Journal of Urban Development in Education
Vol. 1, No. 2, 2025, pp. 35-45



Contents lists available at Journal Nalatama

Journal of Urban Development in Education

Journal homepage: https://journal.nalatama.org/index.php/jude

Visualizing the Solar System with Generative AI: Development and Formative Evaluation of Interactive Multimedia for Lower Secondary Science

Hamidah Khairunnisa[1], Selvi Susanti[1], Taufik Kemas[1]
[1]Universitas Islam Negeri Sultan Syarif Kasim Riau, Pekanbaru, Indonesia

Article Info

Article history:

Received May 21st, 2026
Accepted Oct 10th, 2026

Keyword:

ADDIE model; Astronomy education; ChatGPT; Multimedia learning; Learner acceptance

ABSTRACT

The solar system is an abstract topic that lower secondary learners often find difficult, and generative artificial intelligence (AI) now allows teachers to author rich multimedia without programming. This study developed an interactive multimedia package on the solar system for Grade VII science and formatively evaluated how learners received it. Following the ADDIE model, the authors used Canva, ChatGPT, Gemini, and Pippit AI to produce interactive slides, explanatory texts, quiz items, and narrated videos, hosted a gamified quiz on Wayground, and checked all content during finalization. In a small group trial at an early development stage, five Grade VII learners rated 20 statements on a four point scale covering display, content, interactivity, benefit, and ease of use. The overall mean was 3.41 (85.25% of the maximum score), and no learner disagreed with any statement. Display received the highest rating (M = 3.65) and interactivity the lowest (M = 3.20). These results indicated preliminary learner acceptance rather than learning effectiveness. The study documented a transparent workflow for AI supported authoring and identified expert validation, manipulable simulations, and controlled trials with larger samples as priorities for the next cycle.

© 2025 The Authors. Published by Nalatama.
This is an open access article under the CC BY license
(Creative Commons Attribution 4.0 International License)

Corresponding Author:

Hamidah Khairunnisa,
Universitas Islam Negeri Sultan Syarif Kasim Riau, Indonesia
Email: midah0194@gmail.com

Introduction

The solar system is among the first topics through which lower secondary learners study astronomy systematically, and it is also among the most cognitively demanding. Explaining why day and night alternate, why seasons change, or how the planets differ in size and distance requires learners to coordinate observations made from Earth with a heliocentric model viewed from outside the system, a shift of perspective that depends heavily on spatial reasoning (Plummer, 2014). More than three decades of astronomy education research have documented persistent alternative conceptions about these phenomena, including the belief that seasons result from changes in the distance between Earth and the Sun (Lelliott & Rollnick, 2010). Because such phenomena cannot be manipulated or observed directly within a single lesson, instruction that

relies on verbal explanation, textbook pictures, and static diagrams gives learners little support for building coherent mental models. The problem is pressing in Indonesia, where students aged 15 averaged 383 points in science in PISA 2022, well below the OECD average of 485 points (OECD, 2023). In Indonesian lower secondary schools, namely Sekolah Menengah Pertama (SMP) and Madrasah Tsanawiyah (MTs), the solar system is taught in Grade VII within integrated natural science (Ilmu Pengetahuan Alam, IPA), and local studies have addressed its abstractness through an Android based electronic module for Grade VII (Khaerani, M, & Nurmayanti, 2023), an animated video for Grade V (Rafli, Ali, & Sujana, 2024), and an augmented reality application for Grade VI (Sati, Kurniawan, & Herlambang, 2024).

The cognitive theory of multimedia learning (CTML) explains why well designed multimedia can help learners with such content. CTML assumes that people process verbal and pictorial information through separate channels, that each channel has limited capacity, and that meaningful learning occurs when learners actively select, organize, and integrate words and pictures with prior knowledge (Mayer, 2021, 2024). From this perspective, narrated animations can externalize spatial relations that would otherwise have to be held and transformed in working memory, whereas decorative or redundant elements impose extraneous cognitive load that competes with essential processing (Sweller, van Merriënboer, & Paas, 2019). An overview of systematic reviews confirmed that most CTML design principles have positive effects on learning, although effect sizes vary across principles and contexts (Noetel et al., 2022), and these principles now form a consolidated evidence base for instructional design (Mayer & Fiorella, 2022). Interactivity adds a further layer. In science education, gamification has generally been associated with higher engagement and motivation (Kalogiannakis, Papadakis, & Zourmpakis, 2021), and gamified quizzes with immediate feedback have supported performance in formative assessment (Zainuddin, Shujahat, Haruna, & Chu, 2020), although part of this advantage may reflect a novelty effect that fades over time (Rodrigues et al., 2022).

Generative artificial intelligence (AI) has changed how such materials can be produced. Large language models such as ChatGPT can draft explanations, examples, and assessment items within seconds, and multimodal generators can convert text prompts into images, narrated animations, and short videos (Kasneci et al., 2023; Lo, 2023). Reviews of AI in education distinguish applications that serve learners directly, such as intelligent tutoring, from applications that support teachers in planning, content creation, and assessment (Chiu, Xia, Zhou, Chai, & Cheng, 2023; Holmes & Tuomi, 2022). For teachers, the most immediate promise lies in design work: generative tools can assist lesson planning and the preparation of learning resources, provided that teachers exercise critical judgment over the outputs (Hodges & Kirschner, 2024; van den Berg & du Plessis, 2023). Mishra, Warr, and Islam (2023) argue that this development expands the technological pedagogical content knowledge (TPACK) required of teachers, because deciding what to ask a generative system and evaluating what it returns both depend on sound content and pedagogical knowledge. Early experimental evidence on instructional videos generated with AI is encouraging: among university students, such videos produced learning outcomes comparable to those of videos recorded by human instructors, although learners reported lower enjoyment or a weaker sense of social presence (Netland, von Dzengelevski, Tesch, & Kwasnitschka, 2025; Xu et al., 2025).

These opportunities come with documented risks. Generative models can produce fluent but factually incorrect content, a phenomenon described as hallucination (Ji et al., 2023). In science education, Cooper (2023) found that ChatGPT could generate useful ideas and resources for science teaching but that its outputs required critical evaluation before classroom use. A systematic scoping review of large language models in education identified limited technological readiness, insufficient transparency and replicability, and inadequate attention to privacy and beneficence as persistent challenges (Yan et al., 2024). UNESCO guidance therefore recommends a human centered approach in which educators validate generated content and remain accountable for pedagogical decisions (Miao & Holmes, 2023). A systematic review of empirical research on AI in science teaching and learning likewise highlighted both its potential to support learning and the challenges that accompany its classroom integration (Almasri, 2024).

Taken together, the literature reveals three gaps that motivate the present study. First, solar system media developed in Indonesia have mostly targeted elementary learners or relied on conventional authoring tools such as mobile application builders, animation software, and augmented reality platforms (Khaerani et al., 2023; Rafli et al., 2024; Sati et al., 2024); how generative AI can be orchestrated across an entire instructional design cycle for lower secondary astronomy remains scarcely documented. Second, most empirical work on generative AI in education concerns university students or chatbots used directly by learners, whereas the use of generative tools by teachers as authoring partners for younger learners has received far less attention (Chiu

et al., 2023; Yan et al., 2024). Third, development studies frequently label products as based on AI without distinguishing AI supported authoring from AI driven instruction and without reporting how generated content was checked, which weakens both replicability and trust. Closing these gaps requires a transparent account of what was generated, by which tool, for which instructional function, how it was verified, and how learners responded to the resulting product.

This study responds to these gaps by designing, developing, and formatively evaluating an interactive multimedia package on the solar system for Grade VII science in which generative AI served as an authoring partner rather than as an adaptive tutor. Guided by the ADDIE model (Analysis, Design, Development, Implementation, and Evaluation), the study addressed two research questions: (1) How can generative AI tools be integrated into the ADDIE phases to produce interactive multimedia on the solar system for Grade VII learners? (2) How do Grade VII learners perceive the display, content, interactivity, benefit, and ease of use of the resulting product in a small group formative trial? Consistent with the formative purpose of the trial, the study does not test learning effectiveness. Instead, it offers a transparent development record and early user evidence that can inform expert validation and controlled experimentation in subsequent cycles.

Method

This study adopted a design and development research approach guided by the ADDIE instructional design model, which organizes product development into analysis, design, development, implementation, and evaluation phases and allows revision decisions to be traced to evidence generated in each phase (Branch, 2009; Spatioti, Kazanidis, & Pange, 2022). Evaluation was framed within the formative evaluation framework of Tessmer (1993), which distinguishes expert review, one to one evaluation, small group evaluation, and field testing. The present article reports one complete development cycle and its small group stage. Before the learner trial, the prototype was presented to the course lecturer and peers in an informal walkthrough to obtain initial feedback on its display, content, navigation, videos, and quizzes. Formal expert review with validated instruments and field testing were beyond the scope of this cycle and are treated as limitations. Table 1 summarizes the activities, tools, and outputs of each ADDIE phase.

Table 1. Activities, tools, and outputs of each ADDIE phase

Phase	Main activities	Tools	Output
Analysis	Identified learning needs, characteristics of Grade VII learners, and difficulties in learning the solar system	None	Problem statement and learner profile
Design	Formulated learning objectives, a content map, screen layout, navigation flow, video plans, and a quiz blueprint; assigned each tool to a production task	Tool selection	Design blueprint
Development	Built interactive slides; drafted texts and quiz items; generated narrated videos; built the gamified quiz; integrated audio, voiceover, and navigation; verified content and functions; held an informal walkthrough	Canva (including Canva AI), ChatGPT, Gemini, Pippit AI, Wayground	Prototype multimedia package
Implementation	Small group trial with five Grade VII learners covering the materials, videos, and gamified quiz	Prototype package, Wayground	Observations and quiz completion
Evaluation	Administered a questionnaire of 20 items; performed descriptive analysis; identified revision needs	Questionnaire	Acceptance profile and revision agenda

The small group comprised five Grade VII learners from an Indonesian lower secondary school (SMP/MTs), who at this grade are typically 12 to 13 years old. The learners constituted a small group assembled for formative testing rather than a sample drawn to represent a population. Five learners are too few for inferential analysis or generalization, but the number is consistent with the logic of formative small group evaluation,

whose purpose is to detect problems of clarity, navigation, and appeal before a product is tested with larger groups (Tessmer, 1993). Usability research likewise shows that a small number of users, around five, can reveal a large share of the most frequent usability problems of a product (Nielsen & Landauer, 1993). The trial data were therefore treated as diagnostic evidence for revision rather than as a basis for generalization.

Five platforms were combined, each assigned to a distinct instructional function. Canva served as the main authoring environment for the interactive slide deck and provided templates, animations, icons, color schemes, and hyperlinked navigation buttons. ChatGPT was prompted to draft concise explanations suited to Grade VII learners and to generate multiple choice and true or false practice items. Short narrated videos featuring an animated child astronaut were generated from Indonesian text prompts with the AI features of Canva, with Gemini, and with Pippit AI (Figure 1), and instrumental background audio and AI voiceover were added to sustain attention. Wayground (formerly Quizizz) hosted a gamified quiz that provided immediate feedback and a live leaderboard. All components were assembled into a single package, with the Wayground quiz accessed through a link, so that learners could work through the material at their own pace.

Because generated content can contain errors, AI outputs were treated as drafts rather than as finished material. During finalization, the authors checked the entire content of the package, including the explanations and quiz items drafted by ChatGPT and the generated videos, and revised the display and navigation where needed. This check relied on the authors rather than on independent content experts, and no systematic accuracy audit was performed; its adequacy is examined in the Results and Discussions section.

Learner perceptions were measured with a response questionnaire of 20 statements written in simple Indonesian and grouped into five aspects of four items each: display (items 1 to 4), content (items 5 to 8), interactivity (items 9 to 12), benefit (items 13 to 16), and ease of use (items 17 to 20). The five aspects correspond to the qualities targeted in the design of the package. Display and content items reflect the concern of CTML with the clarity and organization of words and pictures (Mayer, 2021); interactivity items capture the ability of learners to act on and revisit the material; benefit items capture perceived learning value and enjoyment; and ease of use items reflect usability. Responses were given on a four point scale (1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree). The statements were short and concrete, which suits young respondents, because the quality of their responses depends on their cognitive development and on item simplicity (Borgers, de Leeuw, & Hox, 2000). The questionnaire had not undergone formal content validation by experts, and no validity index was computed in this cycle. The observations made by the authors of learner behavior, difficulties, and comments during the trial were used to complement the questionnaire.

Before the trial, all functions of the package, including navigation buttons, video playback, animations, and the Wayground quiz, were checked for errors. In a single session, learners explored the material menus, watched the videos, and completed the gamified quiz while the authors were present to assist, after which the learners completed the questionnaire.

Questionnaire data were analyzed descriptively. For each aspect and for the questionnaire as a whole, the mean (*M*), the standard deviation (*SD*), and the percentage of the maximum attainable score were computed, the latter as P = (obtained score / maximum score) × 100. The proportion of strongly agree responses was reported to indicate the intensity of agreement, and item means were computed to locate specific strengths and weaknesses. Means were interpreted with four equal interval categories whose width of 0.75 was obtained by dividing the scale range of 3 by the number of categories: 1.00 to 1.75 (very negative), 1.76 to 2.50 (negative), 2.51 to 3.25 (positive), and 3.26 to 4.00 (very positive). An item or aspect was therefore classified as positive or very positive when its mean was at least 2.51. Because each aspect comprised 20 responses, a change of one scale point in a single response shifts an aspect mean by 0.05, which was taken into account when aspects were compared. Cronbach's alpha was computed for descriptive completeness only, because reliability coefficients estimated from five respondents are highly unstable and cannot substantiate the reliability of an instrument (Taber, 2018). Observations were used qualitatively to help explain the quantitative pattern.

The trial was conducted with the permission of the school, and parents or guardians were informed about the activity and consented to the participation of their children; the learners also gave verbal assent. Participation was voluntary, did not affect grades, and could be discontinued at any time. Respondents are reported only as R1 to R5, and the names visible on the quiz leaderboard were masked before publication. The characters that appear in the videos were generated by AI and do not depict real persons.

Generative AI tools were part of the object of this study, as described above. In preparing the revised manuscript, an AI assistant (Claude, Anthropic) was used to support language editing, the structuring of the

literature synthesis, and formatting according to the journal template; the authors reviewed and verified all content, analyses, and references and take full responsibility for the article. The multimedia package and the anonymized response data are available from the corresponding author upon reasonable request.

Results and Discussions

The analysis phase showed that learners had difficulty understanding the solar system because instruction was dominated by lectures and textbook reading and because the available media were limited to textbooks and simple physical models. The concepts that learners found hardest to picture were planetary motion, the arrangement of the solar system, and the characteristics of celestial bodies. The analysis therefore concluded that interactive multimedia was needed to present these concepts more concretely, attractively, and comprehensibly. Because progress in astronomy understanding depends largely on the development of spatial thinking (Plummer, 2014), this need implied media that visualize the motion and arrangement of celestial bodies, can be revisited independently, and give feedback on understanding. In the design phase, these needs were translated into learning objectives, a content map of the solar system topic that included the planets and planetary rotation and revolution, a navigation flow organized around menus, video plans, and a quiz blueprint, and each tool was assigned to a specific production task (Table 1).

The development phase produced an interactive package consisting of a main menu, material pages on the solar system, narrated videos generated with AI, images and animations, practice items, and an evaluation quiz. Figure 1 documents how the video segments were produced. Each prompt specified the narrator (a child astronaut), the language (Indonesian), and the target concept: the planet Mercury (Figure 1a), rotation as the cause of day and night (Figure 1b), and revolution as the cause of seasonal change (Figure 1c). This workflow, driven entirely by prompts, allowed short videos with a recurring character to be produced without animation or programming skills, a task that would ordinarily require specialized software and considerable production time. The package was then finalized through checks of content, display, navigation, and playback and was presented to the course lecturer and peers in the informal walkthrough described in the Method section.

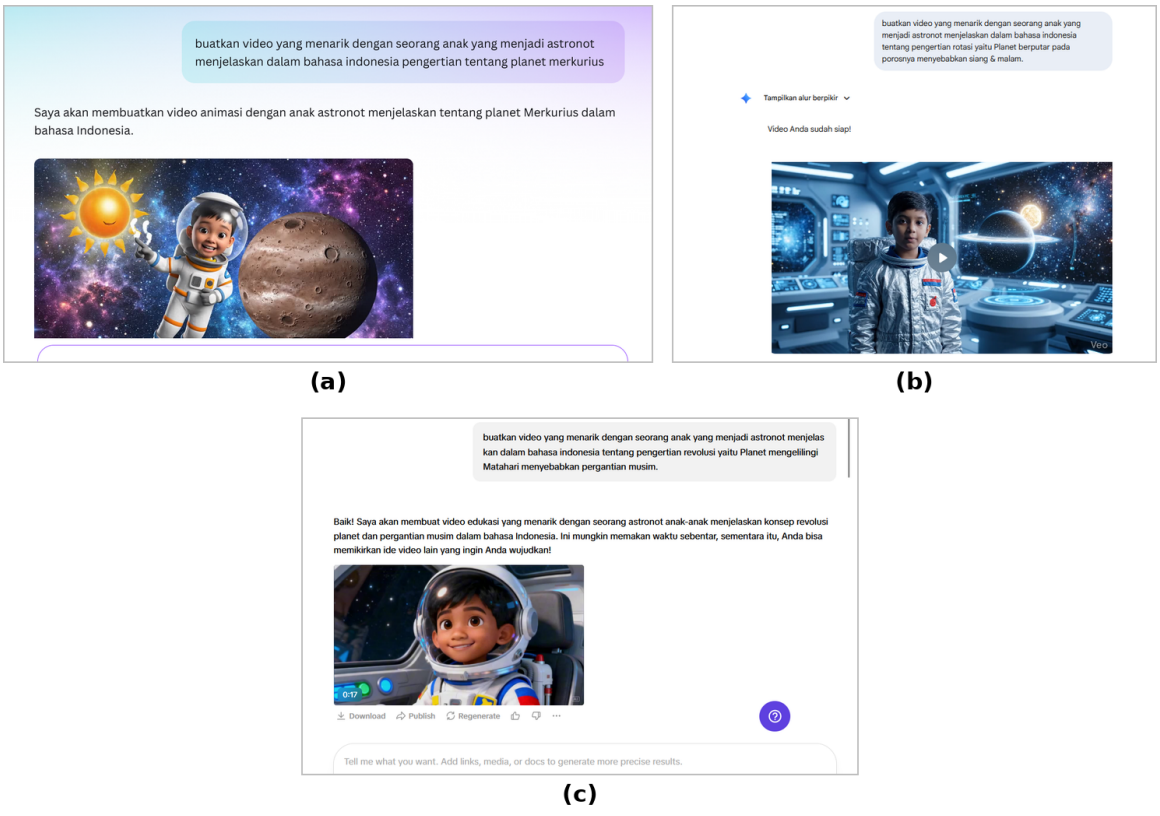


Figure 1. Generation of video segments from Indonesian prompts using (a) Canva AI, (b) Gemini, and (c) Pippit AI

(resource: authors' documentation)

During the small group trial, learners appeared attentive and enthusiastic. They watched the animations and videos closely, explored the available menus, and, as observed by the authors, answered the gamified quiz with more energy than they usually showed in regular lessons. Several learners remarked that the display was attractive and made the solar system easier to understand. Two minor obstacles emerged: some learners needed guidance in using the navigation buttons at the beginning of the session, and some needed time to adapt to the digital quiz format. Both obstacles were resolved through brief assistance. All five learners completed the Wayground quiz (Figure 2). Because leaderboard points combine accuracy with response speed, they were not treated as a measure of achievement.

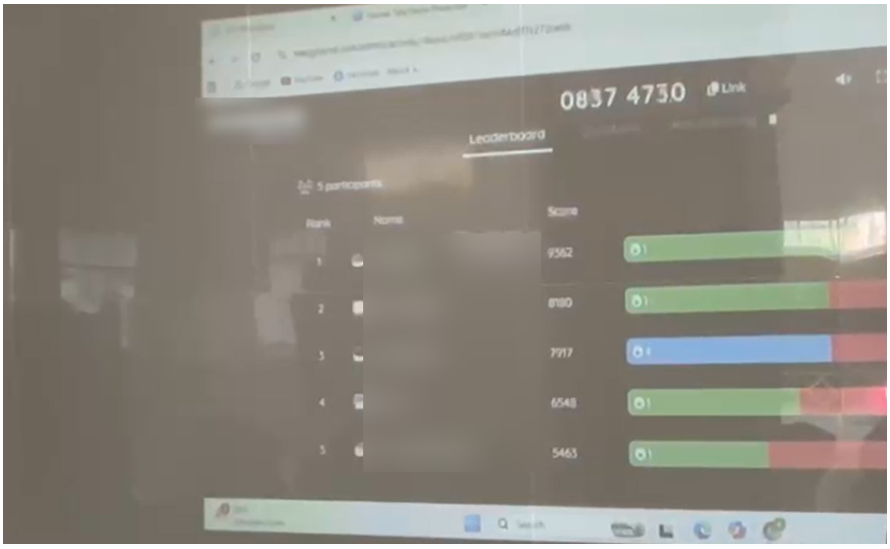


Figure 2. Wayground quiz leaderboard during the small group trial (learner names masked)
(resource: authors' documentation)

Table 2 summarizes learner responses by aspect. The overall mean was 3.41 (*SD* = 0.49), which corresponds to 85.25% of the maximum score and falls in the very positive category. All 100 item responses (five learners × 20 items) were either agree (59%) or strongly agree (41%), and no learner selected a disagreement option for any item. Four of the five aspects were categorized as very positive, whereas interactivity (*M* = 3.20, 80.00%) was categorized as positive. The display aspect received the highest mean (*M* = 3.65, 91.25%), with 65% of its responses at strongly agree.

Table 2. Learner responses by aspect (n = 5)

Aspect	Items	*M*	*SD*	% of max	SA (%)	Category
Display	1 to 4	3.65	0.49	91.25	65	Very positive
Content	5 to 8	3.30	0.47	82.50	30	Very positive
Interactivity	9 to 12	3.20	0.41	80.00	20	Positive
Benefit	13 to 16	3.50	0.51	87.50	50	Very positive
Ease of use	17 to 20	3.40	0.50	85.00	40	Very positive
Overall	**1 to 20**	**3.41**	**0.49**	**85.25**	**41**	**Very positive**

Note. *M* = mean; *SD* = standard deviation of the 20 responses within each aspect (100 responses for the overall row); % of max = percentage of the maximum attainable score; SA = proportion of strongly agree responses.

At the item level (Table 3), the highest means were obtained for the statements that the animations helped learners understand the material and that the text was easy to read (both *M* = 3.80), followed by the attractiveness of the display and three benefit items (*M* = 3.60). The lowest mean was obtained for the statement “I can interact with the multimedia” (*M* = 3.00), for which all five learners chose agree and none chose strongly agree. Five further items, on understanding the material, the clarity of its sequence, the ease of using buttons and menus, the possibility of repeating the material, and remembering the material, each had a

mean of 3.20. Responses varied across learners as well: individual means ranged from 3.20 (R1) to 3.80 (R2), and R2 alone contributed 16 of the 41 strongly agree responses. Because a single response shifts an aspect mean by 0.05, the differences between aspects, which ranged from 0.10 to 0.45, describe a tentative profile rather than a stable ranking. Cronbach's alpha across the 20 items was .82, but the coefficients for the individual aspects, each with four items, ranged from below zero to .86, which illustrates why such values cannot be interpreted as evidence of reliability with five respondents.

Table 3. Item level responses of the five learners (statements translated from Indonesian)

No	Aspect	Statement	R1	R2	R3	R4	R5	*M*
1	Display	The multimedia display is attractive	3	4	3	4	4	3.60
2	Display	Colors and pictures are clearly visible	3	4	3	3	4	3.40
3	Display	The animations help me understand the material	3	4	4	4	4	3.80
4	Display	The text is easy to read	3	4	4	4	4	3.80
5	Content	The solar system material is easy to understand	4	3	3	3	3	3.20
6	Content	The material is presented in a clear sequence	3	4	3	3	3	3.20
7	Content	The examples help my understanding	3	4	3	3	4	3.40
8	Content	The material matches the lessons in class	4	4	3	3	3	3.40
9	Interactivity	I can interact with the multimedia	3	3	3	3	3	3.00
10	Interactivity	The buttons and menus are easy to use	3	4	3	3	3	3.20
11	Interactivity	I can repeat the material at any time	3	4	3	3	3	3.20
12	Interactivity	The quizzes and exercises help my understanding	4	4	3	3	3	3.40
13	Benefit	The multimedia makes learning more interesting	3	4	4	4	3	3.60
14	Benefit	I understand the solar system more easily	3	4	4	4	3	3.60
15	Benefit	I remember the material more easily	3	3	3	3	4	3.20
16	Benefit	Learning feels more enjoyable	4	4	4	3	3	3.60
17	Ease of use	The multimedia is easy to use	3	4	4	3	3	3.40
18	Ease of use	I have no difficulty using it	3	4	4	3	3	3.40
19	Ease of use	The time provided is sufficient	3	3	4	3	4	3.40
20	Ease of use	I would like to use media like this again	3	4	4	3	3	3.40

Note. Scale: 1 = strongly disagree, 2 = disagree, 3 = agree, 4 = strongly agree; *M* = item mean.

With respect to the first research question, the development record shows that generative AI can be integrated into the main production tasks of the development phase while analysis, design decisions, verification, and evaluation remain under human control. The division of labor was deliberate: ChatGPT drafted verbal explanations and items, the video generators produced pictorial and narrated segments, Canva integrated the materials, and Wayground supplied retrieval practice with feedback. This configuration positions the developer, and by extension the teacher, as designer and curator rather than as the producer of every asset, a role shift anticipated in recent analyses of instructional design in the age of generative AI (Hodges & Kirschner, 2024; Mishra et al., 2023). It also clarifies a conceptual point that is often blurred in local development studies: the product relies on AI in its authoring process, not in its behavior during use. Learners interacted with a conventional hyperlinked package and a quiz platform rather than with an adaptive AI system, an arrangement that suits learners of this age, given that UNESCO guidance proposes a minimum age of 13 for the independent use of generative AI tools in classrooms (Miao & Holmes, 2023).

With respect to the second research question, learners responded very positively overall, and the display aspect received the strongest endorsement. The high ratings for animations and readability (both *M* = 3.80) are

consistent with the multimedia and modality principles of CTML, which predict that combining narration with dynamic pictures helps learners connect verbal and pictorial representations of spatial phenomena (Mayer, 2021, 2024). The pattern also aligns with local studies in which an animated video judged highly feasible by experts was followed by gains in understanding among Grade V learners (Rafli et al., 2024), an augmented reality application was followed by improved conceptual understanding among Grade VI learners (Sati et al., 2024), and an electronic module for Grade VII learners was rated as highly practical (Khaerani et al., 2023). Nevertheless, attractiveness is not equivalent to learning. Generated videos easily include cinematic backgrounds, expressive characters, and background music that are appealing but not essential to the concept; according to the coherence principle and cognitive load theory, such elements can consume working memory that should be devoted to the scientific model (Mayer & Fiorella, 2022; Sweller et al., 2019). Future iterations should therefore test whether the decorative richness that generative tools produce so easily supports or distracts from conceptual understanding.

The comparatively lower rating for interactivity (M = 3.20) is the most informative finding for revision, particularly because all five learners chose agree, and none chose strongly agree, for the statement “I can interact with the multimedia”. The interactivity available to learners was largely navigational, consisting of menu buttons, replay, and an external quiz, rather than manipulative, such as changing the orbit of a planet or tilting the axis of Earth and observing the consequences. The response pattern suggests that learners perceived this limitation. Because spatial reasoning is central to progress in astronomy (Plummer, 2014), and because an augmented reality application that presents the solar system as three dimensional models has been followed by gains in conceptual understanding (Sati et al., 2024), the next development cycle should add manipulable simulations and embedded questions that require learners to predict and explain rather than merely recall. Conversational AI helpers could raise interactivity further, but only with safeguards for accuracy, privacy, and child safety and within the age provisions noted above (Miao & Holmes, 2023; Yan et al., 2024).

Benefit was the second highest aspect (M = 3.50), with learners agreeing that the multimedia made learning more interesting, easier, and more enjoyable. The gamified quiz probably contributed to this perception, since gamification in science education and gamified formative quizzes have generally been associated with higher engagement and motivation (Kalogiannakis et al., 2021; Zainuddin et al., 2020). Several alternative explanations must nevertheless be considered before these perceptions are attributed to the design itself. First, a single exposure to an unfamiliar format invites a novelty effect, which in gamified learning has been shown to decline after the initial weeks (Rodrigues et al., 2022). Second, young respondents are prone to response effects such as suggestibility and socially desirable answering (Borgers et al., 2000), a tendency that may be stronger when the developers are present; the complete absence of disagreement responses, together with the concentration of strongly agree responses in one learner, is consistent with acquiescent or socially desirable responding. Third, the benefit item on remembering the material received one of the lowest ratings (M = 3.20), suggesting that learners were more certain about enjoyment than about learning gains. For these reasons, the favorable perceptions reported here should be read as preliminary acceptance, not as evidence of improved understanding, interest, or motivation.

The development record also exposes a risk that is specific to generative authoring. The prompt used for the Pippit AI video (Figure 1c) framed revolution as the cause of seasonal change. This formulation is incomplete, because seasons arise from the tilt of the rotational axis of Earth relative to its orbital plane, which changes the angle and duration of sunlight received by each hemisphere during the annual revolution. Presenting revolution alone as the cause may reinforce the widespread misconception that seasons depend on the distance between Earth and the Sun (Lelliott & Rollnick, 2010). A generative tool renders such a prompt faithfully and attractively, so a conceptual weakness in the prompt is more likely to be amplified than corrected. This observation substantiates concerns that generated science content requires expert scrutiny (Cooper, 2023; Ji et al., 2023) and shows that verification must cover not only AI outputs but also the prompts written by developers. In line with the TPACK perspective (Mishra et al., 2023), the quality of generative authoring depends on the content knowledge of the person who writes the prompt. The narration of this segment should therefore be reviewed so that it states the role of axial tilt explicitly, and validation by content experts must precede wider use of the package.

These findings extend emerging evidence on instructional videos generated with AI. Experimental studies with university students found that such videos can support learning to a degree comparable to that of videos recorded by human instructors, although learners experienced less enjoyment or a weaker social presence (Netland et al., 2025; Xu et al., 2025). The present study adds perception data from younger learners, for whom

an animated child astronaut, a narrator resembling their own age group, did not appear to reduce appeal, given that display and benefit received the highest ratings. Whether such a narrator also promotes learning, as opposed to liking, remains an open question for experimental testing. For practice, the workflow suggests that science teachers could produce multimedia with a recurring character, narration, and feedback using widely accessible tools that require no programming, which is relevant for schools with limited budgets for commercial media. The same accessibility, however, shifts the burden of quality assurance to teachers, reinforcing calls for professional development that combines AI literacy with content expertise (Miao & Holmes, 2023; van den Berg & du Plessis, 2023).

Several limitations bound these conclusions. The small group of five learners, drawn from a single school, precludes generalization and inferential analysis. Formal validation by content and media experts using validated instruments was not conducted, so the scientific accuracy and instructional quality of the package have not yet been independently established; the problem identified in Figure 1c illustrates why this step is indispensable. The questionnaire lacked formal evidence of content and construct validity, and its reliability could not be estimated meaningfully with five respondents. No measure of learning outcomes, such as a conceptual test administered before and after instruction, was included, so no claim of effectiveness can be made. The trial consisted of a single session conducted by the developers, which may have inflated responses through novelty and social desirability. Finally, generative tools evolve rapidly, and identical prompts may yield different outputs in later versions, which limits exact reproducibility.

Future work should proceed through the remaining stages of formative evaluation (Tessmer, 1993). First, the package should be revised on the basis of the present findings and examined by content and media experts using quantified validity indices. Second, a field test with at least one intact class should assess usability with a standardized instrument. Third, a controlled comparison with pretest and posttest measures of conceptual understanding, including items that target known misconceptions about seasons and planetary motion (Lelliott & Rollnick, 2010), together with measures of cognitive load and delayed retention, would allow the effectiveness of multimedia produced with generative AI to be tested rigorously. Finally, comparing versions with and without manipulable simulations would clarify whether the interactivity gap identified here matters for learning.

Conclusions

This study shows that generative AI can be integrated into the development phase of the ADDIE model to produce interactive multimedia on the solar system for Grade VII science, with ChatGPT, Canva, Gemini, Pippit AI, and Wayground each serving a distinct instructional function under the control of the developers. In a small group formative trial, learners responded very positively overall, rating display and benefit highest and interactivity lowest. These results establish preliminary learner acceptance and identify concrete targets for revision, namely manipulable interactivity and rigorous verification of scientific content, including the prompts from which generated media are produced; they do not demonstrate improved understanding, interest, or motivation. The main contribution of the study is a transparent and documented account of generative AI as an authoring partner in lower secondary science, together with a concrete example of how an incomplete prompt can carry a misconception into attractive media. Expert validation, field testing with larger samples, and experimental studies that measure conceptual understanding are required before the package can be recommended for wider classroom use.

References

Almasri, F. (2024). Exploring the impact of artificial intelligence in teaching and learning of science: A systematic review of empirical research. *Research in Science Education, 54*(5), 977–997. https://doi.org/10.1007/s11165-024-10176-3

Borgers, N., de Leeuw, E., & Hox, J. (2000). Children as respondents in survey research: Cognitive development and response quality. *Bulletin de Méthodologie Sociologique, 66*(1), 60–75. https://doi.org/10.1177/075910630006600106

Branch, R. M. (2009). *Instructional design: The ADDIE approach*. New York, NY: Springer. https://doi.org/10.1007/978-0-387-09506-6

Chiu, T. K. F., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence, 4*, 100118. https://doi.org/10.1016/j.caeai.2022.100118

Cooper, G. (2023). Examining science education in ChatGPT: An exploratory study of generative artificial intelligence. *Journal of Science Education and Technology, 32*(3), 444–452. https://doi.org/10.1007/s10956-023-10039-y

Hodges, C. B., & Kirschner, P. A. (2024). Innovation of instructional design and assessment in the age of generative artificial intelligence. *TechTrends, 68*(1), 195–199. https://doi.org/10.1007/s11528-023-00926-x

Holmes, W., & Tuomi, I. (2022). State of the art and practice in AI in education. *European Journal of Education, 57*(4), 542–570. https://doi.org/10.1111/ejed.12533

Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., . . . Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys, 55*(12), 248. https://doi.org/10.1145/3571730

Kalogiannakis, M., Papadakis, S., & Zourmpakis, A.-I. (2021). Gamification in science education. A systematic review of the literature. *Education Sciences, 11*(1), 22. https://doi.org/10.3390/educsci11010022

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., . . . Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences, 103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

Khaerani, K., M, U., & Nurmayanti, N. (2023). Development of science e-modules on solar system materials based on Android for junior high school students. *Indonesian Journal of Instructional Media and Model, 5*(1), 58–68. https://doi.org/10.32585/ijimm.v5i1.4169

Lelliott, A., & Rollnick, M. (2010). Big ideas: A review of astronomy education research 1974–2008. *International Journal of Science Education, 32*(13), 1771–1799. https://doi.org/10.1080/09500690903214546

Lo, C. K. (2023). What is the impact of ChatGPT on education? A rapid review of the literature. *Education Sciences, 13*(4), 410. https://doi.org/10.3390/educsci13040410

Mayer, R. E. (2021). *Multimedia learning* (3rd ed.). Cambridge, England: Cambridge University Press. https://doi.org/10.1017/9781316941355

Mayer, R. E. (2024). The past, present, and future of the cognitive theory of multimedia learning. *Educational Psychology Review, 36*(1), 8. https://doi.org/10.1007/s10648-023-09842-1

Mayer, R. E., & Fiorella, L. (Eds.). (2022). *The Cambridge handbook of multimedia learning* (3rd ed.). Cambridge, England: Cambridge University Press. https://doi.org/10.1017/9781108894333

Miao, F., & Holmes, W. (2023). *Guidance for generative AI in education and research*. Paris, France: UNESCO. https://doi.org/10.54675/EWZM9535

Mishra, P., Warr, M., & Islam, R. (2023). TPACK in the age of ChatGPT and generative AI. *Journal of Digital Learning in Teacher Education, 39*(4), 235–251. https://doi.org/10.1080/21532974.2023.2247480

Netland, T., von Dzengelevski, O., Tesch, K., & Kwasnitschka, D. (2025). Comparing human-made and AI-generated teaching videos: An experimental study on learning effects. *Computers & Education, 224*, 105164. https://doi.org/10.1016/j.compedu.2024.105164

Nielsen, J., & Landauer, T. K. (1993). A mathematical model of the finding of usability problems. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems* (pp. 206–213). New York, NY: ACM. https://doi.org/10.1145/169059.169166

Noetel, M., Griffith, S., Delaney, O., Harris, N. R., Sanders, T., Parker, P., . . . Lonsdale, C. (2022). Multimedia design for learning: An overview of reviews with meta-meta-analysis. *Review of Educational Research, 92*(3), 413–454. https://doi.org/10.3102/00346543211052329

OECD. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. Paris, France: OECD Publishing. https://doi.org/10.1787/53f23881-en

Plummer, J. D. (2014). Spatial thinking as the dimension of progress in an astronomy learning progression. *Studies in Science Education, 50*(1), 1–45. https://doi.org/10.1080/03057267.2013.869039

Rafli, A., Ali, E. Y., & Sujana, A. (2024). Pengembangan video animasi multimedia bertema KATSU untuk meningkatkan pemahaman peserta didik SD kelas V pada materi tata surya [Development of a KATSU themed multimedia animation video to improve the understanding of fifth grade students on the solar system]. *Al-Madrasah: Jurnal Ilmiah Pendidikan Madrasah Ibtidaiyah, 8*(4), 1440–1453. https://doi.org/10.35931/am.v8i4.4074

Rodrigues, L., Pereira, F. D., Toda, A. M., Palomino, P. T., Pessoa, M., Carvalho, L. S. G., . . . Isotani, S. (2022). Gamification suffers from the novelty effect but benefits from the familiarization effect: Findings from a longitudinal study. *International Journal of Educational Technology in Higher Education, 19*, 13. https://doi.org/10.1186/s41239-021-00314-6

Sati, L., Kurniawan, D. T., & Herlambang, Y. T. (2024). Implementasi media pembelajaran Solar System Augmented Reality View (SoLAR View) pada pembelajaran IPA materi sistem tata surya terhadap pemahaman konsep [Implementation of the Solar System Augmented Reality View (SoLAR View) learning medium in science lessons on the solar system and its effect on conceptual understanding]. *INKUIRI: Jurnal Pendidikan IPA, 13*(2), 205–210. https://doi.org/10.20961/inkuiri.v13i2.86472

Spatioti, A. G., Kazanidis, I., & Pange, J. (2022). A comparative study of the ADDIE instructional design model in distance education. *Information, 13*(9), 402. https://doi.org/10.3390/info13090402

Sweller, J., van Merriënboer, J. J. G., & Paas, F. (2019). Cognitive architecture and instructional design: 20 years later. *Educational Psychology Review, 31*(2), 261–292. https://doi.org/10.1007/s10648-019-09465-5

Taber, K. S. (2018). The use of Cronbach's alpha when developing and reporting research instruments in science education. *Research in Science Education, 48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tessmer, M. (1993). *Planning and conducting formative evaluations: Improving the quality of education and training*. London, England: Kogan Page.

van den Berg, G., & du Plessis, E. (2023). ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education. *Education Sciences, 13*(10), 998. https://doi.org/10.3390/educsci13100998

Xu, T., Liu, Y., Jin, Y., Qu, Y., Bai, J., Zhang, W., & Zhou, Y. (2025). From recorded to AI-generated instructional videos: A comparison of learning performance and experience. *British Journal of Educational Technology, 56*(4), 1463–1487. https://doi.org/10.1111/bjet.13530

Yan, L., Sha, L., Zhao, L., Li, Y., Martinez-Maldonado, R., Chen, G., . . . Gašević, D. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology, 55*(1), 90–112. https://doi.org/10.1111/bjet.13370

Zainuddin, Z., Shujahat, M., Haruna, H., & Chu, S. K. W. (2020). The role of gamified e-quizzes on student learning and engagement: An interactive gamification solution for a formative assessment system. *Computers & Education, 145*, 103729. https://doi.org/10.1016/j.compedu.2019.103729