Educazione: Journal of Education and Learning
Vol 03, No 02, 2026 | Page (312-322)
doi: https://doi.org/10.61987/educazione.v3i2.2430

OPEN ACCESS

The Effectiveness of Mathematics Learning Media Developed with Artificial Intelligence (AI) Assistance in Training Basic Arithmetic Skills

Husnu Athoillah, Risma Firda Diana*

Institut Al Fithrah Surabaya, Indonesia

Email Corresponding : rismafirdadiana@alfithrah.ac.id

Abstract

Limited integration of artificial intelligence into elementary mathematics instruction continues to constrain opportunities for students to develop basic arithmetic skills through engaging learning experiences. Addressing this challenge, the present study examined the effectiveness of AI-assisted mathematics learning media developed through the pedagogical integration of ChatGPT and Canva AI for second-grade students. A quantitative pre-experimental approach employing a one-group pretest-posttest design was implemented with 25 participants selected through purposive sampling. Data were collected using expert validation sheets and students' arithmetic achievement tests. The learning media and assessment instruments were first evaluated for validity, after which students completed pretest and posttest assessments. Data analysis was performed using descriptive statistics, normality testing, and the Wilcoxon Signed-Rank Test. Expert evaluation classified both the learning media and the assessment instrument as highly valid, confirming their suitability for classroom implementation. Students' arithmetic achievement improved significantly following the intervention, demonstrating that the developed learning media effectively strengthened basic arithmetic skills. These findings indicate that the educational value of generative AI extends beyond content generation by supporting the design of interactive, visually engaging, and learner-centered mathematics instruction. Integrating ChatGPT and Canva AI therefore represents a promising instructional approach for enhancing foundational numeracy learning in elementary education.

Article History

Received : 03 March 2026
Revised : 19 May 2026
Accepted : 30 June 2026

Keywords:

Artificial Intelligence, Arithmetic Skills, Elementary Mathematics, Artificial Intelligence, Mathematics Learning Media

How to Cite:

Athoillah, H., & Diana, R. F. (2026). The Effectiveness of Mathematics Learning Media Developed with Artificial Intelligence (AI) Assistance in Training Basic Arithmetic Skills. *Educazione: Journal of Education and Learning, 3*(2), 312–322.

INTRODUCTION

The rapid advancement of digital technology has fundamentally transformed contemporary education, requiring schools to adopt instructional approaches that are more adaptive, interactive, and responsive to students' learning needs. Among recent technological innovations, artificial intelligence (AI) has emerged as one of the most promising tools for enriching teaching and learning processes by supporting personalized instruction, generating educational content, and facilitating more engaging learning experiences (Bhutoria, 2022; Ingkavara et al., 2022; Maier & Klotz, 2022). The emergence of generative AI has further expanded these possibilities by enabling teachers to develop instructional materials more efficiently while enhancing the quality of learning resources through intelligent content generation and visual design (Choi et al., 2024; Luo et al., 2025; Squalli Houssaini et al., 2024). In mathematics education, where conceptual understanding and continuous practice are essential for developing numeracy skills, AI-assisted learning media offer new opportunities to present mathematical concepts in ways that are more meaningful and accessible for young learners (Deng et al., 2020; Sein Minn, 2022; Zhang & Zhang, 2024). Consequently, the integration of AI into mathematics instruction should not be viewed merely as a technological innovation but as a pedagogical opportunity to redesign learning environments that promote active participation, strengthen conceptual understanding, and improve students' overall learning experiences in the digital era.

Despite the growing recognition of AI in education, its implementation in elementary mathematics classrooms remains limited, particularly in supporting the development of basic arithmetic skills (Relmasira et al., 2023). In many schools, mathematics instruction continues to rely on conventional teacher-centered approaches characterized by repetitive exercises and textbook-based activities, providing few opportunities for students to interact actively with mathematical concepts (Rich, 2021). This situation is especially concerning because arithmetic competence forms the foundation for subsequent mathematical learning and influences students' ability to solve increasingly complex problems. Preliminary observations and interviews conducted at *MI Al Fithrah Surabaya* revealed that technology had not yet been optimally integrated into classroom instruction, despite the availability of digital resources. Mathematics lessons were predominantly delivered through explanation and paper-based exercises, resulting in relatively low student engagement during learning activities (Wang, 2020). Teachers also reported that many second-grade students experienced difficulties in mastering basic addition and subtraction, particularly when solving contextual arithmetic problems. These conditions indicate a clear need for instructional innovation that not only incorporates emerging technologies but also supports meaningful mathematical learning through engaging and developmentally appropriate learning experiences. Such a need highlights the importance of designing AI-assisted learning media that can bridge the gap between technological advancement and effective classroom practice.

Previous studies have consistently highlighted the potential of artificial intelligence to enhance mathematics learning through adaptive instruction, personalized feedback, and increased student engagement. For example, Schorcht et al. (2024) reported that AI-supported mathematics instruction improved elementary students' learning outcomes by facilitating more interactive learning activities. Similarly, Relmasira et al. (2023) emphasized that generative AI can enrich mathematics education by providing personalized learning experiences and immediate feedback that promote deeper conceptual understanding. Go et al. (2024) also found that AI-assisted instructional technology positively influenced mathematics achievement by creating more engaging classroom environments, while Avdiu (2019) concluded that AI has considerable potential to improve educational quality in elementary schools. Although these studies provide valuable evidence regarding the educational benefits of AI, most focus primarily on conceptual discussions, systematic literature reviews, or the general implementation of AI in education. Empirical evidence demonstrating how generative AI can be translated into practical instructional media for teaching basic arithmetic in authentic elementary classroom settings remains limited (Luo et al., 2025). Studies exploring the collaborative use of multiple generative AI tools to develop pedagogically integrated learning media, particularly within the context of Islamic elementary education. These limitations indicate an important research gap that warrants further investigation.

Addressing these gaps, the present study proposes a different approach by integrating ChatGPT and Canva AI to develop AI-assisted mathematics learning media that combines intelligent content generation with interactive visual design in a game-based learning environment. Unlike previous studies that primarily examined AI as a general instructional technology or focused on its conceptual potential, this study emphasizes the pedagogical integration of two complementary generative AI applications to produce learning media specifically designed to strengthen basic arithmetic skills among second-grade elementary students. The developed media, *Petualangan Matematika*, transforms conventional arithmetic practice into an interactive digital learning experience that encourages students to solve mathematical problems through active participation rather than passive repetition. By embedding AI within an instructional design that aligns with the cognitive characteristics of young learners, this study seeks to demonstrate that the effectiveness of AI depends not merely on technological sophistication but on how technology is meaningfully integrated into classroom pedagogy. This research extends the current literature by providing empirical evidence on the practical application of generative AI in elementary mathematics education within the context of an Islamic elementary school, where studies on AI-assisted instructional media remain relatively limited.

Based on the identified research gap, this study aims to examine the effectiveness of AI-assisted mathematics learning media in improving elementary students' basic arithmetic skills. Specifically, the study evaluates learning media developed through the pedagogical integration of ChatGPT and Canva AI and examines its impact on second-grade students' arithmetic achievement at MI Al Fithrah Surabaya. It is hypothesized that the implementation of the AI-assisted learning media will produce a statistically significant improvement in students' basic arithmetic skills following the instructional intervention. By providing empirical evidence from an authentic classroom setting, this study seeks to strengthen current understanding of how generative AI can be effectively integrated into elementary mathematics instruction through well-designed learning media. The findings are expected to provide evidence-based insights for educators and researchers regarding the implementation of AI-supported instructional innovation in mathematics education, particularly for strengthening foundational numeracy skills among young learners.

RESEARCH METHOD

This study employed a quantitative approach using a pre-experimental method with a one-group pretest-posttest design to examine the effectiveness of artificial intelligence (AI)-assisted mathematics learning media in improving students' basic arithmetic skills (Carter et al., 2021; Chang et al., 2022; Little et al., 2020). This design was selected because it enables the measurement of changes in students' learning achievement before and after the implementation of an instructional intervention within the same group of participants (Little et al., 2020). The study was conducted at MI Al Fithrah Surabaya and involved 25 second-grade students selected through purposive sampling based on their participation in mathematics learning during the research period. The instructional intervention consisted of AI-assisted mathematics learning media collaboratively developed using ChatGPT to generate instructional content and Canva AI to design interactive visual learning materials. Prior to the intervention, students completed a pretest to determine their initial arithmetic competence. Subsequently, they participated in mathematics learning activities using the AI-assisted learning media entitled *Petualangan Matematika*. At the end of the learning sessions, students completed a posttest that measured the same learning indicators as those assessed during the pretest. The comparison between students' pretest and posttest scores served as the primary basis for evaluating the effectiveness of the developed learning media.

Two complementary instruments were employed to collect the research data. First, expert validation sheets were used to evaluate the feasibility of the AI-assisted mathematics learning media before classroom implementation. The validation process involved two experts: a lecturer from the Primary School Teacher Education (*PGMI* (*Pendidikan Guru Madrasah Ibtidaiyah*, Islamic primary school teacher education)) Program at *Institut Al Fithrah Surabaya* and an experienced mathematics teacher from *MI Al Fithrah Surabaya*. The experts assessed the media based on content accuracy, instructional appropriateness, language clarity, visual presentation, and media functionality. Second, students' learning achievement was measured using an arithmetic achievement test administered as both a pretest and a posttest. The test consisted of questions representing the learning indicators of basic addition and subtraction competencies for second-grade elementary students. The pretest measured students' initial arithmetic ability before the intervention, whereas the posttest assessed their achievement after participating in learning activities using the AI-assisted mathematics learning media.

The collected data were analyzed using IBM SPSS Statistics version 27 through several sequential procedures. First, validity data obtained from the expert evaluation sheets were analyzed descriptively by calculating the percentage of the total score relative to the maximum possible score using the following formula:

$$\text{Validity Percentage} = \frac{\text{Obtained Score}}{\text{Maximum Score}} \times 100\%$$

The resulting percentage was subsequently interpreted according to the validity criteria presented in **Table 1**.

Table 1. Interpretation Criteria for Media and Instrument Validity

Percentage	Validity Criteria
21%–40%	Invalid
41%–60%	Less Valid
61%–80%	Valid
81%–100%	Highly Valid

Source: Adapted from Maulana (2026).

Following the validity analysis, descriptive statistics were generated to summarize students' pretest and posttest performance. Before conducting hypothesis testing, the distribution of the data was examined using the Kolmogorov-Smirnov and Shapiro-Wilk normality tests to determine whether the assumption of normality was satisfied (Horváth et al., 2020; Makigusa & Naito, 2020). The normality test results indicated significance values below 0.05 for both the pretest and posttest scores, demonstrating that the data were not normally distributed. Consequently, the Wilcoxon Signed-Rank Test was employed as the primary inferential statistical procedure to examine differences between students' arithmetic achievement before and after the implementation of the AI-assisted mathematics learning media. The research hypotheses were formulated as follows:

H_0: There is no statistically significant difference between students' basic arithmetic skills before and after the implementation of the AI-assisted mathematics learning media.

H_1: There is a statistically significant difference between students' basic arithmetic skills before and after the implementation of the AI-assisted mathematics learning media.

The decision criterion was established at a significance level of 0.05 (Martini et al., 2020). If the Asymp. Sig. (2-tailed) value obtained from the Wilcoxon Signed-Rank Test was less than 0.05, the null hypothesis (H_0) was rejected and the alternative hypothesis (H_1) was accepted, indicating that the AI-assisted mathematics learning media produced a statistically significant improvement in students' basic arithmetic skills (Howard & Pimentel, 2021). Conversely, if the significance value exceeded 0.05, the null hypothesis was accepted, indicating that the implementation of the AI-assisted learning media did not result in a statistically significant improvement in students' arithmetic performance.

RESULT AND DISCUSSION

Result

The analysis in this section reports the effectiveness of the artificial intelligence (AI)-based mathematics learning media in training the basic arithmetic skills of second-grade students at MI Al Fithrah Surabaya. The presentation follows the sequence of the research procedure, beginning with the developed media, followed by the validity of the media and the test instrument, and closing with the hypothesis testing that compares student performance before and after the intervention. The learning media itself was built from a combination of two generative AI technologies, ChatGPT and Canva AI, and was framed as a digital snake and ladder game titled *Petualangan Matematika*. Before it was used in the classroom, the media was validated by a PGMI lecturer at Institut Al Fithrah (IAF) Surabaya and a teacher at MI Al Fithrah, so that the product entering the learning process had already met the standard of feasibility. A preview of the media is shown in **Figure 1**.

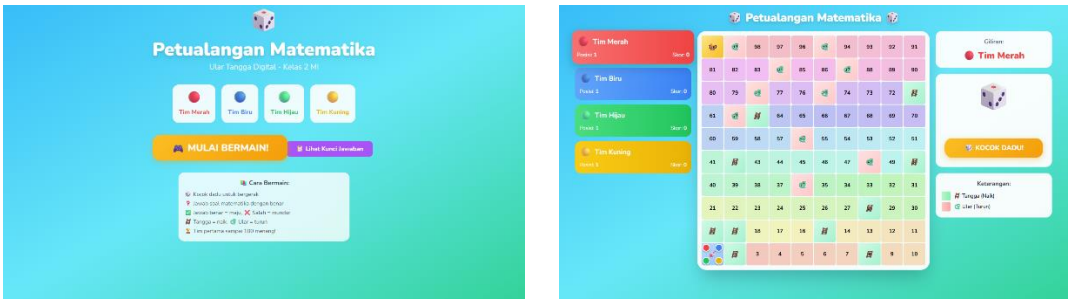

Figure 1. AI-Based Learning Mathematics Materials
Source: Developed by the researchers using ChatGPT and Canva AI (2026)

Figure 1 displays the two main interfaces of the digital board game, namely the opening menu with team selection and the playing board with a numbered grid. The design translates abstract addition and subtraction tasks into a visual, turn based activity, where students roll a dice, move across the board, and resolve arithmetic problems to advance. This format matters because it converts routine drilling into an interactive contest that sustains attention, which is often the missing element when young learners practice basic operations. The bright color coding of the teams and the clear ladder and snake markers also reduce the cognitive load of following the rules, allowing students to spend their effort on the calculation rather than on decoding the interface.

Validity Test

The evaluation of both the instructional media and the test instrument rested on the criterion of validity, understood here as the degree of alignment between the content, presentation, and function of the product and the intended learning objectives. Two validators carried out the assessment: a PGMI lecturer at IAF and a mathematics teacher at MI Al Fithrah Surabaya, ensuring that the judgment combined academic and classroom perspectives. The media validation covered 20 items (P1 to P20) addressing the suitability of the AI-based learning media, each rated on a four-point scale for a maximum total of 80. The recapitulation of the media validation is presented in .

Table 2. Summary of Media Expert Validity Results

Validator	Score	Maximum	Percentage	Criteria
PGMI Lecturer, IAF	72	80	90%	Highly Valid
Teacher, MI Al Fithrah Surabaya	74	80	92%	Highly Valid

Source: Primary data processed by the researchers (2026)

Table 2 shows that the first validator awarded a score of 72 out of 80, or 90 percent, while the second validator awarded 74 out of 80, or 92 percent. Each percentage was obtained by dividing the acquired score by the maximum score and multiplying by one hundred. Both results fall within the 81 to 100 percent band and therefore belong to the Highly Valid category. The close agreement between the two validators, with only a two point gap, indicates that the judgment was consistent rather than dependent on a single reviewer. This convergence gives confidence that the AI based media was sound in its content, its visual presentation, and its functional design before it was brought into the classroom. The same two validators then examined the test instrument, as the accuracy of the effectiveness claim depends on whether the pretest and posttest accurately measure basic arithmetic skills. The assessment focused on five items (P1 to P5) reflecting the suitability of the instrument for capturing students' basic numeracy skills, each rated on a scale of 1 to 4 for a maximum total of 20. The summary of the test instrument validation is presented in **Table 3**.

Table 3. Summary of Expert Test Validity Results

Validator	P1	P2	P3	P4	P5	Score	Maximum	Total	Criteria
PGMI Lecturer, IAF	3	3	3	4	4	17	20	85%	Highly Valid
Teacher, MI Al Fithrah Surabaya	4	3	3	3	4	17	20	85%	Highly Valid

Source: Primary data processed by the researchers (2026)

Table 3 indicates that both validators arrived at an identical total of 17 out of 20, equivalent to 85 percent, which places the instrument in the Highly Valid category within the 81 to 100 percent range. The fact that two independent reviewers reached the same total, despite small differences in their item by item scores, points to a stable and reliable judgment of the instrument. This result confirms that the test items were appropriate in content, clear in their wording, and aligned with the learning indicators for basic arithmetic. With both the media and the instrument established as highly valid, the study could proceed to test the media in actual use, confident that any measured change in scores would reflect genuine learning rather than a flaw in the tools.

Hypothesis Test

Once the validity of the media and the instrument was confirmed, the study examined the effectiveness of the media during the learning process itself. The design employed a one-group pretest-posttest procedure with 25 second-grade students at *MI Al Fithrah Surabaya*. Students first took a pretest to establish their baseline proficiency in basic arithmetic, then engaged in learning activities using the AI-assisted learning media, and finally completed a posttest to capture the change in their ability. The classroom implementation of these pretest and posttest sessions is documented in **Figure 2**.

Figure 2. Documentation of the Pretest and Posttest
Source: Researchers' field documentation at MI Al Fithrah Surabaya (2026)

Figure 2 records the learning activity in progress, showing students working on the arithmetic tasks in the classroom setting where the media was applied. The documentation confirms that the treatment was delivered under ordinary classroom conditions rather than in an artificial testing environment, which strengthens the practical relevance of the findings. It also shows students engaging directly with the material, consistent with the interactive intention behind the game-based design. This authentic setting matters because it situates the measured improvement within the real constraints of a second-grade classroom, making the result more meaningful for teachers who would adopt the media. Before any hypothesis testing, the distribution of the pretest and posttest scores was examined to determine which statistical test would be appropriate. The results of the normality test are presented in **Table 4**.

Table 4. Summary of the Normality Test

Data	Kolmogorov-Smirnov Statistic	df	Sig.	Shapiro-Wilk Statistic	df	Sig.
Pre	0.367	25	<0.001	0.634	25	<0.001
Post	0.409	25	<0.001	0.610	25	<0.001

Source: Primary data processed with IBM SPSS Statistics 27 (2026)

Table 4 reports that both the Kolmogorov-Smirnov and the Shapiro-Wilk tests returned significance values below 0.001 for the pretest and the posttest alike. Because every one of these values falls below the 0.05 threshold, the scores cannot be treated as normally distributed. This outcome has a direct methodological consequence: a parametric test that assumes normality would not be fully justified for establishing significance. The study therefore adopted the Wilcoxon Signed Rank Test, a non-parametric procedure suited to paired samples whose distribution departs from normality, as the primary basis for the hypothesis decision. The result of that test is presented in **Table 5**.

Table 5. Summary of the Wilcoxon Signed Rank Test

Statistic	Post − Pre
Z	−4.669
Asymp. Sig. (2-tailed)	<0.001

Source: Primary data processed with IBM SPSS Statistics 27 (2026)

Table 5 shows a Z value of −4.669 with an asymptotic significance of less than 0.001, computed on the basis of negative ranks. Since this significance value is well below 0.05, the null hypothesis (H0) is rejected and the alternative hypothesis (H1) is accepted. In substantive terms, there is a statistically significant difference between the students' scores before and after they learned with the AI assisted media. The negative rank basis of the statistic further indicates that the shift was overwhelmingly in the direction of improvement, meaning the posttest scores were systematically higher than the pretest scores across the group. This provides the core inferential evidence that the media had a real effect on students' basic arithmetic performance rather than a difference attributable to chance. To describe the magnitude of that change, the paired-samples statistics offer a clear picture of the average performance at each stage. These descriptive figures are presented in **Table 6**.

Table 6. Summary of the Paired Samples Statistics

Pair 1	Mean	N	Std. Deviation	Std. Error Mean
Pre	71.20	25	10.132	2.026
Post	92.80	25	9.798	1.960

Source: Primary data processed with IBM SPSS Statistics 27 (2026)

Table 6 reveals that the mean score rose from 71.20 on the pretest to 92.80 on the posttest, an average gain of 21.60 points for the 25 students. Beyond the increase in the mean, the standard deviation narrowed slightly from 10.132 to 9.798, which suggests that the students' scores became a little more clustered after the intervention. This is a meaningful detail, because it hints that the media did not simply raise the top performers but lifted the group as a whole, tightening the spread of ability. Taken together, the higher average and the reduced dispersion indicate that the improvement was both substantial in size and reasonably even across the class. The magnitude of the difference was further examined through the paired sample comparison of the pretest and posttest results, as presented in **Table 7**.

Table 7 confirms a mean difference of −21.60 between the pretest and posttest, with the entire 95 percent confidence interval, from −24.902 to −18.298, lying below zero. Because the interval does not cross zero, the direction of the effect is unambiguous: scores rose after the intervention, and the true size of that rise is estimated to fall roughly between 18 and 25 points. The significance value of less than 0.05 reinforces the conclusion already reached through the

Wilcoxon test, so the two analyses converge on the same result from different angles. On this combined evidence, the alternative hypothesis (H1) is accepted and the null hypothesis (H0) is rejected, establishing that the AI based mathematics learning media produced a significant improvement in the basic arithmetic skills of second grade students at MI Al Fithrah Surabaya.

Table 7. Summary of the Paired Samples Test

Pair 1	Mean	Std. Deviation	Std. Error Mean	95% CI Lower	95% CI Upper	t	df	Sig. (2-tailed)
Pre − Post	−21.600	8.000	1.600	−24.902	−18.298	−13.500	24	<0.001

Source: Primary data processed with IBM SPSS Statistics 27 (2026)

Discussion

The significant improvement in students' basic arithmetic skills demonstrates that the effectiveness of the developed learning media cannot be attributed solely to the integration of artificial intelligence, but rather to the way AI was purposefully embedded into instructional design. The collaborative use of ChatGPT and Canva AI enabled the development of mathematics learning media that combined contextual arithmetic content with interactive visual presentation, allowing students to engage with mathematical concepts in a more meaningful manner (Madunić & Sovulj, 2024). Rather than relying on conventional drill-based instruction, the *Petualangan Matematika* media transformed arithmetic practice into an active learning experience where students continuously interacted with addition and subtraction tasks through game-based activities. This instructional approach appears to have increased students' attention, motivation, and willingness to participate throughout the learning process, creating repeated opportunities to strengthen both conceptual understanding and procedural fluency (Ding & Yu, 2024; Nadeem et al., 2023; Partovi & Razavi, 2019). Such findings suggest that the observed learning gains were not simply the result of technology use, but of an instructional environment that encouraged students to learn through exploration, interaction, and continuous practice. Consequently, the present study demonstrates that generative AI can support mathematics learning more effectively when it is intentionally integrated into pedagogically meaningful learning experiences rather than functioning merely as a tool for content generation.

The present findings extend the growing body of evidence regarding the educational potential of artificial intelligence in mathematics learning while offering several important distinctions from previous studies. Earlier research has consistently reported that AI can improve students' mathematical achievement by supporting personalized learning, providing immediate feedback, and increasing learning engagement (Maamin et al., 2022). Similarly, Canonigo (2024) demonstrated that AI-assisted instructional technology contributes positively to mathematics learning outcomes through more interactive learning experiences. However, most previous studies have either focused on conceptual discussions, literature reviews, or the general implementation of AI without examining how generative AI can be systematically translated into practical learning media for elementary classrooms. The present study addresses this gap by providing empirical evidence from a real classroom intervention involving second-grade students in an Islamic elementary school (Ramos et al., 2024; Relmasira et al., 2023). More importantly, the novelty of this study lies not merely in the use of artificial intelligence itself, but in the pedagogical collaboration between ChatGPT and Canva AI to develop an interactive mathematics learning medium that integrates instructional content, visual communication, and game-based learning into a single instructional design. This finding therefore broadens existing knowledge by demonstrating that the educational value of generative AI depends less on the technology itself than on how it is intentionally designed to facilitate meaningful learning experiences that actively support students' conceptual and procedural understanding of mathematics.

The effectiveness of the AI-assisted learning media can be understood through several complementary perspectives on learning. From the perspective of the Cognitive Theory of Multimedia Learning, students learn more effectively when verbal explanations are integrated with meaningful visual representations, enabling them to process information through both verbal and visual channels simultaneously. The combination of AI-generated instructional content and visually engaging media developed with Canva AI provided learning experiences that supported this dual-processing mechanism, making arithmetic concepts easier for young learners to understand. At the same time, the game-based design encouraged students to construct mathematical understanding through active participation rather than passive information reception, reflecting the principles of constructivist learning. Instead of repeatedly memorizing computational procedures, students continuously interacted with arithmetic problems while receiving immediate feedback from the learning activities, allowing them to refine their understanding through practice and reflection. The structured presentation of learning tasks appeared to minimize unnecessary cognitive demands, which is consistent with Cognitive Load Theory. By organizing arithmetic exercises into sequential and visually intuitive activities, the media enabled students to allocate greater cognitive resources to understanding mathematical concepts rather than decoding instructional procedures. These complementary learning mechanisms explain why the developed media improved not only students' procedural accuracy but also their engagement and confidence during mathematics learning, indicating that the instructional effectiveness resulted from thoughtful pedagogical design rather than the presence of artificial intelligence alone.

CONCLUSION

This study concludes that AI-assisted mathematics learning media developed through the integration of ChatGPT and Canva AI is effective in improving elementary students' basic arithmetic skills. The developed learning media met the established validity criteria and significantly enhanced students' arithmetic performance after the instructional intervention, indicating that the purposeful integration of generative AI into mathematics instruction can create more engaging, interactive, and meaningful learning experiences for young learners. Nevertheless, this study has several limitations that should be acknowledged. The research was conducted using a one-group pretest-posttest design without a comparison group and involved a relatively small sample drawn from a single Islamic elementary school, which may limit the generalizability of the findings to broader educational contexts. The study evaluated only short-term learning outcomes related to basic arithmetic skills and did not examine students' long-term knowledge retention, higher-order mathematical thinking, or learning motivation. Future studies are therefore encouraged to employ more rigorous experimental designs involving control groups, larger and more diverse participant populations, and longer intervention periods. Further research may also investigate the effectiveness of AI-assisted learning media across different mathematical topics, educational levels, and school settings to provide more comprehensive evidence regarding the sustainable integration of generative AI in mathematics education.

ACKNOWLEDGMENT

The authors express their deepest gratitude to MI Al Fithrah Surabaya for granting permission and providing the facilities that served as the site of this research. The authors further extend their sincere appreciation to all students of MI Al Fithrah Surabaya for their active participation as research respondents, which greatly facilitated the completion of this journal article.

REFERENCES

Avdiu, E. (2019). Game-Based Learning Practices in Austrian Elementary Schools. *Educational Process: International Journal*, 8(3), 196–206. https://doi.org/10.22521/edupij.2019.83.4

Bhutoria, A. (2022). Personalized Education and Artificial Intelligence in the United States, China, and India: A Systematic Review Using a Human-In-The-Loop Model. *Computers and Education: Artificial Intelligence*, 3, 100068. https://doi.org/10.1016/j.caeai.2022.100068

Canonigo, A. M. (2024). Levering AI to Enhance Students' Conceptual Understanding and Confidence in Mathematics. *Journal of Computer Assisted Learning*, 40(6), 3215–3229. https://doi.org/10.1111/jcal.13065

Carter, G., Wilson, C. B., & Mitchell, G. (2021). The Effectiveness of a Digital Game to Improve Public Perception of Dementia: A Pretest-Posttest Evaluation. *PLoS ONE*, 16(10), e0257337. https://doi.org/10.1371/journal.pone.0257337

Chang, S. J., Lee, K. E., Yang, E., & Ryu, H. (2022). Evaluating a Theory-Based Intervention for Improving eHealth Literacy in Older Adults: A Single Group, Pretest–Posttest Design. *BMC Geriatrics*, 22(1). https://doi.org/10.1186/s12877-022-03545-y

Choi, G. W., Kim, S. H., Lee, D., & Moon, J. (2024). Utilizing Generative AI for Instructional Design: Exploring Strengths, Weaknesses, Opportunities, and Threats. *TechTrends*, 68(4), 832–844. https://doi.org/10.1007/s11528-024-00967-w

Deng, L., Wu, S., Chen, Y., & Peng, Z. (2020). Digital Game-Based Learning in a Shanghai Primary-School Mathematics Class: A Case Study. *Journal of Computer Assisted Learning*, 36(5), 709–717. https://doi.org/10.1111/jcal.12438

Ding, A. C. E., & Yu, C. H. (2024). Serious Game-Based Learning and Learning by Making Games: Types of Game-Based Pedagogies and Student Gaming Hours Impact Students' Science Learning Outcomes. *Computers and Education*, 218, 105075. https://doi.org/10.1016/j.compedu.2024.105075

Go, B., Lim, C., & Shin, B. (2024). Development of a Math-AI Convergence Instructional Model Using a Generative AI Chatbot. *Journal of Educational Technology*, 39(3), 1–40. https://doi.org/10.17232/kset.40.1.1

Horváth, L., Kokoszka, P., & Wang, S. (2020). Testing Normality of Data on a Multivariate Grid. *Journal of Multivariate Analysis*, 179, 104640. https://doi.org/10.1016/j.jmva.2020.104640

Howard, S. R., & Pimentel, S. D. (2021). The Uniform General Signed Rank Test and Its Design Sensitivity. *Biometrika*, 108(2), 381–396. https://doi.org/10.1093/biomet/asaa072

Ingkavara, T., Panjaburee, P., Srisawasdi, N., & Sajjapanroj, S. (2022). The Use of a Personalized Learning Approach to Implementing Self-Regulated Online Learning. *Computers and Education: Artificial Intelligence*, 3, 100086. https://doi.org/10.1016/j.caeai.2022.100086

Little, T. D., Chang, R., Gorrall, B. K., Waggenspack, L., Fukuda, E., Allen, P. J., & Noam, G. G. (2020). The Retrospective Pretest–Posttest Design Redux: On Its Validity as an Alternative to Traditional Pretest–Posttest Measurement. *International Journal of Behavioral Development*, 44(2), 175–183. https://doi.org/10.1177/0165025419877973

Luo, T., Muljana, P. S., Ren, X., & Young, D. (2025). Exploring Instructional Designers' Utilization and Perspectives on Generative AI Tools: A Mixed Methods Study. *Educational Technology Research and Development*, 73(2), 741–766. https://doi.org/10.1007/s11423-024-10437-y

Maamin, M., Maat, S. M., & Iksan, Z. H. (2022). The Influence of Student Engagement on Mathematical Achievement Among Secondary School Students. *Mathematics*, 10(1), 41. https://doi.org/10.3390/math10010041

Madunić, J., & Sovulj, M. (2024). Application of ChatGPT in Information Literacy Instructional Design. *Publications*, 12(2), 11. https://doi.org/10.3390/publications12020011

Maier, U., & Klotz, C. (2022). Personalized Feedback in Digital Learning Environments: Classification Framework and Literature Review. *Computers and Education: Artificial Intelligence*, 3, 100080. https://doi.org/10.1016/j.caeai.2022.100080

Makigusa, N., & Naito, K. (2020). Asymptotics and Practical Aspects of Testing Normality with Kernel Methods. *Journal of Multivariate Analysis*, 180, 104665. https://doi.org/10.1016/j.jmva.2020.104665

Martini, V., Favero, F., Corà, D., Varughese, F., Sica, A., & Gennari, A. (2020). Abstract P3-05-09: mRNA Expression Level and Prognostic Significance of Different Immune-Related Biomarkers in Early Breast Cancer. *Cancer Research*, 80(4), P3-05-09. https://doi.org/10.1158/1538-7445.sabcs19-p3-05-09

Nadeem, M., Oroszlanyova, M., & Farag, W. (2023). Effect of Digital Game-Based Learning on Student Engagement and Motivation. *Computers*, 12(9). https://doi.org/10.3390/computers12090177

Partovi, T., & Razavi, M. R. (2019). The Effect of Game-Based Learning on Academic Achievement Motivation of Elementary School Students. *Learning and Motivation*, 68, 101592. https://doi.org/10.1016/j.lmot.2019.101592

Ramos, L., Simões, V., & Franco, S. (2024). "Active Mathematics"—A Classroom-Based Physical Active Learning Intervention in an Elementary School: An Experimental Pilot Study. *Education Sciences*, 14(6), 637. https://doi.org/10.3390/educsci14060637

Relmasira, S. C., Lai, Y. C., & Donaldson, J. P. (2023). Fostering AI Literacy in Elementary Science, Technology, Engineering, Art, and Mathematics (STEAM) Education in the Age of Generative AI. *Sustainability*, 15(18), 13595. https://doi.org/10.3390/su151813595

Rich, K. M. (2021). Examining Agency as Teachers Use Mathematics Curriculum Resources: How Professional Contexts May Support or Inhibit Student-Centered Instruction. *Teaching and Teacher Education*, 98, 103249. https://doi.org/10.1016/j.tate.2020.103249

Schorcht, S., Buchholtz, N., & Baumanns, L. (2024). Prompt the Problem – Investigating the Mathematics Educational Quality of AI-Supported Problem Solving by Comparing Prompt Techniques. *Frontiers in Education*, 9. https://doi.org/10.3389/feduc.2024.1386075

Sein, M. (2022). AI-Assisted Knowledge Assessment Techniques for Adaptive Learning Environments. *Computers and Education: Artificial Intelligence*, 3, 100050. https://doi.org/10.1016/j.caeai.2022.100050

Squalli Houssaini, M., Aboutajeddine, A., Toughrai, I., & Ibrahimi, A. (2024). Development of a Design Course for Medical Curriculum: Using Design Thinking as an Instructional Design Method Empowered by Constructive Alignment and Generative AI. *Thinking Skills and Creativity*, 52, 101491. https://doi.org/10.1016/j.tsc.2024.101491

Wang, Y. (2020). Engagement in PC-Based, Smartphone-Based, and Paper-Based Materials: Learning Vocabulary Through Chinese Stories. *Technology in Language Teaching & Learning*, 2(1), 3–21. https://doi.org/10.29140/tltl.v2n1.319

Zhang, J., & Zhang, Z. (2024). AI in Teacher Education: Unlocking New Dimensions in Teaching Support, Inclusive Learning, and Digital Literacy. *Journal of Computer Assisted Learning*, 40(4), 1871–1885. https://doi.org/10.1111/jcal.12988