ISSN: 2821-9163 (Online)

# International Journal of Science Education and Teaching


IJSET Vol.3 No. 3
(September - December 2024)

Published by
The Science Education Association (Thailand)

*International Journal of Science Education and Teaching*
*Vol. 3, No. 3 (September – December 2024)*

## ABOUT IJSET

**International Journal of Science Education and Teaching (IJSET)** is supported by Science Education Association (Thailand) or SEAT. IJSET seeks articles addressing issues including science education, physics education, chemistry education, biology education, technology education, STEM education, science teacher education, early childhood science education, science curriculum and instruction, and other related science educational fields.

**Peer Review Process**
Articles submitted for publication must never have been previously published or currently under review at another journal. Each paper is reviewed by the editor and, if it is judged suitable for this publication, it is then sent to at least two independent reviewers for double blind peer review. Based on their recommendation, as well as consultation between relevant Editorial Board members the editors then decides whether the paper should be accepted as is, revised or rejected.

**Publication Frequency**
The IJSET provides an academic platform for work in the fields of interdisciplinary education. The IJSET publishes 3 issues annually. These include:

- Issue 1 (January - April)
- Issue 2 (May - August)
- Issue 3 (September - December)

**Issue: IJSET Vol. 3 No. 3 (September – December 2024)**
**Publication Date: December 31, 2024**
**Publisher: Science Education Association (Thailand)**

## IJSET EDITORIAL TEAM

# TABLE OF CONTENTS

*International Journal of Science Education and Teaching*
*Vol. 3, No. 3, pp. 86 - 96, Sep. – Dec. 2024*
*http://doi.org/10.14456/ijset.2024.8*

# Improvement of Students' Critical Thinking Skills by Guided Inquiry with Augmented Reality-based Solar System Learning Media

**Khanifatul Mukaromah[1] and Fidia Fibriana[2*]**
[1]*Student of Science Education Undergraduate Study Program, Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Semarang, Central Java 50229, Indonesia*
[2]*Faculty of Mathematics and Natural Sciences, Universitas Negeri Semarang, Semarang, Central Java 50229, Indonesia*
*Email: fibriana.f@mail.unnes.ac.id

Received: 24 June 2024 Revised: 13 October 2024 Accepted: 15 October 2024

**Abstract.** Critical thinking is an important skill of the 21$^{st}$ century. An intensive observation during the learning process at State Junior High School (SMP Negeri) 8 Semarang found that students had low critical thinking abilities in solar system topics. Thus, a guided inquiry model assisted by augmented reality media on the solar system study material is expected to engage students by allowing them to participate actively in their learning process to improve their critical thinking skills. This research used a quasi-experimental method with a pretest-posttest control group design by applying the guided inquiry model assisted by augmented reality (AR) to improve students' critical thinking skills. The subjects of this research were students in VII F and VII G classes (Academic Year 2024). The research instruments included a multiple-choice test sheet with reasoning and a student response questionnaire. Data analysis employed the N-gain test and t-test. The results of the N-gain analysis showed an increase in students' critical thinking skills in the experimental class by 0.62 and in the control class by 0.27. The t-test results showed a significant difference between the critical thinking abilities of the experimental and control classes with Sig. (2-tailed) of 0.000 < 0.05. The guided inquiry model implementation with AR in improving the students' critical thinking skills is also supported by the results of the student response questionnaire, which showed a very good category (84.99%). This research concludes that learning using the guided inquiry model assisted by augmented reality could improve student's critical thinking skills.

**Keywords:** Guided inquiry model, augmented reality, critical thinking skill.

## INTRODUCTION

The rapid technological development and changes in organizational behavior have impacted the characteristics of the current labor market, such as a rapidly changing work atmosphere. People, especially employees, must have 21$^{st}$-century skills, including 4C (critical thinking, collaboration, communication, and creativity). Critical thinking is the most important among those four skills since it is the highest form of intelligence. The ability to think critically is the ability that a person hopes to face various conflicts in social and personal life. Critical thinking refers to the act or practice of thinking critically by applying reason, questioning assumptions, and evaluating information. It involves analyzing available facts, evidence, observations, and arguments to form rational, skeptical, and unbiased judgments. Critical thinking is essential for problem-solving, discerning biases, and making informed decisions (McGlynn, 2024). Critical thinking in science learning is a systematic process that involves mental activities such as analyzing assumptions and making decisions. Students with critical thinking skills can think when making decisions about certain statements, which is essential for developing students' potential. However, it cannot be mastered instantly within a short period. Therefore, family support, education schools, parents, and teachers are fundamental in infusing and preparing a supportive educational environment to stimulate critical thinking skills in students as early as possible (Thornhill-Miller et al., 2023).

In order to support the educational environment along with science and technology development, various learning media have emerged, especially those that utilize the latest technology, such as computers and mobiles (Punggeti et al., 2024). Teachers can use media to clarify their understanding of science concepts, including solar system study material. Solar system material is one of the materials that require suitable learning media to explain learning material concretely. Solar system study material is very theoretical. Students must be able to describe various celestial bodies found within the solar system. Therefore, special media is needed to help the learning process of the solar system concept and ensure student understanding and critical thinking skills (Rahmawati et al., 2019).

In Indonesia, solar systems are science learning materials studied in junior high school. This material is very close to everyday life, but the material on the process of Earth's rotation and revolution is challenging to understand and tends to be abstract. When students work on questions related to this material, they experience difficulties and have low analytical skills. Based on the preliminary study at State Junior High School (SMP Negeri) 8 Semarang, students still had relatively low critical thinking skills. Low critical thinking skills of students were observed when students were unable to work on questions that referred to higher-level thinking questions. When students are faced with high-level questions, there are still many students whose scores are below average. In addition, students with low critical thinking skills cannot reach arguments and conclude reasonably. According to an interview with the science teachers, it was obtained that students were less interested in participating in classroom learning activities due to the lack of learning innovation. Apart from that, students also quickly procrastinated in doing assignments, and students tended to get bored. They had little desire to look for information related to the material, and students' critical thinking abilities were also lacking, as seen from their low curiosity and lack of critical thinking in science learning. This results in students being less trained in developing analytical and argumentative skills as indicators of critical thinking abilities. Teachers also have not tried to provide questions that can train students' critical thinking skills. The lack of application of supportive learning models causes low critical thinking skills. This result aligns with research showing that students' low critical thinking abilities were caused by a monotonous teaching method and less innovative media (Surayya et al., 2014). Also, the questions given to students to evaluate only reach levels C1 (remember) and C2 (understand), so it is not enough to train students' critical thinking abilities (Baharuddin et al., 2017). Questions used to test thinking abilities critical load level C4. (analysis), (analyzing), C5. (synthesis), and C6 (evaluate) with solar system material studied by students, which contain indicators of critical thinking skills.

One solution is to innovate using learning models and media to improve students' critical thinking skills in solar system material. Guided inquiry learning is a practical approach for enhancing critical thinking skills among students. Guided inquiry learning is a practical approach to improving critical thinking skills among students. The guided inquiry learning process becomes important to apply, especially in science learning, to make the learning process more meaningful. The steps taken direct students to think emancipatory when finding the concept. Indirectly, this will train them to think critically and creatively. In this model, learners actively explore topics, ask questions, and investigate problems, fostering more profound understanding and analytical abilities. A study conducted with 11th-grade students using guided inquiry models found a significant increase in critical thinking skills (29.9% improvement) and cognitive learning outcomes (22.9% improvement). The strong correlation ($r = 0.721$) between critical thinking skills and cognitive learning outcomes highlights the effectiveness of this approach (Rosania et al., 2023).

Apart from learning models, the role of learning media is also essential for the learning process. Learning media can train students' critical thinking skills by stimulating them to argue and answer questions. Media use appropriate learning as stated by Hasnunidah (2011) that a person's expertise teachers in choosing the proper learning media are one of the determining factors for the successful development of students' critical thinking skills. This follows the opinion of Mayer (2002) that students' critical thinking skills can be developed using appropriate learning media. Android-based media is learning media relevant to current conditions; one alternative learning media that can be applied to smartphones is augmented reality (Herlanti et al., 2019; Rahmawati et al., 2019). Augmented Reality (AR) learning media can significantly enhance students' critical thinking ability. By integrating AR technology into educational experiences, students engage with content more interactively and immersively, improving cognitive skills. AR learning media provides an innovative approach to enhancing critical thinking skills by promoting active engagement and more profound understanding (Bakri et al., 2021).

There are few research results on the guided inquiry model using AR media for solar system learning to improve students' critical thinking skills (Table 1). Therefore, based on theoretical and other study analysis, the guided inquiry learning model AR media in solar system study material is expected to enable learners to use authentic context by exploring the visualization using the AR, interact with virtual models, foster curiosity, and encourage critical thinking (Wen et al., 2023).

## RESEARCH OBJECTIVES

Based on the background outlined, the research problem can be formulated as follows: 1) how does implementing the guided inquiry model assisted by augmented reality (AR) on the solar system study material enhance students' critical thinking skills? and 2) what are the students' responses to applying the guided inquiry model assisted by augmented reality in improving their critical thinking skills?

**Table 1: Review of relevant research**

| Objectives and Results | Learning Model | Study Materials | Learning Media | References |
|---|---|---|---|---|
| To apply learning media based on a scientific approach using AR to improve critical thinking skills. It was found that students' critical thinking abilities increased after using learning media based on a scientific approach. | Scientific | Solar System | AR | (Gaol et al., 2022) |
| To improve students' critical thinking skills through the AR-assisted PBL model. The results of this research showed that the application of the AR-assisted PBL model enhanced critical thinking skills. | Project-based Learning | Solar System | AR | (Isatunada et al., 2023) |
| To determine the effect of AR-based learning media on students' critical thinking abilities. The research showed that the AR-assisted discovery learning model affected students' critical thinking abilities. The results of the essential thinking abilities in the experiment class were more significant than those in the control class by 11%. | Discovery Learning | Movement System | AR | (Ashari, 2019) |
| To produce an AR-integrated contextual learning module to improve students' critical thinking skills. This research showed that the integrated AR module improved students' critical thinking skills. | Contextual | Chemical Bond | AR | (Mashami et al., 2021) |
| To determine the validity, practicality, and effectiveness of STEM-based teaching materials using AR to improve students' critical thinking skills. This research showed that STEM-based teaching materials using AR effectively improved students' critical thinking skills. | STEM | Vibration, Wave, and Sound | AR | (Oktaviyanti et al., 2023) |
| To implement the guided inquiry model assisted by augmented reality (AR) on the solar system study material to enhance students' critical thinking skills. | Guided inquiry | Solar System | AR | This study |

**Research Tools**

This research used an experimental method, a quasi-experimental design with a nonequivalent control group. The study consisted of a control class and an experimental class. The learning control class used conventional media (pictures and videos) with a discovery learning model, while the experiment class used digital media of augmented reality (AR) in the Assemblr EDU application (Figure 1) with a guided inquiry model. Figure 2 shows an example of a question that applies the guided inquiry model. The research design is presented in Table 2.

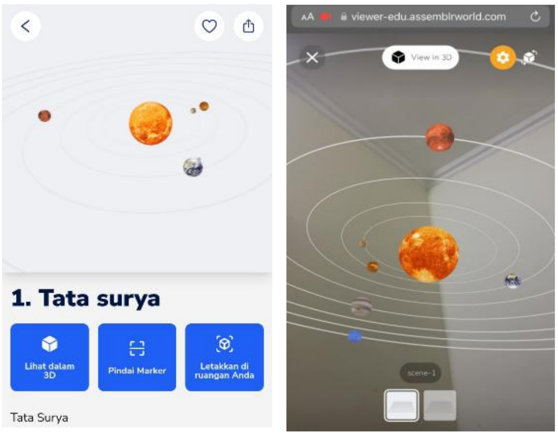
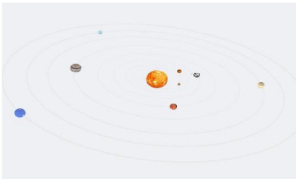



**Figure 1: Assembler EDU application with AR-based learning media on solar system**

Look at the picture below!



A planet is a member of the solar system. Planets make movements, one of which is circling the Sun in a certain trajectory. Based on this image, it can be seen that the planet's trajectory has the shape...

A. Ellipse

B. Circle

C. Parabola

D. Semicircle

Reason:

**Figure 2: Example of a question applied to the Assemblr Edu application**

**Table 2: Research design**

| Group | Pretest | Treatment | Posttest |
|---|---|---|---|
| Experiment class | $O_1$ | X | $O_2$ |
| Control class | $O_3$ | Y | $O_4$ |

Source: (Sugiyono, 2013)

Description:

X : experiment class (guided inquiry and AR media)

Y : control class (discovery learning and pictures-videos media)

$O_1$ : pretest of experiment class

$O_2$ : posttest of experiment class

$O_3$ : pretest of control class

$O_4$ : posttest of control class

### Data Collection

The data collection technique was to determine students' critical thinking abilities before and after being given treatment. A reasoned multiple-choice test and a students' response questionnaire were employed. An example of a reasoned multiple-choice question can be seen in Figure 2, which provides an indicator of thinking ability to focus on a question. The test measured the students' critical thinking abilities regarding solar system material. The test and questionnaire were prepared based on critical thinking indicators (Ennis, 2011) and were given to both samples before and after treatment.

**Data Analysis**

Parametric and non-parametric statistical techniques were used in this study. When the generated data follows a normal distribution and is in interval format, parametric statistics are used. Conversely, non-parametric statistics are employed if the data does not follow a normal distribution and is nominal. The analysis used to determine whether the guided inquiry model assisted by augmented reality could enhance students' critical thinking skills involved a t-test using SPSS Software.

## RESULTS AND DISCUSSION

This quasi-experimental research involved control and experiment classes as research treatments. In both classes, learning was carried out in four meetings on solar system study material. The material taught at each meeting is the solar system, including understanding the Sun, planets, and their satellites, and getting to know the Sun more closely.

Based on the pretest and posttest data on students' critical thinking abilities in all classes, all data were distributed normally and homogeneously. Thus, parametric statistics were applied using a t-test. The results of the average pretest and posttest scores for the critical thinking abilities of experimental and control class students can be seen in Figure 3.

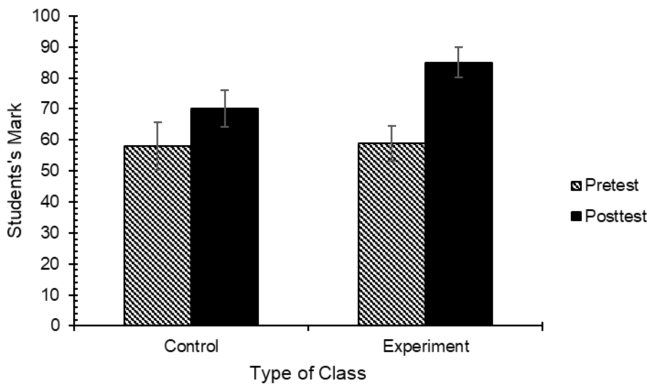
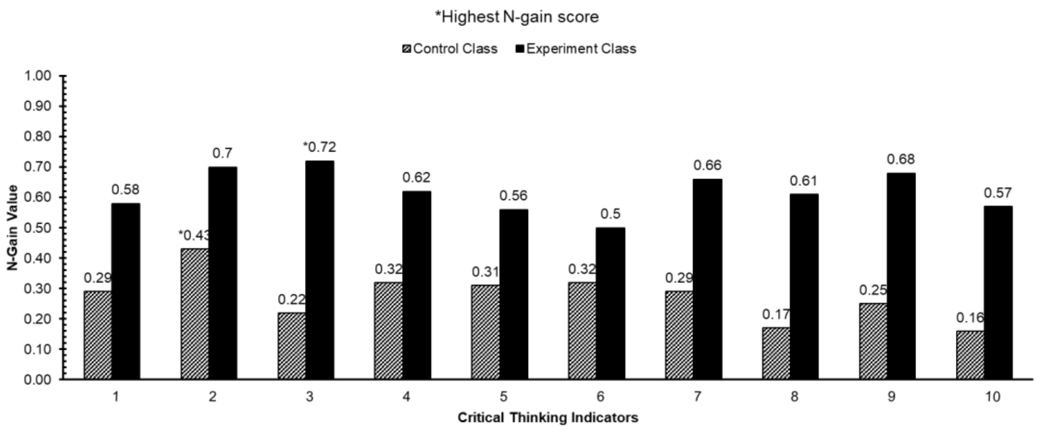



**Figure 3: Pretest and posttest results of control and experiment classes**

Based on Figure 3, the pretest scores for both classes are not significantly different. In contrast to the results of the posttest scores in the experimental class, the results are significantly higher than those in the control class. As for the results of the pretest value testing using the related t-test, the Sig. (2-tailed) value was 0.377 > 0.05, meaning there is no significant difference in the pretest results of the control and experiment classes because the significant value is > 005. Thus, no other factors could influence students' critical thinking abilities in both classes apart from the treatment given. Therefore, testing the presence of treatment given was done using student posttest data.

Based on the test results for the application due to posttest data treatment using the related t-test, a Sig. (2-tailed) value of 0.000 > 0.05 was obtained. Based on the t-test results, there was a significant difference in the posttest results for the experiment and control classes. Based on the t-test results of students' pretest and posttest data, it can be concluded that applying the guided inquiry model assisted by augmented reality can improve students' critical thinking skills regarding solar system material.

Furthermore, significant differences can be observed based on the analysis of pretest and posttest data using the N-gain test. This test was used to determine the effectiveness of the treatment given before and after learning. The N-gain test was carried out twice based on the average and each student's critical thinking ability indicator in the pretest and posttest questions. The N-gain test calculation based on the average in the control class increased by 0.27, while in the experiment class, it increased by 0.62. These results showed that the increase in the experiment class was higher than in the medium category of the control class ($0.3 \leq g \leq 0.7$). Next, the pretest and posttest data were carried out with an N-gain test on each indicator of critical thinking ability used. According to Ennis (2011), ten indicators of student's critical thinking abilities were used and distributed in 15 reasoned multiple-choice questions. The results of the N-gain test for critical thinking ability indicators in the control and experimental classes are presented in Figure 4.




**Description:**
1 : Focus to a something question
2 : Analyze the question
3 : Inquiry and answering the question
4 : Consider a source used
5 : Observe and consider the observation report
6 : Concluding and consider the conclusions deduction
7 : Make and consider inductive arguments and conclusions
8 : Define the terms and consider the definition
9 : Identify various assumptions
10: Deciding on a course of action

**Figure 4: N-gain test for critical thinking ability indicators**

Figure 4 shows increased students' critical thinking abilities as seen in N-gain. The image results show that the N-gain value obtained in the experimental class is higher than the control class. The highest score in the experimental class was 0.72 in the high category, while in the control class, the highest score was 0.43 in the medium category. The results of the N-gain analysis stated significant differences in each indicator of students' critical thinking abilities in the control and experiment classes. This analysis data shows a better improvement in the experimental class than in the control class.

A person is considered to have critical thinking abilities if they meet the indicators of critical thinking abilities (Ennis, 2011). The increase in students' critical thinking skills was measured using N-gain, which was analyzed for each indicator. The critical thinking ability indicators in this research adopt indicators from Ennis (2011), which have five aspects: providing simple explanations, building essential and basic skills, concluding, making further explanations, and organizing strategies and tactics. The first aspect, namely providing a simple explanation, was divided into three indicators: focusing on a question, analyzing the question, and asking and answering questions. Each indicator between the experimental and control classes had a significant difference, where the results of increasing N-gain in the experimental class were higher than in the control class.

The increase in results was influenced by, among other things, the use of the guided inquiry model with the help of AR, which occurred at the orientation or problem formulation stage. During the learning process, the teacher motivated students to ask questions because they were facilitated by jointly observing narratives, learning videos, and AR media in the experimental class, which stimulated students to focus and think to ask questions. It is related to guided inquiry learning, which requires students to think actively, formulate problems, focus, and try to identify questions that arise from observed physical phenomena to develop their critical thinking abilities (Thornhill-Miller et al., 2023).

Furthermore, regarding the second aspect of building basic skills, the experiment class was better than the control class. There are two indicators, i.e., considering a source used and considering observation reports (Koksal & Berberoglu, 2014). The highest increase occurred in the indicator considering the source used because students were asked to reconsider answers in learning activities based on AR media information. It allows students to observe the questions given. A person will carefully consider or determine whether to reject, accept, or delay receiving information with critical thinking skills. Adapting to sources is a person's ability to use existing procedures from trusted sources, such as formulas, statements, and facts, to solve a problem.

The third aspect is concluding, where two indicators were used, i.e., concluding and considering deductive conclusions and making and considering inductive arguments and findings. Of the two indicators above, the indicator of creating and considering inductive arguments and conclusions had the highest N-gain value. Students were trained to answer questions related to AR media and asked to come up with several options by considering these decisions. The existence of interaction activities between students creates an exchange of ideas, where this activity influences the indicators for making an induction. When making an induction, students must actively participate in the discussion to draw careful conclusions from specific problems (Komala et al., 2017).

The fourth aspect is making further explanations. In this aspect, two indicators were used to measure critical thinking skills, i.e., defining terms/considering definitions, and identifying various assumptions. The highest increase occurred in the indicator identifying various assumptions, where there was an increase in this indicator due to implementation using AR media, which trained students in thinking and producing a conclusion based on their assumptions. Critical thinking requires rigorously examining every assumptive knowledge or belief based on supporting evidence and the subsequent conclusions that result from it (Rahmawati et al., 2019; Zuniari et al., 2022).

The fifth aspect is organizing strategy and tactics, where deciding on an action occurs in the guided inquiry model syntax. Designing an experiment requires students to act according to the work method in the student worksheet. The use of AR media during the experiment was not only to make the experiment easier but also to make students interested in learning and gain more in-depth knowledge regarding AR media in student worksheets. In contrast to the control class, which only uses the discovery learning model with the help of learning videos, students felt bored more quickly. They were less enthusiastic about learning activities in class. Apart from that, in this indicator, students were trained to determine the reasons for and actions for decisions through observations on AR media. At the end of each activity on the worksheet, students were required to answer questions. The activities in AR media attempted to guide students in answering problems and determining appropriate actions. Determining an action can be improved by solving questions correctly (Khamhaengpol et al., 2021).

Overall, this research showed that learning using the guided inquiry model assisted by AR media implemented in the experimental class could improve student's critical thinking abilities. The AR-assisted guided inquiry model could train students' critical thinking skills and be more student-centered. This model's stages invited students to think about formulating problems and drawing conclusions. Meanwhile, AR media was used to conduct investigations to gain knowledge independently. This activity made students more active and meaningful. It made obtaining information from various sources easier with the help of the internet or student teaching materials, improving their critical thinking abilities. Meanwhile, it was not optimal in the control class since learning through videos still made them lack understanding of the information. It did not facilitate investigations that could enable students to gain new knowledge and draw conclusions from the investigation results.

The guided inquiry learning model assisted by augmented reality can improve students' critical thinking skills. It was supported by analyzing the results of a questionnaire given to students after the learning process. The questionnaire analysis consists of 10 statement items grouped into five aspects of the student response questionnaire. The results of the data analysis are presented in Table 3.

**Table 3: Percentage of students' responses to the questionnaire**

| Aspects of questionnaire | Percentage (%) |
|---|---|
| Interest in learning models and media | 84.19 |
| Utilization of learning models and media | 84.19 |
| Understanding learning material | 89.33 |
| Developing students' information | 83.82 |
| Generating students' critical thinking attitudes | 83.45 |

Based on Table 3, it can be concluded that the results of the students' response questionnaire in the experimental class in the guided inquiry learning model assisted by augmented reality in every aspect obtained very good responses from students with an average of 84.99%. It shows that the guided inquiry learning model assisted by augmented reality learning media suits solar system study material use.

## CONCLUSION AND IMPLICATIONS

Guided inquiry learning using AR technology media in the context of solar system study material significantly enhanced students' critical thinking skills. Application of augmented reality It has also been proven to give students the courage to express real ideas; it is part of the concepts that exist within them and should be developed and adequately directed (Hermawan & Hadi, 2024). Furthermore, Nugroho (2023) also explained that implementing augmented reality in learning spaces can enhance the experience and bring significant changes in learning. The guided inquiry model assisted by augmented reality received a positive response from students in the very good category—AR-enabled inquiry activities immerse learners in an authentic context. Students can explore the solar system, visualize celestial bodies, and interact with virtual models. This engagement fosters curiosity and encourages critical thinking. AR allows students to collect data beyond the classroom, and their experience promotes analytical thinking and problem-solving. AR also facilitates collaboration among students working together on complex solar system concepts. Effective communication and teamwork are essential for critical thinking. In summary, integrating guided inquiry learning with AR technology in solar system education enhances critical thinking by offering immersive experiences, data-driven exploration, and collaborative problem-solving. To strengthen the findings in this research, the researcher provides suggestions for future researchers to carry out experiments related to using augmented reality learning media to improve other science skills.

## ACKNOWLEDGEMENTS

The Science Education Undergraduate Study Program supported this research. The authors would like to thank the science teachers of SMP Negeri 8 Semarang for assisting with the research work.

## REFERENCES

Ashari, D. H. (2019). *Pengaruh Penggunaan Media Pembelajaran Berbasis Augmented Reality dengan Model Pembelajaran Discovery Learning Terhadap Kemampuan Berpikir Kritis Siswa pada Materi Sistem Gerak*.

Baharuddin, B., Indana, S., & Koestiari, T. (2017). Perangkat pembelajaran IPA berbasis inkuiri terbimbing dengan tugas proyek materi sistem ekskresi untuk menuntaskan hasil belajar siswa SMP. *JIPI (Jurnal IPA Dan Pembelajaran IPA)*, *1*(1), 81–97.

Bakri, F., Vani, N. D., Permana, H., & Muliyati, D. (2021). Textbook with augmented reality technology: Improve critical thinking skill in elasticity concept. *AIP Conference Proceedings*, *2331*(1).

Gaol, A. F. L., Azizahwati, A., & Zulhelmi, Z. (2022). Implementasi Media Pembelajaran Berbasis Pendekatan Saintifik Menggunakan Augmented Reality pada Materi Tata Surya untuk Meningkatkan Keterampilan Berpikir Kritis Peserta Didik Kelas VII SMP. *Jurnal Pendidikan Tambusai*, *6*(2), 14190–14199.

Hasnunidah, N. (2011). Keterampilan Berpikir Kritis Siswa SMP Pada Penggunaan Media Maket Melalui Contextual Teaching and Learning. Prosiding Seminar Nasional Pendidikan MIPA 2011 FKIP Unila.

Hermawan, A., & Hadi, S. (2024). Realitas Pengaruh Penggunaan Teknologi Augmented Reality dalam Pembelajaran terhadap Pemahaman Konsep Siswa. *Jurnal Simki Pedagogia*, *7*(1), 328-340.

Herlanti, Y., Mardiati, Y., Rahmawati, R., Putri, A. M. K., Jamil, N., Miftahuzzakiyah, M., Sofyan, A., Zulfiani, Z., & Sugiarti, S. (2019). Finding learning strategy in improving science literacy. *Jurnal Penelitian Dan Pembelajaran IPA*, *5*(1), 59–71.

Isatunada, A., Indriyani, S., & Dewi, N. R. (2023). Peningkatan Keterampilan Berpikir Kritis Peserta Didik melalui Model Problem Based Learning Berbantuan Augmented Reality. *Proceeding Seminar Nasional IPA*.

Khamhaengpol, A., Sriprom, M., & Chuamchaitrakool, P. (2021). Development of STEAM activity on nanotechnology to determine basic science process skills and engineering design process for high school students. *Thinking Skills and Creativity*, *39*, 100796.

Koksal, E. A., & Berberoglu, G. (2014). The effect of guided-inquiry instruction on 6th grade Turkish students' achievement, science process skills, and attitudes toward science. *International Journal of Science Education*, *36*(1), 66–78.

Komala, T. R., Nurlaelah, I., & Setiawati, I. (2017). Peningkatan Kemampuan Penalaran Siswa Melalui Model Problem Based Learning (PBL) Ditinjau Dari Kemampuan Akademik Di SMA. *Quagga: Jurnal Pendidikan Dan Biologi*, *9*(01).

Mayer, R. E. (2002). Cognitive Theory and The Design of Multimedia Instruction: An Example of The Two-Way street between cognition and nstruction. *New Directions for Teaching and Learning*, *2002*(89), 55–71.

Mashami, R. A., Khaeruman, K., & Ahmadi, A. (2021). Pengembangan modul pembelajaran kontekstual terintegrasi augmented reality untuk meningkatkan keterampilan berpikir kritis siswa. *Hydrogen: Jurnal Kependidikan Kimia*, *9*(2), 67–77.

Nugroho, W. D. (2023). Evaluasi Usabilitas Penggunaan Platform Metaverse Dengan Virtual Reality Menggunakan Metode Use (Usefulness, Satisfaction, Dan Ease Of Use)(Studi Kasus: Penyelenggaraan Acara Mice) (Doctoral dissertation, Universitas Atma Jaya Yogyakarta). https://e-journal.uajy.ac.id/29868/.

McGlynn, A. (2024). Educating for intellectual virtue in a vicious world. *Inquiry*, *67*(2), 784–797.

Oktaviyanti, R., Fatmahanik, U., & Fadly, W. (2023). Pengembangan Bahan Ajar Berbasis STEM dengan Memanfaatkan Augmented Reality dalam Meningkatkan Kemampuan Berpikir Kritis. *Jurnal Tadris IPA Indonesia*, *3*(3), 303–314.

Punggeti, R. N., Rukmini, A., & Intes, A. (2024). The Use of ICT as a Resource and Media for Modern 21st Century Learning in Primary Schools. *JILTECH: Journal International of Lingua & Technology*, *3*(1).

Rahmawati, Y., Ridwan, A., Hadinugrahaningsih, T., & Soeprijanto. (2019). Developing critical and creative thinking skills through STEAM integration in chemistry learning. *Journal of Physics: Conference Series*, *1156*, 12033.

Rosania, R. A., Ibrohim, I., & Handayani, N. (2023). Improvement of critical thinking skills and cognitive learning outcomes through guided inquiry learning models. *AIP Conference Proceedings*, *2569*(1).

Sugiyono, D. (2013). *Metode penelitian pendidikan pendekatan kuantitatif, kualitatif dan R&D*.

Surayya, L., Subagia, I. W., & Tika, I. N. (2014). Pengaruh model pembelajaran think pair share terhadap hasil belajar IPA ditinjau dari keterampilan berpikir kritis siswa. *Jurnal Pendidikan dan Pembelajaran IPA Indonesia*, *4*(1).

Thornhill-Miller, B., Camarda, A., Mercier, M., Burkhardt, J.-M., Morisseau, T., Bourgeois-Bougrine, S., Vinchon, F., El Hayek, S., Augereau-Landais, M., & Mourey, F. (2023). Creativity, Critical Thinking, Communication, and Collaboration: Assessment, Certification, and Promotion of 21st Century Skills for the Future of Work and Education. *Journal of Intelligence*, *11*(3), 54.

Wen, Y., Wu, L., He, S., Ng, N. H.-E., Teo, B. C., Looi, C. K., & Cai, Y. (2023). Integrating augmented reality into inquiry-based learning approach in primary science classrooms. *Educational Technology Research and Development*, *71*(4), 1631–1651.

Zuniari, N. I., Ridlo, Z. R., Wahyuni, S., Ulfa, E. M., & Dharmawan, M. K. S. (2022). The effectiveness of implementation learning media based on augmented reality in elementary school in improving critical thinking skills in solar system course. *Journal of Physics: Conference Series*, *2392*(1), 12010.

*International Journal of Science Education and Teaching*
*Vol. 3, No. 3, pp. 97 - 112, Sep. – Dec. 2024*
*http://doi.org/10.14456/ijset.2024.9*

# Incorporating Computational Thinking in Mathematics through Block-Based Programming: Effects on Students' Problem-Solving Skills

**Ian Cesar P. Balacuit[1*], Amelia T. Buan [1], Joneil B. Medina[1], Alexis Michael B. Oledan[1]**
[1]*Mindanao State University – Iligan Institute of Technology, 9200 Iligan City, Mindanao, Philippines*
*Email: iancesar.balacuit@g.msuiit.edu.ph

Received: 15 October 2024; Revised: 11 November 2024; Accepted: 12 November 2024 2024

**Abstract.** Computational thinking is considered a fundamental skill in the 21st century. It is a vital skill for students to empower their problem-solving skills through the growing presence of computer technology. As a result, this study utilized the Research and Development method with a Backward Curriculum design to develop Block-Based Programming-Integrated Mathematics Problem-Solving activities that aimed to incorporate computational thinking in block-based programming with mathematics learning competencies. The activities were implemented on Grade 7 students to investigate their problem-solving and computational thinking skills. Results suggest that the students generally exhibited the characteristics of problem-solving and computational thinking skills during the implementation. In addition, they also acquired some improvements. For students' problem-solving skills, the improvements were marked by their acquired ability to 1) establish goals in problem understanding, 2) provide mathematical reasoning in solution planning, 3) draw on the application of mathematics concepts in solution execution, and 4) debug program errors in monitoring and evaluation. On the other hand, students' computational thinking improvements were highlighted by their learned ability to 1) apply a rules-based and systematic approach in problem-solving for algorithmic thinking and 2) utilize patterns in generalizing solutions for pattern recognition. Based on the results, it can be concluded that the activities helped the students develop their problem-solving and computational thinking skills.

**Keywords:** Block-based Programming, Block-based Programming - Integrated Mathematics Problem-Solving, Computational Thinking Skill, Problem-Solving Skill

## INTRODUCTION

Problem-solving (PS) is considered one of the central activities in most educational curricula in the modern era (Ozpinar & Arslan, 2023). Researchers have seen this skill as significant to students because of its applicability to real life. Consequently, developing PS skills has been a goal of many educational programs, including the Philippine K-12 Mathematics curriculum (DepEd, 2012).

One strategy for improving the PS skills of students is to introduce Computational Thinking (CT) in the curriculum (Prsala, 2024). CT, which gained prominence in 21st-century education, was first suggested by Wing (2006) to involve "solving problems, designing systems, and understanding human behavior by drawing on the concept fundamental to computer science." This skill enables students to practice PS by employing computer-related concepts and approaches, i.e., through representation, abstraction, decomposition, simulation, verification, and innovation (Sengupta et al., 2018). These components also cater to the PS framework in mathematics proposed by Fraillon et al. (2019), who argue that PS includes analyzing the problem, developing an algorithm, and testing. With this, studies have demonstrated a strong link between CT and PS (e.g., Bati et al., 2018; Israel-Fishelson & Hershkovitz, 2022), and a high CT skill is shown to add confidence to students' PS skills (Wei et al., 2021).

Researchers agree that CT development should focus on guiding students to obtain an appropriate understanding of problems and their solutions with the help of computing and computers (Hurt et al., 2023). The advent of computers implies that CT is essential in fostering students' PS skills (Lin et al., 2021). In an era where technology is an integral part of learning, Jocius et al. (2021) assert that every student should learn CT skills and that the need to introduce it to the K-12 mathematics curriculum arises (Ye et al., 2023).

Meanwhile, programming has been agreed to explicitly teach CT, including its core concepts (Basogain et al., 2018). Programming includes writing, testing, debugging, and maintaining codes, a necessary skill set to foster CT among students (Harimurti, 2019). Computer programming is a mutual strategy for developing CT and other essential skills such as PS and communication (Yusoff et al., 2020). With this, researchers support pedagogies allowing all learners, regardless of grade level, to learn programming (Zeng et al., 2023).

A crucial issue of learning CT through programming, especially among students outside computer science, is its complexity and the burden of following the correct syntax in writing codes (Bala, 2021). As a result, researchers suggest the use of Block-based Programming (BBP) language in engaging K-12 students with programming activities to facilitate CT (e.g., Brennan & Resnick, 2012; Yu et al., 2024). Lye and Koh (2014) claim that BBP language is suitable for incorporating CT in programming contexts in K-12 education since it better facilitates the three dimensions, including CT concepts, practices, and perspectives.

Several studies have already investigated the effects of engaging students in BBP language programming activities. For instance, recent studies reveal that introducing BBP activities to the K-12 context improves students' CT skills (e.g., Totan & Korucu, 2023; Kastner-Hauler et al., 2022). Additionally, studies conducted by Durak (2018) and Kwon & Cheon (2019) show empirical evidence of how BBP fosters students' PS skills.

For these reasons, there is an evident need to incorporate CT through BBP with K-12 learning competencies to foster PS skill development. However, the search for pedagogies that would introduce CT for programming through BBP in Philippine K-12 mathematics education continues. There is still limited attempt to introduce CT concepts formally within the context of the mathematics curriculum. This current state is even though PS is one of the twin goals set in the Department of Education's (DepEd) K-12 mathematics education framework. Therefore, this study was conducted to develop PS activities that incorporate CT through BBP within K-12 mathematics education.

## RESEARCH OBJECTIVES

CT as a PS approach allows students to create or use technology to solve problems (McClelland & Grata, 2018). Teaching CT empowers students' PS skills by allowing them to maximize the presence of computer technology. Researchers agree that integrating BBP in K-12 education fosters students' CT and PS skills development (e.g., Hickmott et al., 2018; De Chenne & Lockwood, 2022).

However, in the Philippine setting, introducing BBP in K-12 education is still in its infancy, particularly in mathematics. Therefore, this study developed PS activities incorporating CT in BBP into mathematics learning competencies and used these activities to investigate students' PS and CT skills. Specifically, it sought to:

a) Develop Block-based programming–integrated Mathematics problem-solving (BBP-IMPS) activities; and
b) Investigate students' PS and CT skills as they do the BBP–IMPS activities.

## METHODOLOGY

This study adopted the research and development (R&D) method of Borg & Gall (1983) in developing BBP-IMPS activities. The R&D approach was utilized to support a systematic and iterative process of analyzing, designing, testing, and refining activities that improved the reliability of the developed intervention. The reasoning is anchored on Gay et al.'s (2012) claim that the purpose of R&D is to develop effective products for use in schools.

Alongside R&D, the principles of Backward curriculum design were followed to ensure that each component of the BBP-IMPS activities is aligned with the desired learning outcomes (Wiggins & McTighe, 2011). Defining the learning outcomes right at the beginning of the process enhanced the coherence of activities. The principle facilitated the alignment of objectives, strategies, and assessments with the competencies in Grade 7 mathematics and concepts in CT.

### Participants and Data Collection

This study was conducted online with six Grade 7 students from one of the schools in Dumingag, Zamboanga del Sur, Philippines. The participants were purposively selected based on their capacity to participate in online learning. None of the participants had any experience in programming activities. This criterion ensured that the BBP-IMPS activities fostered PS and CT skills without relying on prior knowledge.

This study lasted two weeks and was implemented using a synchronous learning modality with Google Meet as the online platform. Throughout the implementation, the researcher recorded all events, including student-to-student interactions and student-to-teacher interactions, using Open Broadcaster Software (OBS). Writing prompts were prepared so that students could write down their thought processes for every activity. The student participants were divided into three pairs for the entire implementation. In addition, three experts evaluated students' PS and CT using observation checklists.

### Data Analysis

The study's sample size is a notable limitation that could influence the generalizability of its findings. To address this, the study employed data triangulation from multiple sources to include both quantitative and qualitative data. Triangulation was done to validate patterns across different data sources. The quantitative data was obtained through experts' ratings on students' PS and CT using the checklists reported using mode.

On the other hand, qualitative data included recordings of students' conversations and interactions, researchers' observations, and writing prompts that elicited students' thought processes as they did every activity. An in-depth narrative analysis supported the quantitative findings for students' PS and CT skills. Moreover, the analysis was reviewed by an expert to ensure reliability.

### Process of the Development of the BBP-IMPS Activities

The development of BBP-IMPS activities followed Borg and Gall's research and development method as recommended by Daulay and Zaman (2012). The R&D method suggested six steps: research and information collection, planning, development of a preliminary product, preliminary testing, revising, and field testing. In the development process, backward curriculum design was embedded in the R&D method. Figure 1 presents a diagram of the development process.

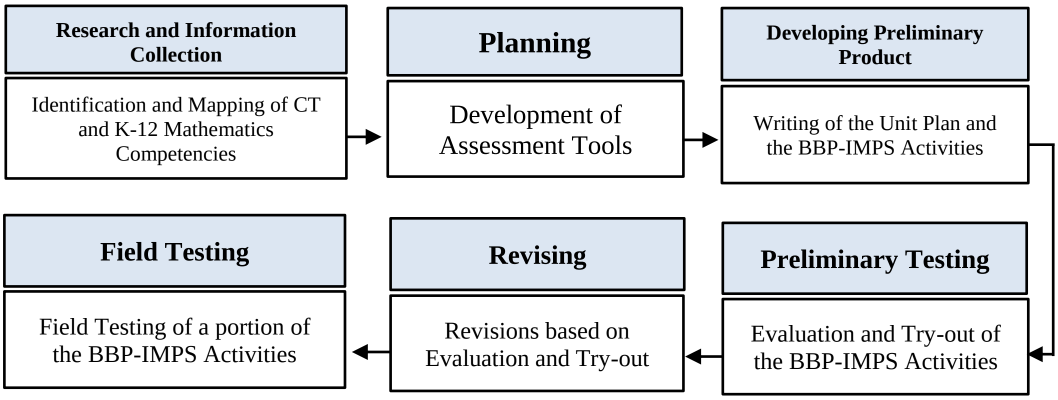



**Figure 1: Process of the Development of the BBP-IMPS Activities**

*Identification and Mapping of K-12 Mathematics Competencies CT Concept and Skills*

Identifying learning competencies in Mathematics and CT concepts was an essential step in ensuring an efficient integration of programming activities in Mathematics. As a result of the literature review, the researchers identified competencies from the Philippine Mathematics K-12 curriculum guide under the content area of Geometry. The chosen competencies were: (44) classify the different kinds of angles and (51) construct triangles, squares, rectangles, regular pentagons, and hexagons. The researchers also identified CT concepts based on the framework of Brennan and Resnick (2012), namely sequence, events, loops, variables, and operators. The CT concepts were mapped with Mathematics learning competencies. In addition, cornerstones of CT including decomposition, algorithmic thinking, and pattern recognition, were also considered by the researchers in the mapping. Table 1 presents the mapping of competencies for the developed BBP-IMPS activities.

**TABLE 1. Mapping of Mathematics and CT Concepts and Skills**

| Topic | Act # | Mathematics Learning Competency | CT Concept | CT Skill |
|---|---|---|---|---|
| Basic Geometric Concept | **1** | classify the different kinds of angles | Sequence; Events | Decomposition; Algorithmic Thinking |
| | **2** | construct triangles and squares | Loops | Algorithmic Thinking; Pattern Recognition |
| Polygons | **3** | construct regular pentagon, hexagons, etc. | Variables; Operators | Decomposition; Pattern Recognition |

*Development of Assessment Tool*

In determining assessment tools for assessing students' output, the researchers determined characteristics from related literature to be included in the PS Skill Observation Checklist and CT Skill Observation Checklist. The checklists' contents were validated based on the results of the field testing conducted. The two checklists have an inter-rater reliability of 87% and 82%, respectively.

*Writing of the Unit Plan and the BBP-IMPS Activities*

A significant consideration in writing the unit plan and the BBP-IMPS activities was the basic information of Grade 7 students. In designing the unit plan, the researchers adopted an established unit planning template from Intel Education's Designing Effective Project resource (2011). The researchers determined the targeted standards of the learning unit and the essential questions that guided the entire unit. The unit plan design allowed the students to construct a regular program generator with game-like elements as their final output. The design was rooted in the principle of game-design-based learning.

On the other hand, in writing the BBP-IMPS activities, the researchers adapted a framework from the Programme for International Student Assessments (PISA) for PS, which are composed of five components: namely, 1) problem recognition; 2) problem understanding and establishment; 3) solution planning; 4) execution of solution; and 5) monitoring and evaluation (2013). Three (3) BBP-IMPS activities were written by the researchers, with each activity composing three main parts: Exploration, Challenge, and Generalization. Exploration includes activities that allow learners to explore the language of Scratch and Mathematics concepts anchored on Lye & Koh's suggestion (2014). The Challenge part presents the PS task related to Mathematics that the learners must solve by applying the CT and mathematics concepts they have explored in the previous part.

*Evaluation of the Unit Plan Design and the Developed BBP-IMPS Activities*

The panel of evaluators rated the unit plan using the Unit Planning Rubric (Intel Education, 2011). The unit plan design was rated according to its Targeted Standard, Curriculum Framing Questions, Assessment Plan, and Procedures with *excellent*, *satisfactory*, *below satisfactory*, and *poor* scales. Based on the mean rating, the unit plan was rated "Excellent" in terms of its adherence to the four categories, namely targeted standards ($\bar{X} = 3.67, SD = 0.33$), curriculum framing questions ($\bar{X} = 3.89, SD = 0.19$), assessment plan ($\bar{X} = 3.67, SD = 0.58$), and procedures ($\bar{X} = 3.78, SD = 0.38$). The grand mean rating ($\bar{X} = 3.75, SD = 0.35$) of the unit plan evaluation implies that the unit plan only needs minor revisions based on the evaluators' suggestions.

On the other hand, the panel of evaluators rated the developed BBP-IMPS activities based on their potential to support the characteristics of five skills: Problem Recognition, Problem Understanding, Solution Planning, Execution of Solution, and Monitoring and Evaluation. The evaluation suggests that all BBP-IMPS activities had the potential to support the development of the five skills as seen in the rating for the BBP-IMPS Activity 1 ($\bar{X} = 1.95, SD = 0.12$), Activity 2 ($\bar{X} = 1.94, SD = 0.14$), and Activity 3 ($\bar{X} = 1.94, SD = 0.12$). Furthermore, the grand mean rating of the evaluation $\bar{X} = 1.94$ ($SD = 0.04$) implies that the characteristics of PS activity are observable in all the developed BBP-IMPS activities.

*Try-out and Revisions of the BBP-IMPS Activities*

The developed BBP-IMPS activities were subjected to a try-out to determine their appropriateness and acceptability concerning the target respondents. It was also conducted to find the aspects of the activities that need possible improvement. During the try-out, the performances of the student participants were carefully observed. The difficulties they encountered throughout the try-out were noted and became the basis of improvement. The researchers, with two mathematics field teachers, observed students' performance during the try-out.

Meanwhile, the BBP-IMPS activities were revised based on the tryout results. Students' responses and interactions, observers' observations, and researchers' observations during the tryout were processed and became the basis for revising the BBP-IMPS activities.

## RESULTS AND DISCUSSION

### Students' PS Process During the Field Testing of BBP-IMPS activities

The developed activities were subjected to field testing to determine the actual PS process of students doing the BBP-IMPS activities. The field testing significantly validated the developed PS and CT Skills Observation Checklists. It was conducted with a different set of Grade 7 students. Figure 2 shows a diagram of the actual PS process of students during the activity try-out.

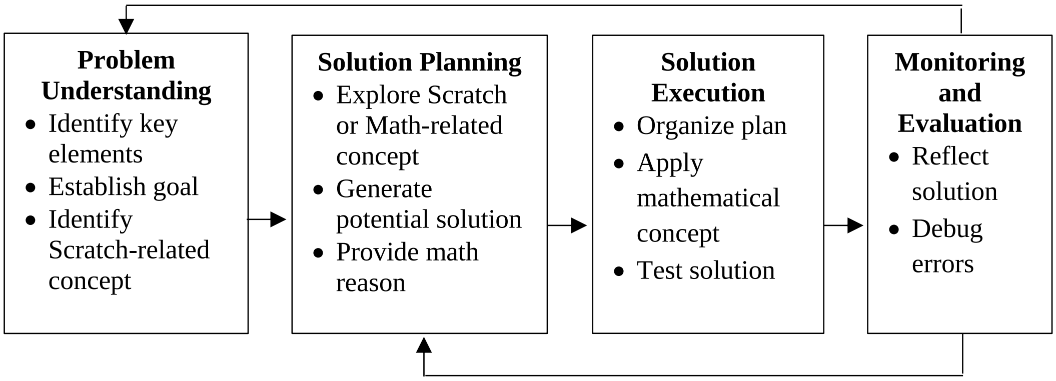



**FIGURE 2. Actual PS Process of Students in the Field Testing**

In the activities, students display their skill in understanding the problem by identifying its essential elements. Their ability to identify the material elements in solving the problem led them to establish goals that guided them in their solution planning. In planning, students started to generate potential solutions by determining possible code block arrangements in their coding screens (Çakiroglu & Mumcu, 2020). In this stage, they explored the Scratch concept to deepen their understanding of the functions of each code block. The understanding they acquired in their exploration was used to support the rationale of their plans. Hence, how the students provided mathematical reasoning in their plans was based on their exploration (Bouck & Yadav, 2020).

In organizing their plans, students drew on the application of geometric concepts. Students attempted to develop algorithms by following mathematical rules, a characterization of algorithmic thinking (Dagiene et al., 2017). Moreover, testing solution plans to validate the correctness of their solutions was a significant aspect of their PS process. It was also done to determine errors in their program codes.

Meanwhile, students evaluated and monitored program solutions after running their respective programs. At this stage, reflection on their solution was observable during the process. This is a crucial stage in their PS process as it led them to reflect on the possible reasons for their errors so that they could proceed to iterate their process (Kim et al., 2018).

### Students' PS Skill During the Implementation of the BBP-IMPS Activities

After analyzing students' processes and program codes, the researchers identified students who displayed variations in constructing their program solutions across activities. Hence, this study's findings intend to present two cases of PS and CT development. Case 1 and Case 2 represent the pairs of students whose produced program codes display consistent variations across implemented BBP-IMPS activities.

*Problem Understanding Skill*

Table 2 summarizes the ratings obtained for students' problem understanding skills in all implemented activities. It can be observed that identifying key elements of the problem and identifying Scratch-related concepts are the two characteristics consistently observed from students in both cases across all activities. This finding is consistent with Oliveri et al.'s (2017) proposition that in solving problems, students start by defining the problem through identifying its central elements. An example of students' interactions reflecting these characteristics in solving problem understanding says, "[$S_1$] *We can use go to [block] here since it can create a slanted line without the need to turn its [sprite's] direction*."

**TABLE 2. Summary of Ratings for Problem Understanding Skill**

| Problem Understanding | Act 1 | Act 2 | Act 3 |
|---|---|---|---|
| | Mode | Mode | Mode |
| ***CASE 1*** | | | |
| Identify key elements | 1 | 1 | 1 |
| Establish goal | 0 | 0 | 1 |
| Identify Scratch related concept | 1 | 1 | 1 |
| ***CASE 2*** | | | |
| Identify key elements | 1 | 1 | 1 |
| Establish goal | 1 | 0 | 1 |
| Identify Scratch related concept | 1 | 1 | 1 |

***Note:*** *The data in the table follows the description: 0- Not Observed and 1- Observed.*

In Case 1, students did not engage in goal establishment, at least for the first and second activities; however, they learned to display the skill in the final activity. The improvement was evident when $S_1$ suggested, "*Let's improve this one; let us use this slider-slider [referring to variable] and operator. But let us put base from our output [in Activity 2].*" This suggestion reflected $S_1$'s desire to enhance their previous code in solving Activity 3 by integrating the CT concept variable.

On the other hand, students in Case 2 engaged in goal establishment in their initial and final activity; however, they did not exhibit the subskill in the second activity. This instance where Case 2 tends to disregard goal establishment in the process of understanding the problem is a potential consequence of their developed familiarity with the problem, as evident in $S_4$'s statement, *"It seems to resemble [the problem] in the previous activity. Let us try solving this one."* Congruent to Priemer et al.'s (2019) claim, developing familiarity with the problem situation directs students' solutions despite being unable to set goals explicitly.

Hence, Case 2 sustained their problem-understanding skills, while Case 1 improved them, manifesting in their learned ability to establish clear PS goals. Problem understanding is essential to their PS skill development as it helps them formulate their plans more effectively (Wei et al., 2021). As a result, both cases displayed more fluency in planning Activity 3 compared to the previous activities, given their established goal. This finding highlights the significance of establishing goals to pursue efficient planning within programming-related problems (Chao, 2016).

*Solution Planning Skill*

Table 3 summarizes the ratings obtained for the students' solution planning skills in all three activities. Students generally generated potential solutions by planning the arrangement of code blocks. In doing so, they were focused on determining the sequence of their program. An example of how students generate their program solution is exemplified by $S_3$ saying, "*Just place [turn after turn] lines, only two of them. That's good. Then, place them inside repeat [block]. Then insert [wait time]*."

**TABLE 3. Summary of Rating for Solution Planning Skill**

| Solution Planning | Act 1 | Act 2 | Act 3 |
|---|---|---|---|
| | Mode | Mode | Mode |
| ***CASE 1*** | | | |
| Explore Math or Scratch Concept | 1 | 1 | 1 |
| Generate potential solution | 1 | 1 | 1 |
| Provide mathematical reasoning | 0 | 1 | 1 |
| ***CASE 2*** | | | |
| Explore Math or Scratch Concept | 1 | 1 | 1 |
| Generate potential solution | 1 | 1 | 1 |
| Provide mathematical reasoning | 0 | 0 | 1 |

***Note:*** *The data in the table follows the description: 0- Not Observed and 1- Observed.*

However, it can be noticed that the students failed to provide mathematical reasoning when planning the first activity. This indicates they lack understanding of the mathematical foundations of their plans during this activity. This further led them to use a trial-and-error strategy to build their program solution. Nevertheless, both cases are already engaged in mathematical reasoning in their final activity. As a result, students already generated their plans based on mathematical rules. It eventually increased the efficiency of their solution process since it lessened their reliance on the trial-and-error method. The finding implies that solutions are more systematic when plans are supported with math-informed reasoning (Cui et al., 2023).

The significance of providing mathematical reasons in planning can be emphasized in the second activity, where this characteristic was observed by students in Case 1 but not in Case 2. In this activity, Case 1 planned fluently and developed a systematic approach to solving the problem more than Case 2. This claim was evident in how well students in Case 1 organized their solution in creating a program that draws a square.

In general, both cases show improvement in solution planning skills. The improvement is marked by students' learned ability to provide mathematical reasons for their plans, particularly in the final activity. The ability allowed them to construct solid plans and helped them increase the efficiency of their solution processes (Çakiroglu & Mumcu, 2020). This conclusion parallels the claim that engaging in mathematical reasoning increases success in solving math-related problems (Ayal et al., 2016).

*Solution Execution Skill*

Table 4 displays the summary of ratings obtained for the students in terms of solution execution skills across activities. Students tend to execute their solutions by organizing what has been discussed among their pairs during the planning stage. They organized the arrangement of the code blocks they had planned and applied the things they had learned during their explorations.

Testing every solution was also integral to the students' PS process, as this was consistently observed from the students' processes across all activities. An example of this claim is reflected when $S_4$ said, "*What was it [discussed during planning]? Reduce the size of [sprite]. Then, place the [sprite] here and then "go to" here. Next is pen, erase all, repeat [block], and move. Then wait [block] and turn left [block]. Let us see!*" The statement reflects how students test the solutions they have conceived in the planning stage.

Meanwhile, both cases vary their solution execution skills regarding how they apply mathematical concepts in their solutions. While Case 1 was consistent in drawing the application of math concepts in their solutions, Case 2 did not apply them in their solution for the first two activities. This instance can be traced back to the latter failing to support their plans with mathematical reasoning during these activities. As a result, their solution process tends to characterize a trial-and-error strategy rather than an application of mathematical rules (Sumartini, 2018). This further affected the success of their solution, particularly in the first activity, where their final program did not draw a perfect match with the figure they were tasked to construct.

**TABLE 4. Summary of Rating for Solution Execution Skill**

| Solution Execution | Act 1 | Act 2 | Act 3 |
|---|---|---|---|
| | Mode | Mode | Mode |
| ***CASE 1*** | | | |
| Organize argument or plans | 1 | 1 | 1 |
| Apply mathematical concept | 1 | 1 | 1 |
| Test solution plan | 1 | 1 | 1 |
| ***CASE 2*** | | | |
| Organize argument or plans | 1 | 1 | 1 |
| Apply mathematical concept | 0 | 0 | 1 |
| Test solution plan | 1 | 1 | 1 |

***Note:*** *The data in the table follows the description: 0- Not Observed and 1- Observed.*

Therefore, data supports that for Case 1, students sustained their skill in solution execution across activities. On the other hand, students in Case 2 improved the said skill, as revealed by how they improved their solution process from the initial to their final activity. The improvement is highlighted by how Case 2 learned to apply mathematical concepts in executing their solution plan. This helped students enhance the fluency and accuracy of their program code (Ran, et al., 2020). It is said that PS skills in the integrated math and programming domain must include the ability to apply mathematical concepts efficiently (Lu & Fletcher, 2009).

*Monitoring and Evaluation Skill*

Table 5 summarizes the ratings obtained for students' monitoring and evaluation skills. Regarding reflecting program solutions, both cases consistently exhibited the said characteristic across all activities. Reflection allowed the students to evaluate the correctness of their solutions. It also helped them identify errors in their programs so that they could be susceptible to fixing (Angeli et al., 2016).

**TABLE 5. Summary of Rating for Monitoring and Evaluation Skill**

| Monitoring and Evaluation | Act 1 | Act 2 | Act 3 |
|---|---|---|---|
| | Mode | Mode | Mode |
| ***CASE 1*** | | | |
| Reflect solution | 1 | 1 | 1 |
| Test and debug program | 1 | 1 | 1 |
| ***CASE 2*** | | | |
| Reflect solution | 1 | 1 | 1 |
| Test and debug program | 0 | 1 | 1 |

***Note:*** *The data in the table follows the description: 0- Not Observed and 1- Observed.*

It can be noticed that the subskill testing and debugging programs were not observable in the first activity for students in Case 2. The inefficiency of Case 2 in debugging the errors of their programs can be attributed to the fact that their planning and execution of solutions need to be anchored on mathematical knowledge. As evidence, their output in the first activity did not accurately match the figure presented in their worksheets. Nevertheless, debugging was already observable in Case 2 during the second and third activities, as evident in the statement "[$S_3$] *We need only to adjust this [variable slider] $S_4$ so that we do not need to deal with other blocks. We'll find ways on how to deal with it.*" This is an essential aspect of students' monitoring and evaluation skills since it allows them to enhance accuracy and maximize their program solutions (Bocconi et al., 2016).

Meanwhile, for Case 1, the students sustain all characteristics of monitoring and evaluating skills across activities. This is essential to their PS skill development, enabling them to revisit program codes to construct a more refined solution (NRC, 2013).

**Students' CT Skill During the Implementation of the BBP-IMPS Activities**

In this study, students' CT skills were measured in terms of their decomposition, algorithmic thinking, and pattern recognition skills, which are the three cornerstones addressed in mapping CT and mathematics learning competencies. Hence, the investigation of CT skills was limited to these three cornerstones.

*Decomposition Skill*

Table 6 summarizes the ratings obtained for students' decomposition skills across activities. In both cases, students sustained their decomposition skills across activities.

**TABLE 6. Summary of Rating for Decomposition Skill**

| Decomposition | Act 1 | Act 3 |
|---|---|---|
| | Mode | Mode |
| ***CASE 1*** | | |
| Make decisions about dividing a task into subtasks with integration in mind, | 1 | 1 |
| Sort subtasks into categories and place them in structured order | 1 | 1 |
| Think solution of problem in terms of its components | 1 | 1 |
| ***CASE 2*** | | |
| Make decisions about dividing a task into subtasks with integration in mind | 1 | 1 |
| Sort subtasks into categories and place them in structured order | 1 | 1 |
| Think solution of the problem in terms of its components | 1 | 1 |

***Note:*** *The data in the table follows the description: 0- Not Observed and 1- Observed.*

Students focused their PS processes on determining parts of the figure tasks and constructed program codes for each identified part. As evidence, some interactions are presented below:

> "[$S_1$] *I've already made one point, then move 10 steps. Now, I have already two points. You can help me in solving the last part…we can address this angle.*"
>
> "[$S_3$] *Let's address this one first. Let's put base code first [referring to previous work] before we proceed to integrate slider-slider [referring to variable block]. Then, let's deal with this one [referring to operator].*"

From the statements above, it can be observed how students determined subparts of the problem, which they addressed separately to solve the whole problem. It can also be observed that students sorted the determined subparts in terms of priorities. According to Andrian & Hikmawan (2021), this technique is an essential aspect of decomposition, which helps students abstract a whole problem.

Thus, in both cases, students managed to exhibit all characteristics essential in problem decomposition. This implies that they banked on the strategy of addressing an entire problem task in terms of its material subparts (Durak, 2018). Such a strategy allowed them to create program solutions more manageably, which is a crucial cognitive strategy in constructing program codes (Kwon & Cheon, 2019).

*Algorithmic Thinking Skill*

Table 7 summarizes ratings obtained for students' algorithmic thinking skills across activities. In solving Activities 1 and 2, students arranged the code blocks needed in the program following the sequence displayed by the sprite's movement in the problem tasks. Thus, the fourth characteristic of algorithmic thinking was observable in both cases across activities. Sentance & Czismadia (2016) considered this aspect of algorithmic thinking crucial in enhancing students' PS skills.

**TABLE 7. Summary of Rating for Algorithmic Thinking Skill**

| Algorithmic Thinking | Act 1 | Act 2 |
|---|---|---|
| | Mode | Mode |
| ***CASE 1*** | | |
| Think in terms of sequence | 1 | 1 |
| Think and solve problems in terms of rules | 0 | 1 |
| Develop a systematic solution to the problem | 0 | 1 |
| Identify steps that can be communicated as instructions, codes or programs to other people or to computing devices | 1 | 1 |
| ***CASE 2*** | | |
| Think in terms of sequence | 1 | 1 |
| Think and solve problems in terms of rules | 0 | 0 |
| Develop a systematic solution to the problem | 0 | 1 |
| Identify steps that can be communicated as instructions, codes or programs to other people or to computing devices | 1 | 1 |

***Note:*** *The data in the table follows the description: 0- Not Observed and 1- Observed.*

In both cases, students did not devise their program solution based on mathematical rules in Activity 1. In effect, students' solutions were not systematic. It means that they resorted more to exploring different possible arrangements of code blocks. However, in Activity 2, Case 1 already exhibited the subskill thinking solutions based on rules. As a result, they exhibited systematic thinking when planning the solution. This finding can be confirmed by how $S_1$ explained their solution:

> "*First is we arrange the code blocks in right sequence. For example, we determine first point [referring to the starting point of the sprite]. Second, we determine direction of the sprite so that it moves in that specific direction. Then, we insert erase all and then pen down for drawing purposes. Then, we put wait [time] before moving 100 steps. The 100-step movement will determine the length of the sides of our square. Then we put 90 degrees for turn [block] so that it forms a specific angle for square. Then, we utilize repeat block to repeat the process four times, and eventually make a square.*"

On the other hand, students in Case 2 were able to develop systematic solutions for the problem in Activity 2; however, rules-based thinking is still not observable. This can be attributed to the fact that they did not apply mathematical concepts in executing their solution (Zeng et al., 2023).

Students in both cases still showed improvement in terms of algorithmic thinking skills across activities. For Case 1, the improvement is emphasized by how students learned to apply rules-based thinking and develop systematic solutions to the problem. Case 2's improvement is marked by how students learned to develop systematic solutions to a problem. Such improvement plays a very crucial role in students' algorithmic thinking skills as it reflects systematic thinking in their solutions (Dogan, 2020).

*Pattern Recognition Skill*

Table 8 presents the ratings for students' pattern recognition skills. Students from both cases exhibited the first and third characteristics essential to pattern recognition across activities. It means that students identified patterns and similarities and used this technique to represent a sequence of codes in the process (Taylor, 2018). For instance, both cases determine the significance of using repeat blocks in their program codes. This recognition of patterns allowed the students to predict the next set of code blocks needed to continue the correct movement of the sprite.

**TABLE 8. Summary of Rating for Pattern Recognition Skill**

| Pattern Recognition | Act 2 | Act 3 |
|---|---|---|
| | Mode | Mode |
| ***CASE 1*** | | |
| Identify patterns and similarities between problems or sub-problems | 1 | 1 |
| Utilize general solution | 1 | 1 |
| Use patterns to describe and represent sequences in data or process | 1 | 1 |
| Make predictions based on arrangement and relationship between parts | 1 | 1 |
| ***CASE 2*** | | |
| Identify patterns and similarities between problems or sub-problems | 1 | 1 |
| Utilize general solution | 0 | 1 |
| Use patterns to describe and represent sequences in data or process | 1 | 1 |
| Make predictions based on arrangement and relationship between parts | 0 | 1 |

***Note:*** *The data in the table follows the description: 0- Not Observed and 1- Observed.*

In doing Activity 2, students in Case 2 failed to recognize the patterns existing in their program solution. This prevented them from generalizing their solution using the repeat block, resulting in a lengthy code (NRC, 2013). On the other hand, students in Case 1 immediately identified patterns in their code as reflected in S1's statement, "*Didn't we repeat the same blocks four times in creating square? Then we repeat the same set of blocks three times for triangle. Can we use slider (variable block) so that the motion of the sprite be just repeated?*" The statement reflects an intent to utilize the repeat block in generalizing the patterns existing in their solution, resulting in the utilization of a shorter code.

In Activity 3, students from both cases exhibited all characteristics essential to pattern recognition. However, what is notable is Case 2's development, which finally utilizes patterns in generalizing their solution using the repeat block. This eventually allowed them to predict the sprite's movement whenever they changed the value of the repeat block (Kalelioglu & Gülbahar, 2014).

The findings suggest that students in Case 1 sustained their pattern recognition skills across activities. The findings also reveal that students in Case 2 improved pattern recognition skills, manifested in how they learned to utilize patterns to build generalized solutions. The improvement shown in Case 2 highlights the significance of pattern recognition in building quick solutions to problems (Dazgupta & Purzer, 2016).

## CONCLUSION AND RECOMMENDATIONS

This study demonstrates the potential of BBP-IMPS activities in incorporating CT effectively within the Grade 7 mathematics education curriculum. Following the research and development model and guided by backward curriculum design principles, the developed activities align block-based programming with mathematics competencies, providing students with opportunities to foster PS and CT skills through hands-on practice. Field experts confirmed that the activities are suitable for meeting needs and standards in mathematics education and are appropriate for classroom implementation. Observations during the study revealed improvement in students' PS and CT skills over the three activities. This finding supports the conclusion that the BBP-IMPS framework contributes to achieving the goals of Philippine K-12 mathematics education.

Additionally, this study contributes to CT instruction, mathematics education, and pedagogy. It proposes a concrete model for integrating CT into mathematics education using BBP that can be adapted for K-12 contexts and is accessible to students without prior programming experience. It also offers an outline of curriculum design that developers can use as a foundation to infuse CT skills within the standards of the country's mathematics education framework. It demonstrates how the integration of CT and mathematics fosters PS while highlighting two relevant activities, i.e., (i) constructing CT outputs while drawing on the concepts of mathematics and (ii) understanding mathematics concepts while participating in CT activities.

Furthermore, this study recommends (1) investigating the effectiveness of BBP-IMPS activities across different grade levels considering large sample size; (2) implementing BBP-IMPS framework using other BBP language to establish its acceptability across languages; (3) conducting longitudinal studies to examine the long-term impact of the sustained use of BBP-IMPS activities on students' PS and CT skills; (4) explore alternative assessments to obtain more comprehensive understanding of students' PS and CT skills; and (5) explore on professional development programs focused on integrating BBP into DepEd's K-12 mathematics curriculum.

## ACKNOWLEDGEMENTS

The authors would like to express their immense gratitude to the Capacity Building Program in Science and Mathematics Education (CBPSME) of the Philippines' Department of Science and Technology – Science Education Institute (DOST-SEI) for the financial funding that made this study possible.

## REFERENCES

Andrian, R., & Hikmawan, R. (2021). The importance of computational thinking to train structured thinking in problem solving. *Jurnal Online Informatika*, *6*(1), 113. https://doi.org/10.15575/join.v6i1.677

Angeli, C., Voogt, J., Fluck, A., Webb, M., Cox, M., Malyn-Smith, J., & Zagami, J. (2016). A K-6 Computational Thinking Curriculum Framework: Implications for Teacher Knowledge. *Educational Technology & Society, 19 (3), 47–57.*

Ayal C., Kusuma, Y., Sabandar, J., & Dahlan, J. (2016). *The enhancement of mathematical reasoning ability of junior high school students by applying mind mapping strategy.* JEP 7, 25-31

Bala, R. B. (2021). Scratch in Teaching Programming: Effect on Problem Solving Skill and Attitude. *International Journal of Quality in Education, 5(2), 63-81*. https://dergipark.org.tr/tr/pub/ijqe

Basogain, X., Olabe, M. Á., Olabe, J. C., & Rico, M. J. (2018). Computational thinking in pre-university blended learning classrooms. *Computers in Human Behavior, 80,* 412-419. https://doi.org/10.1016/j.chb.2017.04.058

Bati K., Yetişir M. I., Çalişkan I., Güneş G., Saçan E. G. (2018). Teaching the concept of time: A steam-based program on computational thinking in science education. *Cogent Education*, 5(1), 1–16. https://doi.org/10.1080/2331186X.2018.1507306

Bocconi, S., Chioccariello, A., Dettori, G., Ferrari, A., Engelhardt, K., Kampylis, P., & Punie, Y. (2016). Developing Computational Thinking in Compulsory Education- Implications for policy and practice. *Publications Office of the European Union*. http://dx.doi.org/10.2791/792158

Borg, W., & Gall, M. (1983). *Educational research: An introduction*. Longman, New York.

Bouck, E. C., & Yadav, A. (2020). Providing access and opportunity for computational thinking and computer science to support mathematics for students with disabilities. *Journal of Special Education Technology, 37(1), 151–160*. https://doi.org/10.1177/0162643420978564

Brennan, K., & Resnick, M. (2012). *New frameworks for studying and assessing the development of computational thinking* [Conference session]. Proceedings of the 2012 annual meeting of the American Educational Research Association, Vancouver, Canada.

Çakiroglu, Ü., & Mumcu, S. (2020). Focus-fight-finalize (3F): Problem-solving steps extracted from behavioral patterns in block based programming. *Journal of Educational Computing Research, 58(7), 1279–1310*. https://doi.org/10.1177/0735633120930673.

Chao, P. Y. (2016). Exploring students' computational practice, design and performance of problem solving through a visual programming environment. *Computers & Education, 95, 202–215*.

Cui, Z., Ng, O., & Jong, M. S.-Y. (2023). Integration of Computational Thinking with Mathematical Problem-based Learning: Insights on Affordances for Learning. *Educational Technology & Society, 26(2), 131-146*. https://doi.org/10.30191/ETS.202304_26(2).0010

Dagienė, V., Sentance, S., & Stupurienė, G. (2017). Developing a two-dimensional categorization system for educational tasks in informatics. *Informatica, 28*(1), 23-44. https://doi.org/10.15388/informatica.2017.119

Dasgupta, A., & Purzer, S. (2016). No patterns in pattern recognition: A systematic literature review. *EEE Frontiers in Education Conference (FIE),* 1–3. https://doi.org/10.1109/FIE.2016.7757676

Daulay, P., & Zaman, B. (2012). PENGEMBANGAN MODEL PENELUSURAN DISKUSI TUTORIAL ONLINE MELALUI APLIKASI FAQ (FREQUENTLY ASK QUESTION). *Jurnal Pendidikan Terbuka Dan Jarak Jauh, 13*(2), 80-93. http://jurnal.ut.ac.id/index.php/jptjj/article/view/413

De Chenne, A., & Lockwood, E. (2022). A Task to connect counting processes to lists of outcomes in combinatorics. *The Journal of Mathematical Behavior, 65, 100932*. https://doi.org/10.1016/j.jmathb.2021.100932

DepEd. (2012). *K to 12 Mathematics Curriculum Guide*. K to 12 BASIC EDUCATION CURRICULUM. http://lrmds.deped.gov.ph/

Doğan, A. (2020). Algorithmic thinking in primary education. *International Journal of Progressive Education, 16*(4), 286–301. https://doi.org/10.29329/ijpe.2020.268.18

Durak, H. (2018). The effects of using different tools in programming teaching of secondary school students on engagement, computational thinking and reflective thinking skills for problem solving. *Technology, Knowledge and Learning, 25*(1), 179-195. https://doi.org/10.1007/s10758-018-9391-y

Fraillon, J., Ainley, J., Schulz, W., Duckworth, D., & Friedman, T. (2019). IEA international computer and information literacy study 2018 assessment framework. *Springer*.

Gay, L. R., Mills, G. E., & Airasian, P. W. (2012). Educational research: Competencies for analysis and applications. *Pearson College Division*.

Harimurti, R., Ekohariadi, E., Munoto, M., Asto B, I. G., & Winanti, E. T. (2019). Analysis of programming skills concept in developing problem solving skills. https://doi.org/10.21831/jptk.v25i1.22638

Hickmott, D., Prieto-Rodriguez, E., & Holmes, K. (2018). A Scoping review of studies on computational thinking in K–12 mathematics classrooms. *Digital Experiences in Mathematics Education, 4(1), 48–69*. https://doi.org/10.1007/s40751-017 0038-8

Hurt, T., Greenwald, E., Allan, S., Cannady, M. A., Krakowski, A., Brodsky, L., Collins, M. A., Montgomery, R., & Dorph, R. (2023). The computational thinking for science (CT-S) framework: operationalizing CT-S for K–12 science education researchers and educators. *International Journal of STEM Education*, *10*(1). https://doi.org/10.1186/s40594-022-00391-7

Intel Education. (2011). *Designing Effective Projects: Intel Teach Development Programs*. Intel Corporation.

Israel-Fishelson., R. & Hershkovitz, A. (2022). Studying interrelations of computational thinking and creativity: A scoping review (2011–2020). *Computers and Education*, 176, 104353. https://doi.org/10.1016/j.compedu.2021.104353

Jocius, R., Ian O'byrne, W., Albert, J., Joshi, D., Robinson, R., & Andrews, A. (2021). Infusing computational thinking into STEM teaching. *Technology & Society*, *24*(4), 166–179. https://www.jstor.org/stable/48629253

Kalelioglu, F., & Gülbahar, Y. (2014). The effects of teaching programming via Scratch on problem solving skills: A discussion from learners' perspective. *Informatics in Education, 13(1),* 33–50. https://doi.org/10.15388/infedu.2014.03

Kastner-Hauler, O., Tengler, K., Sabitzer, B., & Lavicza, Z. (2022). Combined effects of Block-Based Programming and Physical Computing on primary students' computational thinking skills. *Frontiers in Psychology, 13*. https://doi.org/10.3389/fpsyg.2022.875382

Kim, C., Yuan, J., Vasconcelos, L., Shin, M., & Hill, R. B. (2018). Debugging during block-based programming. *Instructional Science*, *46*(5), 767–787. https://doi.org/10.1007/s11251-018-9453-5

Kwon, K., & Cheon, J. (2019). Exploring problem decomposition and program development through block-based programs. *International Journal of Computer Science Education in Schools*, *3*(1), 3-16. https://doi.org/10.21585/ijcses.v3i1.54

Lin, Y., Chen, S., Tsai, C., & Lai, Y. (2021). Exploring computational thinking skills training through augmented reality and AIoT learning. *Frontiers in Psychology*, *12*. https://doi.org/10.3389/fpsyg.2021.640115

Lu, J. & Fletcher, G.H. (2009). Thinking about computational thinking. *ACM SIGCSE Bulletin*, *41*(1), 260-264. https://doi.org/10.1145/1539024.1508959

Lye, S. Y., & Koh, J. H. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12? *Computers in Human Behavior*, *41*, 51-61. https://doi.org/10.1016/j.chb.2014.09.012

McCain, T. (2005). Teaching for tomorrow: Teaching content and problem-solving skills. *Corwin Press*.

McClelland, K., & Grata, L. (2018). A Review of the Importance of Computa tional Thinking in K-12. *The Tenth International Conference on Mobile, Hybrid, and On-line Learning, 32-34*.

National Research Council, Division on Engineering and Physical Sciences, Computer Science and Telecommunications Board, & Committee for the Workshops on Computational Thinking. (2011). *Report of a workshop on the pedagogical aspects of computational thinking*. National Academies Press.

National Research Council. (2013). Education for life and work: Developing transferable knowledge and skills in the 21st century. Division of Behavioral and Social Sciences and Education, Board on Science Education, Board on Testing and Assessment, Committee on Defining Deeper Learning and 21st Century Skills. *National Academies Press*.

OECD. (2013). *PISA 2012 assessment and analytical framework mathematics, reading, science, problem solving and financial literacy: Mathematics, reading, science, problem solving and financial literacy*. OECD Publishing.

Oliveri, M. E., Lawless, R., & Molloy, H. (2017). A literature review on collaborative problem solving for college and workforce readiness. *ETS Research Report Series, 2017(1), 1-27*. https://doi.org/10.1002/ets2.12133

Özpınar, İ., & Arslan, S. (2023). Teacher-based evaluation of students' problem solving skills. *International Journal of Psychology and Educational Studies*, *10*(2), 543–560. https://doi.org/10.52380/ijpes.2023.10.2.1160

Priemer, B., Eilerts, K., Filler, A., Pinkwart, N., R¨ osken-Winter, B., Tiemann, R., & Zu Belzen, A. U. (2019). A framework to foster problem-solving in STEM and computing education.

*Research in Science & Technological Education, 38(1), 105-130*. https://doi.org/10.1080/02635143.2019.1600490
Pršala, J. (2024). The effect of block-based programming activities on the computational thinking skills of pre-service primary school teachers. *Journal of Pedagogical Sociology and Psychology*. https://doi.org/10.33902/jpsp.202426267
Ran, H., Kasli, M., & Secada, W. (2020). A meta-analysis on computer technology intervention effects on mathematics achievement for low-performing students in K–12 classrooms. *Journal of Educational Computing Research*, *59*(1), 119–153.
Sengupta, P., Dickes, A., & Farris, A. (2018). Toward a phenomenology of computational thinking in STEM education. In *Springer eBooks* (pp. 49–72). https://doi.org/10.1007/978-3-319-93566-9_4
Sentance, S., & Csizmadia, A. (2016). Computing in the curriculum: Challenges and strategies from a teacher's perspective. *Education and Information Technologies, 22(2), 469-495*. https://doi.org/10.1007/s10639-016-9482-0
Sumartini, T. S. (2018). Peningkatan Kemampuan Pemecahan Masalah Matematis Siswa melalui Pembelajaran Berbasis Masalah. Mosharafa: *Jurnal Pendidikan Matematika, 5(2), 148–158*. https://doi.org/10.31980/mosharafa.v5i2.270
Taylor, K. (2018). Computational Thinking Framework. Let's Talk Science. Lon don, Ontario, Canada.
Totan, H. N., & Korucu, A. T. (2023). The Effect of Block Based Coding Education on the Students' Attitudes about the Secondary School Students' Computational Learning Skills and Coding Learning: Blocky Sample. *Participatory Educational Research*, *10*(1), 443–461. https://doi.org/10.17275/per.23.24.10.1
Wei X., Lin L., Meng N., Tan W., Kong Kinshuk S. C. (2021). The effectiveness of partial pair programming on elementary school students' computational thinking skills and self-efficacy. *Computers and Education,* 160, 104023. https://doi.org/10.1016/j.compedu.2020.104023
Wiggins, G. P., & McTighe, J. (2011). *The understanding by design guide to creating high-quality units*. ASCD.
Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49(3), 33-35*. https://doi.org/10.1145/1118178.1118215
Wu, T., Asmara, A., Huang, Y., & Hapsari, I. P. (2024). Identification of Problem-Solving Techniques in Computational Thinking Studies: Systematic Literature review. *SAGE Open*, *14*(2). https://doi.org/10.1177/21582440241249897
Ye, H., Liang, B., Ng, O.-L., & Chai, C. S. (2023). Integration of computational thinking in K-12 mathematics education: A systematic review on CT-based mathematics instruction and student learning. *International Journal of STEM Education, 10(1)*. https://doi.org/10.1186/s40594-023-00396-w
Yu, Q., Yu, K., & Li, B. (2024). Effects of Block-Based Visual Programming on K-12 students' learning outcomes. *Journal of Educational Computing Research*. https://doi.org/10.1177/07356331241293163
Yusoff, K. M., Sahari, N., Siti, T., & Mohd, N. (2020). Analysis on the requirements of computational thinking skills to overcome the difficulties in learning programming. *International Journal of Advanced Computer Science and Applications*, *11*(3). https://doi.org/10.14569/ijacsa.2020.0110329
Zeng, Y., Yang, W., & Bautista, A. (2023). Teaching programming and computational thinking in early childhood education: a case study of content knowledge and pedagogical knowledge. *Frontiers in Psychology*, *14*. https://doi.org/10.3389/fpsyg.2023.1252718

*International Journal of Science Education and Teaching*
*Vol. 3, No. 3, pp. 113 - 125, Sep. – Dec. 2024*
*http://doi.org/10.14456/ijset.2024.10*

# Interest And Performance of Grade 10 Students in Science Modular Learning

**Mamdoh S. Laguindab[1*] and Salamah M. Basher [2]**
[1]*RPMD National Science High School, Marawi City, Philippines*
[2]*Mindanao State University, Marawi City, Philippines*
*Email: mamdohlaguindab@gmail.com and salamah.indad@msumain.edu.ph

Received: 13 October 2024; Revised: 15 November 2024; Accepted: 19 November 2024

**Abstract.** This study aimed to investigate the students' performance in science caused by their interest in science and their demographic profile indicated by sex, parents' highest educational attainment, family income, and availability of gadget in a modular learning setting. An interest survey questionnaire which was developed and validated by the researcher and a demographic profile survey were administered to 122 grade 10 students of RPMD National Science High School, Marawi City, Philippines in the first month of the school year 2021-2022. The students were exposed to offline modular learning throughout the school year. The average final grade of the students at the end of the school year served as their performance in science. To support the quantitative data, in-depth interviews were conducted after the final grades became available. The quantitative data were treated and analyzed using descriptive statistics, Pearson correlation coefficient, Chi-square test, and Gamma test while the qualitative data were analyzed thematically. The findings of the study revealed that the interest in science was significantly related to performance in science ($r=.522$, $p<.05$, two tailed test). It also revealed that the sex was significantly associated with the interest in science ($\chi^2=7.149$, $p<.05$, $df=2$) and the performance in science ($\chi^2=8.57$, $p<.05$, $df=2$) in favor of girls. However, parents' highest educational attainment, family income and availability of gadgets were not associated significantly with both interest and performance in science. Meanwhile, the qualitative data revealed that the higher the interest in science of the students, the greater the efforts they exerted in studying the modules and the higher the performance obtained. The study's limitations include a small sample size, the lack of significant associations with parents' education and family income, and its focus on offline modular learning, which may not apply to other educational contexts. Additionally, the timing of interviews after performance assessments may have limited insights into how students' interest developed. Practical implications suggest the need for more engaging teaching strategies, gender-responsive policies, professional development for modular learning, and further research to explore other factors influencing science performance.

**Keywords:** Science Interest, Science Performance, Modular Learning, Demographic Factors, Marawi City Students

## INTRODUCTION

In this era of information technology, countries are in a race to advance their own science and technology to cope with the rapid changes of the world. In fact, different countries are investing efforts in the advancement of their own educational system in order to prepare their young generation to compete on a global level by providing them with the necessary knowledge and skills. In the Philippines, however, the education system underperformed among peers in East Asia and the Pacific (Cordero, 2018) which can be attributed to several factors such as the ineffective teaching-learning practices.

The ineffective teaching-learning practices resulted to the Filipino students' poor performance in many international assessment tests. For example, the 1999 Trends in International Mathematics and Science Study (TIMSS) results revealed a very dismal performance in science where it ranked 36th out of the 38 participating countries. In 2003 TIMSS, it landed 43rd out of 46 participating countries (Raya, 2021). In 2019 TIMSS, the country scored lowest among 58 participating countries in both mathematics and science (Magsambol, 2020). Similarly, the 2018 Program for International Student Assessment (PISA) result revealed that the Filipino students ranked second lowest in both mathematics and science out of the 79 participating countries (San Juan, 2019). In the 2022 PISA, the Philippines ranked second lowest among 81 participating countries (Malipot, 2023). The consistent declining performance of Filipino students in science in the above international assessments mirrored the degenerating academic performance of the students in the classrooms which is alarming that needs immediate actions of all school stakeholders.

### Students' Performance in Science

Academic achievement is a critical aspect of education, reflecting the extent to which students attain their educational goals and demonstrating their knowledge and skills over time (Farooq et al., 2011). According to Narad and Abdullah (2016), academic performance encompasses the knowledge acquired by students, evaluated by educators within a specific timeframe. The significance of students' performance in subjects like science and mathematics extends beyond individual achievement; it serves as a barometer for a nation's socio-economic development. High performance in these areas is indicative of a well-prepared workforce capable of driving economic growth and innovation (Ali et al., 2013). Conversely, poor outcomes in these subjects may signal underlying issues within the educational system and broader socio-economic challenges.

Research consistently identifies a range of factors influencing academic performance, including teaching methodologies and demographic characteristics. Effective student-centered teaching strategies have been shown to enhance engagement and understanding, thereby improving academic outcomes (Ampaso, 2019; Pagayocan, 2018). Additionally, demographic variables such as gender (Joseph et al., 2015; Kohlhass et al., 2010), parental education levels (Edris et al., 2020; Abu Bakar et al., 2017), family income (Machebe et al., 2017), and access to technology (Behnke et al., 2005 as cited in Collison, 2020) significantly impact students' academic success. Understanding these factors is essential for developing targeted interventions that can enhance educational outcomes across diverse student populations.

### Students' Interest in Science

Student interest in learning science is a critical factor influencing academic performance and engagement. Interest can be defined as the desire to be involved with and explore a subject, which in the context of science education, reflects students' curiosity and eagerness to learn about scientific concepts (Mappadang et al., 2022). Hidi (1990) argues that interest acts as a cognitive resource that enhances learning outcomes, leading to improved academic achievement. When students exhibit a high level of interest in science, they are more likely to commit to studying the subject, which can also translate into increased enrollment in STEM-related fields (Hulleman & Harackiewicz, 2009). Thus, fostering a genuine interest in science among students is essential for their educational development and future career paths.

In addition to intrinsic interest, student-centered teaching strategies have been shown to effectively enhance students' engagement and curiosity in science (Ampaso, 2019; Pagayocan, 2018). Research indicates that various demographic factors also play a role in shaping students' interest levels. For instance, studies have linked interest in science to gender differences (Jia et al., 2020; Kang et al., 2018), parental educational attainment (Hacieminoglu, 2015; Dabney et al., 2016; Halim et al., 2018), family income (Conel, 2021; Halim et al., 2018), and access to technology (Syaputri & Usman, 2019). These factors can significantly impact how students engage with scientific content and their overall academic performance. Understanding these influences is crucial for developing effective educational strategies that promote sustained interest and achievement in science.

**Modular Learning**

Modular learning is an educational approach that organizes curriculum content into self-contained units, allowing students to engage with material at their own pace. In the Philippines, this modality has become essential during the COVID-19 pandemic, facilitating continued education through printed self-learning modules (SLMs) and digital resources (Cañete & Potane, 2022). This method promotes self-directed learning, enabling students to take responsibility for their education while receiving minimal direct instruction from teachers (Roque, 2023). Key components of modular learning include well-structured SLMs aligned with essential learning competencies, assessment tools, and feedback mechanisms to monitor student progress (DepEd Order No. 012, s. 2020). However, despite these components being in place, there is a significant gap in understanding how effectively these modules are implemented and received by students. Research indicates a need to explore how demographic factors—such as socio-economic status and parental involvement—affect student engagement and performance in modular learning settings (Mappadang et al., 2022).

Developing an effective modular learning program requires careful planning and collaboration among educators and stakeholders. This includes creating high-quality SLMs relevant to students' needs and ensuring that teachers are adequately trained to facilitate this mode of instruction (Cañete & Potane, 2022). Establishing robust feedback mechanisms is essential for continuous improvement in teaching strategies and module content. Despite significant efforts by the Department of Education (DepEd) to implement modular learning effectively, challenges remain regarding teacher preparedness and resource availability (Villar et al., 2022). Furthermore, there is a notable lack of empirical studies exploring the lived experiences of students within this modality and how these experiences influence their academic outcomes. Addressing these gaps is crucial for enhancing the effectiveness of modular learning in the Philippines and ensuring it meets the diverse needs of all learners.

Per literature review, there are very few studies that investigated both the performance and interest of students in science in a modular learning setting specifically in Marawi City. Previous studies conducted in nearby areas were focused on investigating the effects of some teaching interventions on students’ interest and performance in science (e.g. Ampaso, 2019; Pagayocan, 2018) and no attention has been given to explore the relationship between the two variables--students’ interest and students’ performance in a modular learning setting.

By exploring and understanding both performance and interest in science of students, we can comprehend better how to address the waning performance in science as well as how to enhance the deteriorating interest in science of the students. On the part of the teachers, they can explore more student-centered teaching methods in their classrooms that can ignite students’ interest that will lead to better performance. School administrators can promote more engaging learning activities by capacitating the science teachers through conducting trainings and seminars on both content knowledge and student-centered pedagogy. Learning materials writers can help by promoting more engaging learning activities. And finally, the study can hopefully enlighten future researchers on these crucial educational variables-performance and interest.

The study is anchored on the following theories: Constructivist Learning Theory, Experiential Learning Theory, and Csikszentmihalyi's Flow Theory.

Constructivism views learning as an active process where individuals construct and reconstruct meanings based on their past experiences, emphasizing that students learn by "constructing" knowledge with teachers acting as facilitators rather than traditional lecturers (Walker et al., 2008). In this study, students engaged in module activities designed to assess and build upon their prior knowledge, enabling them to construct new understanding even in the absence of a teacher. Similarly, David Kolb's Experiential Learning Theory focuses on learning through experience, consisting of four stages: concrete learning, reflective observation, abstract conceptualization, and active experimentation. This approach allows students to take responsibility for their learning by engaging in activities that enrich their experiences and apply their knowledge in real-world contexts (Western Governors University Blog, 2020). Lastly, Csikszentmihalyi's Flow Theory highlights the importance of student interest, describing flow as a state of intense focus and enjoyment that occurs when individuals are fully immersed in an activity (Csikszentmihalyi, 1990). In this study, students experienced flow while participating in various activities within the science module, leading to optimal engagement and intrinsic motivation.

Grounded with previous studies and the three theories, the researcher decided to conduct this research to look deeper into the significance relationship between students' interest and performance in science and the effect of some demographic profiles on these two variables-interest and performance in science.

Shown below is the framework of the research.

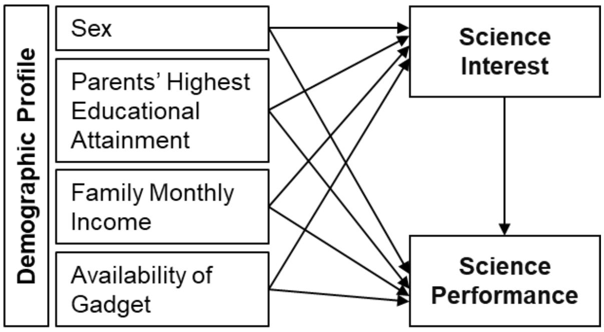



**Figure 1.** The framework of the research

The framework consists of independent variables and one dependent variable. The independent variables namely the students' demographic profile as indicated by sex, parents' highest educational attainment, family monthly income, and availability of gadget, and the students' interest in science are the possible factors that affect the students' performance in science in modular learning setting. However, the aforementioned independent variables are not necessary that they all affect the students' performance in science. In addition, the students' demographic profile as indicated by sex, parents' highest educational attainment, family monthly income, and availability of gadget are also possible factors that affect the students' interest in science. Similarly, not all the demographic profile indicators are necessary that they all affect the students' interest in science.

## RESEARCH OBJECTIVES

The objective of this study was to investigate the students' performance in science in relation to their level of interest in science and their demographic profile in an offline modular learning modality setting. Offline modular learning is a flexible educational approach that utilizes printed self-learning modules (SLMs) to facilitate individualized instruction for students. It is used in this study due to the limited access to internet connectivity in the research locale. Specifically, the study aimed the following: (1) assess the performance in science of grade 10 students in a modular learning setting; (2) assess the interest in science of the grade 10 students; (3) assess the demographic profiles in terms of sex, parents' highest educational attainment, family income and availability of gadget of grade 10 students; (4) determine if there is a significant association between students' levels of interest and performance in science and their demographic profiles, specifically in terms of sex, parents' highest educational attainment, family income, and availability of gadgets; (5) determine if there is a significant relationship between the students' interest in science and their performance in science; and (6) compare the difference of the students' level of interest in science with respect to their academic performance.

## METHODOLOGY

This study employed both descriptive and correlational research design. It is descriptive in the sense that it sought to describe the students' interest and students' performance in science as well as their demographic profile in terms of sex, parents' highest educational attainment, family income and availability of gadget. It is also correlational because it sought to investigate the existing relationship of the students' interest in science and the students' performance in science. Also, it sought to investigate the relationships of these two variables to students' demographic profiles such as sex, parents' highest educational attainment, family income and availability of gadget.

### Participants

The participants of this study consisted of two intact sections of Grade 10 students (N=122) of RPMD National Science High School in Marawi City, Philippines, during 2021-2022 school year. The entire population of Grade 10 students was included in the study, eliminating the need for sampling. The ages of the students ranged from 16 to 19 years old. They represented diverse family backgrounds, including those of government employees, business owners, farmers, and pedicab driver. Additionally, some were orphans who relied on themselves for financial support.

### Research Tools

To determine students' interest in learning science, a science interest survey questionnaire was employed. This instrument comprised 15 items on a 5-point Likert scale, which were consolidated from two existing renowned instruments. Seven of the items were adapted from the Interest in Science Scale of the Test of Science Related Attitudes (Fraser, 1981) and eight items were adapted from the Interest in Science Scale of the Mathematics and Science Attitude Survey (Paciorek, 1997). Following consolidation, the questionnaire underwent evaluations for face validity and content validity by three experts in science education research. After several modifications, the instrument was pilot tested, yielding a reliability coefficient of .816 as measured by Cronbach's alpha. To evaluate students' performance in science, their final grades at the end of the school year were utilized. Additionally, in-depth interviews were conducted with 15 randomly selected respondents from the population at the conclusion of the school year.

### Ethical Considerations

Several ethical considerations were taken into account throughout the study. Prior to data gathering, the participants were informed of the purpose of the study and their right to decline to participate, and their rights to withdraw at any time from participations. In compliance with DepEd Order No. 40, series of 2012, parental consent was sought from the parents or guardians of participants who were under 18 years of age. Out of 122 participants, fifty-seven were below 18, and all complied with this requirement. Participants aged 18 and older were not required to obtain consent according to the aforementioned order. Subsequently, the science interest survey questionnaire as well as the demographic profile were administered. Qualitative data were collected though one-on-one interviews.

**Data Analysis**

The data collected were then subjected to appropriate statistical analyses using Statistical Package for the Social Sciences (SPSS) software. The descriptive statistics, such as the frequency, percentage, mean, and standard deviation, were used to describe the Science Interest Survey, the performance in science, and the demographic profile variables. Pearson r correlation coefficient was used to determine the relationship between the students' interest and the students' performance in science. Chi-square was used to determine if there were significant associations in the following pairs of variables: between the students' level of interest and the students' sex; between the students' level of interest and the availability of gadget; between the students' level of performance and the students' sex; and between the students' level of performance and the availability of gadget. On the other hand, Gamma test was used to determine if there were significant associations in the following pairs of variables: between the students' level of interest and the parents' highest educational attainment; between the students' level of interest and the family monthly income; between the students' level of performance and the parents' highest educational attainment; and between the students' level of performance and the family monthly income. Finally, the qualitative data were analyzed thematically. The researcher began by transcribing the interview data, which was then reviewed by a fellow researcher. The transcriptions were thoroughly read and re-read by the researcher to familiarize themselves with the content and context of the responses. Next, initial codes were generated by highlighting significant phrases or concepts that resonated with the research questions, ensuring a wide range of ideas was captured. After completing the coding process, the researcher grouped these codes into broader themes that reflected common patterns and insights across the interviews, subsequently reviewing and refining these themes to ensure they accurately represented the data. Each theme was clearly defined and named, ensuring that they were distinct and coherent. Finally, the researcher compiled the findings into a comprehensive report.

## RESULTS AND DISCUSSION

The findings of the study revealed that majority (50.82%) of the students had fairly satisfactory performance in science; majority (68.85%) of the students were moderately interested in science; majority of the respondents were female (66.39%), children of parents who finished college (32.79%) and high school (32.79%), belonged to a family with a monthly income of below 20,000 pesos (81.15%), and had gadget available (97.54%).

**Table 1. Association between Demographic Profile and Level of Interest**

| Association | Test Used | Value Obtained | Significance Level | Degrees of Freedom | Remarks |
|---|---|---|---|---|---|
| Sex and Level of Interest in Science | Chi-square Test | 7.149 | .05 | 2 | Significant |
| Parents' Highest Educational Attainment and Level of Interest in Science | Gamma Preliminary Reduction Test | .20 | .05 | | Non-significant |
| Family Monthly Income and Level of Interest in Science | Gamma Preliminary Reduction Test | .086 | .05 | | Non-significant |

| | | | | | |
|---|---|---|---|---|---|
| Availability of Gadget and Level of Interest in Science | Chi-square Test | 3.68 | .05 | 2 | Non-significant |

The association between **sex and level of interest** in science was determined using Chi-square test. The value obtained was 7.149 which is significant at .05 level of significance, df=2. This finding indicated that there were more girls who had higher level of interest in science than boys. This asserted that gender was associated to the level of interest of students in science. The finding was supported by Jia et al. (2020) and Kang et al. (2018) that there were clear gender differences with regards to interest in science in favor of the girls.

The association between the **parents' highest educational attainment and the level of interest** in science was determined using Gamma preliminary reduction test. The value obtained was .20 which was tested at .05 level of significance. The value signified a very weak or no significant association between the parents' highest educational attainment and the students' level of interest in science. Statistically, the value of .20 means that only 20% fewer errors predict students' interest in science using the parents' highest educational attainment. This indicated that parents' highest educational attainment per se does not contribute to their children's interest in science. This finding was in line with Halim et al. (2018) that the positive perceptions and values of parents towards the subject of science, not the educational attainment, propelled parents to cultivate their children's interest in science and any science related careers. However, the finding was contrary to Dabney et al. (2016) and Hacieminoglu (2015) that parents' educational attainment was positively associated to the students' attitude towards science.

The association between the **family monthly income and the level of interest** in science was determined using Gamma preliminary reduction error test. The value obtained was .086 which was tested at .05 level of significance. The value indicated a very weak or no significant association between the family monthly income and the students' level of interest in science. In statistical sense, the value of .086 means that only 8.6% fewer errors are made if the family monthly income is used to predict the students' level of interest in science. In other words, students' variation in terms of economic status did not determine their interest in science. Even children of low-income group can have high interest in science. This could be explained by the perception that parents vary in the extent of supporting their children regardless of their economic status. Even parents of the same economic status may vary in the extent of providing financial support to their children. The finding was affirmed by Conel (2021) that there was a negligible positive correlation between the students' family income and students' interest in science.

The association between the **availability of gadget and the level of interest** in science was determined using Chi-square test. The test revealed a value of 3.68 which was not significant at .05 level of confidence with df=2. This means that students' interest in science was not affected by any availability of gadget. One can be interested in science regardless of whether he/she owns a gadget or not. This finding was supported by Syaputri & Usman (2019) that the use of gadget has a very minimal effect (around 1.1%) on students' interest in learning. On the contrary, the finding was disputed by Djumingin et al. (2021) that a gadget has a positive and a significant effect on students' interest in learning.

**Table 2. Association between Demographic Profile and Level of Performance**

| Association | Test Used | Value Obtained | Significance Level | Degrees of Freedom | Remarks |
|---|---|---|---|---|---|
| Sex and Level of Performance in Science | Chi-square Test | 8.57 | .05 | 2 | Significant |

| | | | | | |
|---|---|---|---|---|---|
| Parents' Highest Educational Attainment and Level of Performance in Science | Gamma Preliminary Reduction Test | .119 | .05 | | Non-significant |
| Family Monthly Income and Level of Performance in Science | Gamma Preliminary Reduction Test | .09 | .05 | | Non-significant |
| Availability of Gadget and Level of Performance in Science | Chi-square Test | .40 | .05 | 2 | Non-significant |

The association between **sex and level of performance** in science was determined using Chi-square test. The value obtained was 8.57 which is significant at .05 level of significance, df=2. The finding indicated that female students had a higher performance than males. This finding also suggested that female students were more likely to have a higher academic performance in science than males. This finding was supported by El Refae et al. (2021) that there was a positive significant relationship between gender and student GPA both in face-to-face learning and in distance learning. The same finding was reported by OECD (2020) that girls outperformed boys in science in PISA 2018.

The association between the **parents' highest educational attainment and the level of performance** in science was determined using Gamma preliminary reduction test. The value obtained was .119 which was tested at .05 level of significance. The value signified a very weak or no significant association between the parents' highest educational attainment and the students' level of performance in science. Statistically, the value of .119 means that only 11.9% fewer errors are made if the parents' highest educational attainment is used to predict the students' level of performance in science. In other words, parents' highest educational attainment was not related with the students' performance in science. This was supported by Darko-Asumadu and Sika-Bright (2021) that the parents' education did not significantly affect the students' academic performance. In fact, a study conducted by Fasasi (2017) found that students from lowly educated parents had better performance than students from highly educated parents.

The association between the **family monthly income and the level of performance** in science was determined using Gamma preliminary reduction error test. The value obtained was .09 which was tested at .05 level of significance. The value indicated a very weak or no significant association between the family monthly income and the students' level of performance in science. In statistical sense, the value of .09 means that only 9% fewer errors are made if the family monthly income is used to predict the students' level of performance in science. In other words, the family income did not contribute to the academic performance of students in science. Indeed, there are students who belong to the low-income family who excel in science. The finding could be attributed to the perception that whilst majority of the parents belong to the low-income earners, they might have prioritized the basic needs of their family, such as food instead of spending for educational materials. This finding was supported by Gobena (2018) and Machebe et al. (2017) that family affluency does not affect students' academic performance.

The association between the **availability of gadget and the level of performance** in science was determined using Chi-square test. The test revealed a value of .40 which indicated no significant association. The finding indicated that gadget ownership is not the main factor that can affect students' performance in science. Indeed, there are students who use gadget frequently, yet, they excel in their classes. Similarly, some students who frequently use gadget have failing grades. Perhaps, it is on how a person uses gadget that can affect his/her performance, but not the mere ownership of a gadget. The finding was supported by Balbaguio (2020) that no significant impact of the electronic gadgets on the academic performance of students. In fact, Othman et al. (2020) found out that students who spent more time on electronic gadget has a high level of dependency towards gadget which led to poor academic achievement.

**Table 3. Correlation Between the Performance and Interest in Science**

| Association | Test Used | Value Obtained | Significance Level | Remarks |
|---|---|---|---|---|
| Performance in Science and Interest in Science | Pearson's Correlation Coefficient Test | .522 | .05, two-tailed test | Significant |

The study hypothesized that a student's interest in science contributes to his/her academic performance in science. To determine the validity of this hypothesis, Pearson's correlation coefficient test was utilized. The r value obtained was .522, which is a value that is significant ($p < .05$, two-tailed test). Hence, the null hypothesis that there is no significant relationship between the respondents' interest in science and their performance in science is hereby rejected. In other words, the students' interest in science is positively related to their academic performance in science. The higher the interest of a student in science is, the higher his/her academic performance would be.

This particular finding is consistent with the motivation theory. One's interest on something is an indication of his/her motivation. Indeed, according to the theory of motivation, one who is properly motivated to do something tends to perform better than those who are not motivated. In the same token, one who is not interested to learn the subject cannot be expected to have higher grades. The finding is in agreement with several studies abroad. Dahliani et al. (2020) found a significant correlation between students learning interest in learning Biology with students learning outcomes. Another recent study conducted by Mappadang et al. (2022) found that students' psychological conditions, as one of academic interest, contribute significantly to improving academic performance. This indicates that the higher the academic interest, the more the students will make an effort in their learning for better results. Moreover, a study conducted by Abaidoo (2018) found that one of the student factors that contribute to an improvement in academic performance is interest in a subject.

Meanwhile, the thematic analysis of the qualitative data revealed the following: (a) the less interested and low performer students were struggling in understanding the content of modules which often led them to just set aside their modules without reading them. They did not have timetable for studying and they were very occupied by leisure activities, such as playing online games and basketball. Hence, the average of their final grade was categorized as low; (b) the moderately interested and moderate performer students were reading their modules more than once depending on their mood. However, their attention on their module was easily diverted to social media and online games. Also, their timetables for studying were not followed. The average of their final grades was categorized as moderate. On the other hand, (c) the highly interested and high performer students read their modules many times until they understood them well. Whenever they encountered unfamiliar or difficult concepts to understand, they sought help from more knowledgeable people or made use of their internet resources. They strictly followed their timetables for studying and they even spent their leisure times in studying their modules. The average of their final grades was categorized as high. To wrap things up, the higher the interest in science, the higher the effort exerted by the student, which resulted to higher performance in the subject.

The study has few limitations including the small sample size of 122 students from a single school which limits the generalizability of the findings to broader populations, the lack of significant associations with parent's educational attainment and family income suggests that other unexamined variables may influence students' interest and performance in science, the focus on offline modular learning may not reflect the dynamics of traditional classroom settings or online education, potentially affecting student engagement differently, the qualitative interviews conducted after performance assessments may limit insights into how students' interest develop throughout the learning process.

Moreover, the variable "availability of gadgets" was operationalized in this study by asking participants about the gadgets they possessed to assess their impact on students' educational experiences. However, simply owning a gadget does not fully capture its effectiveness in enhancing learning outcomes. It is crucial to consider how these devices are used for educational purposes, such as accessing online resources and engaging in interactive learning, which this study did not address. Therefore, future research should focus on both the availability and practical use of gadgets to gain a better understanding of their role in student performance and interest in science education.

## CONCLUSION AND IMPLICATIONS

This study provides valuable insights into how students engage with science, revealing some noteworthy trends. Most strikingly, female students showed greater interest and performed better in science than their male peers. Interestingly, factors like parental education and family income didn't seem to have a significant impact on students' interest or performance. The presence of gadgets also didn't play a meaningful role. Qualitative feedback from students highlighted that those who were more interested in science tended to put in more effort, leading to better results.

These findings carry important implications for educators, parents, and policymakers. First and foremost, there's a clear need for initiatives aimed at boosting interest in science among male students, helping to close the gender gap in both engagement and achievement. Since parental education and income didn't significantly affect outcomes, focusing on developing strong study habits and support systems for students who struggle could be more beneficial.

Creating hands-on, interactive learning experiences can ignite students' passion for science, making it feel relevant and exciting. Additionally, teaching students how to use technology as a helpful learning tool—rather than a source of distraction—can enhance their academic experience. Ultimately, by fostering a supportive environment that caters to students' interests and challenges, we can help them thrive in science and beyond.

Moreover, future research should explore a broader range of variables influencing student performance and interest in science, including, but not limited to, teaching methods, classroom environment, peer interactions, and access to resources. This approach will help develop a more nuanced and comprehensive understanding of the multifaceted factors that contribute to academic success in science. Additionally, research should focus on both the availability and practical use of gadgets to gain deeper insights into their role in student performance and interest in science education.

## ACKNOWLEDGEMENTS

The authors express their heartfelt gratitude to the officials of the City Schools Division of Marawi, the principal of RPMD National Science High School, and all the participants in the study for their steadfast support and valuable contributions.

## REFERENCES

Abaidoo, A. (2018). Factors contributing to academic performance of students in a junior high school. Retrieved from https://www.grin.com/document/450284

Abu Bakar, N., Mamat, I., & Ibrahim, M. (2017). Influence of Parental Education on Academic Performance of Secondary School Students in Kuala Terengganu. International Journal of Academic Research in Business and Social Sciences. 7(8), 2222-6990, 2017.

Ali, S., Haider, Z., Munir, F., Khan, H. & Ahmed, A. (2013). Factors Contributing to the Students' Academic Performance. A case study of Islamia University, Sub-Campus. Science and Education Journal, Vol 1 issue 8, pp1-10.

Ampaso, S. S. (2019). Effects of Computer Assisted Instruction on Grade 9 Students' Conceptual Understanding, Performance, and Interest in Chemistry [Unpublished master's thesis]. Mindanao State University Marawi City.

Balbaguio, L. B. et al. (2020). Effects of Electronic Gadgets in the Academic Performance of Senior High School Students, from https://doi.org/10.2139/ssrn.3736606

Cambridge University Press. (n.d.). Interest. In Cambridge Dictionary. Retrieved [April 28, 2024], from https://dictionary.cambridge.org

Cañete, K. & Potane, J. (2022). Senior High School Students' Assessment of Academic Support on Modular Distance Learning During COVID-19. International Journal of Arts and Social Science (IJASS) ISSN, 2581-7922.

Collison, J. (2020, April 7). 10 Benefits and Uses of Electronic Gadgets in Learning. The Average Guy. https://theaverageguy.tv/2020/04/07/10-benefits-and-uses-of-electronic-gadgets-in-learning/

Conel, E. B. (2021). Factors affecting students' interest in learning science. IOER International Multidisciplinary Research Journal. P – ISSN 2651 - 7701 | E – ISSN 2651 – 771X. Vol. 3, No. 1, March 2021.

Cordero, T. (2018, March 15). PHL education system underperforms in East Asia Pacific —WB report. GMA News Online. Retrieved from: https://www.gmanetwork.com

Csikszentmihalyi, M. (1990). Flow: The Psychology of Optimal Experience. Research Gate. Retrieved from https://www.researchgate.net/publication/224927532

Dabney, K. P., Tai, R. H., & Scott, M. R. (2016). Informal Science: Family Education, Experiences, and Initial Interest in Science. International Journal of Science Education, Part B, 6(3), 263-282, from https://doi.org/10.1080/21548455.2015.1058990

Dahliani, E., Rahmatan, H., & Djufri. (2020). The correlation between students' interest and learning outcomes in biology. Journal of Physics: Conference Series, Volume 1460 2019. DOI: https://doi.org/10.1088/1742-6596/1460/1/012072

Darko-Asumadu, D. A., & Sika-Bright, S. (2021). Parental Involvement and Pupils' Academic Performance in the Cape Coast Metropolis, Ghana. Open Education Studies, 2021; 3: 96-109, from https://doi.org/10.1515/edu-2020-0142

Department of Education (DepEd). (2020). DepEd order no. 012 s. 2020: Guidelines on the implementation of modular distance learning for SY 2020-2021. https://www.deped.gov.ph/wp-content/uploads/2020/06/DO_s2020_012.pdf

Djumingin, S., et al. (2021). The Relationship Between the Use of Gadgets and Students' Interest in Learning Literature at Higher Education in Indonesia. Turkish Journal of Computer and Mathematics Education, 12(14), 5896-5912, 2021. Retrieved from https://turcomat.org/index.php/turkbilmat/article/view/11860

El Refae, G. A., Kaba, A., & Eletter, S. (2021). The Impact of Demographic Characteristics on Academic Performance: Face-to-Face Learning Versus Distance Learning Implemented to Prevent the Spread of COVID-19. Semantic Scholar Org., from, https://doi.org/10.19173/IRRODL.V22I1.5031

Idris, M., Hussain, S., & Ahmad, N. (2020). Relationship between Parents' Education and their children's Academic Achievement. Journal of Arts and Social Sciences, 7(2), 82-92, 2020, from https://doi.org/10.46662/jass-vol7-iss2-2020(82-92)

Farooq, M. S., Chaudhry, A. H., Shafiq, M., & Berhanu, G. (2011). Factors affecting students' quality of academic performance: a case of secondary school level. Journal of Quality and Technology Management, Volume VII, Issue II, December, 2011, Page 01 - 14

Fasasi, R. A. (2017). Effects of ethnoscience instruction, school location, and parental educational status on learners' attitude towards science. International Journal of Science Education, Vol 39, 2017, Issue 5, from https://doi.org/10.1080/09500693.2017.1296599

Fraser, B. J. (1981). Test of science-related attitudes (TOSRA). CYFAR. https://cyfar.org/sites/default/files/Attitude%20Towards%20Science%20(middle%20school%20and%20high%20school).pdf

Gobena, G. A. (2018). Family Socio-Economic Status Effect on Students' Academic Achievement at College of Education and Behavioral Sciences, Haramaya University, Eastern Ethiopia. Semantic Scholar. https://www.semanticscholar.org/paper/Family-Socio-Economic-Status-Effect-on-Students'-at-Gobena/b931bff360d3dc281e76ae5fea3eb8a9b1c83753
Hacieminoglu, E. (2015). Elementary School Students' Attitude toward Science and Related Variables. International Journal of Environmental & Science Education, 11(2), 35-52, from https://doi.org/10.12973/ijese.2016.288a
Halim, L., Abd Rahman, N., Zamri, R., & Mohtar, L. (2018). The roles of parents in cultivating children's interest towards science learning and careers. Kasetsart Journal of Social Sciences, Vol. 39, Issue 2, 2018, Pages 190-196. Retrieved from https://www.sciencedirect.com/science/article/pii/S2452315116301825
Hidi, S. (1990). Interest and its contribution as a mental resource for learning. Review of Educational Research, 60, 549–571.
Hulleman, C., and Harackiewicz, J. (2009). Promoting interest and performance in high school science classes. Science 326, 1410–1412, from https://doi.org/10.1126/science.1177067
Jia, Yang, Qian & Wu (2020). The Gender Differences in Science Achievement, Interest, Habit, and Creativity. Science Education International, Volume 31, Issue 2, 195-202, from https://doi.org/10.33828/sei.v31.i2.9
Joseph, A., John, O., Eric, I., Yusuf, S., & Olubunmi, A. (2015). Effect of Gender on Students' Academic Performance in Computer Studies in Secondary Schools in New Bussa, Borgu Local Government of Niger State. Journal of Education and Practice, 6(33), 2015, 2222-1735.
Kang, Hense, Scheersoi, & Keinonen (2018). Gender study on the relationships between science interest and future career perspectives. International Journal of Science Education, Volume 41, Issue 1, 2019.
Kohlhaas, K., Lin, H., & Chu, K. (2010). Disaggregated Outcomes of Gender, Ethnicity, and Poverty on Fifth Grade Science Performance. RMLE Online Research in Middle Level Education, 33(7), 1940-4476, 2010, from https://doi.org/10.1080/19404476.2010.11462070
Machebe, C. H., Ezegbe, B. N. & Onuoha, E. J. (2017). The Impact of Parental Level of Income on Students' Academic Performance in High School in Japan. Universal Journal of Educational Research, 5(9), 1614-1620, 2017, from https://doi.org/10.13189/ujer.2017.050919
Magsambol, B. (2020, December 9). PH lowest among 58 countries in math, science – global assessment. Rappler. Retrieved from https://www.rappler.com
Malipot, M. H. (2023, December 7). 2022 PISA results a 'clear indication' that PH education system is in 'worst state' --- PBEd. Manila Bulletin. Retrieved from https://mb.com.ph
Mappadang, A., Khusaini, K., Sinaga, M., & Elizabeth, E. (2022). Academic interest determines the academic performance of undergraduate accounting students: Multinomial logit evidence, Cogent Business & Management, 9:1, from https://doi.org 10.1080/23311975.2022.2101326
Narad, A. & Abdullah, B. (2016). Academic Performance of Senior Secondary School Students: Influence of Parental Encouragement and School Environment. Rupkatha Journal on Interdisciplinary Studies in Humanities, (ISSN 0975-2935), Special Issue, Vol. VIII, No. 2, 2016.
OECD (2020), "Girls' and boys' performance in PISA", in PISA 2018 Result (Volume II): Where All Students Can Succeed, OECD Publishing, Paris. https://doi.org/10.1787/f56f8c26-en
Othman, N., Khairuz, M., & Jamaludin, T. (2020). The Impact of Electronic Gadget Uses with Academic Performance among Secondary School Students. Research Gate. https://www.researchgate.net/publication/338791252_The_Impact_of_Electronic_Gadget_Uses_with_Academic_Performance_among_Secondary_School_Students
Paciorek, E. (1997). Mathematics and Science Attitude Survey Instrument. CYFAR. https://cyfar.org/sites/default/files/PsychometricsFiles/Interest%20in%20Science%20%28high%20school%20and%20college%20girls%29_0.pdf
Pagayocan, J. G. (2018). The Effects of Modelling Construction-Oriented Modular Instruction on Students' Conceptual Understanding, Performance and Interest in Biology. Master's Thesis. Mindanao State University Main Campus, Marawi City, Philippines.
Raya, R. R. (2021, July 18). PISA, TIMSS, and quality education. Business World. Retrieved from https://www.bworldonline.com

Roque, J. P. (2023). Modular Distance Learning in the Area of Education During the New Normal: A Systematic Review. AIDE Interdisciplinary Research Journal, 3(1), 253–268. https://doi.org/10.56648/aide-irj.v3i1.67

San Juan, R. (2019, December 4). DepEd welcomes PISA results, recognizes 'gaps' in education quality. The Philippine Star. Retrieved from https://www.philstar.com

Syaputri, S. & Usman, O. (2019). Effect of the Use of Gadget, Parenting Parents, Peer Environment, Teacher Teaching Skills of Learning Interest. Available at SSRN: https://ssrn.com/abstract=3510021 or http://dx.doi.org/10.2139/ssrn.3510021

Villar, M. C. G., Filipinas, J. P., Villanueva, J. B. and Cabello, C. A. (2022). The Transition, Transformation, and Adaptation from Modular-Printed Instruction to Limited Face-To-Face Classes: A Phenomenology. Psych and Education: A Multidisciplinary Journal, 4(9), 932-940, 2022. Retrieved from https://philpapers.org/rec/VILTTT

Walker, J.D., Cotner, S.H., Baepler, P.M., and Decker, M.D. (2008). A delicate balance: Integrating Active Learning into a large Lecture Course. CBE-Life Sciences Education from, https://doi.org/10.1187/cbe.08-02-0004

Western Governors University Blog. (2020, May 30). Five Educational Learning Theories. Western Governors University. https://www.wgu.edu/blog/five-educational-learning-theories2005.html#close

*International Journal of Science Education and Teaching*
*Vol. 3, No. 3, pp. 126 - 138, Sep. – Dec. 2024*
*http://doi.org/10.14456/ijset.2024.11*

# Comparison of ChatGPT and Gemini AI in Answering Higher-Order Thinking Skill Biology Questions: Accuracy and Evaluation

**Thoriqi Firdaus[1*], Siti Aminatus Sholeha[2], Miftahul Jannah[2], and Andre Ramdani Setiawan[2]**
[1]*Natural Science Education, Universitas Negeri Yogyakarya, Indonesia*
[2]*Natural Science Education, Universitas Trunojoyo Madura, Indonesia*
**Email: thoriqifirdaus.2023@student.uny.ac.id*

Received: 25 November 2024; Revised: 30 December 2024; Accepted: 30 December 2024

**Abstract.** AI is becoming increasingly prevalent and advancing over time, yet the accuracy of this intelligent technology remains a subject of scrutiny. This study aims to provide an in-depth evaluation of the capabilities of two platforms, ChatGPT and Gemini AI, by analyzing and comparing their performance, assessing answer accuracy, and offering comprehensive recommendations. A quantitative comparative approach was employed to evaluate the performance of ChatGPT and Gemini AI in answering Higher-Order Thinking Skills (HOTS) questions. The questions utilized were HOTS-based items on the subject of biology. The analysis shows that ChatGPT's accuracy rate (55%) is slightly higher than Gemini AI's (50%). However, Gemini AI's average score (0.5) is higher than ChatGPT's (0.4), meaning Gemini AI gives overall more accurate answers, even though its percentage of correct responses is lower. This difference is likely due to the types of questions and specific cognitive aspects involved. ChatGPT demonstrated strengths in questions requiring analysis and evaluation, while Gemini performed better in creation-based questions. Both systems faced challenges with questions that integrated complex cognitive processes and procedural knowledge, highlighting opportunities for further improvement in their respective knowledge-processing algorithms. The standard deviations for ChatGPT and Gemini are nearly identical, at 0.5026 and 0.5130, respectively, indicating a comparable level of consistency in the responses of both models. The mean standard error for ChatGPT (0.1124) is slightly lower than that of Gemini (0.1147), suggesting that ChatGPT's mean estimates are marginally more stable. This study highlights that ChatGPT and Gemini AI exhibit distinct strengths and weaknesses in answering Higher-Order Thinking Skills (HOTS) questions. ChatGPT excelled in cognitive dimensions involving analysis (C4) and factual knowledge, providing detailed and comprehensive answers. In contrast, Gemini AI demonstrated an advantage in the creation dimension (C6) and tasks requiring concise, straightforward responses, such as producing or planning solutions.

**Keywords:** comparison, ChatGPT, Gemini AI, HOTs, Biology

## INTRODUCTION

Technological advancements have transformed educational paradigms, driving a shift towards the digitalization of learning (Firdaus, 2023). Among these advancements, Artificial Intelligence (AI) has emerged as a key tool for streamlining contemporary learning processes (Nirwani & Priyanto, 2024). AI's ability to provide rapid responses and analyze large amounts of data has made it increasingly prevalent in education. However, concerns about the accuracy and reliability of AI-generated answers persist, especially in educational contexts where incorrect information could have significant implications (Johnson et al., 2023).

The integration of AI in education aligns with 21st-century learning goals, particularly in fostering Higher-Order Thinking Skills (HOTS). These skills, which include analysis, evaluation, and creation, are essential for student success in the Society 5.0 era, where cognitive abilities serve as a critical benchmark (Firdaus, 2022; Latifah & Maryani, 2021). HOTS goes beyond memorization and comprehension, requiring students to apply knowledge creatively and solve complex problems. As such, the effectiveness of AI platforms in supporting these cognitive processes has become a critical area of investigation.

Among the most widely used AI platforms in education are ChatGPT by OpenAI and Gemini AI by Google. Each platform has distinct strengths and applications. ChatGPT excels in speed, text processing, and efficiency, while Gemini AI is better suited for tasks that require deep analysis and comprehension (Rane et al., 2024). Research by Bahil et al. (2024) found that Gemini AI outperformed ChatGPT in terms of accuracy, achieving a rate of 66% compared to ChatGPT's 62%. However, the suitability of these platforms for addressing HOTS questions requires further investigation, particularly in understanding how their differences affect learning outcomes.

The integration of AI into educational contexts also presents challenges. A World Bank report (2024) identified ChatGPT as the most visited AI platform with 2.3 billion visits, far surpassing other platforms such as Gemini AI (123 million visits). Despite their widespread adoption, these platforms often produce answers that may seem accurate but can lead to misconceptions if not critically evaluated (Dalalah & Dalalah, 2023; Karatas et al., 2024). This highlights the need for teacher oversight and robust mechanisms to monitor the accuracy of AI-generated responses.

Thinking is a process that involves the brain's cognitive system and emotions in addressing and resolving problems (Firdaus et al., 2022). In Indonesia, the K13 curriculum prioritizes the development of HOTS as a key learning objective, emphasizing students' abilities to analyze, evaluate, and create (Asrafil et al., 2020). However, the increasing reliance on AI tools in education raises concerns about their potential to undermine critical thinking skills. Misuse of AI-generated content could lead to misconceptions and negatively impact students' understanding of complex concepts (Hidayatullah et al., 2024; Owan et al., 2023). Addressing these challenges requires a comprehensive evaluation of AI platforms to determine their efficacy in supporting HOTS development.

Research indicates that AI's adoption in education has grown significantly over the past decade, with studies on "AI" and "Education" accounting for 70% of related publications since 2010 (Chen et al., 2020). While many of these studies highlight the benefits of AI, its application in solving HOTS-related questions remains underexplored. The need for educators and developers to enhance AI tools to better support learning processes is therefore paramount (Jinhe et al., 2022).

Artificial Intelligence (AI) has become a transformative force across various fields, including education, leveraging advancements such as Natural Language Processing (NLP) to facilitate communication in human languages (Obaigbena et al., 2024). NLP enables seamless interactions between humans and machines, enhancing the accessibility and functionality of AI systems. These capabilities, combined with advances in computer vision, allow AI to recognize and interpret gestures, emotions, and facial expressions, making it a ubiquitous tool in daily life (Obaigbena et al., 2024). Despite these advantages, over-reliance on AI for scientific responses can undermine cognitive processes, especially when the accuracy of such responses is not thoroughly verified (Danry et al., 2023; Johnson et al., 2023).

In the educational sector, AI promotes active student engagement and creates an interactive learning environment. However, its use is not without challenges. Misinterpretation of AI-generated content due to a lack of critical evaluation can lead to misconceptions, negatively affecting students' critical thinking skills (Dilekli & Boyraz, 2024). This issue underscores the need for careful evaluation of AI-generated feedback to ensure its alignment with educational goals. Advanced data analytics, a feature of AI, has also sparked debates about its role in monitoring performance and generating personalized recommendations with consistency and precision (Heaven, 2020).

Two widely recognized AI platforms, ChatGPT by OpenAI and Gemini AI by Google, exemplify the diverse applications of AI in education. Research indicates that ChatGPT excels in tasks requiring contextual intelligence and reasoning, while Gemini AI is preferred for tasks necessitating extensive analysis and deep comprehension (Rane et al., 2024). Studies comparing their performance in various domains provide mixed results. For example, Carla et al. (2024) found that ChatGPT demonstrated superior analytical performance in assisting medical professionals, whereas Gemini AI faced significant limitations, particularly in complex tasks. These findings highlight the platforms' respective strengths and weaknesses, emphasizing the importance of selecting an appropriate tool based on the specific needs of the task.

The integration of AI in education aligns with the demands of 21st-century learning, particularly in fostering Higher-Order Thinking Skills (HOTS). Defined in the revised Bloom's Taxonomy, HOTS encompasses cognitive processes such as analysis, evaluation, and creation, which are critical for addressing complex problems (Jaenuddin et al., 2020; Latifah & Maryani, 2021). In Indonesia, the K13 curriculum emphasizes HOTS as a core objective to enhance students' abilities in solving, evaluating, and creating solutions (Asrafil et al., 2020). As AI becomes more integrated into classrooms, its potential to support HOTS development is increasingly evident. However, concerns remain about its accuracy and the potential for misuse, which could lead to misconceptions and undermine students' understanding (Hidayatullah et al., 2024; Owan et al., 2023).

AI systems' ability to simplify complex material has been widely recognized, with studies showing their positive impact on learning outcomes (Joseph et al., 2013). For instance, research by Wang et al. (2023) indicates that inaccuracies in AI-generated answers often stem from the specificity of the posed questions, leading to varying performance across platforms. ChatGPT demonstrates strengths in data processing and contextualizing responses but struggles with deep understanding in certain domains (Yasmar & Amalia, 2024). Conversely, Gemini AI, with its latest model improvements, has enhanced its reasoning capabilities (Muchlis & Maulida, 2024).

Given the increasing adoption of AI, a comprehensive evaluation of its efficacy in addressing HOTS questions is essential. This study aims to compare the performance of ChatGPT and Gemini AI in answering HOTS questions. By analyzing their respective strengths, weaknesses, and accuracy, this research seeks to provide insights into the suitability of these platforms for fostering higher-order cognitive processes in education.

## RESEARCH OBJECTIVES

This study aims to provide an in-depth evaluation of the capabilities of two prominent artificial intelligence platforms, ChatGPT and Gemini AI, in answering Higher-Order Thinking Skills (HOTS) questions in biology. The primary objective is to analyze and evaluate the performance and accuracy of ChatGPT and Gemini AI in addressing HOTS-based biology questions. This includes a comprehensive comparison of the platforms, focusing on parameters such as accuracy levels, relevance of responses, and alignment with scientific validity criteria and educational content. The analysis seeks to uncover the distinct characteristics of each platform, including ChatGPT's strengths in text processing efficiency and Gemini AI's advantages in deep reasoning and resource integration. Additionally, the objective encompasses identifying limitations, such as potential inaccuracies or misconceptions, that could affect the platforms' effectiveness in HOTS-based learning environments. By integrating performance analysis and accuracy evaluation into a single objective, the study aims to provide a cohesive understanding of these platforms' foundational capabilities.

The second objective is to deliver comprehensive recommendations for using AI platforms in HOTS-based learning. These recommendations will address the needs of students, educators, and AI developers, guiding them in selecting and refining platforms that are more effective, accurate, and aligned with 21st-century educational requirements. Moreover, the findings will offer valuable insights to AI platform providers, enabling them to implement continuous improvements and evaluations of their technologies. By doing so, this study seeks to contribute to the optimal use of AI in education, ensuring these tools effectively enhance HOTS-oriented learning and meet the evolving demands of modern education.

## METHODOLOGY

This study adopts a quantitative comparative approach to evaluate the performance of ChatGPT and Gemini AI in answering Higher-Order Thinking Skills (HOTS) questions. The quantitative comparative method is designed to numerically measure differences between two or more variables, enabling objective statistical analysis. Comparative quantitative research involves comparing two or more groups or variables to identify their differences or similarities. According to Sugiyono (2012), comparative research compares the presence of one or more variables within a single sample, in this case, using different AI platforms. The variables compared in this study are the performances of ChatGPT and Gemini AI in addressing HOTS questions.

The questions used in the study are HOTS-based biology questions adopted from Yuliani's (2017) research. These questions were selected due to their strong reliability, with an estimated coefficient of 0.93, which categorizes the instrument as highly reliable. This confirms that the measurements obtained using this instrument are dependable. The average item logit value of 0.0 indicates that the instrument can assess higher-order thinking abilities. As Bond and Fox (2013) stated, an average item logit of 0.0 represents a random value that reflects a 50:50 probability, indicating a balance between respondents' ability levels and the difficulty of the questions. If the average item logit does not reach 0.0, the instrument is generally considered less effective in accurately measuring the intended abilities.

Summary Of 197 Measured Person

| | Total Score | Count | Measure | Model Error | Infit MNSQ | Infit ZSTD | Outfit MNSQ | Outfit ZSTD |
|---|---|---|---|---|---|---|---|---|
| MEAN | 10.2 | 20.0 | .03 | .52 | .99 | .1 | 1.00 | .1 |
| S.D. | 3.9 | .3 | 1.01 | .10 | .11 | .7 | .27 | .8 |
| MAX. | 17.0 | 20.0 | 1.84 | 1.07 | 1.33 | 2.4 | 2.82 | 3.3 |
| MIN. | 1.0 | 18.0 | -3.22 | .46 | .55 | -1.4 | .14 | -1.0 |

REAL RMSE .54 TRUE SD .86 SEPARATION 1.59 Person RELIABILITY .72
MODEL RMSE .53 TRUE SD .86 SEPARATION 1.64 Person RELIABILITY .73
S.E. OF Person MEAN = .07
Person RAW SCORE-TO-MEASURE CORRELATION = .99
CRONBACH ALPHA (KR-20) Person RAW SCORE "TEST" RELIABILITY = .72

Summary Of 20 Measured Item

| | Total Score | Count | Measure | Model Error | Infit MNSQ | Infit ZSTD | Outfit MNSQ | Outfit ZSTD |
|---|---|---|---|---|---|---|---|---|
| MEAN | 100.6 | 196.5 | .00 | .16 | 1.01 | -.1 | 1.00 | .1 |
| S.D. | 21.0 | .8 | .64 | .02 | .08 | 1.1 | .13 | 1.1 |
| MAX. | 176.0 | 197.0 | .53 | .25 | 1.20 | 2.4 | 1.35 | 2.2 |
| MIN. | 79.0 | 194.0 | -2.50 | .16 | .90 | -2.1 | .85 | -1.6 |

REAL RMSE .17 TRUE SD .62 SEPARATION 3.73 Item RELIABILITY .93
MODEL RMSE .16 TRUE SD .62 SEPARATION 3.81 Item RELIABILITY .94
S.E. OF Item MEAN = .15

UMEAN=.0000 USCALE=1.0000
Item RAW SCORE-TO-MEASURE CORRELATION = -.99
3931 DATA POINTS. LOG-LIKELIHOOD CHI-SQUARE: 4572.60 with 3715 d.f. p=.0000
Global Root-Mean-Square Residual (excluding extreme scores): .4472
Capped Binomial Deviance = .2526 for 3931.0 dichotomous observations

**Figure 1**. Realibility HOTs Question
Source: Yuliani's (2017) research

The accuracy data analysis was conducted using a t-test statistic, employing the following calculation formula:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Meanwhile, the evaluation analysis was performed using descriptive analysis and presented in a data observation table. The results of ChatGPT and Gemini AI responses were compared based on the accuracy of their answers.

## RESULTS AND DISCUSSION

### Analyzing and comparing the capabilities of ChatGPT and Gemini AI

The research outcomes were analyzed to evaluate the performance of two systems, ChatGPT and Gemini AI, in answering questions designed based on cognitive dimensions, cognitive processes, and types of knowledge. This analysis focused on identifying the accuracy levels of both systems' responses compared to the correct answers and exploring performance patterns relative to the characteristics of the questions. The results provide insights into the strengths and weaknesses of each system in handling various levels of complexity and types of knowledge assessed.

**Table 1: Comparison of Responses: ChatGPT vs. Gemini**

| Code | Cognitive Dimension | Cognitive Process | Knowledge Dimension | ChatGPT Answer | Gemini Answer | Correct Answer |
|---|---|---|---|---|---|---|
| A1 | C4 (Analyzing) | Attributing | Factual | E | E | E |
| A2 | C4 (Analyzing) | Organizing | Procedural | D | D | D |
| A3 | C4 (Analyzing) | Organizing | Procedural | D | D | A |
| A4 | C5 (Evaluating) | Examining | Factual | A | C | C |
| A5 | C4 (Analyzing) | Distinguishing | Procedural | E | A | A |
| A6 | C4 (Analyzing) | Organizing | Conceptual | B | B | C |
| A7 | C5 (Evaluating) | Critiquing | Conceptual | B | B | B |
| A8 | C6 (Creating) | Producing | Metacognitive | E | C | A |
| A9 | C5 (Evaluating) | Critiquing | Factual | D | A | D |
| A10 | C6 (Creating) | Producing | Conceptual | C | C | C |
| A11 | C4 (Analyzing) | Distinguishing | Factual | A | E | D |
| A12 | C6 (Creating) | Formulating | Metacognitive | B | B | B |
| A13 | C6 (Creating) | Planning | Factual | A | A | E |
| A14 | C6 (Creating) | Planning | Conceptual | B | B | D |
| A15 | C5 (Evaluating) | Critiquing | Metacognitive | B | B | B |
| A16 | C6 (Creating) | Producing | Conceptual | B | E | E |
| A17 | C6 (Creating) | Planning | Metacognitive | B | B | B |
| A18 | C6 (Creating) | Formulating | Procedural | A | A | C |
| A19 | C5 (Evaluating) | Critiquing | Conceptual | B | B | A |
| A20 | C4 (Analyzing) | Attributing | Factual | C | C | E |

The analysis results indicate that ChatGPT demonstrates a slightly higher accuracy rate than Gemini, with 55% correct answers versus 50% for Gemini. Although the margin is small, ChatGPT has an advantage in understanding and responding to certain questions. Regarding cognitive dimensions, ChatGPT outperformed Gemini in tasks requiring analysis skills (C4), answering 4 out of 8 questions correctly, compared to Gemini's three correct answers. Both systems performed equally well for the evaluation dimension (C5), each providing four correct answers out of 6 questions. However, in the creation dimension (C6), Gemini showed a slight edge, answering 4 out of 6 questions correctly, while ChatGPT managed three correct answers.

Regarding cognitive processes, ChatGPT excelled in questions involving organizing processes, with two correct answers compared to Gemini's 1. However, Gemini demonstrated greater consistency in tasks requiring producing and planning processes, highlighting its ability to handle more complex question types. ChatGPT, on the other hand, struggled more with questions requiring distinguishing or examining, though its overall performance remained satisfactory.

From the perspective of knowledge dimensions, ChatGPT showed superior performance in questions based on factual knowledge. Gemini, meanwhile, delivered nearly comparable results to ChatGPT on questions involving conceptual and metacognitive knowledge. However, both systems faced significant challenges with procedural knowledge-based questions, with higher error rates observed in this dimension. Certain questions, such as A3, A6, A13, and A18, presented a high difficulty level for both systems. On these questions, the responses from both ChatGPT and Gemini diverged from the correct answers, indicating that the combination of complex cognitive processes and procedural knowledge posed significant challenges. Questions A3, A6, A13, and A18 incorporate visual elements such as images and diagrams, which significantly increase the complexity for AI systems to generate accurate responses. Specifically, image-based questions (A3, A6, and A18) and diagram-based questions (A13) require the AI to interpret visual data, a capability that remains a notable limitation for most natural language processing (NLP) models.

In addition to their visual components, these questions engage cognitive processes classified under Bloom's Taxonomy as C4 (analyzing) and C6 (creating). AI systems face challenges with C4 tasks because they require synthesizing information from multiple modalities, including textual explanations, visual data, and contextual knowledge. For example, A3 and A6 necessitate the integration of image interpretation with biological concepts, a task that exceeds the current capabilities of text-based AI models like ChatGPT and Gemini AI. Similarly, questions involving C6 processes, such as A13 and A18, pose difficulties because they demand the generation of novel and creative outputs. These questions often require the AI to not only interpret diagrams but also propose original solutions or construct new concepts based on limited or incomplete information. This highlights a critical gap in the ability of current AI systems to perform tasks that simulate higher-order cognitive skills, especially those requiring creativity and deep reasoning.

Understanding why these questions are challenging provides valuable insights for improving AI systems. Enhancements such as multimodal training, which integrates textual and visual data processing, or more robust algorithms for handling abstract and creative reasoning, could address these limitations. Future AI training models should focus on bridging these gaps to improve performance in tasks that combine visual interpretation and higher-order cognitive skills. Conversely, questions like A1, A2, A7, A10, A12, A15, and A17 demonstrated that both systems could consistently provide correct answers. These questions generally involved factual or conceptual knowledge paired with relatively straightforward cognitive processes.

**Evaluating the accuracy of responses**

A comparison between ChatGPT and Gemini AI in answering HOTS biology questions requires further analysis. This analysis aims to evaluate and compare the performance of the two models based on group statistical results. The data provided includes sample size (N), mean score (Mean), standard deviation (Std. Deviation), and standard error of the mean (Std. Error Mean), as presented in Table 2. Using this data, we can assess the consistency of each model's responses and determine whether there are significant differences between the two models.

**Table 2: Statistical Comparison Results**

| Group Statistics | N | Mean | Std. Deviation | Std. Error Mean |
|---|---|---|---|---|
| **ChatGPT** | 20 | 0.4 | 0.5026246899500346 | 0.11239029738980327 |
| **Gemini** | 20 | 0.5 | 0.512989176042577 | 0.11470786693528086 |

The statistical analysis indicates that ChatGPT and Gemini's performance in answering questions exhibits nearly equivalent characteristics. The comparison is conducted on a balanced dataset based on an identical sample size of 20 for each model. The average score (mean) for Gemini was slightly higher than that of ChatGPT, at 0.5 compared to 0.4. This suggests that Gemini, on average, provides somewhat more accurate responses than ChatGPT. However, the small mean difference of 0.1 necessitates further statistical testing to determine its significance. Regarding performance variation, the standard deviations for ChatGPT and Gemini were nearly identical, at 0.5026 and 0.5130, respectively. This reflects that both models exhibit similar levels of consistency in answering questions. The comparable variation indicates that ChatGPT and Gemini demonstrate similar fluctuations in performance on the tested data. Furthermore, the standard error of the mean for ChatGPT was slightly smaller than that for Gemini, at 0.1124 versus 0.1147. ChatGPT's mean score estimation is marginally more stable than

Gemini's. Although Gemini showed a higher average score, the levels of variation and consistency between the two models are nearly identical. The small difference in average scores suggests that the performance of the two models in answering questions is not significantly different. Advanced statistical analysis, such as a significance test, must confirm whether this difference is statistically meaningful or simply due to random variability in the data.

**t-Test Analysis**

A t-test is necessary to determine the statistical significance of the differences between the two groups, ChatGPT and Gemini, regarding their average scores. The table includes key metrics from the t-test analysis, such as:

- t-Statistic
- Degrees of Freedom (df)
- Two-tailed Significance (Sig. 2-tailed)
- Mean Difference
- Standard Error of Difference (Std. Error Difference)

By evaluating these values, the analysis will establish whether the observed mean difference between ChatGPT and Gemini is statistically significant or merely attributable to chance variations within the dataset.

**Table 3:** Uji T ChatGPT dan Gemini dalam Menjawab Soal

| **t-statistic** | **df** | **Sig. (2-tailed)** | **Mean Difference** | **Std. Error Difference** |
|---|---|---|---|---|
| -1.0000002 | 19 | 0.3298768009211 | -0.0999999999998 | 0.09999999999998 |

The t-test results reveal a t-statistic value of -1.0 with degrees of freedom (df) equal to 19. This indicates a minor difference between the mean scores of ChatGPT and Gemini. The significance value (Sig. 2-tailed) is 0.330, which exceeds the standard significance threshold of 0.05. This means the observed difference in mean scores between the two groups is not statistically significant. The mean difference between the two groups is -0.1, suggesting that ChatGPT's average score is slightly lower than Gemini's. However, with a standard error of difference of 0.1, this discrepancy is too small to be considered meaningful. The negative value in the mean difference simply indicates the direction of the difference ChatGPT scoring lower than Gemini but does not imply any substantive or significant disparity. This analysis concludes that although there is a slight difference in mean scores between ChatGPT and Gemini, this difference is not statistically significant. In other words, the performance of both groups can be considered equivalent in the context of this test. Further research with a larger sample size or an alternative test design may be necessary to explore performance differences between these models further.

**Evaluation of ChatGPT and Gemini AI Responses**

Based on the analysis of AI usage (ChatGPT and Gemini AI) in solving higher-order thinking Skill (HOTS) questions, here is a detailed narrative based on cognitive dimensions, answer accuracy, explanation quality, consistency, as well as the strengths and weaknesses of each AI. The analyzed questions involve cognitive dimensions such as analyzing (C4), evaluating (C5), and creating (C6) and include knowledge dimensions spanning factual, conceptual, procedural, and metacognitive domains. Both AIs could provide responses aligned with the required cognitive levels. However, variations were observed in the depth of explanation, particularly for questions requiring extensive exploration. Regarding answer accuracy, both AIs generally provided answers that matched the correct key, although differences in delivery were noted. For example, in Question 1, ChatGPT and Gemini AI accurately explained moss's role in the ecosystem.

In Question 2, both provided correct answers regarding the functions of antipodal and synergid cells, but ChatGPT's explanation was more detailed than Gemini AI's.

Regarding explanation quality, ChatGPT delivered structured and in-depth explanations. For example, besides answering the question, ChatGPT elaborated on related processes or mechanisms, such as the ecosystem impact of moss on the carbon cycle. In contrast, Gemini AI provided concise and direct answers, focusing on the main points. While these answers were relevant to the questions, Gemini AI often lacked the deep elaboration necessary for a more comprehensive understanding. Regarding response consistency, ChatGPT provided more consistent, detailed answers to high-difficulty questions. At the same time, Gemini AI tended to perform optimally on questions based on factual or simple conceptual knowledge. For metacognitive questions, Gemini AI showed weaknesses in delivering exploratory responses.

As for strengths and weaknesses, ChatGPT's strength lies in its detailed and comprehensive explanations, making it suitable for fostering deep understanding. However, its lengthy responses can make it difficult for users to pinpoint the main answers quickly. Gemini AI's strength is providing concise and focused answers, ideal for tasks requiring straightforward responses. However, it lacks the depth for questions requiring extensive exploration and elaboration.

ChatGPT and Gemini AI were assessed in solving higher-order thinking Skill (HOTS) questions across various categories, focusing on accuracy, explanation quality, and analytical capabilities.

- **Question Code A1:** Analyzing the role of moss in life. ChatGPT exhibited strong analytical skills, detailing the benefits of moss, such as preventing erosion, retaining water, and aiding soil formation. It also provided relevant ecological context. In contrast, Gemini AI only gave a core response, stating the importance of moss in ecosystems without elaboration.
- **Question Code A2:** Functions of antipodal cells and synergids. ChatGPT provided an in-depth explanation, describing their location and role in double fertilization. It mentioned antipodal cells as nutrient providers and synergids as chemical signalers guiding pollen tubes to the ovum. Gemini AI’s response was correct but concise, lacking a detailed mechanism.
- **Question Code A3:** Identifying ovule location in a Gymnosperm reproduction diagram. ChatGPT identified the correct location and explained the ovule as the site of sperm and ovum fusion, enhancing biological understanding. Gemini AI gave the correct location but without context or explanation.
- **Question Code A4:** Evaluating errors in moss metagenesis. ChatGPT elaborated on the stages, explaining that the zygote develops into sporogonium before producing spores and detailing the functions of archegonia and antheridia. Gemini AI answered correctly but without a comprehensive explanation.
- **Question Code A5:** Optimal temperature for moss growth based on a graph. ChatGPT provided a comprehensive response, explaining how 22–28°C supports moss metabolism and how extreme temperatures reduce physiological activity. Gemini AI only stated the optimal temperature without linking it to biological mechanisms.
- **Question Code A6:** Identifying errors in statements about Gymnosperms and Angiosperms. ChatGPT analyzed differences in seed protection, reproductive tools, and seed structure, aiding conceptual understanding. Gemini AI only mentioned core differences without further discussion.
- **Question Code A7:** Determining corn plant characteristics based on an image. ChatGPT detailed monocot features, such as fibrous roots, parallel leaves, and separate flowers, providing rich context. Gemini AI answered correctly but omitted specifics.

- **Question Code A8:** Identifying a fruit not classified as a Dicot. ChatGPT correctly identified coconut as a Monocot and melinjo as a Gymnosperm, explaining scientific classifications. Gemini AI provided the correct answer but lacked elaboration.
- **Question Code A9:** Identifying a fruit not classified as a Dicot. ChatGPT accurately described the coconut as a Monocot and melinjo as a Gymnosperm, including Monocot characteristics and Gymnosperm seed properties. Gemini AI's answer was correct but lacked depth.
- **Question Code A10:** Designing an experiment to distinguish Monocots from Dicots using fruits. ChatGPT outlined a structured procedure, including seed observation, to identify cotyledons. The explanation connected methods to outcomes. Gemini AI's response was correct but lacked logical reasoning and procedural details.
- **Question Code A11:** Determining if a flower is complete and perfect. ChatGPT explained the criteria for completeness (presence of peduncle, receptacle, petals, stamens, and pistils) and perfection (both reproductive organs). Gemini AI stated the result without elaborating on the supporting parts.
- **Question Code A12:** Identifying ferns based on characteristics like segmented stems and small spiral leaves. ChatGPT identified the plant as *Equisetum sp.*, adding habitat details, such as moist mountainous areas. Gemini AI provided the species name without additional context.
- **Question Code A13:** Explaining differences between tropophyll (sterile) and sporophyll (fertile) leaves. ChatGPT linked their functions to photosynthesis and spore production, integrating their roles in the fern life cycle. Gemini AI provided the correct answer but lacked depth.
- **Question Code A14:** Classifying and identifying benefits of Pteridophytes. ChatGPT included classifications such as *Adiantum* (ornamental), *Lycopodium* (herbal medicine), and *Azolla* (green fertilizer), connecting benefits to plant classes. Gemini AI listed classifications without discussing uses.
- **Question Code A15:** Explaining haploid and diploid stages in Pteridophyte metagenesis. ChatGPT explained that the gametophyte arises from spores via meiosis (haploid), while the sporophyte arises from a zygote (diploid), linking meiosis to the plant life cycle. Gemini AI's response, though accurate, lacked detailed connections.
- **Question Code A16:** Outlining a practical experiment to observe *Azolla pinnata* and *Sphagnum sp.* using eosin solution. ChatGPT gave a detailed procedure, including preparation, observation duration, and analysis. Gemini AI provided a brief response without technical specifics.
- **Question Code A17:** Identifying tools for anatomical observation of *Sphagnum sp.* ChatGPT listed tools like microscopes, slides, pipettes, tweezers, and cutters, explaining each tool's purpose. Gemini AI mentioned only basic tools, omitting functional details.
- **Question Code A18:** Formulating a hypothesis for eosin absorption experiments. ChatGPT proposed a theory based on physiological differences between *Azolla* and *Sphagnum*, such as vascular tissue efficiency, supported by biological reasoning. Gemini AI offered a simpler hypothesis without elaboration.
- **Question Code A19:** Evaluating incorrect statements about moss and fern morphology. ChatGPT explained the gametophyte dominance in mosses and sporophyte dominance in ferns, highlighting additional distinctions like true roots, stems, and leaves. Gemini AI only identified generational dominance differences.
- **Question Code A20:** Determining germination types in Monocots and Dicots. ChatGPT provided examples of epigeal germination (cotyledons above ground) and hypogeal germination (cotyledons underground). Gemini AI answered correctly but omitted comparative details.

## Policy Recommendations

In advancing higher-order thinking skills (HOTS) based learning, the indispensable role of teachers remains at the forefront, even amidst the growing integration of AI tools such as ChatGPT and Gemini AI. While these tools hold considerable potential to enrich the educational process, their efficacy hinges on deliberate and well-structured implementation strategies. Teachers must be active facilitators and validators with actionable methodologies to incorporate AI into lesson planning, assessment, and classroom management.

Teachers can use ChatGPT to craft analytical scenarios (C4) or in-depth evaluative materials during lesson planning. At the same time, Gemini AI is well-suited for designing creative tasks that involve planning and production (C6). In assessment, teachers play a pivotal role in validating AI-generated responses to ensure accuracy and alignment with learning objectives. For instance, ChatGPT's comprehensive and detailed answers can serve as a foundation for facilitating in-depth classroom discussions encouraging critical and analytical thinking. Conversely, Gemini AI's concise responses can be utilized as comparative tools for exploring alternative solutions or initiating talks.

AI can also enhance classroom management when strategically integrated into learning activities. ChatGPT excels in providing elaborate explanations to help students grasp complex concepts, whereas Gemini AI is ideal for quick, formative assessments like quizzes. However, direct teacher intervention is crucial for tasks that necessitate intricate cognitive processes or procedural knowledge. Teachers can guide students through complex problem-solving steps, ensuring they understand the logical connections between each phase and the expected outcomes.

To maximize the utility of AI, teachers require a structured framework for validating AI-generated responses. This framework should include aligning AI outputs with curricular standards, posing follow-up questions to deepen students' comprehension, and engaging students in critically evaluating AI-provided answers as an exercise in analytical reasoning. Such measures enhance learning accuracy and reinforce students' critical, creative, and problem-solving skills.

Consequently, policies governing the use of AI in education must unequivocally establish teachers as the primary arbiters of these tools. Teachers must ensure that AI serves as a complement to, rather than a replacement for, their pedagogical expertise. By maintaining this balance, AI technologies such as ChatGPT and Gemini AI can become powerful instruments in creating an effective, meaningful, student-centered learning environment that fosters enduring understanding.

# CONCLUSION

This study demonstrates that ChatGPT and Gemini AI exhibit distinct strengths and weaknesses in addressing questions about higher-order thinking skills (HOTS). ChatGPT excels in cognitive dimensions involving analysis (C4) and factual knowledge, offering detailed and comprehensive responses. In contrast, Gemini AI performs better in creation (C6) tasks and processes requiring concise, direct answers, such as production or planning. However, both AIs encounter challenges with procedural knowledge-based questions and complex cognitive processes. The average accuracy difference between the two is not statistically significant, indicating that their overall performance is relatively comparable, with each AI excelling in different areas. The consistency and quality of their responses vary depending on the complexity of the questions and the type of knowledge being assessed. ChatGPT and Gemini AI can significantly contribute to learning, particularly in HOTS-based questions. Nevertheless, it is essential to emphasize that the successful implementation of AI in education depends heavily on how teachers utilize these technologies to enhance learning rather than as substitutes for their roles in guiding and supporting students.

## IMPLICATIONS

ChatGPT, with its ability to provide detailed and comprehensive responses, is well-suited for supporting analytical and evaluative tasks that require in-depth exploration. Its more elaborate answers can serve as a foundation for class discussions, encouraging students to think critically and develop a deeper understanding of concepts. On the other hand, Gemini AI, with its concise and direct answers, is ideal for quick learning activities such as quizzes or tasks requiring time efficiency. These differences highlight the potential for both AIs to be used complementarily to address diverse learning needs.

However, the study also identifies limitations in both AIs, particularly with questions involving procedural knowledge and more complex cognitive processes. These challenges underscore the critical role of teachers in validating AI-generated answers and providing direct guidance to students. Teachers are responsible for ensuring that students receive correct answers and grasp the underlying reasoning. In this context, AI responses can serve as starting points for discussion or as tools to explain concepts, but teachers remain the primary agents in fostering students' understanding.

Furthermore, integrating AI into learning must be strategically designed to maximize its effectiveness. ChatGPT can support tasks requiring in-depth elaboration, while Gemini AI can be utilized in simpler, more straightforward contexts. The use of AI should complement, not replace, the essential human interactions that are central to the educational process. With the right approach, these technologies can enhance learning quality, help students develop higher-order thinking skills, and support teachers in creating richer and more diverse educational experiences.

## ACKNOWLEDGEMENTS

We extend our gratitude to KOPI ALINEA (Komunitas Peneliti Akademi Literasi Sains dan Budaya) for facilitating this research process. The support provided, including access to discussions, references, and feedback, has been instrumental in developing ideas and completing this study. The presence of KOPI ALINEA as a space for collaboration and learning has been a vital part of this research journey. We hope this community continues to grow and provide valuable benefits to other researchers in the future.

## REFERENCES

Asrafil., Retnawati, h., & Retnowati, E. (2020). The Difficult of Students when Solving HOTS Problem and the Description of Students Cognitive Load After Given Worked Example as a Feedback. *International Conference on Science Education and Technology*. 1511(1). 1-11.

Bond, T. G., & Fox, C. M. (2013). *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press.

Carla, M. M., et al. (2024). Large Language Models as Assistance for Glaucoma Surgical Cases: a ChatGpt VS Google Gemini Comparison. *Graefe's Archive for Clinical and Experimental Ophthalmology.* 2945-2959.

Chen, L., Chen, P., & Lin, Z. (2020). Artificial Intelegence in Education: A Rivew. *Ieee Access.* (8). 75264-75278.

Dalalah, D., & Dalalah, O. M. (2023). The false positives and false negatives of generative AI detection tools in education and academic research: The case of ChatGPT. *The International Journal of Management Education*, *21*(2), 100822.

Danry, V., et al. (2023). Don't Just Tell Me, Ask Me: AI Systems that Intelligently Frame Explanations as Qestions Improve Human Logical Discernment Accuracy Over Causal AI Explanations. *Proceedings of the 2023 CHI Converence on Human Factors in Computing Systems.* 23(352). 1-13.

Dilekli, Y., & Boyraz, S. (2024). From “Can AI think” to “Can AI help thinking deeper?”: Is use of Chat GPT in higher education a tool of transformation or fraud?. *International Journal of Modern Education Studies.* 8(1). 49-71.

Firdaus, T. (2022). PENERAPAN MODEL DIRECT INSTRUCTION BERBASIS SETS PADA PEMBELAJARAN IPA UNTUK MENINGKATKAN KETERAMPILAN BERPIKIR KRITIS SISWA. *Natural Science Education Research (NSER)*, *5*(1), 119-134. https://doi.org/10.21107/nser.v5i1.15759

Firdaus, T., Ahied, M., Qomaria, N., Putera, D. B. R. A., & Sutarja, M. C. (2022). Jurnal Pembelajaran Sains. *Jurnal Pembelajaran Sains*, *6*(1).

Firdaus, T. (2023). Representative platform cyber metaverse terkoneksi BYOD sebagai upaya preventive urgensi digital pada sistem pendidikan Indonesia. *Jurnal Integrasi dan Harmoni Inovatif Ilmu-Ilmu Sosial*, *3*(2), 123-131. https://doi.org/10.17977/um063v3i2p123-131

Joseph, G. V., et al. (2024). Impact of Digital Literacy, Use of AI tools and Peer Collaboration on AI Assisted Learning: Perceptions of the University Students. *Digital Education Review.* 45. 43-49.

Karataş, F., Abedi, F. Y., Ozek Gunyel, F., Karadeniz, D., & Kuzgun, Y. (2024). Incorporating AI in foreign language education: An investigation into ChatGPT’s effect on foreign language learners. *Education and Information Technologies*, 1-24.

Latifah, A., & Maryani, I. (2021). Developing HOTS Questions for the Materials of Human and Animals Respiratory Organs for Grade V of Elementary School. *Jurnal Prima Edukasia.* 9(2). 179-192.

Muchlis & Maulida, R. (2024). Komparasi Respons ChatGPT dan Gemini terhadap Command Pattern Identik dengan Metode Black Box. *Jurnal Teknik Informatika STMIK Antar Bangsa, 10*(2). 68-71. *https://ejournal.antarbangsa.ac.id*

Nirwani, N., Priyanto. (2024). Integrasi Artifical Intelegence dalam Pembelajaran Bahasa di SMP. *Jurnal Pendidikan Bahasa dan Sastra.* 7(1). 31-38.

Obaigbena, A., et al. (2024). AI and Human-Robot Interaction: A Riview of Recent Advances and Challenges. *GSC Advanced Research and Reviews.* 18(02). 321-330.

Owan, V. J., et al. (2023). Exploring the Potential of Artificial Intellegence Tools in Educational Measurement and Assessment. *Journal of Mathematics, Science and Technology Education.* 19(8). 1-15.

Rane, N. L., Choudhary, S. P., & Rane, J. (2024). Gemini or ChatGPT? Efficiency, Performance, and Adaptability of Cutting-Edge Generative Artificial Intelligence (AI) in Finance and Accounting. 1-12.

Rane, N. R., Choudhary, S.P., Rane, J. (2024). Gemini versus ChatGPT: applications, performance, architecture, capabilities, and implementation. *Journal of Applied Artificial Intelligence, 5*(1). 69-93. *https://doi.org/10.48185/jaai.v5i1.1052*

Sugiyono. (2012). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung: Alfabeta.

Wang, Y., Shen, S., & Lim, B. (2023). RePrompt: Automatic Prompt Editing to Refine AI Generative Art Towards Precise Expressions. ArXiv, abs/2302.09466. *https://doi.org/10.1145/3544548.3581402*

World Bank. (2024). *Who on Earth Is Using Generative AI?*

Yasmar, R., & Amalia, D. R. (Analisis SWOT Penggunaan Chat GPT dalam Dunia Pendidikan Islam. *Fitrah Jurnal Studi Pendidikan*, 15(1). https://doi.org/10.47625/fitrah.v15i1.668

Yuliani, E. (2017). Pengembangan Manual Test berbasis Higher Order Thinking Skill (HOTs) Serta Implementasinya di SMA Unggul Negeri 8 Palembang. (Disestation, Universitas Islam Negeri (UIN) Raden Fatah Palembang).

SEAT

# IJSET

Published December 31, 2024 by
Science Education Association (Thailand)
Sukhumvit 23, Bangkok, 10110, THAILAND
Tel: 66-2204-2528 Fax 66-2204-2528