# Development and Validation of an AI-Assisted E-Comic to Support Deep Learning and Critical Thinking in Senior High School

**Aris Rudi Purnomo[1], Prima Vidya Asteria[2]*, Muhamad Ro'is Abidin[3], Suhailee Sohnui[4], Irhamna Nirbhaya Carreca[5]**

[1]Department of Natural Science Education, Universitas Negeri Surabaya, Indonesia

[2]Department of Indonesian Language and Literature Education, Universitas Negeri Surabaya, Indonesia

[3]Department of Visual Communication Design, Universitas Negeri Surabaya, Indonesia

[5]Faculty of Vocational Studies, Universitas Negeri Surabaya

[4]Department of Eastern Language, Chiang Mai University, Thailand

*Corresponding author: primaasteria@unesa.ac.id

**HOW TO CITE:**
Aris Rudi Purnomo, Prima Vidya Asteria, Muhamad Ro'is Abidin, Suhailee Sohnui, Irhamna Nirbhaya Carreca (2026), Development and Validation of an AI-Assisted E-Comic to Support Deep Learning and Critical Thinking in Senior High School

International Journal of Special Education, 41(21s), 344 - 356



**ABSTRACT**

Indonesian students consistently perform below the OECD average in reading and scientific literacy, a gap that underscores the need for instructional media that make abstract, conceptually dense topics accessible while cultivating higher-order thinking. This study developed and evaluated an artificial intelligence (AI)-assisted electronic comic (e-comic) designed to operationalise the deep learning paradigm (mindful, meaningful, and joyful learning) for senior high school students. Using the ADDIE (Analysis, Design, Development, Implementation, Evaluation) research and development model, the e-comic was produced through generative AI tools (a large-language-model chatbot for narrative and prompt design, and AI-assisted illustration platforms for character-consistent artwork) around three narratives addressing biology and language topics that teachers had identified as difficult to teach. Content and media validity were assessed by expert validators using a four-point Likert rubric; practicality and reliability were assessed through a 17-item student questionnaire (Cronbach's $\alpha = 0.82$); and learning impact was assessed with a pretest–posttest design ($N = 30$). All content-validity indicators reached the maximum score (4/4, 'very valid'), and media validity averaged 3.67/4. The questionnaire showed a modal item score of 3 ('valid/practical') on 16 of 17 items, with an overall mean of 3.16 ($SD = 0.53$) on a 4-point scale. Students' comprehension and critical-thinking scores rose from a pretest mean of 46.40 ($SD = 12.61$) to a posttest mean of 82.10 ($SD = 8.54$), a statistically significant gain, $t(29) = 17.69$, $p < .001$, Cohen's $d = 3.31$, corresponding to a moderate-to-high normalised gain ($\langle g \rangle = 0.67$). Qualitative feedback indicated that students valued the clarity of illustrations and the accessibility of the narratives, while some requested additional chapters, colour illustration, and reduced reliance on AI-generated art style. The findings suggest that AI-assisted e-comics are a valid, reliable, and

practical medium for strengthening critical thinking within a deep learning framework, while highlighting specific directions, visual diversification and curriculum alignment for iterative refinement.

**Keywords:** artificial intelligence; digital comics; deep learning; critical thinking; instructional media; research and development

## Introduction

Preparing learners who are intellectually capable, emotionally resilient, and adaptive to a rapidly digitising world remains a central challenge for the Indonesian education system. National participation in the Programme for International Student Assessment (PISA) has repeatedly shown that Indonesian 15-year-olds score well below the average of the Organisation for Economic Co-operation and Development (OECD) in reading, mathematics, and science; in the most recent 2022 cycle, Indonesia's mean reading score was 359 points against an OECD average of 476, with only about a quarter of students reaching the minimum reading proficiency level attained by roughly three-quarters of students across OECD countries (OECD, 2023). Earlier cycles produced a similarly wide gap in reading and in general problem-solving and critical-thinking competencies (OECD, 2018, 2019). These figures point to a persistent, structural need for instructional approaches that are simultaneously more engaging and more cognitively demanding than conventional transmission-based teaching.

One paradigm increasingly advocated as a response to this need is deep learning, understood here not as the machine-learning technique but as a pedagogical orientation that couples meaningful understanding, critical reflection, and authentic application of knowledge (Kovač et al., 2023; Quinn et al., 2021). The approach is often summarised through three interlocking principles: learning should be mindful (learners are actively and reflectively engaged), meaningful (content connects to learners' lived experience), and joyful (the process itself is intrinsically motivating) (Fullan & Langworthy, 2014; Quinn et al., 2021). Comparative and leadership-focused studies suggest that schools which successfully embed these principles report gains in students' capacity to analyse, evaluate, and transfer knowledge across contexts (Harris et al., 2024). Embedding deep learning within Indonesian secondary education is also consistent with the country's commitments under Sustainable Development Goal 4 on inclusive, quality education (Alm et al., 2022; Ambariyanto & Utama, 2020), and with national policy attention to strengthening twenty-first-century competencies among graduates.

Artificial intelligence (AI) has emerged as a widely discussed lever for operationalising such pedagogical ambitions at scale. Systematic reviews of AI applications in education report consistent, if uneven, gains in personalisation, engagement, and efficiency of understanding (Strielkowski et al., 2024; Zawacki-Richter et al., 2019), and generative AI tools in particular are reshaping how instructional content text, image, and multimodal narrative can be produced (Anantrasirichai & Bull, 2022; Antony & Huang, 2024). At the same time, national survey data indicate that a majority of Indonesian schools and teachers still lack adequate technological access and confidence to exploit these tools (Kemendikbudristek, 2023), which points to a practical need for concrete, classroom-ready applications of AI rather than abstract policy aspiration alone.

Educational comics sequential, narrative-driven visual media are a promising vehicle for this integration. Prior research shows that comics can support conceptual understanding, motivation, and critical thinking across science and language subjects, from primary to secondary and tertiary levels (Damopolii et al., 2021; Kurnia et al., 2020; Mahanal et al., 2023; Safitri et al., 2021). In physics education specifically, comic-based interventions have been linked to measurable

gains in students' critical-thinking scores and conceptual representation (Badeo & Ong Kian Koc, 2021; Maghfiroh & Kuswanto, 2022), while comic-supported instruction in language and mathematics has been associated with improved motivation and creative-thinking outcomes (Mariani et al., 2023; Pakpahan & Sinaga, 2023). The cognitive mechanism underlying these effects is thought to involve the integration of visual and verbal processing that narrative comics afford (Cohn, 2020), together with the emotional engagement and contextualisation that narrative more broadly provides for complex or discipline-specific content (Mawasi et al., 2021; Robin, 2008; Soares et al., 2023; Sun, 2022). Digital storytelling in science education, in particular, has been shown to help learners connect abstract mechanisms with everyday experience and to stimulate scientific curiosity (Teixeira et al., 2023; Zhang & Bowman, 2012, 2022).

Generative AI adds a further layer of relevance to comic-based instruction. Where earlier comic-based interventions relied on manual illustration, a process that is time- and skill-intensive for teachers AI image generators and conversational chatbots now allow instructional designers to iteratively prototype consistent characters, scenes, and visual styles from natural-language prompts (Anantrasirichai & Bull, 2022; Antony & Huang, 2024; Li & Ironsi, 2024). This lowers the barrier for teachers and researchers to produce discipline-specific narrative media tailored to locally identified curriculum difficulties, rather than relying on generic, off-the-shelf comics. However, the literature on AI-assisted (rather than AI-themed) comics as an instructional medium, particularly at the secondary-school level and validated through a structured research and development process, remains limited. Most existing comic-based studies employ manually drawn illustrations, and few report a complete validity–practicality–effectiveness evaluation cycle for AI-generated visual narratives.

This study addresses that gap by developing and evaluating an AI-assisted e-comic intended to strengthen critical thinking within a deep learning framework for senior high school students. The comic was built around curriculum topics that teachers across nine partner schools identified as both difficult to teach and under-served by existing media, cellular biology and neurodegenerative disease (framed through an Alzheimer's-disease narrative), environmental pollution and waste management (framed through an organic-waste-to-biogas narrative), and figurative language in Indonesian-language literature (framed through a forest-adventure narrative). Story content, prompt engineering for image generation, and layout were developed iteratively using a chatbot-based large language model and AI image-generation platforms, then refined through expert validation and classroom trial. The novelty of the study lies in (a) treating generative AI as a production tool for a validated instructional medium rather than as the object of study, (b) grounding comic content directly in a needs assessment across multiple schools rather than a single convenience sample, and (c) reporting a full R&D evaluation cycle, content and media validity, internal-consistency reliability, student-reported practicality, and pretest–posttest learning gains for the resulting product. Specifically, the study addresses three research questions: (1) How is an AI-assisted e-comic developed to strengthen critical thinking within a deep learning framework for senior high school students? (2) How valid is the resulting e-comic, in terms of content and media quality, as judged by expert validators? (3) How practical and effective is the e-comic when implemented in the classroom, as evidenced by student-reported practicality and pretest–posttest performance?

## 2. Materials and Methods

### 2.1. Research Design

This study used a research and development (R&D) design following the ADDIE model, Analysis, Design, Development, Implementation, and Evaluation (Branch, 2009) which is widely applied for developing and validating instructional products where the goal is a usable, evidence-based artefact rather than a generalisable causal claim (Sugiyono, 2019; Van den Akker et al., 2006). Each phase produced a

defined output and success indicator, summarised in Table 1.

**Table 1. ADDIE phases, activities, and outputs for e-comic development**

| Phase | Main activity | Output / success indicator |
|---|---|---|
| Analysis | Needs assessment via questionnaire distributed to teachers in nine partner schools; literature review on deep learning and AI-based media | Documented list of difficult-to-teach topics; needs-analysis report |
| Design | Story development with a chatbot-based large language model; storyboard and prompt design for character/scene consistency | Storyboards for three narratives; structured image-generation prompts |
| Development | Illustration generation with AI image platforms; layout design; expert content and media validation | Beta version of the e-comic; validated instrument scores |
| Implementation | Classroom trial in senior high schools; data collection via test, questionnaire, and observation | Field-trial dataset (test scores, questionnaire responses, qualitative feedback) |
| Evaluation | Analysis of validity, reliability, practicality, and learning outcomes; revision of the e-comic | Final e-comic version; evaluation report |

### 2.2. Needs Analysis and Story Development

A needs-assessment questionnaire was distributed to biology and Indonesian-language teachers across nine senior high schools in East Java, Indonesia. Teachers rated the difficulty of specific curriculum topics and the criteria they prioritised when selecting instructional media. In biology, teachers rated cellular concepts as the most difficult to teach (67% of respondents), followed by environmental pollution (65%), the transport system (55%), and the immune system (42%); in Indonesian-language instruction, poetry and figurative language were rated most difficult (75%), followed by scientific writing (65%), traditional narrative/hikayat (47%), proposal writing (45%), drama (39%), and debate (35%). When asked which criteria mattered most in selecting instructional media, teachers most frequently cited visual appeal and creativity (85.7%) and ease of use (78.6%), ahead of cost and accessibility (28.6–35.7%), indicating that pedagogical engagement, rather than logistical convenience, was the primary driver of media adoption in this sample. These findings directly informed the selection of an illustrated, narrative-driven medium and guided the three story lines developed for the e-comic: a story on cellular neurodegeneration modelled on Alzheimer's disease, chosen to integrate cell biology, the immune system, and scientific-writing skills; a story on converting organic waste into biogas, chosen to integrate environmental-pollution content with argumentative and debate skills; and a forest-adventure story built around figurative language, chosen to address the language-arts gap identified by teachers.

### 2.3. AI-Assisted Comic Production

Narrative structure and dialogue were drafted collaboratively with a chatbot-based large

language model, which was also used to translate each narrative beat into structured image-generation prompts. To maintain visual consistency across panels, prompts specified five recurring attributes for each character: gender, hair colour and style, facial expression, clothing colour, and overall art style (monochrome manga-influenced line art was used throughout). Illustrations were produced with AI image-generation platforms and composed into comic layouts using standard design software. Two versions of visual style (a Japanese manga-influenced style and a Western comic style) were prototyped before the manga-influenced style was selected for full production, based on informal preference feedback from the design team and partner teachers.

### 2.4. Participants

Expert validation involved subject-matter and media validators holding a minimum of Sinta-2-accredited (nationally accredited) publication records in learning media/technology, language, or science, consistent with the credentialing criterion used for validator selection. The classroom implementation and field-trial data reported here involved 30 senior high school students (two class groups) who used the e-comic during a limited classroom trial and completed the comprehension/critical-thinking test and the practicality questionnaire. Participation followed standard institutional research-ethics procedures for classroom-based educational research, including informed consent from participating schools, teachers, and students/guardians.

### 2.5. Instruments

Three instruments were used. First, a content-validation rubric assessed whether the comic represented essential concepts, stimulated higher-order thinking aligned with the analyse–evaluate–create levels of the revised Bloom's taxonomy (Anderson & Krathwohl, 2001), was grounded in authentic, student-relevant problems, prompted reflection and metacognition, and encouraged collaboration and discussion. Second, a media-validation rubric assessed pedagogical, technical, and aesthetic quality, informed by universal-design-for-learning principles regarding accessibility and multimodal representation (Meyer et al., 2014). Both rubrics used a four-point Likert scale (1 = not appropriate to 4 = very appropriate) completed by expert validators. Third, a 17-item student questionnaire measured perceived practicality of the e-comic (clarity of visuals and text, ease of access, engagement, narrative coherence, and support for understanding), also scored on a four-point Likert scale, together with two open-ended items asking students to describe what they liked and what could be improved. Fourth, a pretest–posttest instrument assessed students' comprehension of the story content and associated critical-thinking indicators (identifying problems, interpreting information, analysing arguments, evaluating and drawing conclusions, proposing solutions, reflecting on multiple perspectives, and formulating critical questions), administered before and after the classroom trial.

### 2.6. Data Analysis

Quantitative validation and practicality data were analysed descriptively, following the modal-score convention commonly used in Indonesian R&D studies, whereby an instrument or product is categorised as valid/practical when the aggregate percentage of favourable ratings exceeds 70% (Damopolii et al., 2021). Internal consistency of the 17-item questionnaire was estimated using Cronbach's alpha, interpreted according to conventional thresholds in which $\alpha \geq .70$ indicates acceptable reliability (Cronbach, 1951; Nunnally & Bernstein, 1994; Tavakol & Dennick, 2011). Pretest and posttest scores were compared using a paired-samples t-test, and learning gain was additionally expressed as a normalised gain, $\langle g \rangle$ = (posttest − pretest) / (100 − pretest), following the convention in which $\langle g \rangle \geq 0.7$ is classified as high, $0.3 \leq \langle g \rangle < 0.7$ as moderate, and $\langle g \rangle < 0.3$ as low (Hake, 1998). Effect size was reported as Cohen's d for paired samples. Open-ended qualitative responses (student-reported strengths and limitations) were analysed using conventional qualitative content analysis, involving coding of responses, identification of

recurring themes, and narrative synthesis (Sugiyono, 2019).

## 3. Results

### 3.1. Content and Media Validity

Expert validators rated all five content-validity indicators at the maximum score of 4 out of 4 ('very appropriate'), summarised in Table 2. The comic was judged to represent core curriculum concepts clearly, to stimulate analysis, evaluation, and synthesis consistent with higher-order cognitive processes, to be grounded in problems relevant to students' lives, to stimulate critical questioning and self-reflection, and to encourage collaboration and discussion among students.

Table 2. Expert validation results for content quality

| No. | Content-validity indicator | Score (max 4) |
|---|---|---|
| 1 | Represents core curriculum concepts/topics | 4.00 |
| 2 | Stimulates analysis, evaluation, or synthesis (Bloom's C4–C6) | 4.00 |
| 3 | Grounded in authentic problems relevant to students' lives | 4.00 |
| 4 | Stimulates critical questioning and self-reflection | 4.00 |
| 5 | Encourages interaction, collaboration, or discussion | 4.00 |

Media-quality validation, summarised in Table 3, produced a mean score of 3.67 out of 4 across nine indicators spanning pedagogical, technical, and aesthetic dimensions. All three technical indicators (readability of images/text, ease of access, and cross-device compatibility) received the maximum score, while two pedagogical indicators (curriculum-level alignment and stimulation of active engagement) and one aesthetic indicator (age-appropriate visual appeal) received a score of 3, indicating scope for further refinement in these areas.

Table 3. Expert validation results for media quality (M = 3.67/4)

| Aspect | Indicator | Score (max 4) |
|---|---|---|
| Pedagogical | Presents clear and directed learning objectives | 4 |
| Pedagogical | Content aligned with curriculum and students' developmental level | 3 |
| Pedagogical | Stimulates active student engagement in learning | 3 |
| Technical | Image and text quality is clearly readable | 4 |
| Technical | Easily accessible to teachers and students | 4 |
| Technical | Usable across different devices/media | 4 |
| Aesthetic | Visual appearance is engaging and age-appropriate | 3 |
| Aesthetic | Storyline is coherent and easy to follow | 4 |

| Aspect | Indicator | Score (max 4) |
|---|---|---|
| Aesthetic | Colour, font, and visual elements support comprehension | 4 |

### 3.2. Instrument Reliability

The 17-item student practicality questionnaire (reduced to 13 items after excluding four items with weak item-rest correlation, retaining Items 1–9, 12, and 14–16) demonstrated good internal consistency, Cronbach's α = 0.82. Item-rest correlations ranged from 0.32 to 0.65, and dropping any individual item would not have meaningfully improved the overall coefficient (all 'if item dropped' α values remained between 0.79 and 0.82), indicating an internally coherent instrument (Table 4).

Table 4. Item reliability statistics for the practicality questionnaire (scale Cronbach's α = 0.817)

| Item | M | SD | Item-rest r | α if dropped |
|---|---|---|---|---|
| Item 1 | 3.07 | 0.25 | 0.40 | 0.811 |
| Item 2 | 3.17 | 0.38 | 0.54 | 0.800 |
| Item 3 | 3.20 | 0.41 | 0.54 | 0.799 |
| Item 4 | 3.13 | 0.43 | 0.35 | 0.813 |
| Item 5 | 3.17 | 0.38 | 0.64 | 0.792 |
| Item 6 | 3.23 | 0.43 | 0.65 | 0.790 |
| Item 7 | 3.20 | 0.41 | 0.33 | 0.814 |
| Item 8 | 3.10 | 0.48 | 0.34 | 0.815 |
| Item 9 | 3.23 | 0.43 | 0.44 | 0.807 |
| Item 12 | 3.27 | 0.45 | 0.32 | 0.816 |
| Item 14 | 3.53 | 0.51 | 0.42 | 0.809 |
| Item 15 | 3.43 | 0.50 | 0.62 | 0.790 |
| Item 16 | 3.37 | 0.49 | 0.44 | 0.807 |

### 3.3. Student-Reported Practicality

Across all 17 items and 30 respondents, the questionnaire produced an overall mean of 3.16 (SD = 0.53) on the four-point scale, equivalent to 78.9% of the maximum possible score, which exceeds the 70% threshold conventionally used to classify a medium as practical (Damopolii et al., 2021). The modal response was 3 ('appropriate/practical') for 16 of the 17 items

and 3–4 (bimodal) for one item; no item had a modal response below 3. Item-level means ranged from 2.97 to 3.53, with the highest-rated items concerning the clarity of storyline and the accessibility of language, and the lowest-rated (though still above the neutral midpoint) concerning visual style preference, consistent with the qualitative feedback discussed below.

### 3.4. Learning Outcomes: Pretest–Posttest Comparison

Students' comprehension and critical-thinking scores improved substantially after using the e-comic (Table 5). The mean pretest score was 46.40 (SD = 12.61, range 23–70) and the mean posttest score was 82.10 (SD = 8.54, range 65–98). A paired-samples t-test indicated that this improvement was statistically significant, $t(29) = 17.69$, $p < .001$, with a very large effect size (Cohen's $d = 3.31$). The normalised gain, $\langle g \rangle = 0.67$, falls at the upper boundary of the moderate-gain category (Hake, 1998), indicating a substantial improvement relative to the maximum possible gain given each student's starting score.

Table 5. Descriptive statistics for pretest and posttest scores

| Measure | N | M | SD | Min | Max |
|---|---|---|---|---|---|
| Pretest | 30 | 46.40 | 12.61 | 23 | 70 |
| Posttest | 30 | 82.10 | 8.54 | 65 | 98 |

At the level of specific critical-thinking indicators assessed during implementation (Table 6), students scored highest on formulating critical questions (M = 4.3/5) and identifying the central problem in the narrative (M = 4.2/5), and lowest—though still above the midpoint—on reflecting on personal versus alternative perspectives (M = 3.7/5) and evaluating information to draw logical conclusions (M = 3.8/5).

Table 6. Mean scores on classroom-observed critical-thinking indicators

| No. | Critical-thinking indicator | M (1–5) |
|---|---|---|
| 1 | Identifying the central problem | 4.2 |
| 2 | Interpreting information and relating it to prior knowledge | 4.1 |
| 3 | Analysing arguments or ideas | 3.9 |
| 4 | Evaluating and drawing logical conclusions | 3.8 |
| 5 | Proposing alternative solutions to the problem faced by characters | 4.0 |
| 6 | Reflecting on personal views and alternative perspectives | 3.7 |
| 7 | Formulating critical questions from the narrative | 4.3 |

### 3.5. Qualitative Student Feedback

Open-ended responses converged on two dominant themes of appreciation: the clarity and

accessibility of the explanations embedded in the narrative, and the visual appeal of the illustrations. Several students specifically noted that the story format made an otherwise difficult topic (Alzheimer's disease and its cellular mechanisms) easier to follow than a conventional text explanation, and that the visuals helped them connect the narrative to the underlying science. A smaller number of responses referenced the novelty of an AI-assisted production process as a point of interest in itself. Suggestions for improvement clustered around three areas: requests for additional chapters or a longer story arc; requests for colour illustration, since the published version used monochrome line art; and, from a minority of respondents, a preference for less visibly AI-generated artwork or a more locally grounded (rather than manga-influenced) visual style. These patterns are broadly consistent with the quantitative pattern in Section 3.1, where the aesthetic indicator on age-appropriate visual appeal received the lowest (though still acceptable) expert score.

## 4. Discussion

### 4.1. Development Process

The ADDIE-based development process anchored the e-comic's content directly in a documented instructional need curriculum topics that teachers across multiple schools independently rated as difficult rather than in researcher assumption. This is consistent with recommendations in the instructional-design literature that effective media development begin from a systematic needs analysis (Branch, 2009; Van den Akker et al., 2006). The use of a chatbot-based large language model for narrative and prompt design, paired with AI image-generation platforms for character-consistent illustration, reflects a broader shift in how generative AI is being incorporated into instructional-design workflows (Anantrasirichai & Bull, 2022; Antony & Huang, 2024; Li & Ironsi, 2024). Framing difficult, discipline-specific content within narrative also aligns with the theoretical rationale for digital storytelling in science and language education, which holds that narrative structure supports emotional engagement and simplifies complex mechanisms by anchoring them to relatable events (Mawasi et al., 2021; Robin, 2008; Sun, 2022; Zhang & Bowman, 2022).

### 4.2. Validity

The maximum content-validity scores obtained here indicate that expert validators judged the narrative to successfully embed higher-order cognitive demand (Bloom's C4–C6; Anderson & Krathwohl, 2001) within an authentic, student-relevant problem context, consistent with the deep learning principle that meaningful learning requires grounding in real-world problems (Fullan & Langworthy, 2014; Quinn et al., 2021). The strong technical-validity scores full marks on readability, accessibility, and cross-device compatibility support the practical feasibility of e-comics as classroom-ready media, consistent with universal-design-for-learning guidance on multimodal, flexible access to content (Meyer et al., 2014). The comparatively lower (though still acceptable) scores on curriculum alignment, active-engagement stimulation, and age-appropriate visual appeal converge with the qualitative feedback reported in Section 3.5, and point to two concrete refinement priorities: tighter mapping of narrative beats onto specific curriculum indicators, and diversification of visual style for example, offering a colour version or a less overtly 'AI-generated' aesthetic to address the minority of students who expressed a preference in this direction.

### 4.3. Practicality and Reliability

The practicality questionnaire's internal consistency ($\alpha = 0.82$) falls within the range conventionally described as good reliability (Nunnally & Bernstein, 1994; Tavakol & Dennick, 2011), and the exclusion of four weak-loading items during instrument refinement illustrates the routine trade-off between scale length and internal consistency that alpha estimation is known to be sensitive to (Cronbach, 1951; Tavakol & Dennick, 2011). Substantively, an overall practicality score of 78.9% of the maximum, comfortably above the 70% threshold used in prior Indonesian R&D studies (Damopolii et al., 2021), together with a predominant modal rating of 'practical' across

items, supports the conclusion that students found the e-comic usable and engaging in its intended classroom context. This finding parallels prior comic-based interventions in physics and mathematics education that similarly report favourable student practicality ratings alongside measurable learning gains (Badeo & Ong Kian Koc, 2021; Maghfiroh & Kuswanto, 2022; Pakpahan & Sinaga, 2023).

### 4.4. Learning Outcomes

The statistically significant, large-magnitude improvement from pretest to posttest ($d = 3.31$), together with a normalised gain at the upper end of the moderate range ($\langle g \rangle = 0.67$), indicates a substantial classroom effect associated with e-comic use, though the absence of a control or comparison group means this result should be interpreted as evidence of pre–post improvement under the intervention rather than as a causal claim isolating the e-comic's unique contribution relative to standard instruction. The item-level pattern strongest performance on identifying problems and formulating critical questions, comparatively weaker performance on reflective evaluation and drawing logical conclusions mirrors findings from other comic-based critical-thinking interventions, in which narrative and problem-identification skills tend to develop more readily than evaluative and reflective skills within a single instructional cycle (Kurnia et al., 2020; Mahanal et al., 2023). This suggests that a single exposure to the e-comic may be sufficient to strengthen surface-level engagement and question-generation, while deeper evaluative and reflective competencies may require structured follow-up activities for instance, open-ended post-reading discussion prompts or structured debate, both of which the source narratives were designed to support but which were not systematically implemented during this limited trial.

### 4.5. Limitations

This study has several limitations. First, the field trial involved a relatively small sample ($n = 30$) from a limited number of classrooms, and the pretest–posttest design lacked a control group, which restricts causal inference regarding the e-comic's effect relative to conventional instruction. Second, the practicality questionnaire's four excluded items suggest that the original 17-item instrument may have included constructs not fully aligned with the retained 13-item structure; future work should consider a confirmatory factor analysis to better establish the instrument's dimensionality. Third, expert validation involved a small number of validators, consistent with common R&D practice but limiting the generalisability of validity scores. Finally, because the e-comic was produced with generative AI tools, questions of visual-style bias, character representativeness, and the pedagogical implications of AI-authored narrative content warrant further, dedicated investigation beyond the scope of this study.

## 5. Conclusion

This study developed and evaluated an AI-assisted e-comic intended to strengthen critical thinking within a deep learning framework for Indonesian senior high school students, following a systematic ADDIE-based research and development process grounded in a multi-school needs analysis. The resulting product achieved maximum expert-rated content validity, strong media validity ($M = 3.67/4$), good instrument reliability ($\alpha = 0.82$), and high student-reported practicality (78.9% of the maximum scale score). Classroom implementation was associated with a statistically significant and large improvement in comprehension and critical-thinking scores from pretest to posttest ($d = 3.31$, $\langle g \rangle = 0.67$). These findings support AI-assisted e-comics as a valid, reliable, and practical instructional medium for engaging students with conceptually difficult, discipline-specific content, while also identifying concrete refinement priorities visual-style diversification, tighter curriculum alignment, and structured post-reading reflection activities for future iterations. Future research should extend this work with a controlled comparison design, a larger and more diverse sample, and closer examination of how specific AI-generation choices (style, representativeness, narrative structure) shape both learning outcomes and student reception.

## References

Alm, K., Beery, T. H., Eiblmeier, D., & Fahmy, T. (2022). Students' learning sustainability – implicit, explicit or non-existent: A case study approach on students' key competencies addressing the SDGs in HEI program. International Journal of Sustainability in Higher Education, 23(8), 60–84. https://doi.org/10.1108/IJSHE-12-2020-0484

Ambariyanto, A., & Utama, Y. J. (2020). Educating higher education institutions to support SDGs: Indonesian case. E3S Web of Conferences, 202, 02015. https://doi.org/10.1051/e3sconf/202020202015

Anantrasirichai, N., & Bull, D. (2022). Artificial intelligence in the creative industries: A review. Artificial Intelligence Review, 55(1), 589–656. https://doi.org/10.1007/s10462-021-10039-7

Anderson, L. W., & Krathwohl, D. R. (Eds.). (2001). A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives. Longman.

Antony, V. N., & Huang, C. M. (2024). Id.8: Co-creating visual stories with generative AI. ACM Transactions on Interactive Intelligent Systems, 14(3), 1–29.

Badeo, J. M. O., & Ong Kian Koc, B. C. U. (2021). Use of comic-based learning module in physics in enhancing students' achievement and motivation. Science Education International, 32(2), 131–136. https://doi.org/10.33828/sei.v32.i2.6

Bezemer, J., & Kress, G. (2016). Multimodality, learning and communication: A social semiotic frame. Routledge.

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

Cohn, N. (2020). Your brain on comics: A cognitive model of visual narrative comprehension. Topics in Cognitive Science, 12(1), 352–386. https://doi.org/10.1111/tops.12421

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. Psychometrika, 16(3), 297–334. https://doi.org/10.1007/BF02310555

Damopolii, I., Lumembang, T., & İlhan, G. O. (2021). Digital comics in online learning during COVID-19: Its effect on student cognitive learning outcomes. International Journal of Interactive Mobile Technologies, 15(19), 33–47. https://doi.org/10.3991/ijim.v15i19.23395

Daryanto, Agung, I., & Siswantari. (2020). MGMP teacher organization empowerment in improving students' problem solving ability. Journal of Educational and Social Research, 10(1), 152–160. https://doi.org/10.36941/jesr-2020-0014

Facione, P. A. (2015). Critical thinking: What it is and why it counts. Insight Assessment.

Fullan, M., & Langworthy, M. (2014). A rich seam: How new pedagogies find deep learning. Pearson.

Hake, R. R. (1998). Interactive-engagement versus traditional methods: A six-thousand-student survey of mechanics test data for introductory physics courses. American Journal of Physics, 66(1), 64–74. https://doi.org/10.1119/1.18809

Harris, A., Jones, M., Azorín, C., Southern, A., Griffiths, J., & Ástvaldsdóttir, I. (2024). Leading for deeper learning: International perspectives. Journal of Educational Administration, 62(1), 122–137. https://doi.org/10.1108/JEA-02-2023-0042

Kemendikbudristek. (2023). Laporan nasional: Digitalisasi pendidikan di Indonesia. Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi.

Kovač, V. B., Nome, D. Ø., Jensen, A. R., & Skreland, L. Lj. (2023). The why, what and how of deep learning: Critical analysis and additional concerns. Educational Inquiry, 15(3), 1–17. https://doi.org/10.1080/20004508.2023.2194502

Kurnia, Y. P., Sumarti, S. S., & Utomo, U. (2020). Implementation of problem-based learning assisted with science comic books to improve critical thinking skill of elementary students. Journal of Primary Education, 9(2), 186–192.

Li, Z., & Ironsi, C. S. (2024). The efficacy of generative artificial intelligence in developing science education preservice teachers' writing skills: An experimental approach. Journal of Science Education and Technology. Advance online publication. https://doi.org/10.1007/s10956-024-10148-2

Lin, S.-F., Lin, H., Lee, L., & Yore, L. D. (2015). Are science comics a good medium for science communication? The case for public learning of nanotechnology. International Journal of Science Education, Part B, 5(3), 276–294. https://doi.org/10.1080/21548455.2014.941040

Maghfiroh, A., & Kuswanto, H. (2022). Benthik android physics comic effectiveness for vector representation and critical thinking students' improvement. International Journal of Instruction, 15(2), 623–640. https://doi.org/10.29333/iji.2022.15234a

Mahanal, S., Zubaidah, S., & Suarsini, E. (2023). A systematic review of learning media to improve critical thinking skills in science education. Journal of Baltic Science Education, 22(1), 60–75. https://doi.org/10.33225/jbse/23.22.60

Mariani, R., Novita, R., & Sari, S. M. (2023). Pemanfaatan komik sebagai media pembelajaran Bahasa Indonesia dalam meningkatkan motivasi dan hasil belajar siswa sekolah dasar. Indo-MathEdu Intellectual Journal, 4(3), 2881–2887. https://doi.org/10.54373/imeij.v4i3.664

Mawasi, A., Nagy, P., Finn, E., & Wylie, R. (2021). Narrative-based learning activities for science ethics education: An affordance perspective. Journal of Science Education and Technology, 31(1), 16–26. https://doi.org/10.1007/s10956-021-09928-x

Meyer, A., Rose, D. H., & Gordon, D. (2014). Universal design for learning: Theory and practice. CAST Professional Publishing.

Nunnally, J. C., & Bernstein, I. H. (1994). Psychometric theory (3rd ed.). McGraw-Hill.

OECD. (2018). Preparing our youth for an inclusive and sustainable world: The OECD PISA global competence framework. OECD Publishing.

OECD. (2019). PISA 2018 results: What students know and can do (Vol. I). OECD Publishing.

OECD. (2023). PISA 2022 results (Volume I and II): Country notes – Indonesia. OECD Publishing.

Pakpahan, E. I. S., & Sinaga, B. (2023). The application of comic-based learning assisted by Pixton to improve mathematical creative thinking skill of class IX SMPN 8 Percut Sei Tuan. Asian Journal of Applied Education, 2(3), 471–480. https://doi.org/10.55927/ajae.v2i3.5004

Prasetyo, R. A., & Kuswardani, R. (2021). Developing students' reading ability using visual novel for high school students. Journal of Research on English and Language Learning, 2(2), 132–139. https://doi.org/10.33474/j-reall.v2i2.11336

Quinn, J., Gardner, M., Drummy, M., & Fullan, M. (2021). Ten priorities to activate deep learning and lift from loss. New Pedagogies for Deep Learning.

Robin, B. R. (2008). Digital storytelling: A powerful technology tool for the 21st century classroom. Theory Into Practice, 47(3), 220–228. https://doi.org/10.1080/00405840802153916

Safitri, Y., Mailizar, M., & Johar, R. (2021). Students' perceptions of using e-comics as a media in mathematics learning. Infinity Journal, 10(2), 319–330. https://doi.org/10.22460/infinity.v10i2.p319-330

Soares, S., Gonçalves, M., Jerónimo, R., & Kolinsky, R. (2023). Narrating science: Can it benefit science learning, and how? A theoretical review. Journal of Research in Science Teaching, 60(9), 2042–2075. https://doi.org/10.1002/tea.21851

Strielkowski, W., Grebennikova, V., Lisovskiy, A., Rakhimova, G., & Vasileva, T. (2024). AI-driven adaptive learning for sustainable educational transformation. Sustainable Development. Advance online publication. https://doi.org/10.1002/sd.3221

Sugiyono. (2019). Metode penelitian kuantitatif, kualitatif, dan R&D. Alfabeta.

Sun, L. (2022). Integrating digital storytelling into science and mathematics instruction. School Science and Mathematics, 122(8), 387–389. https://doi.org/10.1111/ssm.12560

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. International Journal of Medical Education, 2, 53–55. https://doi.org/10.5116/ijme.4dfb.8dfd

Teixeira, I., Lima, D., & Vasconcelos, C. (2023). Storytelling as an interdisciplinary strategy in geoscience education. Proceedings of INTED2023, 1–9. https://doi.org/10.36315/2023v2end117

Van den Akker, J., Gravemeijer, K., McKenney, S., & Nieveen, N. (Eds.). (2006). Educational design research. Routledge.

Wei, X., Wang, L., Lee, L. K., & Liu, R. (2025). The effects of GenAI on collaborative problem-solving and team creativity performance in digital story creation: An experimental study. International Journal of Educational Technology in Higher Education, 22(1), Article 23.

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education – where are the educators? International Journal of Educational Technology in Higher Education, 16(1), 39. https://doi.org/10.1186/s41239-019-0171-0

Zhang, J., & Bowman, M. A. (2012). Using comics to teach difficult concepts in science. Science Scope, 35(8), 38–43.

Zhang, L., & Bowman, D. (2022). Exploring effect of level of storytelling richness on science learning in interactive and immersive virtual reality. In Proceedings of the ACM International Conference on Interactive Media Experiences (pp. 19–32). ACM. https://doi.org/10.1145/3505284.3529960