more consistent and accurate scoring outcomes than those based solely on standardized rubrics. These findings suggest that cross-disciplinary AI models have the potential to provide teachers with data-informed insights for evaluating and guiding student learning across subjects [19,20].

Step 1: Data Collection and Preparation

The data for this study were drawn from publicly available chemistry projects in the Taiwan National Science Fair, which annually announces approximately 10–15 reports, all of which are submitted by winners of local competitions in each county and city. In Taiwan's ranking system, awards are granted to the first, second, and third place winners, followed by honorable mentions, referred to locally as Merit Awards. From this pool, we collected 38 chemistry reports from 2021 to 2024, as shown in Table 1. The sample included the top three national award-winning reports for each year (if multiple reports shared the same rank, all were included) as well as non-awarded projects. Merit Awards were excluded because they represent an intermediate ranking between winners and non-winners, which could introduce classification ambiguity in system analysis.

To confirm feasibility, the model was trained on reports from 2021–2023 and tested on those from 2024. Because the ChatGPT-4o version used in this study (June–July 2025) allowed a maximum of ten files to be uploaded per run, the analysis was conducted in batches of up to ten reports. This design consisted of three independent rounds, with awarded and non-awarded works compared within the same year. During data preparation, cover pages and any content revealing award status were removed to ensure complete blinding. In addition, persistent conversation history was disabled during analysis to avoid influence from prior data. This design minimizes evaluation bias and ensures replicability for future research.

Table 1 2021-2024 National Chemistry Exhibition Reports

Number of entries(Number of selected entries)	Top three(First, second, and third place)	honorable mentions	Non-awarded	Total selection
2021	3 (randomly selected from 5)	4 (Not used)	7 (randomly selected from 11)	10
2022	3(randomly selected from 5)	3 (Not used)	7 (all 7 Selected)	10
2023	3 (randomly selected from 4)	3 (Not used)	6(all 6 Selected)	9
2024	3(randomly selected from 4)	3 (Not used)	6(all 6 Selected)	9

Step 2: Model Training

Reports were entered into ChatGPT-4o for analysis of their strengths and weaknesses (e.g., research motivation, methodology, data processing, and writing quality). To improve output quality, the prompts were iteratively refined within the same session.

Because the ChatGPT-4o version used in the experimental tests (June–July 2025) allowed a maximum of ten files to be uploaded per run, the training was conducted in three cumulative rounds:

Round 1: Training with up to 10 reports from 2021 only, analyzed according to the official National Science Fair scoring criteria (research motivation, method design, data processing, and report writing).

Round 2: Training with up to 10 reports from 2021 and 2022, incorporating the 2022 data into the previous round and establishing a revised feature set of award-winning standards.

Round 3: Training with up to 10 reports from 2021, 2022, and 2023, further extending the dataset to refine and validate the feature model.

For each round, the system was asked to:

1. Compare the advantages and disadvantages of award-winning reports and non-award-winning reports from the 2021–2023 in the following aspects, as specified in the official evaluation standards:

A. Research motivation

B. Method design

C. Data processing

D. Report writing

2. Summarize the main writing features that appeared in the award-winning reports.

In Rounds 2 and 3, two additional system prompts were added: (a) repeat the results of the previous round before analyzing the new set of reports, and (b) after completing all provided reports, re-identify and summarize the key writing features of the award-winning reports across the entire dataset.

Step 3: Building the Award-Winning Report Feature Model

Based on the results of the three training rounds, key writing features that consistently appeared in award-winning reports were identified. These features included strengths in research motivation, method design, data processing, and report writing.

The recurring characteristics were then synthesized to form a preliminary “award-winning report feature model." This model served as the reference framework for evaluating new reports in the subsequent prediction and validation stage.

Step 4: Prediction and Validation

Using the feature model constructed from the 2021–2023 reports, we conducted a prediction test on the award-winning reports from the 2024 National Chemistry Fair (a total of nine reports).

This step aimed to determine whether the award-winning report feature model could accurately identify winning projects in a 2024 dataset, while also assessing the reliability of the system under repeated testing conditions.

The direct human evaluation of students' science fair reports was conducted strictly in accordance with the official scoring criteria announced by the National Taiwan Science Education Center [5]. In contrast, the AI model was independently developed by refining the official rubric into four analytical dimensions—scientific reasoning, experimental design, data interpretation, and presentation clarity—derived from the core components of the original framework. Because the written reports did not include presentation-related elements, the modeling process focused solely on the content-based aspects of evaluation. This approach enabled the model to preserve alignment with the official standards while enhancing its analytical precision and generalizability.

3.Results

3.1 Prediction Accuracy

Using science fair reports from 2021–2023 to build a predictive model yielded stronger results for forecasting 2024 award-winning projects. To ensure blinding, the cover pages and acknowledgments of award-winning reports were removed before analysis. Due to system constraints (maximum of 10 reports per run), the training process was conducted year by year. In each round, the model was re-instructed to analyze all available reports according to the four official evaluation criteria: (A) research motivation, (B) method design, (C) data processing, and (D) report writing. The evaluation of students' science fair reports was conducted strictly in accordance with the official scoring criteria announced by the National Taiwan Science Education Center [5]. Building on this framework, the model further refined the criteria into four analytical dimensions to enhance the consistency and interpretability of AI-based evaluation, in line with established principles of performance-based and formative assessment [19,20].

Table 2 summarizes the predictive accuracy under three evaluation conditions, all conducted in June 2025. When using only the official evaluation rubric (Test 1), the model correctly identified 2 out of 3 award-winning reports (success rate: 67%). In contrast, when predictions were based on the round 3 trained model (Test 2), accuracy improved to 100% (3/3). Finally, when the round 3 model was re-established and reapplied (Test 3), the system again achieved 100% success (3/3). Across all tests, the total number of reports considered remained nine.

Table 2. AI Prediction Accuracy Across Different Evaluation Conditions(June 2025)

Evaluation Condition– All conducted in June 2025	Predicted 2024 award-winning reports (Success Rate)	Total Reports
Test 1 – Using official rubric only	2/3 (67%)	9
Test 2 – Using round 3 trained model	3/3 (100%)	9
Test 3 – Re-establishing round 3 model	3/3 (100%)	9

3.2 Recurrent Features of Award-Winning Reports

In addition to higher predictive accuracy, the third round of training revealed recurring features that consistently characterized award-winning reports:

1. Creative questions inspired by daily life – Projects often originated from real-world problems or observations.
2. Well-planned experiments with clear variables and creative tools – Designs included control and test groups, frequently using self-made devices or innovative methods.
3. Strong data analysis with clear explanations and real-life applications – Data were presented transparently, results explained logically, and applications linked to daily life.
4. Clear reports with strong writing and visuals – Reports were well-organized, easy to read, and supported with charts or pictures.
5. Creative integration of ideas from multiple fields – Knowledge from different subjects was combined to generate novel insights or propose improvements.

Overall, these results indicate that while the official evaluation rubric provides a useful baseline for prediction, its accuracy was limited when applied directly. In contrast, the trained AI model was able to capture additional latent features and patterns from prior years' reports that were not explicitly represented in the rubric. The iterative training process allowed the system to refine its recognition of nuanced characteristics—such as innovative approaches, depth of analysis, and clarity of presentation—that often distinguish award-winning projects. This highlights the potential of AI-assisted evaluation to complement traditional rubrics by detecting subtle, multidimensional qualities beyond explicit scoring guidelines.

4. Discussion and Conclusion

4.1 Discussion

This study demonstrates that constructing a **feature-based evaluation model** enhances AI performance in recognizing and predicting award-winning science fair reports. Compared with direct application of the official rubric, the trained model achieved higher accuracy and reliability. The iterative process allowed the AI to detect nuanced elements—such as coherence, creativity, and contextual reasoning—that are difficult to quantify but crucial in authentic student research.

Repetition tests indicated that GPT-4o maintained a high level of response consistency within three identical prompt iterations, aligning with prior findings on LLM stability [18]. Beyond this threshold, slight variations emerged due to stochastic generation processes. These fluctuations underscore the need for clear and detailed prompts when employing AI for evaluative purposes.

A limitation observed in this study was the AI's occasional overreliance on visual proxies (layout consistency, image quality) rather than genuine visual interpretation. Because GPT-4o primarily processed text, its inference of "clear visuals" likely resulted from statistical associations rather than semantic understanding. Future models integrating vision–language architectures may better capture multimodal aspects of report quality.

4.2 Educational Implications

For educators, the findings suggest that AI can serve as a diagnostic and formative assessment assistant rather than a grading authority. By identifying characteristic features of exemplary reports, teachers can design more effective scaffolding for student research projects. The AI-generated feature summaries may also assist in professional development by illustrating how high-quality reports demonstrate creativity, logical reasoning, and coherence.

4.3 Limitations and Future Work

Although the feature model performed consistently within the tested dataset, generalization remains limited by the small sample size and subject specificity. Expanding the corpus to include other disciplines (e.g., biology, physics) and multimodal data (e.g., oral defenses, presentation slides) could improve robustness. Additionally, longitudinal studies could examine how AI-supported feedback influences students' inquiry competence and teachers' assessment literacy.

4.4 Conclusion

This research provides empirical evidence that AI-based evaluation, when structured around feature extraction and iterative refinement, can effectively identify the characteristics of award-winning science fair reports. The study confirms the feasibility and educational potential of AI-assisted assessment in elementary science education. While human judgment remains indispensable, integrating AI as a complementary tool can enhance fairness, reduce teacher workload, and promote sustained engagement in inquiry-based learning.

Reference

[1] Wu, C.-S., J. Natl. Taiwan Normal Univ.: Educ., 55(2), 1–34 (2010).

[2] National Research Council, A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas, National Academies Press, Washington, DC (2012).

[3] Ministry of Education, National Primary and Secondary School Science Fair Implementation Guidelines, Ministry of Education, Taipei, Taiwan (2022).

[4] Ministry of Education (Taiwan), Guidelines for the National Primary and Secondary School Science Fair (revised on Jan 16, 2025), Ministry of Education, Taipei (2025).

[5] Bogden, M.; Wilkerson, N., "A Teacher's Guide to Science Fair," Vivify STEM Blog, Houston, TX (2023).

[6] Edutopia, "Connecting Across Disciplines in Project-Based Learning," Edutopia, San Rafael, CA (2024).

[7] Simply Psychology, "Zone of Proximal Development," Simply Psychology (2025).

[8] Vygotsky, L. S., Mind in Society: The Development of Higher Psychological Processes, Harvard University Press, Cambridge, MA (1978).

[9] Holmes, W.; Bialik, M.; Fadel, C., Artificial Intelligence in Education: Promises and Implications for Teaching and Learning, Center for Curriculum Redesign, Boston, MA (2019).

[10] Luckin, R.; Holmes, W.; Griffiths, M.; Forcier, L. B., Intelligence Unleashed: An Argument for AI in Education, Pearson, London, UK (2016).

[11] Monteith, B., "AI: The Future of Personalized Mentorship in Science Fairs," Medium, San Francisco, CA (2024).

[12] Shneiderman, B., Human-Centered AI, Oxford University Press, New York, NY (2022).

[13] Kong, X.; Fang, H.; Chen, W.; Xiao, J., Humanit. Soc. Sci. Commun., 12, 821 (2025).

[14] OpenAI, "What is the Message Cap on ChatGPT Team?" OpenAI Help Center, San Francisco, CA (2024).

[15] Buchanan, N., "Microsoft-Backed OpenAI Unveils Most Capable AI Model, GPT-4o," Investopedia, New York, NY (2024).

[16] OpenAI, “ChatGPT Free Tier — Usage Limits,” OpenAI Help Center, San Francisco, CA (2025).
[17] OpenAI, “What is ChatGPT Plus? — Usage Limits,” OpenAI Help Center, San Francisco, CA (2025).
[18] Mondal, H.; Mondal, S.; Mittal, P., Perspect. Clin. Res., 15(4), 178–182 (2024).
[19] Brookhart, S. M., How to Create and Use Rubrics for Formative Assessment and Grading, ASCD, Alexandria, VA (2013).
[20] Wiggins, G., Educative Assessment: Designing Assessments to Inform and Improve Student Performance, Jossey-Bass, San Francisco, CA (1998).

Designing and Developing Science Tabletop Games: Engaging Senior High School Students as Designers to Foster Design Thinking Skills

JONG Jingping

New Taipei Municipal Jinhe High School, New Taipei City, Taiwan

porphyrin@jhsh.ntpc.edu.tw

Abstract

Design thinking is a human-centered, cross-disciplinary approach to problem-solving that emphasizes an iterative process for developing effective solutions. This study investigated a school-based course that integrated design thinking into science tabletop game development, aiming to cultivate 11th-grade students' design thinking skills and support them in creating their own science tabletop games. Using a five-stage framework—empathy, define, ideate, prototype, and test—students took on the role of designers, incorporating key chemistry concepts, along with biology, physics, and earth science into engaging game mechanics. The students' design thinking skills, encompassing the skills required across the stages of design thinking, were assessed using a semi-structured questionnaire. Results from a paired-sample t-test revealed significant improvements across all categories. Students described the course as challenging and rewarding, reporting growth in creativity, teamwork, communication, and problem-solving. Furthermore, 95% of participants recommended the course to their peers. The teacher facilitated the design thinking process by linking activities to students' tabletop game experiences, designing stage-specific tasks, and providing worksheets for guidance and progress tracking. These findings highlight the potential of integrating design thinking into science tabletop game design to foster innovation and create more engaging, interactive, and cross-disciplinary learning experiences.

Keywords: design thinking, tabletop games, design thinking skills, students as designers

Introduction

Student learning often relies on established thinking models, which serve as a foundation for analogical reasoning in similar situations. Design thinking is a practical process embodying a creative thinking model based on divergent and convergent thinking (Stanford d. School, 2010). In this process, designers confront problems from a human-centered perspective, developing final products or proposing effective solutions through a cycle of prototyping and iteration.

A tabletop game is a genre of game played on a flat surface, such as a table or board, typically involving physical tokens or cards. Science tabletop games are designed around scientific concepts or topics, aiming not only for entertainment but also to facilitate the acquisition of scientific knowledge through their core design and mechanics. Research indicates that such games can provide students with alternative pathways for engaging in science learning (Cheng et al., 2019; Ladachart et al., 2022).

However, the existing literature has predominantly focused on the outcomes of playing these games, such as the impact on students' learning of specific concepts and their motivation (Cardinot & Fairfield, 2019; Lin et al., 2019). Less attention has been paid to positioning students as active agents in the game design process itself. Granting students the opportunity to participate directly in the planning and creation of science games holds the potential not only to cultivate their design thinking skills but also to foster their creative abilities.

Therefore, the purpose of this study is to investigate the process of empowering students to become the primary designers of their own edutainment science tabletop games. By engaging in this authentic design experience, students can navigate a complete problem-solving cycle and, in doing so, manifest their creativity.

Literature Review and Theoretical Framework

Design thinking and related research

Design thinking is a human-centered problem-solving strategy that emphasizes a developmental process of empathy, collaboration between team members and designers, experimentation, and iterative refinement. It involves understanding and defining the problem, generating ideas, prototyping solutions, and then testing and improving them based on user feedback. The goal is to develop innovative and effective solutions that meet the needs of users (Stanford d. school, 2010; Tu et al., 2018; Yu et al., 2024).

A primary focus of design thinking is the gradual establishment of effective problem-solving strategies through a process of divergent and convergent thinking. Many researcher have developed diverse strategies to help learners master the design thinking process (Sung & Kelly, 2019; Stanford d. school, 2010). Sung and Kelly (2019) identified that design thinking includes steps such as analyzing the situation, defining the problem, modeling ideas, designing prototypes, predicting outcomes, questioning unexpected results, and managing the design process. Zhu et al. (2025) implemented a six-stage design thinking model to improve elementary students' innovation skills through two science projects. The process guided students through discover, focus, imagine, prototype, try, and reflect & share stages, moving from initial knowledge acquisition to hands-on creation and evaluation. The results found that students' design thinking skills, particularly after the second intervention, increased significantly.

The Stanford d. school proposed a five-step, non-linear process for design thinking that includes: empathize, define, ideate, prototype, and test (Stanford d. school, 2010). These steps are described as follows:

- **Empathize**: Focus on understanding the user's needs, which can be achieved through methods like questionnaires and direct observation.
- **Define**: Synthesize the information gathered during the empathy phase to articulate a problem statement and confirm the user's core needs.
- **Ideate**: Brainstorm innovative solutions or product concepts from various perspectives.
- **Prototype**: Begin to implement the proposed solutions or create tangible prototype products based on the ideas generated.
- **Test**: Have users test the solutions or prototypes to gather feedback for improvement and refinement.

At its core, design thinking emphasizes a holistic mindset and balances three essential elements: desirability (creating a suitable product for the user), viability (ensuring the product can be profitable or sustainable), and feasibility (confirming that the innovative idea can be realized with current technology). Design thinking is not only applied in general product design and science education (Ladachart et al., 2022) but also in developing strategies for sustainable development (Shapira et al., 2015). Many studies have applied d. school model into innovation of product. Samadhiya and Agrawal (2022) applied the d. school's five-phase design thinking framework to innovate a traditional handloom used by weavers in India, addressing key usability and productivity issues. The research demonstrated how the design thinking process could be effectively implemented for a traditional product. The findings indicate that the resulting redesigned loom was successfully adopted by the weavers, and the study confirm that design thinking is a valuable methodology for technological intervention and fostering stakeholder collaboration.

Science tabletop games and related research

Science tabletop games are defined as tabletop games that incorporate scientific concepts or their mechanisms. Their purpose extends beyond entertainment, aiming for players to understand the relevant scientific content through gameplay (Cheng et al., 2019). For example, numerous commercially available tabletop games are related to scientific concepts like evolution. Research shows that players can construct a foundational understanding of concepts related to evolution while learning the game's rules (Eterovic & Santos, 2013).

The design of science tabletop games is similar to that of general tabletop games, primarily consisting of a thematic background, rules and procedures, physical components, and other secondary elements. The thematic background explains the game's narrative, making it easier for players to immerse themselves in a specific context. The rules and procedures include elements such as initial setup, use of objects, turn sequence, phase mechanics, and victory conditions. Physical components are the tangible objects that players interact with, which may include cards, a game board, chance or event cards, or dice (Cheng et al., 2019).

Research indicates that through participation in science tabletop games, students engage in immersive interaction in a gamified manner, which not only promotes their learning of scientific concepts but also enhances their interest and motivation (Cardinot & Fairfield, 2019; Lin et al., 2019; Jong et al., 2017; Peppler et al., 2013). Cheng et al. designed a tabletop game about water resource management where, through the game's mechanics and player interaction, players' attitudes toward resource management shifted from an initial profit-oriented perspective to one focused on the public interest. Throughout the process, players not only enhanced their knowledge of water resource management but also developed a better understanding of systems thinking, environmental responsibility, and public value (Cheng et al., 2019). Existing research has explored the use of design thinking to have adolescent participants design board games, through which they learn relevant scientific concepts and systems thinking (e.g., Parekh et al., 2021). Parekh et al. engaged teens in a workshop to design a board game about environmental issues. The research found that this hands-on process served as a powerful educational tool. By creating a game, students were prompted to model complex ecological interactions, which deepened their understanding of environmental importance and nurtured the development of their systems thinking skills (Parekh et al., 2021). However, these studies have primarily focused on the participants' acquisition of scientific concepts and motivation, with less emphasis on how to cultivate students into becoming proactive designers in order to develop their design thinking skills.

Research framework

The design thinking process includes the non-linear steps of empathize, define, ideate, prototype, and test (Stanford d. school, 2010). Through these steps, students can begin to devise solutions for specific problems. It is anticipated that after engaging in this process, students will gradually internalize these steps, forming a cognitive model that constitutes their design thinking skills.

Educators believe that learners can acquire design thinking skills with the assistance of scaffolding (Razzouk & Shute, 2012; Stanford d. school, 2010). However, despite its widespread use, design thinking skills still lacks a single, unified definition (Aflatoony et al., 2018; Razzouk & Shute, 2012). Razzouk and

Shute (2012) proposed a hierarchical model that presents design thinking skills across dimensions such as use resources, iterate diagrams, and innovative design. Specific indicators within this model include identifying needs and goals, identifying resources and generating arguments, breaking down systems or creating innovative models, and testing, refining, and evaluating models to make decisions. Stempfle and Badke-Schaube (2002) suggested that the fundamental cognitive elements of design thinking are similar to general problem-solving skills, comprising generate, explore, compare, and select. Generation and exploration expand the problem space to create innovative solutions, while comparison and selection narrow the problem space to test and refine those solutions. Design thinking skills encompasses setting cognitive goals, considering user needs, generating innovative ideas, designing and implementing models, and continuously refining and evaluating those models.

In light of this, the present study, referencing the Stanford d. school's stages of design thinking (Stanford d. school, 2010) to conceptualize design thinking skills. These skills are defined as the learner's skills to complete effective and innovative solutions through the design thinking process. It is broken down into the following components: empathizing with others, defining the problem, ideating solutions, creating prototypes, and testing. This involves assessing how learners empathize with user needs, how they define a solvable problem from a wide range of user needs, how they use divergent and convergent thinking to find an innovative yet appropriate solution, how they then create a tangible prototype to visualize their ideas, and finally, how they test the effectiveness of the final product. This framework also serves as the basis for assessing students' design thinking skills (see Figure 1).

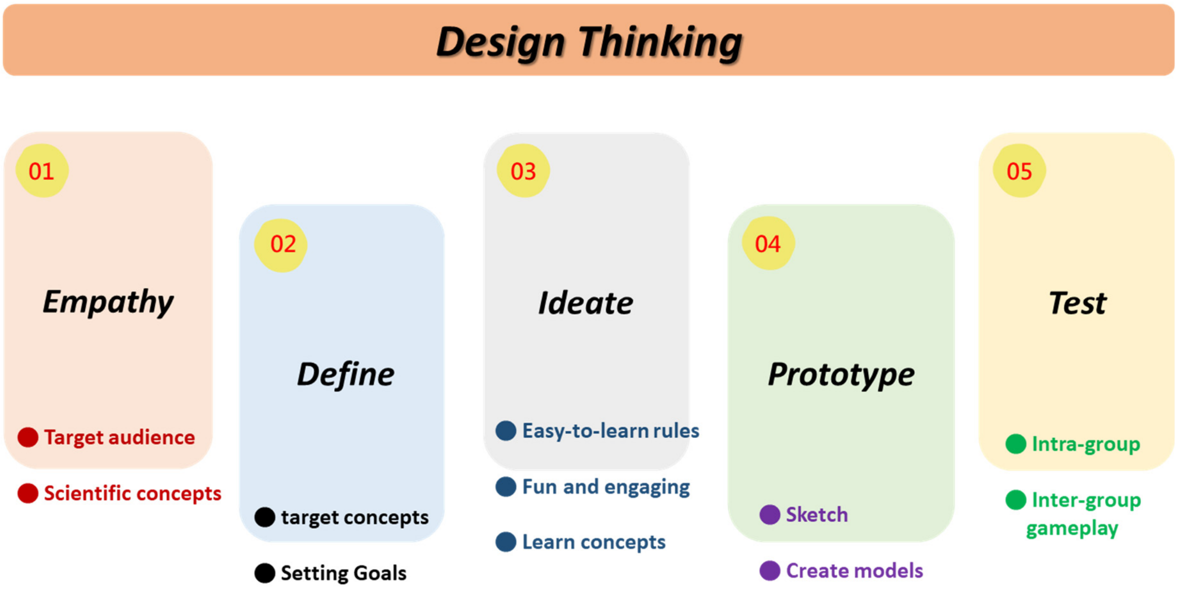


Figure 1. The theoretical framework of design thinking
(Adapted from Standford d. school, 2010)

The ideation phase of design thinking typically begins with divergent thinking to broadly gather ideas, followed by evaluation and review to form convergent thinking. Through this continuous process of refining ideas, innovative solutions are conceived. This process of divergence and convergence is precisely where creativity is manifested. In contrast, previous research has primarily focused on teacher-led initiatives, where educators design science tabletop games to facilitate student learning (e.g., Cheng et al., 2019). This course, however, positions students as the central agents, guided by the instructor to become the main designers. The goal is for them to experience the process of design thinking, and through continuous thinking, planning, designing, and creating, to produce a science tabletop game that elementary, junior, or senior high school students would be willing to engage with. By experiencing the steps of design thinking, students can cultivate their design thinking skills and thereby enhance their creative abilities.

Based on the literature review and discussion above, the primary purpose of this study is to develop a teaching module for science tabletop game design. The module emphasizes how to promote students' learning of design thinking elements within the design thinking framework, with the aim of cultivating their design thinking skills. The research questions are as follows:

1. What are the differences of 11th -grade students' design thinking skills before and after the teaching intervention?
2. How do 11th -grade students refine their designed science tabletop games?
3. After the instruction, what are the students' perceived gains from and evaluations of the course content?

Methods

Participants and Context

This study employed a single-group, pretest-posttest design to investigate the development of design thinking skills in 79 eleventh-grade students from two classes (aged from 16-17) in New Taipei City. The intervention consisted of a teaching module that spanned 18 weeks, with two 50-minute sessions per week. Adopting the five-stage framework of empathize, define, ideate, prototype, and test, students assumed the role of designers. Their task was to create engaging tabletop games by incorporating key concepts from chemistry, biology, physics, and earth science into the game mechanics.

Instructional design and instruments

The primary instrument for this study was a pre-test and post-test questionnaire designed to measure students' design thinking skills. Student-generated data, such as design diagrams, presentation videos, physical artifacts, and written feedback, were collected as supporting evidence.

Design thinking skills questionnaire and the rubrics

A questionnaire assessing design thinking skills was developed based on the design thinking framework. The questionnaire uses a specific product or service as a central theme to measure students' familiarity with and proficiency in the design thinking framework before and after the instruction.

The questionnaire's dimensions align with the five stages of design thinking: empathize, define, ideate, prototype, and test. To ensure the instrument's quality, a two-stage validation process was conducted. First, the questionnaire was reviewed by three 11th-grade students from a non-participating class to ensure item clarity, with revisions made according to their feedback. Subsequently, the instrument was validated for content by two experts in science education. The complete questionnaire is provided in the Appendix.

A corresponding rubrics was developed to assess student performance across the previously mentioned five dimensions (Table 1). Each dimension is evaluated on three performance levels (0-2). For example, within the define dimension, level 0 indicates a non-response or an irrelevant answer; level 1 reflects the ability to describe one way of identifying the problem and its potential challenges; and level 2 reflects the ability to describe two or more ways. The rubrics were also validated by two experts in science education and revised based on their suggestions. To establish inter-rater reliability, the primary researcher and a second rater independently scored a sample of four questionnaires. After discussing and resolving initial discrepancies, they scored the remaining questionnaires, achieving an inter-rater agreement of 0.94.

Table1 The rubrics of design thinking skills

	Level 0	Level 1	Level 2
Empathy	No response or irrelevant answer	Identifies one user-related need connected to the product but lacks depth or clarity	Clearly identifies two or more user-related needs connected to the product, demonstrating an understanding of user perspectives.
Define	No response or irrelevant answer	Mentions the need for a new product but provides limited or unclear description of relevant constraints.	Clearly and thoroughly defines the problem, including its constraints and considerations based on user characteristics.
Ideate	No response or irrelevant answer	Briefly explains the brainstorming and narrowing process but does not clearly connect it to the defined problem.	Provides a detailed explanation of the brainstorming and idea-narrowing process, directly addressing the defined problem and user needs.
Prototype	No response or irrelevant answer	Sketches a simple diagram but does not highlight specific functional parts or explain how they meet user needs.	Creates a detailed diagram, identifies specific functional parts, and provides clear explanations of how they meet user needs and solve the defined problem.
Test	No response or irrelevant answer	Mentions the testing and revision process but fails to connect it to the problem definition or user feedback.	Clearly and comprehensively explains the testing and revision process, incorporating insights from problem definition and user feedback to improve the product.

Instructional design

The instructional design is primarily based on the design thinking process, guiding students to develop a science tabletop game. The instructional process began with a one-week introductory phase where the teacher outlined the learning objectives and required deliverables while explaining the core concepts of design thinking. Following this, students moved into a three-week exploration phase to broaden their perspectives by playing various existing science tabletop games, with the objective of identifying different game mechanics. Subsequently, they worked in groups to discuss, brainstorm, and formulate an initial design plan. The development stage then commenced with a two-week period where each group presented their initial design direction, an interactive session that allowed the teacher and peers to provide crucial feedback for refinement. After incorporating this feedback, groups spent the next two weeks in further discussion and revision, during which the teacher provided timely technical assistance and monitored their progress. Once their plans were revised, students conducted a second presentation over another two weeks, allowing for a further round of peer and teacher refinement. With a well-refined concept in hand, students then entered a four-week creation phase to visualize their ideas and build their initial prototypes. This was followed by a two-week period of internal playtesting, where groups self-evaluated their game's flow and mechanics, identified areas for improvement, and made necessary revisions. Finally, the process culminated in a two-week showcase where the groups presented their completed products using a world café format, allowing students from different groups to rotate, play, and score each other's games.

Data collection

The data collected in this study were categorized into two types: qualitative and quantitative. The qualitative data included students' in-class design thinking questionnaires and the feedback questionnaires. The quantitative data were primarily derived by scoring the students' qualitative responses according to a scoring rubric, and these scores were then analyzed using a paired samples t-test on the pre-test and post-test results. Additionally, materials such as in-class tabletop game design worksheets, student presentation slides, and classroom video recordings served as supplementary data to provide supporting evidence for the teaching process.

Results

The instructional intervention significantly improved students' design thinking skills

To understand the differences in students' design thinking skills before and after the instructional intervention, the researcher scored student responses on a scale of 0, 1, or 2 according to the scoring rubric. A paired samples t-test was then conducted on the total score (maximum of 24) and on the individual constructs: empathize (max 4), define (max 6), ideate (max 4), prototype (max 6), and test (max 4). The results are summarized in Table 2. The findings indicate that there were statistically significant differences between the pre-test and post-test scores for the total score ($p = .000 < .05$), empathize ($p = .000 < .05$), define ($p = .001 < .05$), ideate ($p = .000 < .05$), prototype ($p = .000 < .05$), and test ($p = .000 < .05$).

Table 2. Paired samples t-test for design thinking skills (N = 79)

Item	Test	Mean (SD)	95% CI [Lower, Upper]	t-value	*p*	Cohen's *d*
Total Score	Pre	14.90 (3.68)	3.00, 4.46	10.18	.000***	0.93
	Post	18.63 (4.35)				
Empathize	Pre	2.01 (0.91)	0.62, 1.16	6.52	.000***	0.90
	Post	2.90 (1.07)				
Define	Pre	3.73 (1.32)	0.19, 0.70	3.42	.001**	0.34
	Post	4.18 (1.33)				
Ideate	Pre	2.23 (1.00)	0.63, 1.09	7.40	.000***	0.84
	Post	3.09 (1.04)				
Prototype	Pre	4.48 (1.46)	0.46, 1.09	4.85	.000***	0.55
	Post	5.25 (1.32)				
Test	Pre	2.44 (0.94)	0.54, 1.00	6.62	.000***	0.83
	Post	3.21 (0.92)				

Note: *$p < .05$; **$p < .01$; ***$p < .001$

In the pre-test, students were generally able to reach level 1 in each dimension, indicating that the 11th-grade students possessed a foundational level of design thinking skills. Among the five constructs, their initial performance was highest in prototype, showing that students were capable of conceiving a product based on their own ideas, creating a simple sketch, and explaining its innovative features. The lowest-performing construct in the pre-test was empathize, where students either described only one way to empathize with the audience or provided an irrelevant answer about who they intended to empathize with. In the post-test, the average score for each of the five constructs was nearly at or above level 1.5. Prototype remained the highest-performing area, while empathize remained the lowest, although both showed a statistically significant improvement compared to the pre-test. These results suggest that the instructional intervention was effective in enhancing the design thinking skills of students who already had a basic foundation.

Guidance, feedback, and iteration drove the refinement of student game designs

The science tabletop game design course was structured according to the stages of design thinking. As illustrated in the process diagram (Figure 2), although the students had experience playing tabletop games, they were initially unclear about how to design and produce a finished product. Therefore, the process began with students playing and evaluating various commercial science games to identify their strengths and weaknesses. This activity helped establish a foundational understanding of what a science tabletop game entails and the key elements to consider during the design phase.

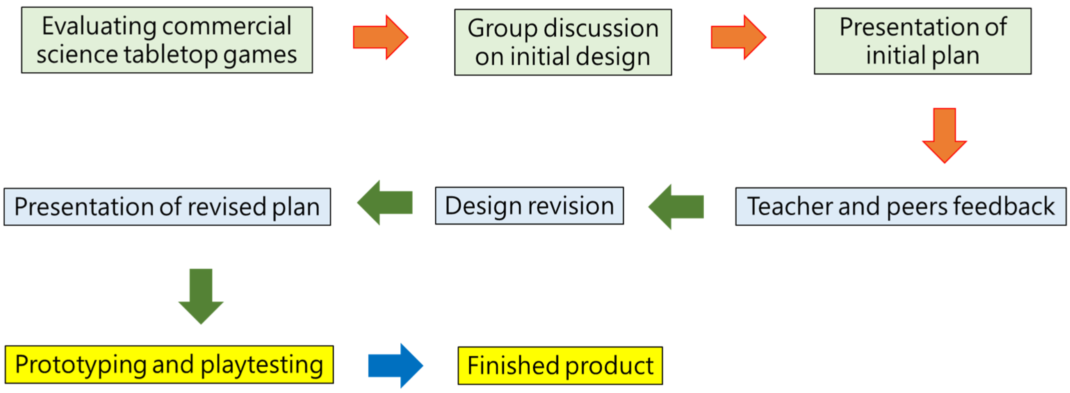


Figure 2. The process of students refining their designed science tabletop games

Following this, student groups began the design process, guided by worksheets and presentation prompts provided by the teacher. These prompts included guiding questions such as: "What is the core scientific concept?", "Why is this concept considered difficult to learn?", and "How can the game mechanics facilitate the learning of this concept?". This structured approach enabled students to develop a complete initial design.

During the initial presentations, peers provided feedback using structured forms, while the teacher also shared insights on the scientific concepts and game design to help students enhance their work. Subsequently, students revised their designs based on the feedback from the first presentation and then conducted a second round of sharing and revision. After two cycles of revision and confirming the integrity of their refined concepts, the groups began the physical production of their tabletop games. Finally, a world café format was used for the final showcase, where groups interacted with, playtested, and scored each other's games.

Students self-reported learning growth and evaluation for the course

Hardship, fun, and a sense of achievement Students described the game design process as a challenging yet ultimately rewarding journey that fostered a profound sense of achievement. The difficulty of creating a product from scratch was a common theme, with one student stating, "Making the tabletop game ourselves was super tiring" (A-25), reflecting the arduous process of translating scientific concepts into playable game mechanics. Despite this, the experience was also filled with enjoyment and collaborative spirit, as another student noted, "Although it was tough, we were full of laughter in the end" (B-03). This journey culminated in a strong sense of accomplishment, best summarized by a student who expressed, "I felt a huge sense of accomplishment after finishing. It was an unforgettable and rewarding memory" (A-33), highlighting the lasting, positive impact of seeing their creation enjoyed by others.

Teamwork The collaborative nature of the game design enhanced students' teamwork, teaching them to consider user needs while fostering internal team coordination. Students learned to approach the design from multiple standpoints, with one student noting that the process taught them "how to consider others' needs from different perspectives, how to make a team work together smoothly, and how to make our product better fit the theme" (A-17). This collective effort underscored the importance of coordination in joint problem-solving. As another student stated, "by cooperating and coordinating with my group members, we found the best solutions to problems, which also strengthened our team cohesion" (B-21). Ultimately, the project required students to work together closely to find solutions, thereby improving their skills to function as a cohesive and effective team.

Communication Students reported that the process highlighted the importance of communication, both for gameplay and for formal presentations. They learned that effective communication was essential for the game itself to function, with one student noting, "When explaining the rules and gameplay to the teacher or other groups, if they couldn't understand what you were saying, the game couldn't proceed. So, I think communication skills are very important" (B-21). Furthermore, students felt that the multiple presentation cycles enhanced their public speaking skills. As another student expressed, "Through so many presentations, I became better at maintaining my composure on stage...spoke clearly without stuttering, and could highlight the main points" (A-25). Thus, the course was perceived to build communication skills in both interpersonal and presentational contexts.

High recommendation rate and student satisfaction Student feedback questionnaires revealed that the vast majority of students would strongly recommend this course. The data shows that 95% of enrolled students (75 out of 79) would recommend that their junior peers take this course. Students were generally satisfied with the course content and learning outcomes, believing it had a positive impact on their learning and development. Students also rated the semester-long course on a scale of 1 to 10, with 10 being the highest. The average overall rating for the course was 8.01. This is a considerably high score, indicating a high level of overall student satisfaction. Among the ratings, two students gave a low score of 1. A closer look revealed their dissatisfaction stemmed from the large amount of time required for thinking and production.

Discussion

Fostering Interdisciplinary Skills Through Science Game Design

The development of a science tabletop game is a highly complex, interdisciplinary challenge. It requires students to set clear objectives, integrate relevant scientific concepts, and design engaging gameplay mechanics. Furthermore, the process involves iterative cycles of testing, refinement, and clear communication of rules to ensure the final product is both educational and appealing. This course provided students with the unique opportunity to navigate this entire product development cycle, offering them a meaningful and holistic learning experience.

To guide students through this complexity, the course was structured around the design thinking framework. The hands-on, collaborative process fostered a positive learning environment and, consistent with prior research (e.g., Tu et al., 2018), enhanced student interaction and communication. Students themselves reported marked improvements in teamwork and collaborative problem-solving as they worked to bring their ideas to fruition.

This study extends previous findings by demonstrating the unique potential of design thinking within a deeply integrated, product-development context. Rather than applying design thinking to a single academic discipline, the core task of creating a science game required a sophisticated synthesis of diverse skills. Therefore, this course serves as a practical model for a truly integrated, multidisciplinary learning experience, aligning with calls for design thinking to be taught as a comprehensive framework rather than a supplemental activity (Yu et al., 2024). The research provides a valuable case study on leveraging a complex design challenge to foster interdisciplinary collaboration and innovation.

Cultivating design thinking skills through science game design

Design thinking is fundamentally an interdisciplinary mindset. The process of developing products or innovative services requires the integration of cross-disciplinary skills, including empathizing with others' needs, reframing problems, continuously diverging and converging on ideas, visualizing concepts, and understanding that success is the result of continuous refinement. By using science tabletop game development as its core, this study allows students to experience the authentic process of product development.

The core objective of the present study is different from that of the research by Parekh et al. (2021), although both successfully utilize tabletop game design as an educational tool. Parekh et al. focused on game design as a medium to enhance understanding of a specific topic—environmental science—and to foster systems thinking. In contrast, the present study centers on cultivating the design thinking process itself, aiming to develop students into proactive innovators. Our research implemented and measured the formal five-stage design thinking framework within a school-based course, whereas the study by Parekh et al. used a workshop format.

Assessing design thinking skills for problem-solving, not product evaluation

This course utilizes the design thinking framework to cultivate students to create a science tabletop game. The process guides students to understand player needs, brainstorm rules, integrate scientific concepts, and iteratively prototype and test their games. This practical framework also fosters key skills, including teamwork and communication.

A fundamental difference from the study by Zhu et al. (2025) lies not only in the project's complexity but also in the assessment methodology. The present study immersed senior high school students in a complex product development challenge, requiring the integration of chemistry, biology, physics, and earth science into game mechanics. In contrast, Zhu et al. engaged elementary students with more structured tasks. Critically, our assessment focused on the students' problem-solving thought processes during the design tasks. In contrast, Zhu et al. evaluated the final product's attributes using a detailed rating scale with dimensions like novelty, technicality, and aesthetic. This distinction highlights our study's emphasis on cultivating the cognitive skills of a designer, rather than solely grading the output.

Balancing Challenge with Appropriate Scaffolding

The findings also showed that some students gave negative comments because of a large time demand for thinking and production. Such comments are important for pedagogical reflection. This apparent “difficulty” is not necessarily a defect in the design of the course but rather a testament to its authenticity. It simulates the messiness and difficulty of real-world product development and constitutes a kind of “productive struggle,” which is an essential element for project-based or STEM learning (e.g., Bolyard, 2024). The amount of work the students had to put in to go through the whole design process from the beginning was almost as great as the sense of fulfillment they described upon completing it.

But it is also a cautionary note to instructors to more effectively scaffold, but not “dumb down”, the challenge of the class. In future terms, the course could address this by possibly dividing the project into a series of smaller bang-for-the-buck milestones and/or dedicating more structured in-class time for hands-on production to decrease after-class demands. In addition, instructors can frame this challenge proactively early in the course, to convey that such struggles are an expected and valuable part of the innovation process. This would help temper student expectations and promote greater resilience.

Conclusion

This study demonstrates the successful implementation of an interdisciplinary course where senior high school students designed and developed science tabletop games. By providing a creative learning environment, the course effectively guided students in applying the design thinking framework to a tangible product development cycle, significantly enhancing their design-related skills.

Following the five-stage design thinking model—empathize, define, ideate, prototype, and test—students engaged in a comprehensive, iterative process. This course began with an analysis of commercial tabletop games and initial brainstorming, progressed through multiple cycles of peer and teacher feedback, and culminated in the production and playtesting of a finished game. The quantitative results confirm the effectiveness of this approach, showing significant improvements in students' design thinking skills across all five dimensions from pre-test to post-test.

Beyond these metrics, qualitative data from student feedback revealed a rich, multifaceted learning experience. Students described the process as both challenging and highly rewarding, fostering a strong sense of achievement upon completion. They reported significant growth in crucial skills, including teamwork and communication. The overwhelmingly positive student evaluations, with 95% recommending the course to their junior peers, underscore its value and impact.

The primary implication of this research is that it offers a validated, practical framework for integrating design thinking into the science curriculum. This model serves as an effective pedagogical strategy for fostering design thinking skills and applying scientific knowledge in an engaging, project-based manner. For future research and practice, the course could be enhanced by incorporating modules on the commercial aspects of product development. For instance, introducing topics such as market analysis, intellectual property, and the patent application process would further empower students, providing them with a more complete understanding of how a scientific concept evolves into a market-ready product. This would not only deepen their learning but also better equip them for future innovation challenges.

Acknowledgments

This manuscript was prepared with the assistance of the generative AI tool, ChatGPT-4o, for translation, grammatical proofreading, and language refinement.

References

Aflatoony, L., Wakkary, R., & Neustaedter, C. (2018). Becoming a design thinker: Assessing the learning process of students in a secondary level design thinking course. *The International Journal of Art & Design Education, 37*(3), 438-453. https://doi.org/10.1111/jade.12139

Bolyard, J., Curtis, R., & Cairns, D. (2024). Learning to struggle: Supporting middle-grade teachers' understanding of productive struggle in STEM teaching and learning. *Canadian Journal of Science, Mathematics and Technology Education, 23*(4), 687–702. https://doi.org/10.1007/s42330-023-00302-0

Cardinot, A., & Fairfield, J. A. (2019). Game-based learning to engage students with physics and astronomy using a board game. *International Journal of Game-Based Learning, 9*(1), 42–57. https://doi.org/10.4018/IJGBL.2019010104

Cheng, P. H., Yeh, T. K., Tsai, J. C., Lin, C. R., & Chang, C. Y. (2019). Development of an issue-situation-based board game: A systemic learning environment for water resource adaptation education. *Sustainability, 11*(5), 1341. https://doi.org/10.3390/su11051341

Eterovic, A., Santos, C. M. D. (2013). Teaching the role of mutation in evolution by means of a board game. *Evolution: Education and Outreach, 6*, 22. https://doi.org/10.1186/1936-6434-6-22

Jong, J. P., Fang, J. J., Wei, C. C., & Lu, Y. T. (2017, July). A novel chemical table game: Devoting students to learn organic compounds. Paper presented at the 7th International Conference of Network for Inter-Asian Chemistry Educators (NICE), July 26-28, Seoul, Korea.

Ladachart, L., Radchanet, V., & Phothong, W. (2022). Design thinking mindsets facilitating students' learning of scientific concepts in design-based activities. *Journal of Turkey Science Education, 19*(1), 1-16. https://doi.org/10.36681/tused.2021.106

Lin, Y. L., Huang, S. W., & Chang, C. C. (2019). The impacts of a marine science board game on motivation, interest, and achievement in marine science learning. *Journal of Baltic Science Education, 18*(6), 907-923. https://doi.org/10.33225/jbse/19.18.907

Parekh, P., Gee, E., Tran, K., Aguilera, E., Pérez Cortés, L. E., Kessner, T., & Siyahhan, S. (2021). Board game design: An educational tool for understanding environmental issues. International Journal of Science Education, 43(13), 2148–2168. https://doi.org/10.1080/09500693.2021.1956701

Peppler, K.; Danish, J. A.; Phelps, D. (2013). Collaborative gaming: Teaching children about complex systems and collective behavior. *Simulation & Gaming*, *44*(5), 683–705. https://doi.org/10.1177/10468781135014

Razzouk, R., & Shute, V. (2012). What is design thinking and why is it important? *Review of Educational Research, 82*(3), 330-348. https://doi.org/10.3102/0034654312457429

Samadhiya, A., & Agrawal, R. (2022). Developing a handloom through d. schools design thinking approach. *Technology in Society, 71*, 102134. https://doi.org/10.1016/j.techsoc.2022.102134

Shapira, H, Ketchie, A., & Nehe, M. (2015). The integration of design thinking and strategic sustainable development. *Journal of Cleaner Production, 140*(1), 277-287. https://doi.org/10.1016/j.jclepro.2015.10.092

Stanford d. school. (2010). *An introduction to design thinking—Process guide.* Hasso Plattner Institute of

Design at Stanford University. https://web.stanford.edu/~mshanks/MichaelShanks/files/509554.pdf
Sung, E., & Kelly, T. R. (2019). Identifying design process patterns: A sequential analysis study of design thinking. *International Journal of Technology and Design Education, 29*(2), 283–302. https://doi.org/10.1007/s10798-018-9448-1
Tu, J. C., Liu, L. X., & Wu, K. Y. (2018). Study on the learning effectiveness of Stanford design thinking in integrated design education. *Sustainability, 10*(8), 2649. https://doi.org/10.3390/su10082649
Yu, Q., Yu, K., & Lin, R. (2024). A meta-analysis of the effects of design thinking on student learning. *Humanities and Social Sciences Communications, 11,* 742. https://doi.org/10.1057/s41599-024-03237-5
Zhu, L., Shu, L., Tian, P., Sun, D., & Luo, M. (2025). Facilitating students' design thinking skills in science class: An exploratory study. *International Journal of Science Education, 47*(1), 23-44. https://doi.org/10.1080/09500693.2024.2309658

Appendix

A group of passionate young individuals aims to design an innovative product or service. Please answer the following questions step by step.

(1) What is the planned "product or service to be developed"?
(2) Who are the intended "target users" for this product or service?
(3) Before designing, how will you "empathize" to identify the "characteristics of the target users"?
(4) Before designing, how will you ensure that the product meets the "needs of the target users"?
(5) Among the various needs of the target users, how will you "define the problem you want to solve"?
(6) During the design process, how will you ensure that the problem you aim to solve aligns with the "needs of the target users" rather than being based solely on your own imagination?
(7) During the design process, what kind of "difficulties or limitations" might you encounter?
(8) During the design process, what methods will you use to facilitate team brainstorming?
(9) During the design process, describe how you will "narrow down and refine your ideas."
(10) (a) Please roughly sketch a prototype of the initial product.
(b) Explain the specific functional components.
(c) How do the specific functional components meet the needs of the users?
(11) How will you "test and revise" the initial product to make the design more suitable for the target users' needs?
(12) How will you "respond to user feedback" to make the design more suitable for the target users' needs? Please provide examples.

Translanguaging as a Scaffold in Bilingual Science Classrooms: Supporting Students' Construction of the Particle Model of Matter and Language Learning

Jing-Yi Liu[1], Chen-Yu Chen[1,2], Chiu-Wen Wang[1,3], Jing-Wen Lin[1*]

[1]*Department of Science Education, National Taipei University of Education, Taiwan*

[2]*Min-An Elementary School, New Taipei City, Taiwan*

[3]*Yong-Shun Elementary School, Taoyuan City, Taiwan*

jwlin@mail.ntue.edu.tw

Abstract

In response to the dual challenges posed by Taiwan's Curriculum Guidelines of 12-Year Basic Education and the "Bilingual 2030" policy, this study designed a science curriculum integrating Model-Based Instruction and bilingual education, with "translanguaging strategies as a scaffold in the bilingual science classroom" as its core perspective. It explores how sixth-grade students employed translanguaging strategies when constructing the Particle Model of Matter (PMM) and how these strategies were distributed across different instructional designs. Participants were 40 sixth-grade students from an elementary school in Taipei City, matched by pretest performance and randomly assigned to four instructional contexts. A qualitative analysis identified five major strategy types: direct imitation; diagram-plus-short-explanation; self-selected language, gestures, or diagrams to convey ideas; switching between forms; and fully integrated multimodal explanations. Among these, "diagram-plus-short-explanation" was the most common, while fully integrated multimodal explanations were the least frequent. Bilingual groups exhibited a greater variety of expressions and more frequent mode-switching, particularly in the bilingual modeling group, where students addressed both conceptual and linguistic challenges and connected everyday experiences to microscopic explanations. The findings highlight the feasibility of combining modeling with bilingual instruction to support the learning of abstract scientific concepts. Integrating modeling with translanguaging scaffolds effectively links macroscopic phenomena to microscopic explanations, expands students' multilingual and multimodal expressive resources, and fosters progression from single representations to integrated explanations. This approach not only deepens students' conceptual understanding but also offers practical design directions for bilingual science instruction, transforming language use from mere vocabulary input into a key tool for supporting scientific reasoning and knowledge construction.

Keywords: Translanguaging Strategies, Bilingual Science Classroom, Model-Based Instruction, Particle Model of Matter

1. Introduction

In recent years, Taiwan has actively promoted the "Bilingual Nation 2030" policy to enhance citizens' English communication skills and global competitiveness. The Ministry of Education in Taiwan further encourages using English as the medium for certain subjects. In classroom practice, bilingual education is not only about language integration but also about helping students understand and express science concepts in ways that support disciplinary learning.

Taiwan's Curriculum Guidelines of 12-Year Basic Education first listed "constructing models" as a science learning performance for elementary students, extending microscopic concepts, such as "matter consists of tiny particles in constant motion," to upper grades [1]. For topics such as the Particle Model of Matter (PMM), this change has prompted the need for teaching designs that bridge students' everyday experiences and abstract chemical concepts.

In bilingual science classes, students must acquire scientific knowledge while switching languages, creating a dual burden. One practical approach for supporting such learning is to integrate bilingual explanations with multiple representations, so that students can link new scientific terms with visual and experiential references. When dealing with microscopic concepts, students often rely on language and complementary representations to reason with models. These multimodal transformations can help reduce language barriers and enhance students' ability to apply concepts in both science and language learning contexts [2].

This paper reports a classroom-based study in which model-based instruction and bilingual elements were integrated into teaching PMM. The focus here is on how the instructional designs were implemented, what kinds of multimodal and cross-language practices were observed, and what implications these offer for chemistry teaching in elementary settings.

2. Literature Review

2.1 Challenges in Students' Learning of the Particle Model of Matter

In Taiwan's Curriculum Guidelines of 12-Year Basic Education [1], the concept that matter is composed of particles was introduced at the elementary school level. However, this concept involves multiple dimensions: including particle volume, shape, motion, and interactions, whose highly abstract nature poses challenges for both students and teachers. Chen and Lin [3] noted that even experienced teachers often hold vague understandings of the particle-related concepts covered in the curriculum.

The PMM is a scientific concept at the microscopic level, which cannot be directly observed and is not easily constructed naturally from students' everyday experiences [4]. Lin [5] proposed eight core particle propositions to help clarify teaching priorities and strengthen connections among concepts. Chen and Lin [3] also pointed out the high potential of analogical modeling in PMM instruction. Through the mapping of familiar situations onto particle-level ideas, students can more easily connect prior experiences with new chemistry-related concepts.

2.2 Application of Translanguaging in Science Education

While model-based instruction supports conceptual understanding, students may still encounter difficulties when learning in a bilingual setting. In such contexts, teachers and students often combine languages, visuals, gestures, and other modes to make meaning. García and Lin [6] note that these practices help address language needs, clarify concepts, and enhance comprehension.

García and Li [7] define translanguaging as the flexible use of language resources, not limited to code-switching between languages but also including multimodal integration. Cenoz and Gorter [8] emphasize that it is learner-centered and promotes meaning-making. Lu and So [9] identified functions such as guiding inquiry, giving praise, and explaining content. Lemmi and Pérez [10] likewise recognize its supportive role in science learning. In chemistry teaching, such strategies can help students connect symbolic, particulate, and macroscopic representations.

Lemke [11] stressed that language is a key medium for constructing concepts. In bilingual science education, integrating purposeful translanguaging with visual and hands-on resources can create an environment where language use directly supports chemistry concept learning.

2.3 Multimodal Learning in Science Education

Multilingual and multimodal resources can promote students' participation in inquiry and facilitate scientific knowledge construction [9]. For PMM, this includes linking particle diagrams, animations, and everyday analogies to chemical phenomena. Kress et al. [12] pointed out that beyond verbal language, teachers use oral explanations, images, gestures, physical models, and writing to convey meaning.

Ainsworth [13] noted that multiple representations can deepen understanding if closely aligned with the learning task. For abstract concepts such as particle motion or energy transformation, multimodality can enhance motivation and comprehension [14]. Pierson et al. [15] pointed out that multimodal resources are essential for engaging in scientific practices, especially when abstract or microscopic concepts are involved. Bolger et al. [16] found that model construction itself is a multimodal learning process.

Xue and Sun [2] emphasized that models and modeling offer significant benefits in chemistry learning, playing important roles in describing phenomena and developing scientific knowledge. Analogies can reduce the difficulty of modeling instruction by connecting prior experiences to new chemical concepts, making microscopic processes more tangible for students.

3. Research Methods

3.1 Research Design and Participants

This study employed a qualitative classroom-based approach to examine how different instructional designs influenced students' participation and expression when learning the PMM, a topic that often challenges elementary learners due to its microscopic and abstract nature. The intervention was implemented across four class periods, with some groups experiencing model-based activities and others receiving more traditional instruction, and with language modes varying between bilingual and Chinese-only contexts.

In the first class, a pre-test was conducted to assess students' baseline PMM understanding, science-related language skills, and familiarity with modeling activities. The "tiny mover" analogy was introduced in a diffusion scenario, together with selected English vocabulary related to the PMM, as a bridge to later instructional activities.

The instructional activities were developed by crossing two factors: teaching strategy (Modeling vs. Traditional) and language mode (Bilingual vs. Chinese), producing four groups:

- Bilingual Analogical Modeling Group (BA): Bilingual instruction with modeling activities, guiding students to construct concepts through a variety of language and visual supports.
- Bilingual Traditional Group (BT): Bilingual instruction using inquiry and lecture-based methods.
- Chinese Analogical Modeling Group (CA): Modeling activities conducted entirely in Chinese.
- Chinese Traditional Group (CT): Traditional inquiry and lecture-based teaching in Chinese.

All lessons were delivered by the same trained team of teachers to ensure consistency. They followed lesson scripts, observed each other's instruction, and engaged in post-lesson reflections to review classroom implementation.

The participants were forty sixth-grade students from a public elementary school in Taipei with bilingual learning experience but no prior exposure to modeling instruction. These students were selected from a larger cohort based on comparable pre-test performance, ensuring similar starting points across the four groups.

3.2 Data Collection

Multiple sources of qualitative data were gathered to document classroom processes and student participation:

(1). Classroom Video and Audio Recordings: All lessons were recorded and transcribed to capture teacher–student interactions and the ways students expressed their ideas.
(2). Instructional Materials and Worksheets: Lesson materials and student work were collected for evidence of concept representation and language use.
(3). Classroom Observation Records: Notes were taken during lessons, aligned with the corresponding pages of instructional materials.
(4). Team Meetings: The research team met regularly to ensure lesson fidelity and reflect on classroom implementation.
These data were coded by lesson type (BA, BT, CA, CT) and linked to student identifiers for later analysis. For example, a student with the identifier 60119 in the Bilingual Modeling group was labeled as BA60119.

3.3 Data Analysis

This study adopted Codebook Thematic Analysis as the primary qualitative data analysis method. The thematic analysis combined predetermined analytical dimensions with inductive theme generation, aligning with analytical principles while facilitating collaborative operations and cross-checking among multiple researchers. In practice, the research team first conducted repeated readings and open coding of all students' written responses in the learning sheets and the verbatim transcripts of the recorded instructional process. Based on theoretical foundations and pilot studies, four preliminary analytical dimensions were established: analogy type, language style, modality integration, and regulation strategies. These four main codes formed the first-level coding framework. Subsequent steps involved defining detailed coding categories, identifying representative statements, constructing classification logic, and performing cross-checking to enhance the reliability and reproducibility of the analysis. Our coding categories are Analogy type, Language style, Mode integration, and Regulation strategies.

This coding framework provided the logical foundation for thematic analysis, supporting both the generation and comparison of themes, as well as serving as the basis for interpreting the relationship between translanguaging strategy types and students' learning performance. All modifications were documented in detail, with complete records of the analytical procedures and corresponding original excerpts to ensure transparency, validity, and reliability of interpretation [17].

To enhance coding consistency and reliability, the research team conducted training and example-based discussions for each category in the codebook. Two researchers independently coded a sample dataset from the Bilingual Traditional group, using the teaching material from the observation records as the primary unit of analysis, supplemented with video recordings, teaching materials, and students' worksheets. Coding was performed for each student, and the results were cross-checked. The Kappa values for all major codes exceeded .78, indicating high inter-coder agreement [18]. Any discrepancies were resolved through negotiation with a third researcher. This framework thus served as the basis for developing, reviewing, and refining preliminary themes.

4. Results

4.1. What translanguaging strategies did students use when learning the PMM?

When learning the PMM, students displayed a range of ways to combine language, visuals, and other resources to explain abstract and microscopic ideas. Five main patterns of expression were observed:

- ●Pattern 1: Direct imitation of the teacher's or peers' wording or diagrams.
- ●Pattern 2: Combining diagrams with brief written or spoken explanations to illustrate a phenomenon.
- ●Pattern 3: Actively using self-selected words, gestures, or diagrams to convey an idea.
- ●Pattern4: Switching between different forms of expression when encountering difficulty in explanation.
- ●Pattern 5: Integrating multiple modes and personal analogies to create a more comprehensive explanation.

In terms of frequency, the second pattern, combining diagrams with brief explanations, was the most common and accounted for 51% of all instances. It was followed by the third pattern, using self-selected language, gestures, or diagrams to convey ideas, at 32%, Pattern 1, direct imitation accounted for 12%, Pattern 4, switching between modes for 3%, and Pattern 5, integrating multiple modes into a personalized explanation, for 2%.

For example, in the Bilingual Traditional group, a student (BT60415) was asked, "*Do hot particles move faster than cold particles?*" The student wrote "*YES*" in English and drew two cups (Figure 1), one labeled "*hot*" and the other "*cold.*" Wavy lines were added to each cup to represent food coloring spreading in water, with shorter, denser lines in the cold cup and more spread-out lines in the hot cup. This visual-text combination clearly illustrated the faster and wider diffusion in hot water, showing how students could use a simple blend of scientific vocabulary and diagrams to convey a chemical concept.

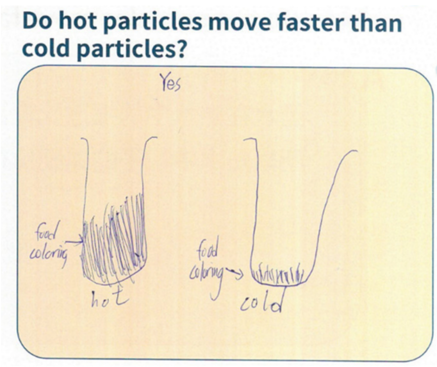


Figure1. BT60415's response in the instructional material

4.2 Under different instructional designs, what patterns do translanguaging strategies exhibit?

The four instructional designs appeared to create different opportunities for students to combine language, visuals, and other resources in explaining PMM concepts. The results were as Table 1. While all groups made use of diagrams and short explanations, the extent and variety of these combinations varied notably.

In the Bilingual Analogical Modeling group (BA), students produced the widest range of expression types, including direct use of scientific vocabulary, switching between Chinese and English, and drawing analogies from everyday experiences to illustrate particle behavior. This group also showed more instances where diagrams, gestures, and verbal explanations were integrated in the same response.

The Bilingual Traditional group (BT) also displayed variety, though most responses combined textbook diagrams with brief bilingual explanations. Occasional analogies and language switching were observed, but these were less frequent than in the Bilingual Modeling group.

In the Chinese Analogical Modeling group (CA), most responses involved diagrams plus short written explanations in Chinese, sometimes enriched with analogies related to everyday phenomena. The range of expression was narrower than in bilingual settings, but the visual explanations were often clear and accurate.

The Chinese Traditional group (CT) relied heavily on reproducing diagrams from the materials, with short labels or captions. This approach conveyed the basic concept but offered fewer opportunities for students to extend or personalize their explanations.

Overall, bilingual instruction appeared to encourage more varied forms of expression, while modeling activities, whether in bilingual or monolingual settings, tended to prompt students to create their own diagrams and connect them to explanations of particle behavior. The richest combinations of modes and language were observed when bilingual instruction and modeling were combined, suggesting that these elements together can create a more dynamic space for students to articulate their understanding of chemical concepts.

Table1. Distribution of translanguaging strategies under different instructional designs

Strategy	BA	BT	CA	CT
Pattern 1	★★	★★		
Pattern 2	★★	★★★	★★★★★★★★★★	★★★★★★★★★★
Pattern 3	★★★★★★	★★★★		
Pattern 4	☆	☆		
Pattern 5	☆			

Note: ★ indicates the proportion of translanguaging strategy use across the four instructional design groups.
☆ indicates a proportion below 10%.

5. Conclusions and Implications

This study examined how students in different instructional settings expressed their understanding of the PMM, focusing on the ways they combined language, visuals, and other resources. Across the four groups, five main patterns of expression were observed: direct imitation, diagram-plus-short-explanation, self-selected language, gestures, or diagrams to convey ideas, switching between forms, and fully integrated multimodal explanations. These patterns appeared with different frequencies—diagram-plus-short-explanation was by far the most common, while fully integrated multimodal explanations were rare.

The distribution of these patterns varied across instructional designs. Bilingual settings encouraged greater variety, with students more often switching between languages, drawing on analogies, and combining multiple modes. Modeling activities, whether conducted in bilingual or monolingual contexts, can encourage students to independently produce diagrams and relate them to explanations of particle behavior, sometimes integrating everyday experiences to enrich their microscopic-level interpretations. The richest and most flexible combinations occurred in the Bilingual Modeling group, where students navigated both conceptual and linguistic challenges.

From these findings, several teaching insights emerge:

Use modeling to anchor abstract concepts

Incorporating analogy-based modeling can prompt students to work with both visual and verbal representations, helping them connect macroscopic observations (e.g., diffusion in hot vs. cold water) to microscopic particle behavior.

Leverage bilingual elements to expand expressive resources

Alternating between Chinese and English in a supportive way can encourage students to reframe and refine their explanations, fostering more precise use of scientific terms.

Encourage movement across modes

Providing opportunities to draw, gesture, and write about chemical phenomena can deepen understanding, especially for concepts that cannot be directly observed.

Support gradual progression toward integrated explanations

Teachers can prompt students to move from reproducing diagrams or repeating terminology toward combining multiple resources into personalized, coherent explanations.

In chemistry teaching, particularly for topics like PMM, integrating modeling with bilingual scaffolding can create a learning environment where students actively construct both conceptual and linguistic understanding. Such integration shifts bilingual science instruction from simply adding English terms to strategically using language as a tool for deepening disciplinary learning.

References

[1] Ministry of Education in Taiwan: *Curriculum Guidelines of 12-Year Basic Education for Elementary School, Junior High and General Senior High Schools: The Domain of Natural Science;* 2018. https://cirn.moe.edu.tw/Upload/file/38227/104346.pdf

[2] Xue, S.; Sun, D.: Integrating Analogy into Scientific Modeling for Students' Active Learning in Chemistry Education. In *Active Learning: Research and Practice for STEAM and Social Sciences Education; Ortega-Sánchez, D., Ed.; IntechOpen: London,* 2022. https://doi.org/10.5772/intechopen.105454

[3] Chen, C. Y.; Lin, J. W.: Unveiling Elementary School Teachers' Mental Models: Utilizing the Particulate Nature of Matter to Explain Water's Three States and Constructing Analogical Models for Their Students. Presented at the 27th IUPAC International Conference on Chemistry Education, Pattaya, Thailand, July 15–19, 2024.

[4] Harrison, A. G.; Treagust, D. F.: Secondary Students' Mental Models of Atoms and Molecules: Implications for Teaching Chemistry. Sci. Educ. 1996, 80 (5), 509–534. https://doi.org/10.1002/(SICI)1098-237X(199609)80:5<509::AID-SCE2>3.0.CO;2-F

[5] Lin, J. W.: Developing Assessment and Instruction of Analogy-Based Modeling Competence to Explore Elementary School Students' Analogy-Based Modeling Competence on the Particle Model of Matter. National Science and Technology Council , 2023 (in Chinese).

[6] García, O.; Lin, A. M. Y.: Translanguaging in Bilingual Education. In *Bilingual and Multilingual Education, Encyclopedia of Language and Education*; García, O., Lin, A. M. Y., May, S., Eds.; Springer: Switzerland, 2016; pp 117–130.

[7] García, O.; Li, W.: *Translanguaging: Language, Bilingualism, and Education;* Palgrave MacMillan:

New York, 2014.

[8] Cenoz, J.; Gorter, D.: *Pedagogical Translanguaging;* Cambridge University Press: Cambridge, 2021. https://doi.org/10.1017/9781009029384

[9] Lu, C.; So, W. W. M.: Translanguaging in Scientific Practices: A Study of High School Teachers in English Medium Instruction Inquiry-Based Science Classrooms. *Int. J. Sci. Educ.* 2023, 45 (10), 850–871. https://doi.org/10.1080/09500693.2023.2175628

[10] Lemmi, C.; Pérez, G.: Translanguaging in Elementary Science. *Int. J. Sci. Educ.* 2024, 46 (1), 1–27. https://doi.org/10.1080/09500693.2023.2185115

[11] Lemke, J. L.: *Talking Science: Language, Learning, and Values;* Ablex Publishing: Norwood, NJ, 1990. https://files.eric.ed.gov/fulltext/ED362379.pdf

[12] Kress, G.; Jewitt, C.; Ogborn, J.; Tsatsarelis, C.: *Multimodal Teaching and Learning: The Rhetorics of the Science Classroom;* RoutledgeFalmer: London, 2001. https://newlearningonline.com/literacies/chapter-12/kress-on-multimodality-in-the-science-classroom

[13] Ainsworth, S.: DeFT: A Conceptual Framework for Considering Learning with Multiple Representations. *Learn. Instr*. 2006, 16 (3), 183–198. https://doi.org/10.1016/j.learninstruc.2006.03.001

[14] Tang, K. S.; Delgado, C.; Moje, E. B.: An Integrative Framework for the Analysis of Multiple and Multimodal Representations for Meaning - Making in Science Education. *Sci. Educ.* 2014, 98 (2), 305–326. https://doi.org/10.1002/sce.21099

[15] Pierson, A. E.; Clark, D. B.; Brady, C. E.: Scientific Modeling and Translanguaging: A Multilingual and Multimodal Approach to Support Science Learning and Engagement. *Sci. Educ.* 2021, 105 (4), 776–813. https://doi.org/10.1002/sce.21622

[16] Bolger, M. S.; Osness, J. B.; Gouvea, J. S.; Cooper, A. C.: Supporting Scientific Practice through Model-Based Inquiry: A Students'-Eye View of Grappling with Data, Uncertainty, and Community in a Laboratory Experience. *CBE—Life Sci. Educ.* 2021, 20 (4), ar59. https://doi.org/10.1187/cbe.21-05-0128

[17] Nowell, L. S.; Norris, J. M.; White, D. E.; Moules, N. J.: Thematic Analysis: Striving to Meet the Trustworthiness Criteria. *Int. J. Qual. Methods* 2017, 16 (1). https://doi.org/10.1177/1609406917733847

[18] Landis, J. R.; Koch, G. G.: The Measurement of Observer Agreement for Categorical Data. *Biometrics* 1977, 33 (1), 159–174. https://doi.org/10.2307/2529310

Invisible Ink, Visible Risk: A STEAM Exploration of Packaging Ink Transfer onto Food

Kaifang peng [1],Hsin-Hui Wang [2],Chiu, Juei-Yu [1] ,Cheng Ming Lin [3]

[1]*National Pingtung University of Science and Technology,*

[2]*National Tsing Hua University,* [3]*Ph.D,Aidmics BioTech.CO.,LTD.*

kaifangpeng0808@gmail.com

In Taiwan, hot soy milk is often packaged in printed plastic bags, raising food safety concerns about ink migration. This study aims to design a visual ink transfer experiment to guide students in recognizing and reflecting on potential risks in food packaging, thereby encouraging them to voluntarily reduce the use of plastic bags for food storage. We simulate ink transfer through physical friction and immersion in different solvents, investigating factors including water temperature, alcohol concentration, and the effects of solvents with varying polarities on ink migration. This serves as a pilot study for developing exploratory educational materials. The experiment utilized water at varying temperatures (20–95℃), ethanol at different concentrations (20–95%), and diverse food items such as rice wine, soy milk, and cooking oil. A smartphone microscope paired with the ColorAnlz app analyzed hue and saturation changes in ink transferred onto filter paper. Results indicate that ink migration significantly increases with rising water temperature or alcohol concentration. Migration also occurs in rice wine and soy milk, likely due to enhanced ink dissolution from elevated temperature and concentration, the polar extraction effect of alcohol, and pigment release facilitated by hydrogen bonding and hydrophobic-hydrophilic interactions with soy milk proteins. Edible oil exhibited extremely low migration, indicating the ink's polar nature. Even without liquid exposure, repeated contact caused physical peeling. This study serves as an authentic and educational STEAM teaching case, adaptable to the 6E instructional model, aligning with the core principles of SDG 3 (Good Health and Well-being) and SDG 12 (Responsible Consumption and Production)..

Keywords: Food packaging safety, Ink migration, Printed plastic packaging , Smartphone microscope

1. Introduction

Amid growing global concerns over sustainability and food safety, the safety of packaging materials in direct contact with food has become a focal point of international discussion. According to EU Regulation EC No. 1935/2004, Switzerland's Printing Ink Ordinance, and recommendations from the European Food Safety Authority (EFSA, 2024), food contact materials must not release substances harmful to human health under normal conditions of use [1]; Specifically, the migration of low molecular weight substances (<1000 Da) must be below 10 ppb. However, in Taiwan and East Asia, packaging such as printed plastic bags, cup lids, and beverage pouches is still commonly used to directly hold high-temperature beverages like hot soy milk, steamed buns, and milk tea. While this practice is widespread, it has long lacked inspection and attention, harboring overlooked food safety risks. In its 2024 review report, the Food Risk Assessment Agency noted that residual monomers and small-molecule photoinitiators in ink constitute one of the primary migration risks. Their toxicity may include carcinogenicity, endocrine disruption, and increased metabolic burden on the liver and kidneys [1]. International brands like Nestlé and Tetra Pak have long established internal regulations prohibiting the use of BPA-containing developers or non-low-migration systems.

Printing inks are inherently composite materials, typically containing pigments, resins, solvents, and functional additives such as acetate esters, long-chain alcohols, vegetable oil-modified resins, monomers (e.g., styrene), photoinitiators (e.g., benzophenones), dyes (e.g., leuco dyes), and developers (e.g., bisphenol A) [2]. According to solubility theory, polar liquids (e.g., ethanol, soy milk) can dissolve or permeate polar or amphoteric molecules within inks. Additionally, elevated temperatures accelerate polymer chain motion and disrupt microcapsule structures, further promoting the release of ink components—a phenomenon termed “ink migration” [3].

Although the use of “low migration inks” is gradually becoming more widespread, practical understanding of these issues among educators and consumers remains limited. There is also a lack of intuitive methods to explain the scientific principles and potential risks of ink leaching to students or the public. Therefore, this study designed an inquiry-based experiment combining authenticity and educational value. Using commercially available printed plastic bags as samples, we compared four common liquids (hot water, rice wine, soy milk, cooking oil) under varying conditions: temperatures (20°C–95°C), alcohol concentrations, and contact frequencies (single vs. repeated exposure). Through observations and documentation using a mobile app and smartphone microscope, we analyzed the interaction effects of three factors—liquid type, temperature, and solubility—on ink migration.

Key features of this study include: (1) the first systematic investigation of packaging ink migration in everyday Taiwanese contexts; (2) the use of visual experiments combined with mobile technology to make “invisible” risks visible to the naked eye; (3) Establishing an educational module applicable to food safety, materials science, and STEAM education to cultivate students' inquiry skills and risk awareness. We aim not only to alert society to potential hazards in everyday packaging but also to provide a practical model for translating scientific research into educational practice.

2. Experimental Methods

This study utilized rice wine containing 22% alcohol (polar liquid), soy milk containing protein components (exhibiting both polar and amphiphilic properties), blended oil and salad oil (as non-polar controls), and peanut soup (a composite everyday beverage). Additionally, pure water at 20°C (room temperature), 30°C, 65°C, 75°C, and 95°C was used. Alcohol concentration groups were set at 20%, 60%, 75%, and 95% at the same temperature (30°C). Additionally, contact modes were compared between single and multiple exposures (2, 3, 4, 5 wipes) to simulate effects from friction or pressure during packaging handling and storage.

Experimental materials included commercially available printed garbage bags as printed plastic bag samples, filter paper (110 mm diameter, matching petri dish size), petri dishes and lids, heating devices (hot water bottle or induction cooker), temperature-controlled incubator, thermometer, smartphone microscope, and the ColorAnlz color analysis app (for analyzing hue and saturation).

In the experiment, filter paper was first cut or placed flat to fit the petri dish. Printed plastic bags cut to the same size were placed on the filter paper with the printed side facing down. Pure water was heated to the specified temperature. For the alcohol group, solutions were prepared at the designated concentrations. Test liquids (soy milk, rice wine, cooking oil, peanut soup, etc.) were also prepared. The test liquid was gently poured into the petri dish to fully saturate the plastic sheet. The dish was then placed in an incubator or temperature-controlled chamber to maintain the set temperature for 24 hours. After the reaction concluded, the plastic sheet was carefully removed, and the pigment residue on the filter paper was observed.

The degree of ink migration was documented using a smartphone microscope, and ColorAnlz application for quantitative analysis. Hue and Saturation values were measured at three fixed positions on each image, with averages calculated. Hue variations indicated shifts in color components, while Saturation changes reflected pigment concentration alterations. Finally, data variations under different contact frequencies and liquid conditions were compared, with cumulative effects presented graphically.

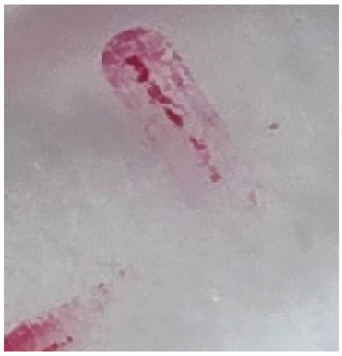


Figure 1: Migrating pigment particles and impurities observed under a smartphone microscope

Observation using a smartphone microscope (Figure 1) reveals distinct pigment particles and impurities released within the migration zone. By capturing hue and saturation data with the ColorAnlz App and conducting ΔE color difference analysis, it was confirmed that ΔE values significantly increased under high temperatures and polar liquids, indicating heightened color variability.

3. Results and Discussion

This study conducted ink migration experiments on printed plastic bags using four common liquids (hot water, rice wine, soy milk, and cooking oil), investigating the combined effects of three factors—temperature, solvent polarity, and contact frequency—on migration outcomes.

(1). Ink Transfer Experiments at Different Temperatures

In the pure water group experiment, as temperature increased from 20℃ to 95℃, pigment migration became significant above 65℃, with high-density pigment deposits and color spots visible on the filter paper surface. As shown in Figure 1, the horizontal axis represents treatment temperature(℃), while the vertical axis displays hue (H) and saturation (S) values. As temperature increased, the Hue value rose sharply after 65℃, reflecting a shift in color components. Saturation values also increased significantly under high-temperature conditions, indicating enhanced pigment concentration and adhesion [5]. This phenomenon relates to accelerated molecular motion due to heat.

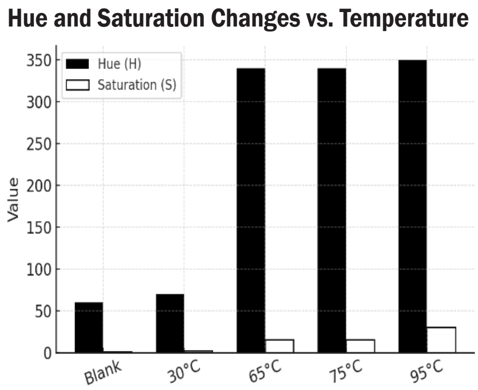


Figure 2: Changes in Hue and Saturation of Ink Migration in the Pure Water Group at Different Temperatures

Elevated temperatures soften polymer chains in both the plastic substrate and ink layer, reducing structural stability [6,7] and loosening the ink layer. Concurrently, microcapsules encapsulating pigments rupture due to expansion, releasing contained dyes. This ultimately causes substantial pigment migration and diffusion onto the filter paper surface [8]. (Figure 2)

(2). Transfer Printing Experiments with Inks of Different Alcohol Concentrations

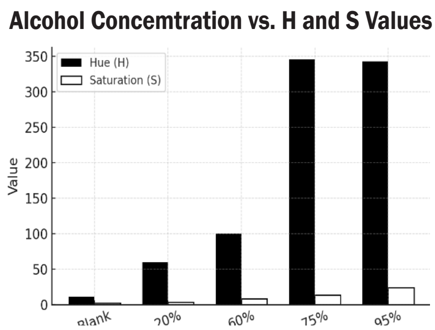


Figure 3: Changes in Hue and Saturation of Ink Migration Across Different Alcohol Concentration Groups

Significant ink migration was observed in the alcohol group (treated with 20%-75% ethanol immersion) (Figure 3). The Hue value increased markedly with concentration, reaching saturation at approximately 350 beyond 75%. This result correlates with ethanol's high polarity and low molecular weight, enabling rapid penetration of the ink layer. This process weakens the interfacial forces between the resin binder and plastic substrate while effectively dissolving dyes and other polar components [6.7.9].

(3). Other Daily Food Items

Although peanut soup samples exhibited some saturation variation (approximately 20%), their Hue values remained low (<50). This indicates that despite increased pigment content, the tonal composition showed minimal variation. This phenomenon may be attributed to the adsorption of pigments by sugars and solids. These components likely encapsulate portions of the pigments via hydrogen bonds, causing pigment entrainment and structural alterations. Consequently, this leads to color migration and abnormal saturation [8.9].

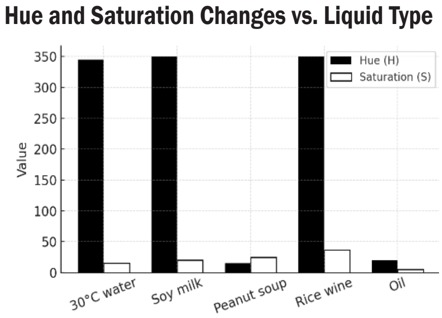


Figure 4: Changes in Hue and Saturation of Ink Migration for Different Test Liquid Groups

The Hue value of the rice wine (22% ethanol) group was close to that of high-concentration alcohol at 350, enabling rapid release of polar dyes onto the filter paper surface. The Hue value of the soy milk group was also near 350, but its Saturation was approximately 15% lower than the alcohol group. This disparity is likely attributed to uneven distribution of ink molecules caused by proteins and emulsifying components in soy milk [9]. The soy milk group exhibited high Hue and low Saturation, presumably due to incomplete dispersion resulting from its protein content and viscosity.

(4). Ink Fading Test Based on Number of Friction times

Based on observations of ink transfer marks on filter paper during experiments, color intensity and area changes serve as evaluation criteria. Repeated contact (2 to 5 times) causes physical peeling and transfer of ink due to friction, even without added liquid, resulting in color fading and diffusion. This indicates that repeated contact or compression between the bag and food during daily use may also cause physical ink transfer. This phenomenon relates to mechanical friction risks and aligns with the EU's technical guidance on ink abrasion resistance [6.7.8] (Figure 5).

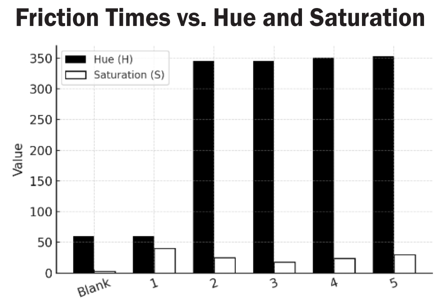


Figure 5: Changes in Hue and Saturation of Ink Migration Caused by Physical Peeling at Different Contact Frequencies (Without Liquid)

4.6E Teaching Model

In this study, experimental procedures and results were further transformed into a STEAM curriculum based on the 6E instructional model, enabling students to learn through hands-on engagement. The 6E process is divided into six stages.

In the “Engage” stage, students may initially be prompted with the question, “Have you ever observed ink transfer from plastic bags stored in the refrigerator?” Such real-life examples can be employed to capture students’ attention and stimulate discussion on potential food safety risks.

In the “Explore” stage, students work in groups to investigate the effects of different liquids (e.g., alcohol, soy milk, oil), varying temperatures, and contact frequencies, while observing the color changes.

In the “Explain” stage, students are guided to interpret the experimental results in order to explain why ink transfers. A smartphone microscope and analysis app are employed to observe the details, analyzing how ink transfers is influenced by polarity, thermal properties, and solvent type.

In the “Engineer” phase, students construct simple observation tools using filter paper, plastic film, and microscopes, create color intensity charts, and use simple instruments such as smartphone microscopes to facilitate observation.

In the “Enrich” phase, the experimental results are connected to broader food safety issues (such as BPA and the 10-ppb limit) and to the significance of SDG 3 and SDG 12, thereby facilitating discussion on global packaging regulations.

In the final “Evaluate” phase, students are required to interpret the data, answer conceptual questions, and use the observed images and data to support their ideas.

5. Conclusion

This study confirms that high temperatures, polar liquids, and contact pressure significantly increase the risk of ink migration. Educationally, this experimental design enables students to visualize potential risks in food packaging and reflect on the safety of daily plastic bag usage, aligning with the core principles of SDG 3 (Good Health) and SDG 12 (Responsible Consumption).

Commercially available printed plastic bags were placed over toast surfaces. Through simulated daily contact scenarios involving friction and heating, ink migration onto the toast surfaces was induced. Subsequently, a smartphone microscope was used for magnified observation, capturing high-resolution images. These images were imported into an image analysis software app for color analysis, yielding data on saturation and hue. The findings were ultimately compiled into visual representations (Figure 6) serving as experimental evidence for this study.

Figure 6: Schematic illustration of ink migration directly from packaging to food [11]

6. References

[1] Eller and Heckman LLP. (2025, January 19). Swiss printing inks ordinance updated. Packaging Law.com https://www.packaginglaw.com/news/swiss-printing-inks-ordinance-updated

[2] Rožić, M., Vukoje, M., et al. Acta Graphica, 28, 137–147-(2017)

[3] Desie, G., & Van Roost, C. Journal of Imaging Science and Technology, 50, 294–303(2006).

[4] Gupta, R. K., Pipliya, S., Karunanithi, S., Eswaran, G. M., Kumar, S., Mandliya, S., Srivastav, P. P., Suthar, T., Shaikh, A. M., Harsányi, E., & Kovács, B. Foods, 13, 3125(2024).

[5] Hunt, R. W. G., & Pointer, M. R. (2011). Measuring Colour(4th ed.). Wiley.

[6] Fried, J. R. (2014). Polymer Science and Technology(3rd ed., pp. 203-272). Pearson Education.

[7] Cussler, E. L. (2009). Diffusion: Mass Transfer in Fluid Systems(3rd ed., pp.1-50). Cambridge University Press

[8] Ghosh, S. K. (2006). Functional coatings and microencapsulation: A general perspective. In Functional Coatings: by Polymer Microencapsulation (pp. 1–28). Wiley-VCH.

[9] Rosen, M. J., & Kunjappu, J. T. (2012). Surfactants and Interfacial Phenomena(4th ed.,pp. 235-245). Wiley.

[10] Ghosh, S. K. (2006). Functional coatings and microencapsulation: A general perspective. In Functional Coatings: by Polymer Microencapsulation (pp. 1–28). Wiley-VCH..

[11] OpenAI. (2023). ChatGPT (Mar 14 version) [Large language model]. https://chat.openai.com/chat

An Exploratory Study on How Generative AI Supports Students' Visual Construction of the Particle Model of Matter

Jing-Yi Liu[1], Chen-Yu Chen[1,2], Chiu-Wen Wang[1,3], Jing-Wen Lin[1*]

[1]*Department of Science Education, National Taipei University of Education, Taiwan*

[2]*Min-An Elementary School, New Taipei City, Taiwan*

[3]*Yong-Shun Elementary School, Taoyuan City, Taiwan*

jwlin@mail.ntue.edu.tw

Abstract

This study explores the application of Generative AI in teaching the Particle Model of Matter (PMM) to sixth-grade students, focusing on how students construct and evaluate visual models by transforming linguistic descriptions through AI. A qualitative approach was adopted. Two PMM-themed lessons were designed, integrating the Generation (G), Evaluation (E), Modification (M), G.E.M. modeling process with three evaluation criteria: Plausibility (P), Alignment with experience and imagination (A), and Detail (D). Participants included three groups of sixth graders. The study examined how students developed initial linguistic models, interacted with AI to iteratively generate visual models, and used the P.A.D. criteria to evaluate and refine them. Data sources included AI dialogue records and classroom transcripts, analyzed thematically. Findings show that students' initial models were often intuitive and macroscopic, with mental models mostly classified as Descriptive (D) or Mixed-consistent (M-c). Through AI-assisted visualization, students identified conceptual gaps, revised their input, and gradually incorporated microscopic ideas, eventually constructing scientifically plausible Basic Particle Models of movement (B-m). The P.A.D. criteria played complementary roles, helping students assess visual models from multimodal perspectives and demonstrating metacognitive engagement. This study presents a practical modeling and evaluation process combining Generative AI with science instruction. It highlights AI's potential to provide instant feedback and visual representations, enabling students to refine internal models and deepen their conceptual understanding. The findings offer practical insights into elementary modeling instruction and emphasize the educational value of P.A.D. criteria in promoting metacognition and scientific learning.

Keywords: Particle Model of Matter, Generative AI, Mental Model, Evaluation Criteria

1. Introduction

With the rapid advancement of artificial intelligence (AI) technologies, Generative AI is gradually entering classrooms and becoming an emerging tool to support both teaching and learning. Educational practices around the world have begun to explore the integration of AI into various contexts such as language learning, creative expression, and scientific modeling, forming a new type of interactive model known as "AI co-learning" [1].In science education in particular, Generative AI has the potential to serve as a visual mediator that helps students concretize abstract concepts through image generation technologies, thereby supporting their understanding and construction of scientific models [2].

Taiwan's Curriculum Guidelines of 12-Year Basic Education officially included "modeling" as one of the expected learning performances in the science domain at the elementary level, and introduced microscopic concepts—such as "matter is composed of tiny particles that are constantly moving" into the upper grades of elementary science education for the first time [3]. The Particle Model of Matter (PMM) is a fundamental concept in the natural sciences. However, due to its microscopic and unobservable nature, it is often considered difficult to teach by educators and challenging to learn for students. For elementary school students who are still in the concrete operational stage, understanding such abstract concepts poses a considerable challenge [4].

Against this backdrop, this study focuses on how sixth-grade elementary students, while learning the PMM, engage with Generative AI to visualize their initial language-based models, evaluate the AI-Generated outputs, and revise them accordingly. The goal is to enhance students' understanding of particle concepts and support the development of metacognitive skills. The research questions are as follows:

1. What are the students' initial linguistic models and the AI-Generated final visual models after interacting with Generative AI?
2. How do students evaluate the AI-Generated visual models?

2. Literature Review

Taiwan's Curriculum Guidelines of 12-Year Basic Education emphasize the importance of students being able to explain natural phenomena through modeling [3]. This highlights the need for teachers to move beyond traditional observation-based instruction and guide students in constructing abstract scientific concepts and ways of thinking. Meanwhile, the rise of Generative AI is reshaping classroom interactions, with growing attention on its potential as a learning mediator. In response, this study reviews the literature from the following three perspectives.

2.1 The Role of Modeling in Elementary Science Education and Its Implementation Challenges

The particulate nature of matter (PNM) is a fundamental concept in science, positing that all matter is composed of tiny, discrete particles (such as atoms or molecules) that are in constant motion and exert attractive or repulsive forces on each other [4]. Chiu et al. [5] noted that educators should not only focus on teaching the content of particle theory but also on how to present it in ways that effectively connect with students' prior knowledge and cognitive developmental stages, thereby facilitating their understanding of abstract concepts. Treagust et al. [6] emphasized that scientific models serve as mediators linking abstract concepts with students' experiences. Models are not merely learning tools, but in some cases, one of the only means to help students grasp abstract scientific theories such as particle theory. Xue and Sun [7] pointed out that scientific models and modeling offer significant benefits in chemistry education, playing an important role in describing natural phenomena and developing scientific knowledge.

Modeling lies at the heart of scientific thinking and practice. As one of the core forms of multimodal learning, it supports science teaching that often relies on multiple modes of representation. Schwarz et al. [8] identified modeling as a core scientific practice encompassing four main processes: constructing, using, evaluating, and revising models. They further stressed the importance of "metamodeling knowledge," which refers to students' understanding of the nature, function, limitations, and revision criteria of models, as essential to deepening scientific literacy. However, research has shown that students often perceive models as static images, overlooking their role as tools for reasoning, explanation, and prediction. This limits their ability to view models as revisable constructs. Without model evaluation skills, students may struggle to determine whether a model effectively represents a scientific phenomenon or to revise it appropriately, thereby reducing its learning potential [9]. In line with the vision of the Next Generation Science Standards (NGSS), modeling is considered a key approach to engaging students in scientific practices and knowledge construction [10]. Treagust et al. [6] called for providing students with more opportunities to manipulate, generate, and reflect on models. Such processes, transforming hypotheses into tangible representations, are essential for fostering depth of understanding and learner autonomy in science education [11,12].

At the elementary level, modeling helps students translate abstract scientific concepts into concrete mental structures. Yet, students often face challenges in language expression and multimodal integration, particularly in aligning images, text, and symbols into coherent scientific meaning. In this process, teachers act as mediators of learning [13], highlighting the need for targeted pedagogical support and well-designed strategies to help students model effectively—an endeavor that presents demands for both teachers and learners [14,15].

2.2 Challenges in Learning the Particle Model of Matter

The PMM is a central concept in science, yet its highly abstract nature often creates considerable difficulty for both teaching and learning. PMM involves microscopic ideas that cannot be directly observed and are difficult to build from students' everyday experiences [16]. Lee et al. [17] noted that students frequently rely on macroscopic intuitions to explain changes in matter. Such experience-based interpretations can lead to deeply rooted misconceptions, which in turn hinder the development and understanding of more advanced concepts such as atoms and molecules at the secondary level. If a correct particle view is not established in the elementary years, students are likely to encounter greater cognitive gaps in subsequent science learning. Chen and Lin [18] found that teachers still hold a vague understanding of the particle concepts introduced in Taiwan's 12-Year Curriculum, and even experienced teachers may not achieve a complete grasp of the scientific framework. This indicates that both teachers and students experience varying degrees of difficulty when dealing with this content. To address these issues, Lin [19] synthesized curriculum guidelines and research to propose eight core particle propositions that help clarify PMM teaching priorities and strengthen conceptual connections. Building on this foundation, Chen and Lin [18] integrated Merritt and Krajcik's [20] classification to develop a PMM mental model framework (Table 1), which provides both teachers and students with clear directions for conceptual adjustment and revision, and also demonstrates the potential of analogical modeling in PMM instruction. However, many students, even when possessing suitable intuitive analogies, often struggle to clearly articulate the logic and corresponding structures of their models. This makes it difficult for teachers to accurately interpret their conceptual understanding [21].

To overcome this limitation, the present study introduces Generative AI's image generation capability. After students express their analogy ideas in language, the system assists in visualizing their models. This approach not only reduces the linguistic barrier for students but also provides teachers with more visible and analyzable outputs as a basis for instructional feedback.

Table 1. PMM mental model framework

Mental Models			Descriptions
Scientific model		S	A complete particle model explains how particles and their interactions determine matter's macroscopic properties and behavior.
quasi- Scientific model		qS	Particles are separated by a vacuum, with gas particle spacing much greater than in solids and liquids, but without the concept of interparticle attraction.
Basic particle model	BS_distance	B-d	The relationship between the random motion of matter and distance.
	BS_movement	B-m	Particles undergo constant random motion
Mixed model	Mixed-consistent	M-c	Use microscopic particle and macroscopic descriptive perspectives to consistently explain phenomena of different substances.
	Mixed-inconsistent	M-ic	Use microscopic particle and macroscopic descriptive perspectives, but show inconsistency when explaining phenomena of different substances.
Descriptive model		D	Describing matter based on its visible appearance.

2.3 Potentials and Challenges of Generative AI in Supporting Students' Analogical Modeling

Gilbert [2] emphasized that visualization plays a crucial role in science learning and that students need to develop metavisual capability in order to effectively translate between different modes of representation and construct models. He further stressed that teachers should explicitly indicate the purposes and stages of model construction; otherwise, students may perceive modeling merely as a drawing activity and overlook its core function in conceptual understanding and scientific reasoning. Within this context, the application of Generative AI offers a new tool for visualization support. Generative AI can quickly produce a variety of visual models and images based on students' descriptions, concepts, or analogical inputs, helping them explore diverse representational forms. It can also strengthen students' understanding and application of analogy-based concepts [22]. Building on this potential, Generative AI can serve as a visual mediator, assisting students in externalizing linguistic descriptions into images and engaging in model construction and revision through interaction with AI-Generated visuals, thereby supporting scientific understanding.

By enabling rapid iteration in analogical modeling and image generation, Generative AI allows students to quickly explore, test, and modify their ideas, deepening their understanding of how knowledge is constructed and validated [23]. Students' evaluation of images during the modeling process is a key aspect of metacognitive engagement [24,25]. Morris [26] argued that Generative AI can serve as an innovative educational tool, supporting well-designed learning activities that stimulate students' metacognition. The combination of visual mapping and structured tasks can effectively promote and assess students' conceptual understanding [27]. Loeckx [27] further highlighted that AI can be an effective learning aid, reducing the workload for both teachers and students while enhancing the overall learning experience.

However, from a cognitive processing perspective, over-reliance on AI tools that quickly provide solutions may hinder learners from developing higher-order thinking skills [29]. Therefore, making effective use of Generative AI as a visual mediator in instruction, ensuring that technology serves as a learning partner in the classroom, will be an important challenge for teachers and educators in the future.

3. Research Methods

3.1 Research Design

This study adopted a qualitative research approach, referencing Khan's [30] model development process: Generation (constructing a model), Evaluation (assessing a model), and Modification (revising a model). Based on the G.E.M. framework, two lessons on the PMM were designed. The lessons were taught by a teacher with training and experience in modeling instruction to minimize errors arising from unfamiliarity with the pedagogy.

In the first lesson, students constructed initial models (G stage) by observing the diffusion of ink in hot and cold water. They then watched a video and completed a worksheet to compare their models with the observed phenomenon, identifying gaps or errors (E stage). Finally, a formative assessment guided students to reflect and revise their models (M stage), resulting in the first lesson's linguistic model. Figure 1 shows the assessment structure.

In the second lesson, students used the linguistic models from the first lesson and employed Generative AI to create analogy-based visual models of the ink diffusion phenomenon (G stage). They then evaluated the generated images (E stage) and provided explanatory instructions to the AI for modifications (M stage). Through repeated interaction with the AI, students refined both their conceptual understanding and visual representations, completing the final models. The entire lesson was video recorded and transcribed verbatim for analysis.

We drop **ink particles** into a cup of **hot water**.
When **thermal energy** is added, it causes the **ink particles** to move **faster/slower**.
Just like:
___ goes into ___ ___ ,
___ causes ___ to move **faster/slower**.
So, ______________________________
(fill in your answer directly).

Figure1. Structure of the assessment tasks

3.2 Participants

The participants of this study were nine sixth-grade students from an elementary school in Taoyuan City, selected from the school's autonomous learning team and science club. The students were divided into three groups for the experiment. They were chosen because they possessed strong oral communication skills and prior experience in discussion, enabling them to engage actively in thinking and dialogue. These characteristics aligned with the study's focus on the evaluation process and metacognitive aspects of the modeling process.

3.3 Data Collection

Data were collected from two main sources: (1) student–AI dialogues and (2) video and audio recordings of the lessons.

(1) Student–AI dialogues: Students expressed, in written form, the linguistic models of the "ink particle diffusion in water" phenomenon they developed in the first lesson to the Generative AI. The AI then produced corresponding images. The recorded dialogues between students and the AI provided insights into how students articulated their ideas, modified visual representations, and engaged in the processes of model construction, evaluation, and revision.

(2) Video and Audio Recordings: The entire instructional process was video- and audio-recorded. These recordings were transcribed verbatim. The video data, in particular, captured authentic peer interactions, students' evaluation criteria for the generated images, and their reasoning behind model modifications.

3.4 Data Analysis

(1). Image Judgment for Research Question 1: A qualitative analysis was conducted to determine students' mental models based on the PMM concept learning framework proposed by Chen and Lin [18].

- Initial Linguistic Model: Derived from the models students developed in the first lesson.
- AI-Generated Final Visual Model: Derived from the models in the second lesson. Students interacted with Generative AI using their initial linguistic models from Lesson 1 to produce visual representations. They then proposed conceptual modifications to the images. When a student responded to an image with remarks such as "That's it," "It's fine," or "It's close enough" and no further images were generated, the model was identified as the final model.

(2). Evaluation Criteria for Research Question 2: A qualitative content analysis approach was adopted. Drawing on Lemke's [31] systemic functional linguistics perspective, which posits that each mode has unique functions and cannot fully replace another, three types of meaning were used as the main coding framework. Their definitions and correspondence to this study are as follows:

- Presentational Meaning: Focuses on students' understanding and evaluation of the AI-Generated image content itself, such as whether the image aligns with scientific principles, facts, or phenomena. In this study, it corresponds to Plausibility(P).
- Orientational Meaning: Concerns students' subjective stance, prior imagination, attitudes, and emotions, reflecting personal perspectives and value judgments about the AI images. In this study, it corresponds to Alignment with experience and imagination(A).
- Organisational Meaning: Analyzes students' evaluation of image details, structures, and element composition, exploring how they interpret and organize information within the image. In this study, it corresponds to Detail(D).

After reaching a consensus on these definitions, two researchers independently coded the qualitative data for one group of students. The Kappa values for all three categories exceeded .85, indicating high agreement [32]. For any coding discrepancies, the research team engaged in further discussion to ensure that the interpretation of students' evaluation criteria was as accurate as possible. These three meaning categories ultimately became the three evaluation criteria for students' assessments of the AI-Generated visual models, as shown in Tables 2 and 3.

Table 2. Evaluation criteria for AI-Generated visual representations

Criteria	Coding Description	Example
Plausibility(P)	Scientific accuracy of the AI-Generated image based on student input.	Dialogue includes scientific concepts or meaningful scientific terms: particles, heat diffuses faster, sweets make ants move faster, analogy diagram…
Alignment(A)	Match between the image and the student's experience	Dripping ink particles into a beaker; switch to Angry Birds; Ah! I didn't tell it that it was raining; Yes, the ants did move faster…
Detail(D)	Specific visual elements students requested (e.g., color, quantity, motion, atmosphere).	Why are there so few ants? This should have a big explosion; It's not even in the water; Ah! It gave me a bunch of black birds…

Table 3. Data identification and coding descriptions

Item	First Code (Target)		Second Code (Modeling Stage)	Third Code (Data Source)		
	Group	Image Sequence	Modeling Process	Generative AI Dialogue	Transcription	
Code Description	T1~T3	0~11 image number	G: Generation E: Evaluation M: Modification	sequential numbering of dialogue turns	S1–S3: three students in each group	Evaluation Code P.A.D
Example	T1.6_E_S3.A : Target: Group 1, Image #6, Modeling Process: Evaluation stage, Data source: statement from Student 3 regarding evaluation criteria for the image. T2.3_M_AI.4: Target: Group 2, Image #3, Modeling Process: Modification stage, Data source: fourth dialogue turn with AI.					

4. Results

4.1 What are the students' initial linguistic models and the AI-Generated final visual models after interacting with Generative AI?

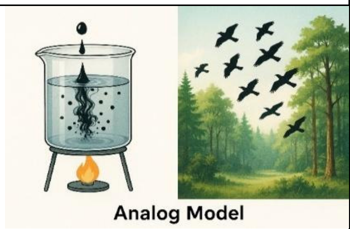
Through the two-lesson analogy-based modeling tasks, we collected students' initial linguistic models and their final AI-Generated visual models. Using the PMM concept learning mental model classification proposed by Chen and Lin [18], we found that students' initial linguistic models were classified as Descriptive (D) and Mixed-consistent (M-c). However, the AI-Generated final Visual Model from the second lesson all reached the level of the Basic Particle Model – Movement (B-m), which reflects a scientifically basic particle movement model. Table 4 provides detailed descriptions of the model characteristics for the three groups.

(1) Initial Linguistic Models: After completing the modeling task in the first lesson, analysis of the students' linguistic models revealed that, within the lesson's modeling framework, all students were able to clearly identify the concept of an "energy source" and describe the "movement" behavior. This finding indicates that the structured modeling framework was necessary and beneficial in helping students grasp the concepts intended by the teacher. It also echoes Gilbert's view [2] that teachers need to make the purpose and stages of modeling explicit; otherwise, students may perceive modeling merely as drawing or surface-level manipulation.

In the lesson design of this study, the step-by-step guidance allowed students to follow the intended process. However, in the initial linguistic models, most analogies from all three groups still described phenomena from a macroscopic perspective without incorporating particle concepts. As a result, the mental models of Groups T1 and T3 were classified as Descriptive (D), while Group T2 showed awareness of the need to address particle concepts but still described them macroscopically, leading to its classification as Mixed-consistent (M-c).

(2) AI-Generated Visual Models: In the stage where the initial linguistic models were transformed into analogy-based images through Generative AI, the representations in the linguistic models were converted into visual forms. Students examined the AI-Generated images and supplemented any missing concepts. When the images matched their intended mental representation, they were identified as the AI-Generated final visual models.

Table 4. Mental Models of the three groups in the PMM modeling process

	T1	T2	T3
Lesson 1 Initial Linguistic Model	A black bird enters a sunny forest. The improved weather speeds up the bird's movement, so it becomes more active.	The angry bird enters a fortress game. A big explosion speeds up its movement, just like particles move faster in hot water.	The ant enters a candy world. The sweets make the ant move faster.
Mental model	Descriptive/D	Mixed-consistent /M-c	Descriptive/D
Lesson 2 AI-Generated final Visual Model			
Mental model	Basic particle model_movement /B-m	Basic particle model_movement /B-m	Basic particle model_movement /B-m

As an example, the process for Group 1 is described below:

Initial Linguistic Model: *Black birds enter a sunny forest; the good weather makes the black birds move faster, so they become more active.* (T1.0_G_AI.1)

Figure 2. The first image generated by the Generative AI

"Why did it give me a bunch of black birds?" (T1.1_E_S2.AD)

"No analogy? Tell it to produce an analogy image." (T1.1_M_S1.PA)

This group noticed that the AI-generated image lacked concepts of quantity and movement, and they attributed this to the absence of the analogy concept in their original description. They then revised their prompt to the Generative AI by adding: *"Help me produce an analogy model"* at the end:

Black birds enter a forest; the good weather makes the black birds move faster, so they become more active. Help me produce an analogy model. (T1.1_M_AI.2)

Subsequently, the group continued interacting and gave multiple instructions to the AI. Examples of these interactions are as follows:

...We drip ink particles into a beaker of hot water. (T1.2_M_S3.D)

Heating makes the ink particles move faster... (T1.2_M_S2.PD)

Black birds enter the forest... move faster and become more active. (T1.2_M_S3.AD)

We drip ink particles into a beaker of hot water; heating causes the ink particles to move faster, just like [black birds entering a sunny forest, where the good weather makes the black birds move faster and more active]. Please produce an analogy model image. (T1.3_M_AI.4)

Finally, the group obtained a satisfactory AI-generated visual model, as shown in Figure 3.

Figure 3. AI-Generated final visual model of Group 1

During the interaction process, students were able to return to the framework of their initial linguistic models for alignment, indicating that in their dialogues with the AI, they responded to relevant concepts through the visualization of images. At this stage, the formative assessment framework provided in the first lesson served its linking function, helping students revisit the phenomenon for further alignment. As shown in Table 4, all three groups emphasized the correspondence of "movement" behaviors during this phase, breaking away from the macroscopic framework of the initial linguistic models and shifting toward a scientifically Basic particle model-movement (B-m) that incorporates microscopic concepts.

The initial linguistic models from Lesson 1 clearly lacked an understanding of the "microscopic particle concept," with mental models generally falling into the Descriptive (D) and Mixed-consistent (M-c) categories. With the assistance of Generative AI, students transformed the representations in their linguistic models into images, which enabled them to better identify the concepts needed in their models. This process strengthened their ability to evaluate and revise models, and in doing so, activated their metacognitive processes and helped them establish concepts. As Grapin et al. [10] noted, multimodality can provide students with rich resources for meaning-making and communication, effectively supporting their engagement in scientific inquiry. The next section will describe how students carried out their evaluations.

4.2 How Did Students Evaluate the AI-Generated Visual Models?

With the assistance of AI-Generated, students transformed the representations in their linguistic models into images, which enabled them to better identify the concepts that needed to be supplemented. This process strengthened their model evaluation and revision. To examine how students evaluated the images to determine whether their models were complete, we analyzed verbatim transcripts of classroom recordings and the compiled prompts students gave to the AI.

The data were first segmented according to the three stages of modeling—Generation (G), Evaluation (E), and Modification (M). Then, following Lemke's [31] functional linguistic framework, each relevant utterance was coded into one of three evaluation criteria: Plausibility(P), Alignment (A), and Detail(D). The results indicate that these three evaluation criteria played different roles across the three stages of the modeling process—Generation (G), Evaluation (E), and Modification (M). The statistical analysis is presented in Table 5.

Table 5. Analysis of P.A.D. usage

<table>
<tr><td colspan="2">Group</td><td colspan="4">T1</td><td colspan="4">T2</td><td colspan="4">T3</td><td></td></tr>
<tr><td colspan="2" rowspan="2">Modeling Stage</td><td colspan="13">Evaluation Criteria</td></tr>
<tr><td>P</td><td>A</td><td>D</td><td>total</td><td>P</td><td>A</td><td>D</td><td>total</td><td>P</td><td>A</td><td>D</td><td>total</td><td>total</td></tr>
<tr><td rowspan="3">Mode -ling Stage</td><td>G</td><td>5</td><td>6</td><td>7</td><td rowspan="3">66
G:18
E:25
M:23</td><td>6</td><td>8</td><td>9</td><td rowspan="3">72
G:23
E:34
M:15</td><td>1</td><td>1</td><td>1</td><td rowspan="3">61
G:3
E:27
M:31</td><td>44</td></tr>
<tr><td>E</td><td>5</td><td>10</td><td>10</td><td>3</td><td>14</td><td>17</td><td>5</td><td>10</td><td>12</td><td>86</td></tr>
<tr><td>M</td><td>4</td><td>4</td><td>15</td><td>4</td><td>4</td><td>7</td><td>4</td><td>13</td><td>14</td><td>69</td></tr>
<tr><td colspan="2">Frequency (%)</td><td>14
21%</td><td>20
30%</td><td>32
49
%</td><td>G:27%
E:52%
M:48%</td><td>13
18%</td><td>26
36%</td><td>33
46%</td><td>G:32%
E69%
M31%</td><td>10
16%</td><td>24
39%</td><td>27
45%</td><td>G:5%
E47%
M53%</td><td rowspan="2">199</td></tr>
<tr><td colspan="2">Overall Total (%):</td><td colspan="12">P37(19%) A70(35%) D92(46%)</td></tr>
</table>

(1). Overall frequency and distribution of P.A.D. evaluation criteria during the modeling process:

Across the entire modeling process, students applied the evaluation criteria 199 times. Detail of visual representation (D) was most frequent (92 times, 46%), followed by Alignment with experience and imagination (A) (70 times, 35%). Plausibility (P) was least frequent (37 times, 19%). This indicates that students showed high sensitivity to visual detail and subjective imagination, consistent with[14], who noted students' preference for concrete, visible features. However, their relative weakness in understanding and articulating abstract concepts limited the plausibility and completeness of their models.

(2). Dynamic development of P.A.D. criteria in the modeling process:

Model Generation (G) stage – Students entered their initial linguistic models into Generative AI. Although based on the first lesson's question framework, they refined wording repeatedly, indicating early self-evaluation. Group 2 showed high interaction (P6, A8, D9) when shaping imagined models and adjusting details. Their final input command to the AI was:

"We want you to create an image of a particle model based on our analogy: Ink particles dropped into a beaker of hot water. Heat makes them move faster, like Angry Birds entering a fortress. The explosion makes the birds faster, so particles in hot water move faster." (T2.0_G_AI.1)

Although no final image was yet produced, evaluative language was already present, suggesting this process fostered self-checking of initial models.

Model Evaluation (E) and Modification (M) stages – Upon receiving the AI-generated image, students showed strong engagement, discussing plausibility, alignment with prior experience, and image details, while revising prompts. This active participation counters concerns that over-reliance on AI might hinder higher-order thinking [29]. The E and M stages were intertwined, with higher P.A.D. usage than in G, strengthening and refining application of the criteria.

(3). P.A.D. criteria activating metacognition: Overall, the criteria appeared throughout the process, forming a cyclical pattern that reinforced concept building [14]. Detail (D) was used most often (46%), followed by Alignment (A) (35%), showing students' reliance on sensory impressions. Though used less (19%), Plausibility (P) served as a key checkpoint for scientific accuracy, often integrated after detail and imagination to align models with disciplinary knowledge. For example, Group 3's initial linguistic model reflected only a descriptive (D) level, but the AI visualization linked to personal understanding, triggering metacognitive behavior. When a student remarked, *"The ants look like they're in the sky?"* (T3.2_E_

S3.P.A.D.), it showed self-assessment against the original input and adjustment based on AI output. This reflects not only model evaluation but also awareness of one's thought process, aligning with Treagust et al. [6], who emphasized that manipulating, generating, and reflecting on models enables students to take ownership of modeling and achieve deep learning.

5. Discussion

Based on the findings, this study argues that the visual mediating role provided by Generative AI has the potential to enhance students' modeling competence and activate their metacognitive processes.

5.1 Generative AI as a means to extend linguistic models and construct scientific concepts

From Research Question 1, we found that the linguistic models constructed by the three groups of students were mostly limited to D- and M-c-type mental models, reflecting an intuitive understanding of macroscopic phenomena and difficulty in concretizing abstract concepts such as particle motion and microscopic structures. When the modeling process incorporated interaction with AI, the visualized images enabled students to identify missing concepts in their linguistic models. Consequently, the macroscopic framework of the initial linguistic model was elevated to the B-m type, which incorporates microscopic concepts. This progression not only demonstrates students' construction of the concept of particle motion, but also highlights the potential of AI-based visual generation tools in supporting the modeling process.

5.2 Insights into Students' Thinking Tendencies and Metacognition through the P.A.D. Evaluation Criteria

During the modeling process, students predominantly relied on Alignment (A) with personal experience and imagination, and Detail (D) of visual representation as their main evaluation criteria. This indicates a preference for sensory experience and subjective speculation, reflecting that the initial construction of their models was based more on intuitive perception than on conceptual reasoning. Although Plausibility (P) appeared less frequently, it was present in every modeling stage and played a crucial supporting role, guiding students to remain focused on scientific accuracy rather than being distracted by visually appealing images.

Students' ongoing revisions in response to visual feedback represented not only explicit model refinement but also the activation of metacognitive processes. In this study, students repeatedly evaluated AI-generated images and modified their verbal descriptions, exemplifying the metamodeling process emphasized by Schwarz and aligning with the principle that "models change as understanding evolves" in the modeling learning trajectory. These P.A.D.-based evaluation criteria not only provided researchers with a lens to examine how students judged images, but also revealed the cognitive shift in students' modeling—from intuitive perception toward conceptually grounded scientific reasoning.

6. Conclusion and Recommendations

This study aimed to explore how students develop models within the G.E.M. framework and to further analyze their evaluation behaviors and criteria when assessing AI-generated images during the modeling process. Based on the findings, the following conclusions and instructional recommendations are proposed as a reference for integrating Generative AI into analogy-based modeling in future science classes.

6.1 Generative AI Facilitates Model Revision and Scientific Concept Construction

Students' initial linguistic models were mostly confined to the level of macroscopic phenomena and intuitive analogies. The use of Generative AI to produce images helped students recognize the conceptual incompleteness of their models, triggering revisions and knowledge transfer. This process gradually enabled them to establish more scientifically accurate concepts and reach higher-level mental models.

6.2 The Mediating Role of Generative AI's Visual Feedback in Stimulating Evaluative Thinking and Metacognition

During the modeling process, the visual feedback provided by Generative AI enabled students to flexibly apply evaluation criteria and progressively develop metacognitive skills. Serving as a "visual feedback mediator" in multimodal learning contexts, Generative AI promoted students' reflection and adjustment of models, thereby deepening the construction of scientific meaning.
Instructional and Research Recommendations:

Instructional practice: Science teachers are encouraged to incorporate Generative AI tools into modeling instruction and integrate explicit evaluation criteria. This approach provides students with a structured framework for multi-dimensional model assessment, enhancing modeling competence while fostering scientific concept understanding.

Future research: Future studies could expand the sample size and further analyze how students' linguistic models are transformed into discipline-specific visual models with the assistance of Generative AI, in order to develop more generalizable instructional strategies.

References

[1] Luckin, R.; Holmes, W.; Griffiths, M.; Forcier, L. B. *Intelligence Unleashed: An Argument for AI in Education;* Pearson Education, 2016. Available at: https://discovery.ucl.ac.uk/id/eprint/1475756

[2] Gilbert, J. K., "Visualization: A metacognitive skill in science and science education, *in Visualization in Science Education,* Gilbert, J. K., Ed., Springer,2005; pp. 9–27. https://doi.org/10.1007/1-4020-3613-2_2

[3] Ministry of Education in Taiwan: *Curriculum Guidelines of 12-Year Basic Education for Elementary School, Junior High and General Senior High Schools: The Domain of Natural Science* (in Chinese); 2018. Available at: https://cirn.moe.edu.tw/Upload/file/38227/104346.pdf

[4] Harrison, A. G.; Treagust, D. F.: The particulate nature of matter: Challenges in understanding the submicroscopic world. In *Chemical education: Towards research-based practice;* Springer,2002; pp. 189–212.

[5] Chiu, M.-H.; Wu, W.-L.; Jong, S.-L.; Lee, S.-P.: Exploring the development of cross-grade students' mental models of ideal gases using concept evolution trees. *J. Sci. Educ.* 2013, 21 (2), 135–162 (in Chinese). http://doi.org/10.6173/CJSE.2013.2102.01.

[6] Treagust, D. F.; Chittleborough, G.; Mamiala, T. L.: Students' understanding of the role of scientific models in learning science, *Int. J. Sci. Educ.*2002, 24(4), 357–368 (2002). https://doi.org/10.1080/09500690110066485

[7] Xue, S.; Sun, D.: Integrating analogy into scientific modeling for students' active learning in chemistry education. In *Active Learning: Research and Practice for STEAM and Social Sciences Education,* Ortega-Sánchez, D., Ed., IntechOpen, 2022. https://doi.org/10.5772/intechopen.105454

[8] Schwarz, C. V.; Reiser, B. J.; Davis, E. A.; Kenyon, L.; Achér, A.; Fortus, D.; Krajcik, J.:Developing a learning progression for scientific modeling: Making scientific modeling accessible and meaningful for learners, *J. Res. Sci. Teach.* 2009, 46(6), 632–654 . https://doi.org/10.1002/tea.20311

[9] Schwarz, C. V.; White, B. Y.: Metamodeling knowledge: Developing students' understanding of scientific modeling, *Cognition Instr.*2005, 23(2), 165–205. https://doi.org/10.1207/s1532690xci2302_1

[10] Grapin, S. E.; Haas, A.; Llosa, L.; Wendel, D.; Pierson, A.; Lee, O.: Multilingual learners' epistemologies in practice in the context of computational modeling in an elementary science classroom, *J. Res. Sci. Teach.*2023, 60(9), 1998–2041. https://doi.org/10.1002/tea.21850

[11] Cheng, M. M. W.; Danielsson, K.; Lin, A. M. Y.: Resolving puzzling phenomena by the simple particle model: examining thematic patterns of multimodal learning and teaching, *Learn.: Res. Pract.*2020, 6(1), 70–87.https://doi-rg.metalib.lib.ntue.edu.tw/10.1080/23735082.2020.1750675

[12] Harrison, A. G.; Treagust, D. F.: A typology of school science models, *Int. J. Sci. Educ.*2000, 22(9), 1011–1026. https://doi.org/10.1080/095006900416884

[13] Halloun, I. A.: Mediated modeling in science education, *Sci. Educ.*2007, 16, 653–697 . https://doi.org/10.1007/s11191-006-9004-3

[14] Gilbert, J. K.: Models and modelling: Routes to more authentic science education, *Int. J. Sci. Math. Educ.*2004, 2, 115–130. https://doi.org/10.1007/s10763-004-3186-4

[15] Jackson, J.; Dukerich, L.; Hestenes, D.: Modeling Instruction: An Effective Model for Science Education. *Sci. Educ.*2008, 17 (1), 10–17. https://files.eric.ed.gov/fulltext/EJ851867.pdf

[16] Harrison, A. G.; Treagust, D. F.: Secondary Students' Mental Models of Atoms and Molecules: Implications for Teaching Chemistry. *Sci. Educ.*1996, 80 (5), 509–534. https://doi.org/10.1002/(SICI)1098-237X(199609)80:5<509::AID-SCE2>3.0.CO;2-F

[17] Lee, O.; Eichinger, D. C.; Anderson, C. W.; Berkheimer, G. D.; Blakeslee, T. D.: Changing Middle School Students' Conceptions of Matter and Molecules: A Longitudinal Study. *J. Res. Sci. Teach.*1993, 30 (3), 249–270. https://doi.org/10.1002/tea.3660300304

[18] Chen, C. Y.; Lin, J. W.: Unveiling Elementary School Teachers' Mental Models: Utilizing the Particulate Nature of Matter to Explain Water's Three States and Constructing Analogical Models for Their Students, Oral presentation at 27th IUPAC International Conference on Chemistry Education, Pattaya, Thailand, July 15–19 , 2024.

[19] Lin, J. W.: Developing Assessment and Instruction of Analogy-Based Modeling Competence to Explore Elementary School Students' Analogy-Based Modeling Competence on the Particle Model of Matter. National Science and Technology Council , 2023 (in Chinese).

[20] Merritt, J.; Krajcik, J.: Learning Progression Developed to Support Students in Building a Particle Model of Matter. In *Concepts of Matter in Science Education,* Tsaparlis, G., Sevian, H., Eds.; Springer: Dordrecht, 2013; pp 11–45. https://doi.org/10.1007/978-94-007-5914-5_2

[21] Harrison, A. G.; Coll, R. K., Eds.: *Using Analogies in Middle and Secondary Science Classrooms: The FAR Guide – An Interesting Way to Teach with Analogies,* Corwin Press: Thousand Oaks, CA, USA, 2008.

[22] Lin, J. W.; Chao, H. Y.: Developing an Instrument to Examine Students' Analogical Modeling Competence: An Example of Electricity. *Sci. Educ.* 2024, 108 (1), 63–85. https://doi.org/10.1002/sce.21828

[23] Windschitl, M.; Thompson, J.; Braaten, M.: Beyond the Scientific Method: Model - Based Inquiry as a New Paradigm of Preference for School Science Investigations. *Sci. Educ.* 2008, 92 (5), 941–967. https://doi.org/10.1002/sce.20259

[24] Murphy, D.; Duncan, R. G.; Chinn, C. A.; Danish, J.; Hmelo - Silver, C. E.; Zhou, J.; Ryan, Z.: Elementary Students' Metacognitive Knowledge of Epistemic Criteria. *J. Res. Sci. Teach.* 2025. https://doi.org/10.1002/tea.22030

[25] Pluta, W. J.; Chinn, C. A.; Duncan, R. G. Learners' Epistemic Criteria for Good Scientific Models. *J. Res. Sci. Teach.* 2011, 48 (5), 486–511. https://doi.org/10.1002/tea.20415

[26] Morris, M.: Scientists' perspectives on the potential for Generative AI in their fields. *arXiv*2023,abs/2304.01420 . https://doi.org/10.48550/arXiv.2304.01420

[27] Jain, G. P.; Gurupur, V. P.; Schroeder, J. L.; Faulkenberry, E. D.: Artificial intelligence-based student learning evaluation: A concept map-based approach for analyzing a student's understanding of a topic, *IEEE Trans. Learn. Technol.*2014, 7 (3), 267–279. https://doi.org/10.1109/TLT.2014.2330297

[28] Loeckx, J.: Blurring boundaries in education: Context and impact of MOOCs. *Int. Rev. Res. Open Distrib. Learn.* 2016, 17 (3). https://doi.org/10.19173/irrodl.v17i3.2395

[29] Yen, J.-C.: A review of the impact of Generative AI on learning from a cognitive processing Perspective. *Taiwan Educ. Rev. Mon.* 2024, 13 (3), 144–153 (in Chinese) http://www.ater.org.tw/journal/article/13-3/free/09.pdf

[30] Khan, S.: Model - based inquiries in chemistry. *Sci. Educ.* 2007, 91 (6), 877–905. https://doi.org/10.1002/sce.20226

[31] Lemke, J.: Multiplying meaning: Visual and verbal semiotics in scientific text. In *Reading Science: Critical and Functional Perspectives on Discourses of Science;* Martin, J. R., Veel, R., Eds.,Routledge: London, 1998; pp 87–113.

[32] Landis, J. R.; Koch, G. G.: The Measurement of Observer Agreement for Categorical Data. *Biometrics* 1977, 33 (1), 159–174. https://doi.org/10.2307/2529310.

Long-term Water Quality Trends (1994–2024) in the Tamsui River: Application of the Mann-Kendall Test and Sen's slope Estimator

JONG Jingping[1*], **CHUNG Chunwei**[2]

[1] *New Taipei Municipal Jinhe High School, New Taipei City, Taiwan*

[2] *Nanshan High School, New Taipei City, Taiwan*

**porphyrin@jhsh.ntpc.edu.tw*

Abstract

Long-term water quality monitoring and trend analysis are crucial for effective river basin management. This study employed the Mann-Kendall test and Sen's slope estimator to analyze long-term trends in water quality in the Tamsui River Basin, northern Taiwan, from 1994 to 2024. The River Pollution Index (RPI) and four key water quality parameters—dissolved oxygen (DO), five-day biochemical oxygen demand (BOD_5), suspended solids (SS), and ammonia nitrogen (NH_3-N)—were used to assess water quality conditions. The results indicate a significant improvement in overall water quality within the Tamsui River basin. The proportion of monitoring stations classified as "unpolluted/slightly polluted" and "mildly polluted" has increased significantly, while the proportion of stations categorized as "moderately polluted" and "severely polluted" has declined. In terms of specific parameters, DO concentrations have increased, whereas BOD_5, SS, and NH_3-N concentrations have decreased. However, despite these improvements, monitoring stations located in highly urbanized and densely populated areas remain at moderate pollution levels, as indicated by the River Pollution Index. Therefore, implementing more effective water quality management strategies in urban regions is recommended to sustain and further enhance water quality conditions in the Tamsui River basin.

Keywords: Trend analysis, Mann-Kendall test, Sen's slope estimator, River Pollution Index, the Tamsui River

Introduction

The Tamsui River basin is located approximately between 24°40' to 25°15' N and 121°10' to 122°00' E, covering the northern region of Taiwan. Its primary tributaries include the Xindian River, Tamsui River, and Keelung River, flowing through urban areas such as Taipei City, New Taipei City, and Taoyuan City (Chen et al., 2019). Due to the basin's passage through densely developed urban zones, the river water quality has been long affected by industrial wastewater, domestic sewage, and agricultural pollution (e.g., Huang, 2021; Jang, 2016), leading to the accumulation of pollutants and subsequent impacts on water quality and ecosystems.

To improve the water quality of the Tamsui River basin, the government in Taiwan implemented the River Restoration Plan, which involves constructing wastewater treatment plants (e.g., Taipei City Neihu Wastewater Treatment Plants) and executing sewage interception projects to reduce pollutant discharge into the river (He-Tuo Planning and Design Consulting Co., Ltd., 2012). Despite years of remediation efforts, whether the water quality of the Tamsui River has significantly improved remains to be further examined. Consequently, water quality monitoring has become a critical basis for evaluating the effectiveness of river restoration measures. The core objective of river water quality monitoring is to understand the current pollution status and long-term trends, providing a scientific foundation for water quality management and pollution prevention strategies. This ensures the continuous improvement of the Tamsui River's water quality and promotes the sustainable development of the aquatic environment.

Literature Review

River Pollution Index: a quantitative assessment of water quality status

Water Quality Index serves as a comprehensive indicator that integrates multiple water parameters along with their respective weights to provide a quantitative assessment of water quality status (Chidiac et al., 2023; Miah et al., 2025). In Taiwan, the Ministry of Environment (MOE) widely employs the River Pollution Index (RPI) to evaluate the degree of river pollution. The calculation of RPI is based on the concentrations of four key water quality parameters: dissolved oxygen (DO), biochemical oxygen demand over five days (BOD_5), suspended solids (SS), and ammonia nitrogen (NH_3-N) (Ministry of Environment, Taiwan, 2025).

SS refers to suspended solids in water. Higher SS values indicate a greater concentration of undissolved solid particles in the water, which reduces the light penetration rate of the water body. This, in turn, affects the photosynthesis of aquatic plants and the transparency of the water. BOD_5 is used to measure the concentration of organic pollutants in the water. An increase in BOD_5 values signifies that micro-organisms require more oxygen to break down organic matter. This may lead to a decrease in dissolved oxygen levels in the water, further impacting the living environment of aquatic organisms. NH_3-N primarily originates from the fermentation and decomposition of nitrogen-containing organic matter. When its concentration rises, it often indicates organic pollution in the water, which can cause toxicity in aquatic organisms, eutrophication of the water body, and deterioration of water quality. DO reflects the amount of oxygen present in the water. Adequate dissolved oxygen helps reduce odors produced by the decomposition of NH_3-N and other organic matter (World Health Organization, 2003). Therefore, higher DO concentrations, accompanied by lower BOD_5, SS, and NH_3-N values, generally indicate purer water quality, suitable for the survival of aquatic organisms and human recreational activities. Since RPI is calculated based on indicators such as DO, BOD_5, SS, and NH_3-N, trend analysis can reveal long-term patterns of RPI changes, further assessing improvements or deteriorations in river pollution levels (Jang, 2016).

The formula for RPI is expressed as: $RPI = 1/4 \sum S_i$, where Si represents the pollution score for each parameter, *i* refers to the water quality parameter, and RPI ranges between 1 and 10. For instance, at the monitoring station near the Tamsui River estuary in December 2003, the observed data were as follows: DO at 3 mg/L, BOD_5 at 2.2 mg/L, SS at 13.6 mg/L, and NH_3-N at 1.54 mg/L. As shown in Table 1, the corresponding pollution scores for these parameters were 6, 1, 1, and 6, respectively. Averaging these scores yields an RPI value of 3.5, which falls within the range of 3.1–6.0, indicating moderately polluted (Ministry of Environment, Taiwan, 2025).

Table 1. The criteria of River Pollution Index (Adapted from Ministry of Environment, Taiwan, 2025)

Water Quality / Parameter	Unpolluted /Slightly Polluted	Mildly Polluted	Moderately Polluted	Severely Polluted
DO (mg/L)	$DO \geq 6.5$	$6.5 > DO \geq 4.6$	$4.5 \geq DO \geq 2.0$	$DO < 2.0$
BOD_5 (mg/L)	$BOD_5 \leq 3.0$	$3.0 < BOD_5 \leq 4.9$	$5.0 \leq BOD_5 \leq 15.0$	$BOD_5 > 15$
SS (mg/L)	$SS \leq 20$	$20 < SS \leq 49.9$	$50 \leq SS \leq 100$	$SS > 100$
NH_3-N (mg/L)	$NH_3\text{-}N \leq 0.50$	$0.50 < NH_3\text{-}N \leq 0.99$	$1.0 \leq NH_3\text{-}N \leq 3.0$	$NH_3\text{-}N > 3.0$
Score(s)	1	3	6	10
Pollution Index Score (S)	$S \leq 2.0$	$2.0 < S \leq 3.0$	$3.1 \leq S \leq 6.0$	$S > 6.0$

Note: DO = Dissolved Oxygen; BOD_5 = Biochemical Oxygen Demand;
SS = Suspended Solids; NH_3-N = Ammoniacal Nitrogen

Trend Analysis of Water Quality

Trend analysis is a method used to explore changes in data over a specific time range (Antonopoulos et al., 2001). Water quality trend analysis enables the examination of long-term data to observe changes in specific water quality parameters within a defined timeframe, determining whether the values of water quality parameters show an upward or downward trend during the monitoring period. This can help predict future water quality conditions (Hashim et al., 2021; Tabari et al., 2011) and provide references for water resource management and decision-making (Modi et al., 2024).

To effectively evaluate long-term trends in water quality changes, researchers widely employ statistical methods to identify the significance and direction of changes. Among these methods, the Mann-Kendall test and Sen's slope estimator have become primary tools for trend analysis due to their applicability to non-normally distributed data. The Mann-Kendall test detects the significance of data trends, while Sen's slope estimator quantifies the magnitude of changes. Together, they provide a more precise evaluation of water quality changes (Hashim et al., 2021; Modi et al., 2024; Tabari et al., 2011).

Mann-Kendall test and Sen's slope estimator for time-series data

The Mann-Kendall test, a non-parametric method, was employed to detect monotonic trends (either increasing or decreasing) in the time-series data. To further quantify the rate of change in trends, Sen's Slope method was applied. This method estimates the median slope of change in time-series data and is less sensitive to the influence of outliers.

Many studies have applied the Mann-Kendall test and Sen's slope estimator to analyze long-term water quality changes in various regions (Hashim et al., 2021; Modi et al., 2024; Tabari et al., 2011). Modi et al. (2024) analyzed changes in water quality across seven monitoring stations along the Godavari River in India over the period from 1981 to 2005. Their study examined parameters including total alkalinity (Alk), calcium (Ca^{2+}), chloride (Cl^-), total hardness (Hard), magnesium (Mg^{2+}), sodium (Na^+), and sulfate (SO_4^{2-}) ions. Using the Mann-Kendall test and Sen's slope estimator, the results showed that Alk, Ca^{2+}, and Hard exhibited upward trends, Cl^-, Na^+, and SO_4^{2-} showed downward trends, while Mg^{2+} displayed mixed changes. Tabari et al. (2011) investigated 16 water quality parameters at four monitoring stations within the Maroon River basin over the period from 1989 to 2008. Employing the Mann-Kendall test, Sen's slope estimator, and linear regression for trend analysis, their findings revealed that the concentrations of water quality parameters increased during spring and winter, while decreasing during summer and autumn, primarily due to river dilution effects. Parameters such as calcium, magnesium, sodium adsorption ratio, pH, and turbidity showed significant changes. Additionally, most water quality parameters were negatively correlated with river flow, demonstrating the significant impact of flow variations on water quality. Hashim et al. (2021) analyzed water quality changes in the upstream region of the Bernam River basin in Malaysia from 1998 to 2018, focusing on six water quality indicators: DO, BOD_5, COD, NH_3-N, TSS (total suspended solids), and pH. Using the Mann-Kendall test and Sen's slope estimator, the results showed that most monitoring stations exhibited a downward trend in water quality index (WQI), while DO, BOD_5, NH_3-N, and pH increased, and COD and TSS decreased, reflecting changes in water quality. Due to significant land-use changes upstream, this study serves as a reference for water pollution prevention and water resource management.

The purpose of this study

Although studies have employed the Mann-Kendall test and Sen's slope estimator to analyze water quality changes in different regions, most focus on specific water quality parameters, such as total alkalinity, hardness, or dissolved oxygen, with less attention given to comprehensive water quality evaluation indicators like the RPI. Particularly in Taiwan, existing studies lack a complete analysis of long-term water quality trends in the Tamsui River basin and a systematic evaluation of water quality changes over the past 30 years.

This study aims to fill the research gap by using trend analysis to evaluate long-term water quality changes in the Tamsui River basin from 1994 to 2024, focusing on the River Pollution Index (RPI) and related water quality parameters. By applying the Mann-Kendall test and Sen's slope estimator, this study will analyze the changes in the number of monitoring stations with varying pollution levels over different years, assess the trends in RPI and water quality parameters, test their statistical significance, and quantify the rate of change. The goal is to deeply explore the trends and corresponding influencing factors of water quality changes in the Tamsui River, providing a more comprehensive reference for sustainable water resource management.

Methods

This study utilized 31 years of monitoring data (1994–2024) for the Tamsui River basin, sourced from Taiwan's National Environmental Water Quality Monitoring Information Network. The dataset included the RPI and its constituent parameters: DO, BOD_5, SS, and NH_3–N. The analysis proceeded in two main stages: data processing, followed by trend analysis with statistical testing.

Data processing

The data collected from 1994 to 2024 for the Tamsui River basin were organized and processed using Microsoft Excel. Monitoring stations were categorized into different pollution levels based on the criteria of RPI, DO, BOD_5, SS, and NH_3-N values (See Table 1).

Trend analysis and statistical testing

Mann-Kendall test The Mann-Kendall test was conducted using the R programming language to evaluate the statistical significance of trends in RPI and water quality parameters over the period from 1994 to 2024.

Sen’s slope estimation Sen’s slope estimation method was applied in R to quantify the rate of change during the same period, providing a detailed assessment of the extent of water quality variation in the Tamsui River basin.

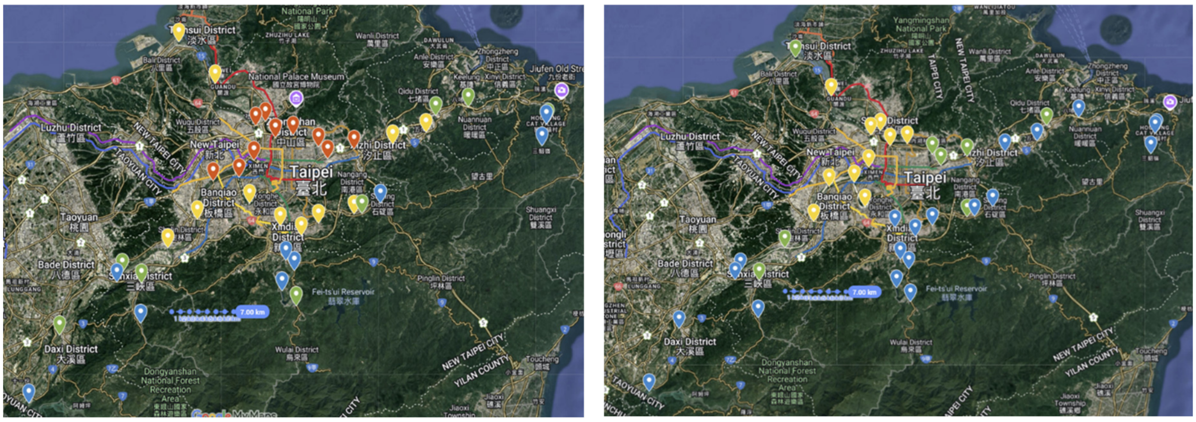
Code refinement The R script was developed through an iterative process guided by our literature review and research questions. We began by generating baseline code for the Mann-Kendall test and Sen's slope with ChatGPT. This script then underwent a cycle of refinement, where it was tested in R and subsequently improved with assistance from ChatGPT to optimize its performance for analyzing the RPI and its constituent parameters.

Results

Analysis of RPI variations across different years in the Tamsui River basin

Visualizing the differences between monitoring stations in 1994 and 2024 As shown in Figure 1, the different levels of RPI are represented by blue, green, yellow, and red, indicating not (or slightly) polluted, mildly polluted, moderately polluted, and severely polluted, respectively. The figure also illustrates the spatial distribution of RPI across monitoring stations for the years 1994 and 2024.

From the figure, notable improvements in water quality can be observed. In 1994, 27% of the monitoring stations (10 out of 37) were marked in red, indicating severely polluted. By 2024, this proportion decreased to 0% (0 out of 37 stations), with most of these stations transitioning to yellow, indicating moderately polluted (22%, 8 out of 37 stations), or green, indicating mildly polluted (5%, 2 out of 37 stations). Additionally, the proportion of stations marked in blue, indicating not (or slightly) polluted, increased significantly from 24% (9 out of 37 stations) in 1994 to 54% (20 out of 37 stations) in 2024. These results highlight some improvements in the water quality of the Tamsui River basin over the past 30 years, as evidenced by the decreasing proportion of stations classified as severely polluted and the increasing proportion of stations classified as not (or slightly) polluted.



Note: Blue represents RPI values of 1–2, indicating not (slightly) polluted; green represents RPI values of 2–3, indicating mildly polluted; yellow represents RPI values of 3–6, indicating moderately polluted; red represents RPI values above 6, indicating severely polluted

Figure 1. RPI of monitoring stations in 1994 (left) and in 2024 (right)

Statistical differences between monitoring stations in 1994 and 2024 The results of the Mann-Kendall trend test and Sen's slope estimates for different levels of RPI are presented in Figure 2 and Table 2. The Mann-Kendall test statistics reveal significant trends across all pollution levels.

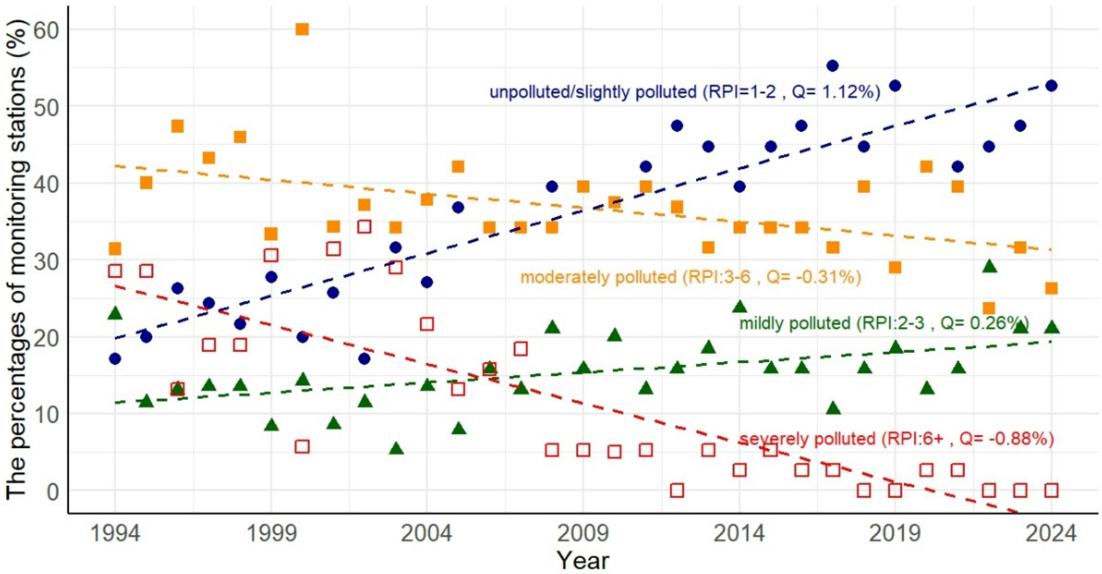


Figure 2. Trends in the percentage of monitoring stations at different pollution levels in the Tamsui River (1994–2024), analyzed using the Mann-Kendall test and Sen's slope estimator.

For unpolluted/slightly polluted, the test revealed a strong increasing trend ($\tau = 0.75$, $S = 342$, $Z = 5.82$, $p < .001$) with a Sen's slope of 1.12% per year (95% CI [0.89, 1.32] %). Similarly, for mildly polluted, a significant positive trend was observed ($\tau = 0.36$, $S = 160$, $Z = 2.73$, $p = .006$) with a Sen's slope of 0.26% per year (95% CI [0.08, 0.46] %).

In contrast, a significant decreasing trend was detected for moderately polluted ($\tau = -0.33$, $S = -148$, $Z = -2.52$, $p = .012$). The Sen's slope estimate was -0.31% per year (95% CI [-0.58, -0.01] %), indicating a decline in moderately polluted levels over time. For severely polluted, the strongest negative trend was observed ($\tau = -0.71$, $S = -317$, $Z = -5.42$, $p < .001$) with a Sen's slope of -0.88% per year (95% CI [-1.23, -0.53] %), suggesting a significant reduction in severely polluted levels.

Overall, the findings indicate a positive improvement in water quality, characterized by decreasing trends in moderately polluted and severely polluted levels, alongside increasing trends in unpolluted/slightly polluted and mildly polluted levels.

Table 2. Statistical results for different levels of RPI obtained through the Mann-Kendall test and Sen's slope estimator

Pollution Level	τ	S	Z	*p*	Slope (%)	95% CI Lower (%)	95% CI Upper (%)	Trend
Unpolluted /Slightly Polluted	0.751	342	5.82	0.000	+1.119	0.887	1.318	significant increase
Mildly Polluted	0.358	160	2.73	0.006	+0.260	0.075	0.456	significant increase
Moderately Polluted	-0.329	-148	-2.52	0.012	-0.314	-0.583	-0.005	significant decrease
Severely Polluted	-0.711	-317	-5.42	0.000	-0.877	-1.228	-0.527	significant decrease

The changes in RPI parameters over different years in the Tamsui River basin

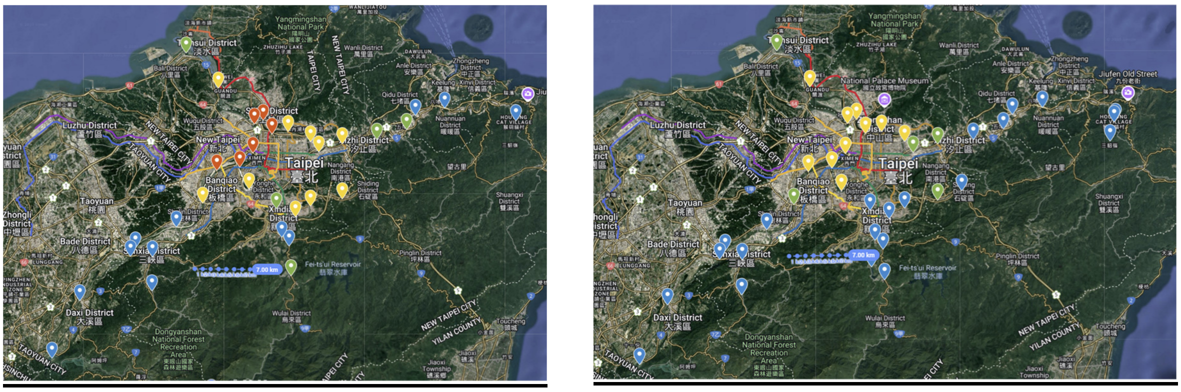
Visualizing the differences between monitoring stations in 1994 and 2024 As illustrated in Figures 3, in 1994, DO levels were generally low, particularly in urbanized areas, with approximately 30% of monitoring stations (11 out of 37) showing severely polluted oxygen conditions (DO < 2 mg/L). By 2024, the proportion of stations with severely polluted DO levels decreased dramatically to 0% (0 out of 37 stations), and most monitoring stations achieved good DO levels (DO ≥ 6.5 mg/L), accounting for 60% of total stations (22 out of 37).

The changes in BOD_5 are illustrated in Figures 4. In 1994, about 5% of monitoring stations (2 out of 37) recorded BOD_5 values exceeding 15 mg/L, and about 59% of monitoring stations (22 out of 37) achieved BOD_5 values within 5 to 15 mg/L, indicating organic pollution at the time. After 30 years of pollution control, the 2024 data shows that the proportion of high BOD_5 stations (> 5 mg/L) decreased to 3% (1 out of 37 stations), while the proportion of stations with acceptable BOD_5 levels (< 5 mg/L) increased from 35% (13 out of 37 stations) in 1994 to 97% (36 out of 37 stations) in 2024.

The distribution of SS is presented in Figure 5. In 1994, approximately 32% of the monitoring stations (12 out of 37) recorded SS levels exceeding 50 mg/L, with the majority of these stations concentrated in downstream river areas. By 2024, the proportion of stations with SS levels exceeding 50 mg/L had decreased to 0% (0 out of 37 stations). Furthermore, most monitoring stations in 2024 reported SS levels below 20 mg/L, accounting for 70% of the total stations (26 out of 37).

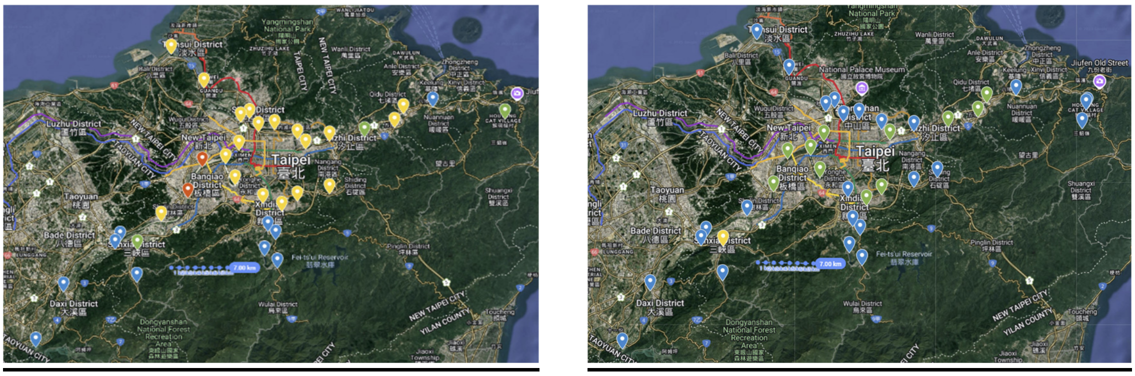
The changes in NH_3-N are shown in Figures 6. In 1994, nearly 46% of monitoring stations (17/37 stations) recorded NH_3-N exceeding 3 mg/L, reflecting insufficient domestic wastewater treatment at the time. After 30 years of sewerage system development, the 2024 data shows that the proportion of high ammonia nitrogen stations (>3 mg/L) decreased to 0% (0 out of 37 stations), while the proportion of stations with good NH_3-N levels (< 1 mg/L) increased from 32% (12 out of 37 stations) in 1994 to 73% (27 out of 37 stations) in 2024.

These results demonstrate some improvements in water quality across the Tamsui River basin over the past 3 decades, particularly in terms of increased dissolved oxygen levels and reduced organic pollutants.



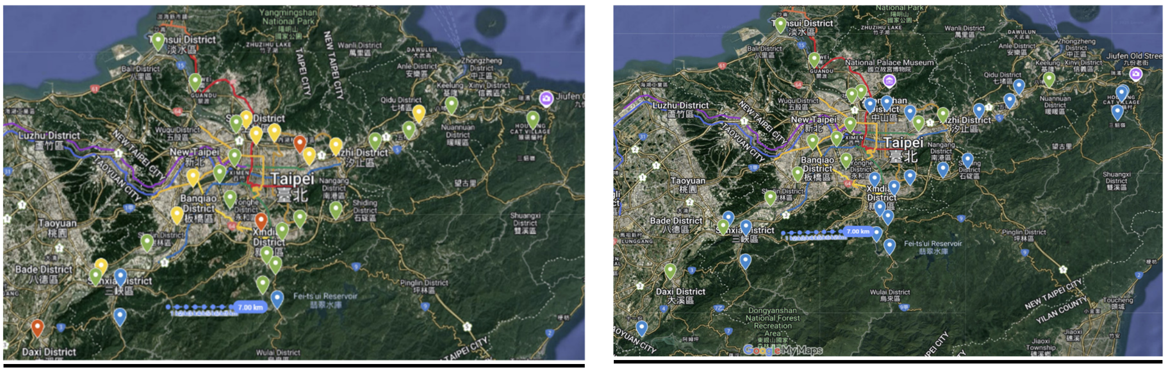
Note: Blue represents DO values above 6.5 ppm, indicating not (slightly) polluted; green represents DO values of 4.6–6.5 ppm, indicating mildly polluted; yellow represents DO values of 2–4.5 ppm, indicating moderately polluted; red represents DO values below 2 ppm, indicating severely polluted

Figure 3: DO of monitoring stations in 1994 (left) and in 2024 (right)



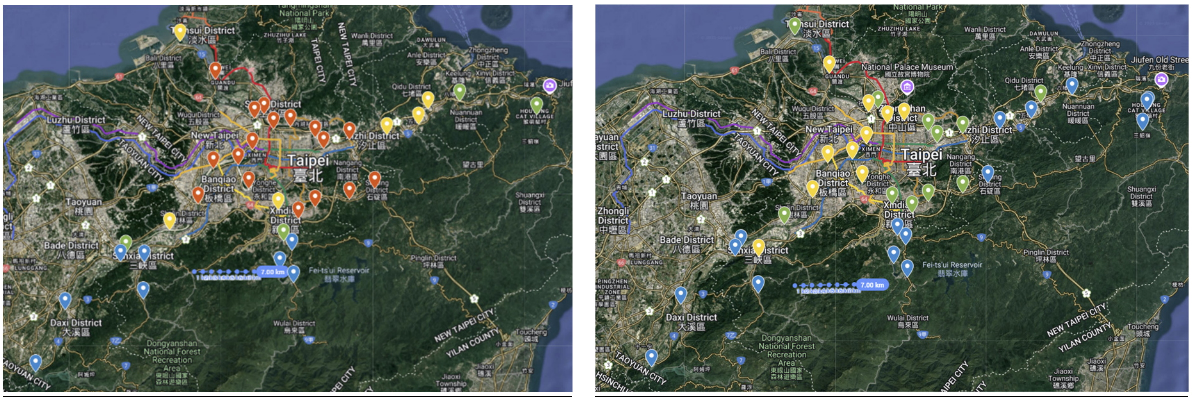
Note: Blue represents BOD_5 values below 3 ppm, indicating not (slightly) polluted; green represents BOD_5 values of 3–4.9 ppm, indicating mildly polluted; yellow represents BOD_5 values of 5–15 ppm, indicating moderately polluted; red represents BOD_5 values above 15 ppm, indicating severely polluted

Figure 4: BOD_5 of monitoring stations in 1994 (left) and in 2024 (right)



Note: Blue represents SS values below 20 ppm, indicating not (slightly) polluted; green represents SS values of 20–49.9 ppm, indicating mildly polluted; yellow represents SS values of 50–100 ppm, indicating moderately polluted; red represents SS values above 100 ppm, indicating severely polluted

Figure 5: SS of monitoring stations in 1994 (left) and in 2024 (right)



Note: Blue represents NH_3-N values below 0.5 ppm, indicating not (slightly) polluted; green represents NH_3-N values of 0.5–0.99 ppm, indicating mildly polluted; yellow represents NH_3-N values of 1.0–3.0 ppm, indicating moderately polluted; red represents NH_3-N values above 3.0 ppm, indicating severely polluted

Figure 6. NH_3-N of monitoring stations in 1994 (left) and in 2024 (right)

Statistical differences in corresponding parameters from 1994 to 2024 The trend in DO levels over the period from 1994 to 2024 was analyzed using the Mann-Kendall test and Sen's slope estimator. As illustrated in Figure 7 and Table 3, the Kendall's tau (τ) for the DO levels was 0.485, indicating a moderate positive trend over the 31-year period. The S statistic was 225, with a corresponding Z value of 3.81. The *p*-value for the Mann-Kendall test was highly significant ($p < 0.001$), confirming that the observed trend is statistically significant.

Sen's slope for the dissolved oxygen levels was calculated to be 0.071, indicating an average increase in oxygen levels of 0.071 mL per year. The 95% confidence interval for the Sen's slope ranged from 0.043 to 0.096, further supporting the robustness of the positive trend observed. These results suggest a significant upward trend in the DO levels over the study period. The analysis confirms that DO levels in the study area have significantly increased over the past three decades.

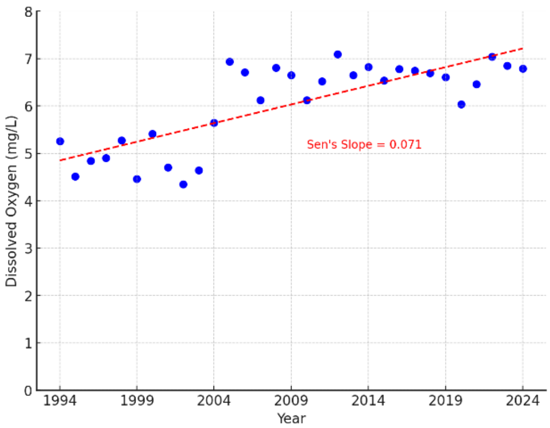


Figure 7. The trend of annual mean DO from 1994 to 2024.

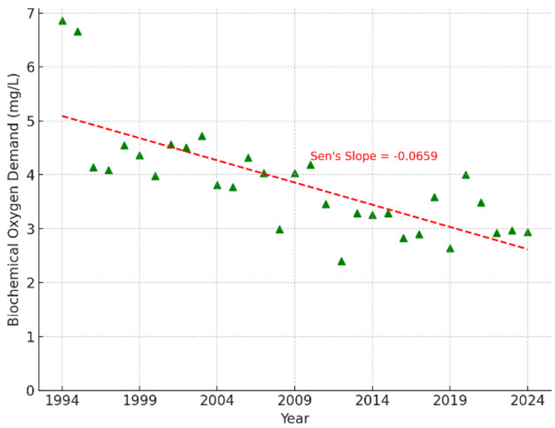


Figure 8. The trend of annual average BOD_5 from 1994 to 2024.

As shown in Figure 8 and Table 3, according to the results of the Mann-Kendall test, BOD_5 exhibited a significant negative trend ($\tau = -0.584$, $p < 0.001$), indicating a notable decline in BOD_5 over time. The results of the Sen's slope method showed a slope of -0.066 for BOD_5, with a 95% confidence interval of [-0.096, -0.044], suggesting an average annual decrease in BOD_5 of -0.066. This change was statistically significant ($p < 0.001$), further confirming the declining trend in BOD_5 values during the study period.

The trend analysis of BOD_5 indicates a significant decline, which is statistically significant. This negative correlation with pollution levels likely reflects an improvement in water quality.

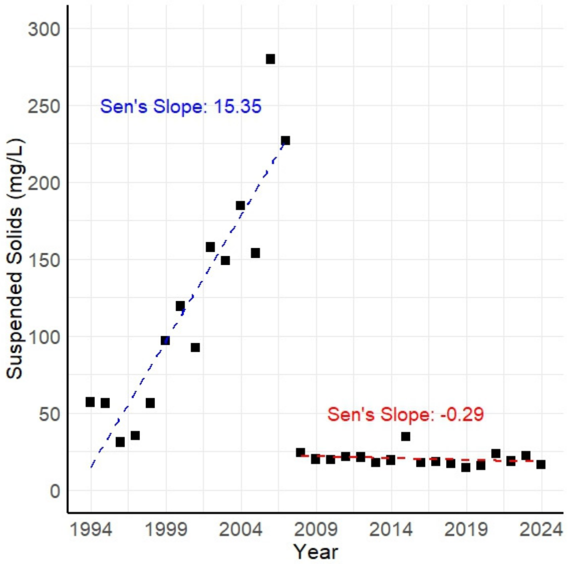


Figure 9. The trend of annual mean SS from 1994 to 2024.

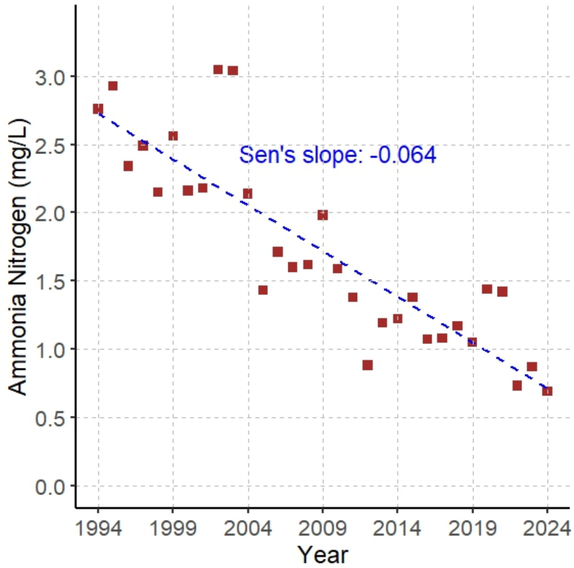


Figure 10. The trend of annual mean NH_3-Nfrom 1994 to 2024

The trend analysis of SS is divided into two periods: 1994-2007 and 2008-2024. As shown in Figure 9 and Table 3, the first segment (1994-2007) shows an increasing trend with a Sen's slope estimate of 15.35 mg/L per year. In contrast, the second segment (2008-2024) presents a decreasing trend with a Sen's slope estimate of -0.29 mg/L per year. These trends are depicted with separate dashed regression lines, each colored differently to distinguish the two periods. The statistical significance of the trends was evaluated using Kendall's Tau, with the earlier period displaying a statistically significant upward trend, while the later period indicates a weaker downward trend.

This analysis provides insight into the temporal variations in SS concentration, potentially reflecting changes in environmental conditions, land use, or pollution control measures over the decades.

As shown in Figure 10 and Table 3, The analysis revealed a significant negative trend in NH_3-N concentrations, with a Kendall's tau (τ) of -0.715, indicating a moderate inverse relationship between the years and NH_3-N levels ($Z = -5.63$, $p < 0.001$). The Sen's slope estimate for the trend was -0.064, suggesting a consistent decrease in NH_3-N concentrations over time. The 95% confidence interval for the Sen's slope ranged from -0.076 to -0.053, confirming the robustness of the observed downward trend. Given the *p*-value of 1.84e-08, the trend was statistically significant.

These results suggest a notable reduction in NH_3-N concentrations over the study period, and this trend was statistically significant, supporting the hypothesis of a decreasing environmental impact related to NH_3-N.

Table 3. Kendall's tau and Sen's slope for DO, BOD_5, SS, and NH_3-N

Water quality parameters	τ	*S*	*Z*	*p*	Slope	95% CI Lower	95% CI upper	trends
DO	0.485	225	3.81	0.000	0.071	0.043	0.096	significant increase
BOD_5	-0.584	-271	-4.59	0.000	-0.066	-0.096	-0.044	significant decrease
SS (1994-2007)	0.736	67	3.61	0.000	15.351	11.977	21.130	significant increase
SS (2008-2024)	-0.324	-44	-1.77	0.077	-0.290	-0.636	0.088	non-significant
NH_3-N	-0.715	-332	-5.63	0.000	-0.064	-0.076	-0.053	significant decrease

Discussion

Consistency with global practices in trend analysis

This study employed the Mann-Kendall test and Sen's slope estimator, two methods widely used for long-term water quality analysis, to investigate pollution trends in the Tamsui River. The results revealed significant trends in the RPI and in the concentrations of DO, BOD_5, SS, and NH_3-N. This methodological approach is consistent with recent research in the field. For example, Modi et al. (2024) used the same statistical tools to analyze various water quality parameters in the Godavari River, as did Hashim et al. (2021) for the Bernam River basin in Malaysia and Tabari et al. (2011) for the Maroon River in Iran. The preference for these non-parametric methods across different geographical regions and river systems highlights their robustness and suitability for environmental time-series data, which is often non-normally distributed and may contain outliers. This shared analytical framework provides a common basis for discussing and contextualizing water quality trends from diverse international river basins.

Holistic improvement in trends in contrast to specific parameter variations

A key finding of this study is the significant and consistent improvement across the comprehensive RPI and its constituent parameters in the Tamsui River over the past 31 years. The results show a significant increase in the proportion of unpolluted/slightly polluted and mildly polluted monitoring stations, coupled with a significant decrease in moderately polluted and severely polluted stations. This contrasts with findings from other studies where water quality trends were more varied. For instance, Modi et al. (2024) observed mixed trends in the Godavari River, with some parameters like alkalinity and hardness increasing while others like chloride and sodium decreased. Similarly, Hashim et al. (2021) found that while the overall water quality index in the Bernam River basin showed a downward trend at most stations, some individual parameters like DO, BOD_5, and pH actually increased.

Conclusion

An analysis of approximately three decades of monitoring data reveals a significant improvement in the overall water quality of the Tamsui River. This is evidenced by a substantial decrease in stations classified as moderately or severely polluted by the RPI, alongside a significant increase in those rated unpolluted/slightly or mildly polluted. In terms of specific parameters, DO concentrations have increased, whereas BOD_5, SS, and NH_3-N concentrations have decreased. However, despite these improvements, monitoring stations in highly urbanized and densely populated areas remain at moderate pollution levels, as indicated by the RPI. To sustain and further enhance water quality conditions in the Tamsui River basin, implementing more effective water quality management strategies in urban regions is recommended, such as establishing or expanding sewage treatment plants, adding small decentralized sewage treatment facilities in highly polluted areas of New Taipei City and Taipei City, formulating stricter discharge standards, and strengthening the regulation of pollution sources.

Acknowledgments

This manuscript was prepared with the assistance of the generative AI tool, ChatGPT-4o, for translation, grammatical proofreading, and language refinement.

References

Antonopoulos, V., Papamichail, D., & Mitsiou, K. (2001). Statistical and trend analysis of water quality and quantity data for the Strymon River in Greece. *Hydrology and Earth System Sciences, 5*(4), 679–691. https://doi.org/10.5194/hess-5-679-2001

Chen, S. K., Jang, C. S., & Chou, C. Y. (2019). Assessment of spatiotemporal variations in river water quality for sustainable environmental and recreational management in the highly urbanized Danshui River basin. *Environmental Monitoring and Assessment, 191*(100). https://doi.org/10.1007/s10661-019-7246-1

Chidiac, S., El Najjar, P., Ouaini, N., El Rayess, Y., & El Azzi, D. (2023). A comprehensive review of water quality indices (WQIs): History, models, attempts, and perspectives. *Reviews in Environmental Science and Biotechnology, 22*(2), 349–395. https://doi.org/10.1007/s11157-023-09650-7

Gilbert, R. O. (1987). *Statistical methods for environmental pollution monitoring*. John Wiley & Sons.

Hashim, M., Nayan, N., Setyowati, D. L., Mat Said, Z., Mahat, H., & Saleh, Y. (2021). Analysis of water quality trends using the Mann-Kendall test and Sen's estimator of slope in a tropical river basin. *Pollution, 7*(4), 933–942. https://doi.org/10.22059/POLL.2021.325794.1118

Heto Planning & Design Consulting Co., Ltd. (2012). Final report on the commissioned project for integrating urban development and basin disaster prevention in the Tamsui River basin in response to global warming and climate abnormalities (Chapter 3: Basic investigation and analysis of the Tamsui River basin). Taipei City: Urban Development Bureau, Taipei City. (In Chinese)

Huang, C. Y. (2021). Who is still polluting the Tamsui River? Revealing the three major sources of pollution｜2021 investigative report on doing something for the Tamsui River #1. CSR@CommonWealth. https://csr.cw.com.tw/article/44029 (In Chinese)

Jang, C. S. (2016). Using probability-based spatial estimation of the river pollution index to assess urban water recreational quality in the Tamsui River watershed. *Environmental Monitoring and Assessment, 188*(36). https://doi.org/10.1007/s10661-015-5040-2

Miah, O., Anik, A. H., Sorker, R., Parvin, F., Shammi, M., & Tareq, S. M. (2025). Impacts of rapid urbanization on long-term water quality of the peripheral River of Dhaka, Bangladesh. *Water and Environment Research.* https://doi.org/10.1002/wer.70000

Modi, P., & Chintalacheruvu, M. (2024). Investigating river water quality assessment through non-parametric analysis: A case study of the Godavari River in India. *The Quality of Environment, 34*(4), 239-264. https://doi.org/10.1002/tqem.22117

Tabari, H., Marofi, S., & Ahmadi, M. (2011). Long-term variations of water quality parameters in the Maroon River, Iran. *Environmental Monitoring and Assessment, 177*(1–4), 273–287.

Ministry of Environment, Taiwan (2025). *Environmental Water Quality Information.* https://wq.moenv.gov.tw/EWQP/en/Default.aspx

World Health Organization (2003). *Guidelines for safe recreational water environments: Vol. 1. Coastal and fresh waters.* World Health Organization.

Bilingual Analogical Modeling and Epistemic Engagement: Exploring Elementary Students' Reasoning in Learning the Particle Model of Matter

Chiu-Wen Wang[1,2], Jing-Yi Liu[1], Chen-Yu Chen[1,3], Jing-Wen Lin[1]*

[1]*Department of Science Education, National Taipei University of Education, Taiwan*

[2]*Yong-Shun Elementary School, Taoyuan City, Taiwan*

[3]*Min-An Elementary School, New Taipei City, Taiwan*

jwlin@mail.ntue.edu.tw

Abstract

This study investigates how bilingual analogical modeling influences elementary students' scientific modeling practices and epistemic reasoning. Grounded in research on analogical reasoning and bilingual science discourse, we designed a curriculum that integrated multimodal scaffolds in bilingual and monolingual formats to support students' understanding of the particle model of matter (PMM). Two cohorts of sixth-grade students (N = 20) participated in a design-based implementation comparing the two instructional conditions, which highlighted the interactions among language, multimodality, and modeling. Classroom data revealed that monolingual learners flexibly used linguistic resources to construct and refine scientific models, displaying higher cognitive engagement but also revealing conceptual misconceptions. In contrast, bilingual learners experienced higher cognitive load, yet some benefited from analogical language scaffolds that enabled more sophisticated reasoning. The findings suggest that both language and analogy function as double-edged pedagogical tools. However, the purposeful integration of bilingual instruction and analogical modeling can provide chemistry teachers with practical strategies to scaffold students' reasoning about abstract and invisible scientific concepts such as PMM, thereby enhancing both conceptual understanding and classroom practice.

Keywords: Analogical Modeling, Bilingual Instruction, Particle Model of Matter (PMM), Modeling Competence, Translanguaging

1.Introduction

In recent years, science education has increasingly emphasized the role of modeling as a central epistemic practice through which students construct, test, and revise representations of scientific phenomena[1][2]. At the elementary level, however, fostering students' modeling competence remains challenging due to the abstract nature of many scientific concepts and students' developing cognitive and linguistic resources. This is especially true in the context of the Particle Model of Matter (PMM), which requires learners to reason across macroscopic experiences and microscopic explanations—an epistemic move that often exceeds the intuitive grasp of young learners [3][4]. To address these challenges, science educators have turned to analogical modeling—the process of using familiar source domains to construct models of unfamiliar scientific targets—as a promising strategy for bridging students' everyday knowledge and formal scientific understanding [5][6]1, 2. While analogical modeling has been shown to support conceptual change and explanation generation, its implementation in classrooms is uneven, particularly when students lack guidance in mapping, evaluating, and revising their analogies in epistemically productive ways [7][8].

Although bilingual contexts may increase cognitive load and present additional challenges, when combined with analogical modeling and supported by structured analogical scaffolds, they may also create opportunities for deeper epistemic engagement. The increasing adoption of CLIL (Content and Language Integrated Learning) and translanguaging pedagogies [9][10] offers potential avenues for integrating language and content learning. Yet, their operationalization in inquiry-based science practices—especially in model-based reasoning tasks—remains under-theorized and under-documented[11]. Drawing on the DEAR modeling cycle (Development, Evaluation, Application, Revision) [6] 2.This study addresses the following research questions:

RQ1. What key modeling practices do students engage in when constructing scientific models of the PMM through analogical reasoning?

RQ2. How do students in bilingual and monolingual analogical modeling groups differ in their engagement across these modeling practices?

RQ3. In what ways do bilingual and monolingual analogical modeling scaffolds mediate students' conceptual transformation during the modeling of PMM?

2.Literature Review

2.1 Modeling as an Epistemic Practice in Science Education

Science is a process of constructing, describing, and explaining predictive conceptual models of natural phenomena [12].Within this process, models are used to organize theoretical entities and processes, explain and predict patterns in data, and are often integrated into theoretical frameworks [13]. Learning science thus requires strengthening both students' understanding of the nature of scientific knowledge and their ability to engage in scientific practices and discourse. Modeling is a core scientific practice, a fundamental means for constructing and communicating scientific knowledge, and a vital component of scientific literacy [12][14].Modeling competence essential for fostering students' abilities to understand natural phenomena, as well as to explain and predict them [6].During the modeling process, researchers have identified several steps, such as constructing, validating, applying, evaluating, and revising scientific models during modeling practice [6][12][13].

2.2 Analogical Reasoning as a Mechanism for Conceptual Change and Model Construction

Scientists often employ diverse analogies and mental simulations to generate and evaluate solutions, thereby developing theories and shifting scientific paradigms. These processes involve not only the direct application of knowledge but, more importantly, the creative reorganization and transformation of existing knowledge to discover new solutions [2][15]. Further, these results corroborate prior evidence that a structural alignment process underlies analogical comparison—even a brief analogical comparison task can confer relational insight, and a structural alignment process underlies such comparisons[15].3.When students establish object correspondences and relational correspondences between the base and target domains. Finally, they use the transferred inferences to test whether the model matches actual observations, thereby facilitating model revision [16].

Students, like scientists, can also demonstrate informal reasoning abilities in science learning, using analogical reasoning to overcome misconceptions and develop new understandings. However, elementary students often face constraints in abstraction and linguistic expression during model construction [2][18]. Young learners tend to rely on intuitive analogies, but require teacher scaffolding to advance toward structure-mapping analogies [5]1.This approach assists students in transforming everyday experiences into scientific concepts [7][17].Employing multiple analogies as scaffolds can link students' prior understanding of everyday events to knowledge in scientific domains[5][1].

2.3 Language, Scaffolding, and Translanguaging in Bilingual Science Modeling

Translanguaging facilitates model construction and conceptual deepening. Students may generate and reason about models in their first language (L1) and then present and express them scientifically in the target language (e.g., English). Teachers can allow students to switch languages during group discussions to ensure depth of conceptual understanding, while progressively transitioning toward disciplinary academic language for formal presentations. Such cross-linguistic comparison not only promotes conceptual restructuring but also helps students grasp the semantic nuances of scientific vocabulary across languages—an especially critical skill for chemistry teachers who must bridge macroscopic phenome.

3.Methodology

3.1 Research Design

This study employed a qualitative design-based research (DBR) approach to explore how bilingual and monolingual analogical modeling scaffolds mediated elementary students' engagement in scientific modeling practices and conceptual transformation regarding the particle model of matter (PMM). The intervention was iteratively designed and implemented across two contrasting instructional groups—bilingual analogical modeling (BA) and monolingual analogical modeling (CA)—each composed of approximately ten sixth-grade students from comparable classrooms in a public elementary school in Taiwan.

3.2 Instructional Intervention and Scaffolding Design

The intervention was structured around three core scaffolding components:

(1) **Concept Mapping Scaffolds:** Each unit incorporated a concept map linking everyday language, macroscopic phenomena, and microscopic particle explanations. These maps activated prior knowledge and bridged everyday and scientific language.

(2) **Mover Analogy and Structural Mapping Table:** A pre-designed "mover" analogy scaffold (小精靈類比) supported students in mapping attributes of the analogy to scientific features of PMM, accompanied by a three-part structural mapping table.

(3) Four-Cycle DEAR Modeling Process: Instruction followed the DEAR cycle—Development, Evaluation, Application, and Revision [6]—guiding students to generate, test, and refine analogical models through teacher scaffolds and peer dialogue.

3.3 Data Collection

Data sources included classroom video and audio recordings, instructional materials, student worksheets, and observation notes. Classroom activities were fully transcribed to capture students' verbal and gestural modeling behaviors. Student interviews were initially planned but not conducted due to logistical constraints.

3.4 Data Analysis

A codebook-based thematic analysis [20] was used to examine transcripts, worksheets, and observation logs. Four analytical dimensions were defined:

● Analogical Strategies (R1–R4: from no analogy to self-generated microscopic analogies).

- Language Use (L1–L4: from monolingual non-academic to stable bilingual code-switching).
- Multimodal Integration (M1–M4: from single modality to integrated multimodal explanation).
- Regulatory Strategies (C1–C4: from no strategy to active modality switching).

Comparative analysis was conducted between BA and CA groups to examine how scaffolds mediated students' modeling practices and conceptual transformations (RQ2, RQ3). All transcripts, worksheets, and visual artifacts were chronologically reorganized and cross-referenced with observation notes to verify alignment with specific scaffolding elements. Analysis proceeded in three phases:

- Phase 1 – Concept Mapping: Identification of cross-linguistic term matching and semantic translation between macro and micro explanations.
- Phase 2 – Mover Analogy Tasks: Analysis of analogical strategies (R2–R4) and students' language use in mapping.
- Phase 3 – DEAR Modeling Cycle: Tracing students' engagement in Development, Evaluation, Application, or Revision to document conceptual transformation trajectories.

To ensure trustworthiness and reliability, the research team conducted trial coding, iterative refinement of categories, and maintained a complete audit trail of coding decisions and quotations.

4.Results

4.1 Students' Key Modeling Practices through Analogical Reasoning

Thematic analysis revealed three recurring practices across both bilingual and monolingual groups:

(1) **Semantic Translation and Conceptual Mapping:** Students translated PMM-related terms across languages and aligned emerging conceptual understanding with scientific vocabulary.

(2) **Analogical Construction and Model Mapping:** Students drew on everyday experiences (e.g., melting, evaporation) to construct analogies and connect observable phenomena with particle behaviors.

(3) **Linguistic Construction and Model Revision:** Students revised their models by identifying mismatches between analogical expressions and observed phenomena, using language as a reflective tool.

These findings suggest that students' analogical modeling followed a cumulative trajectory of mapping, aligning, and revising. However, deeper revision toward microscopic reasoning was not automatic and required sustained scaffolding.

4.2 Differential Engagement in Modeling Practices across Bilingual and Monolingual Groups

This section compares how students in bilingual and monolingual analogical modeling groups engaged differently in the key modeling practices identified in RQ1—(T1) Semantic Translation and Conceptual Mapping, (T2) Analogical Construction and Model Mapping, and (T3) Linguistic Construction and Model Revision. Three claims summarize the contrasting patterns of engagement.

Claim 1. Semantic Translation and Conceptual Mapping: Monolingual students interpreted; bilingual students decoded.

Monolingual students more easily understood and clearly articulated the connection between macroscopic phenomena and particle-level explanations, using their native language to refine and adjust conceptual meanings.For example, A monolingual student(TP14-CA60112) initially used the analogy of "操場變大"[the playground becoming larger"］ to represent particle diffusion. When the teacher played a video on thermal expansion and contraction of gases and asked, "你的模型可以解釋這個現象嗎?"[Can your model explain this phenomenon?〕, the student responded, “No,” and revised it to, “粒子之間空隙變小，是因為冷了不想動”[The spaces between particles become smaller because they don’t want to move when it’s cold.〕The monolingual student attempted to explain the phenomenon using a model but did not deeply align it with the underlying scientific concept.

Claim 2. Analogical Construction and Model Mapping: Monolingual students invented; bilingual students aligned strategically.

Monolingual students created analogies spontaneously based on personal experiences and expressed original mappings between real-world contexts and particle behaviors. For instance, monolingual student’ analogies that compared particles to “chicken eggs” or “quail eggs” (TP09-CA60113), and “blue-shelled bird eggs” (TP09-CA60115), reflect creative and spontaneous analogical thinking. However, some analogies failed to align with core scientific concepts and lacked the potential to develop into productive scientific models.

Claim 3. Linguistic Construction and Model Revision: Monolingual students tended to remain anchored in surface-level features; Bilingual students revised their models through scaffolded alignment between language and concept.

In the modeling revision phase, bilingual students (BA) demonstrated more coherent and micro-level reasoning, particularly when supported by teacher-mediated scaffolds that guided them to align linguistic features with underlying model meanings. For example, a monolingual student (CA60115) used everyday temperature experience as a spontaneous analogy—heating a bird egg (as shown in the figure)—during the model revision stage. Although the description was fluent in language, it remained at the descriptive level, without further elaboration or coordination with particle behavior. The conceptual change stayed at the macroscopic level, making it impossible to carry out a meaningful revision of the initial model.

4.3 Modeling Scaffolds as Mediators of Conceptual Transformation

Analogical modeling scaffolds mediated conceptual transformation differently across groups. Bilingual students benefitted from structured linguistic and visual scaffolds that supported representational shifts, while monolingual students relied more on self-generated analogies, which sometimes reinforced misconceptions.

Table 1 Modeling Scaffolds & Processes as Mediators that Facilitate Conceptual Transformation

Modeling Scaffolds	**Monolingual students**	**Bilingual students**
Concept map	Step-by-step mapping– with both pre-designed and spontaneous analogies	Step-by-step mapping– with pre-designed analogies
Scaffolded reinterpretation of prior analogies (小精靈類比＝mover analogy)	Monolingual students often generate spontaneous analogies due to their linguistic familiarity, making pre-designed scaffolds less necessary—though such analogies may risk reinforcing misconceptions	Pre-designed analogies serve as scaffolding to support bilingual students in evaluating the appropriateness of models.
Four-Phase Cyclical Modeling Process (DEAR)	Model construction integrated with multimodal representations (BA60105)	Accurate language mapping and causal reasoning (CA60115)

4.3.1. Concept map

Structural mapping tables supported alignment of macroscopic phenomena with microscopic explanations. Bilingual students relied more on preset analogies, while monolingual students flexibly combined preset and spontaneous analogies.

4.3.2 Scaffolded Reinterpretation of Prior Analogies

Monolingual students often favored their own spontaneous analogies, which sometimes conflicted with scientific concepts and required strong teacher intervention. Bilingual students relied on pre-designed analogies to evaluate appropriateness, reducing the risk of misconceptions.

4.3.3 Four-Phase DEAR Modeling Cycle

The DEAR cycle promoted iterative development, evaluation, application, and revision. Peers collaborated in reasoning and model critique, while teacher scaffolds guided alignment between language and concept. Bilingual students particularly benefitted from structured prompts to reach microscopic reasoning.

5.Conclusions and Teaching Implications

This study examined how elementary students engaged in analogical modeling to learn the particle model of matter (PMM), comparing bilingual and monolingual instructional conditions. Findings across the three research questions provide insights into students’ modeling practices, the role of language, and the scaffolds mediating conceptual transformation.

5.1 Analogical modeling follows a cumulative trajectory but requires scaffolding

Students engaged in three recurring practices—semantic translation and conceptual mapping, analogical construction, and model revision. This trajectory suggests that deep conceptual transformation is not automatic but requires multi-lesson sequences with checkpoints for evaluation and revision (e.g., the DEAR cycle). One-off analogy activities are insufficient, as sustained scaffolding is necessary for students to progress from surface analogies to particle-level reasoning.

5.2 Distinct pathways for bilingual and monolingual learners

Bilingual students relied on structured supports (concept maps, word walls, mapping tables) to progressively decode and align analogy structures with particle-level concepts. Monolingual students, while fluent in generating creative analogies, often remained at the macroscopic level without explicit guidance. Teachers should adopt differentiated scaffolding:

(1) For bilingual learners: emphasize structural mapping tables, cross-language prompts, and bilingual sentence frames to lower cognitive load and foster abstraction.

(2) For monolingual learners: encourage creative invention, but follow up with guiding questions that press for causal reasoning and microscopic alignment.

5.3 Scaffolding mediates conceptual transformation differently across groups

Three scaffolds were especially influential: (1) bilingual concept maps, (2) the mover analogy structural mapping table, and (3) the DEAR cycle for sustained revision. These mediated learning in different ways: bilingual learners benefited most from translation-based scaffolds, while monolingual learners needed prompts that restructured intuitive analogies into coherent particle models. Across groups, teachers should integrate visual and multimodal supports to connect observable phenomena with the invisible particle world, a central challenge in chemistry education.

Reference

1. Gilbert, J. K.; Justi, R., Models of modelling. In Modelling-based teaching in science education, Springer: 2016; pp 17-40.
2. Clement, J. J., Creative model construction in scientists and students. Springer: 2008.
3. Harrison, A. G.; Treagust, D. F., The particulate nature of matter: Challenges in understanding the submicroscopic world. In Chemical education: Towards research-based practice, Springer: 2002; pp 189-212.
4. Papageorgiou, G.; Johnson, P., Do particle ideas help or hinder pupils' understanding of phenomena? International Journal of Science Education 2005, 27 (11), 1299-1317.
5. Chiu, M. H.; Lin, J. W., Promoting fourth graders' conceptual change of their understanding of electric current via multiple analogies. Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching 2005, 42 (4), 429-464.
6. Chiu, M.-H.; Lin, J.-W., Modeling competence in science education. Disciplinary and Interdisciplinary Science Education Research 2019, 1 (1), 12.
7. Richland, L. E.; Holyoak, K. J.; Stigler, J. W., Analogy use in eighth-grade mathematics classrooms. Cognition and instruction 2004, 22 (1), 37-60.
8. Gentner, D.; Colhoun, J., Analogical processes in human thinking and learning. In Towards a theory of thinking: Building blocks for a conceptual framework, Springer: 2009; pp 35-48.
9. Coyle, D.; Hood, P.; Marsh, D., Clil. Cambridge: Cambridge University Press: 2010.
10. Li, A.-H.; Liu, P. P.; Villarreal, F. J.; Garcia, R. A., Dynamic changes in myocardial matrix and relevance to disease: translational perspectives. Circulation research 2014, 114 (5), 916-927.

11. Pierson, A. E.; Clark, D. B.; Brady, C. E., Scientific modeling and translanguaging: A multilingual and multimodal approach to support science learning and engagement. Science Education 2021, 105 (4), 776-813.
12. Schwarz, C., Developing preservice elementary teachers' knowledge and practices through modeling-centered scientific inquiry. Science Education 2009, 93 (4), 720-744.
13. Passmore, C.; Stewart, J.; Cartier, J., Model-based inquiry and school science: Creating connections. School Science and Mathematics 2009, 109 (7), 394-402.
14. Duschl, R., Science education in three-part harmony: Balancing conceptual, epistemic, and social learning goals. Review of research in education 2008, 32 (1), 268-291.
15. Smith, L. A.; Gentner, D. In Using spatial analogy to facilitate graph learning, International Conference on Spatial Cognition, Springer: 2012; pp 196-209.
16. Clement, J. J., Creative model construction in scientists and students. Springer: 2008.
17. Gentner, D., Structure-mapping: A theoretical framework for analogy. Cognitive science 1983, 7 (2), 155-170.
18. Lehrer, R.; Schauble, L., Cultivating model-based reasoning in science education. na: 2006.
19. Lin, J.-W.; Chiu, M.-H., Evaluating Multiple Analogical Representations from Students' Perceptions. In Multiple representations in physics education, Springer: 2017; pp 71-91.
20. Nowell, L. S.; Norris, J. M.; White, D. E.; Moules, N. J., Thematic analysis: Striving to meet the trustworthiness criteria. International journal of qualitative methods 2017, 16 (1), 1609406917733847.

1. Chiu, M. H.; Lin, J. W., Promoting fourth graders' conceptual change of their understanding of electric current via multiple analogies. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching* 2005, *42* (4), 429-464.
2. Chiu, M.-H.; Lin, J.-W., Modeling competence in science education. *Disciplinary and Interdisciplinary Science Education Research* 2019, *1* (1), 12.
3. Smith, L. A.; Gentner, D. In *Using spatial analogy to facilitate graph learning*, International Conference on Spatial Cognition, Springer: 2012; pp 196-209.
4. Nowell, L. S.; Norris, J. M.; White, D. E.; Moules, N. J., Thematic analysis: Striving to meet the trustworthiness criteria. *International journal of qualitative methods* 2017, *16* (1), 1609406917733847.