

Article

Designing GenAI-Mediated Mathematics Storybooks: Process Patterns and Tool Affordance Uptake Among Pre-Service Early Childhood Teachers

Yau Yu Chan ¹  and Kwok Man Keith Ho ^{2,*} ¹ Faculty of Education, The University of Hong Kong, Hong Kong; chanyaayu2@gmail.com² School of Education and Languages, Hong Kong Metropolitan University, Hong Kong

* Correspondence: kho@hkmu.edu.hk

Abstract

Digital storytelling is widely used in education as a multimodal design practice that foregrounds creator agency and audience awareness. Generative AI (GenAI) expands the speed and range of story production but also raises questions about how educators retain pedagogical control over generated content. This exploratory qualitative study used cross-case workflow analysis to examine how 27 pre-service early childhood educators in Hong Kong designed GenAI-mediated mathematics storybooks. Data included video-recorded process presentations, records of interactions with ChatGPT (GPT-5) and Krea 2, and completed storybooks. Analysis traced participant decisions across planning, generation, curation, and integration, focusing on GenAI-related affordance uptake, recurring workflow patterns, and tensions among technical innovation, ethical accountability, and cultural impact. Findings show that participants commonly used GenAI to steer generation through pedagogical constraints; expand narrative and visual alternatives; and select, reject, or revise generated materials. Two recurring process patterns were identified: constraint-guided planning-to-generation, in which pedagogical requirements were established early and used to guide subsequent decisions, and exploration-to-selection, in which participants generated and compared alternatives before progressively stabilizing a preferred direction. Across both patterns, tensions were particularly evident during curation and integration, when participants balanced efficiency with sustained oversight, creative variation with pedagogical coherence, and generic outputs with local contextual fit. The findings show that GenAI-mediated storybook creation is not simply a process of generating text and images, but a pedagogical workflow in which pre-service teachers exercise agency through the critical selection, revision, and integration of AI-generated materials. Implications are discussed for teacher education, AI literacy, and the design of scalable support for teachers' use of GenAI.



Academic Editors: José
Fernández-Cerero and Marta
Montenegro-Rueda

Received: 29 May 2026

Revised: 16 July 2026

Accepted: 16 July 2026

Published: 21 July 2026

Copyright: © 2026 by the authors.
Licensee MDPI, Basel, Switzerland.
This article is an open access article
distributed under the terms and
conditions of the [Creative Commons
Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

Keywords: generative AI; digital storytelling; pre-service teachers; teacher education

1. Introduction

Digital storytelling has received growing research attention in recent decades. With the emergence of generative AI (GenAI), producing digital stories has become more accessible, as GenAI can generate multimodal resources with minimal specialized equipment or technical expertise. However, these lowered production barriers also intensify concerns central to a design- and agency-oriented understanding of AI-mediated storytelling. This

is because AI systems can shift or obscure authorial control, encourage uncritical acceptance of machine-generated outputs, and introduce less visible yet consequential influences on narrative coherence, representation, and meaning (Amershi et al., 2019; Long & Magerko, 2020). In educational settings, these issues are particularly salient because teachers remain accountable for the pedagogical intent, developmental appropriateness, and cultural implications of the stories they bring into classrooms, even when parts of the content are AI-generated (Cukurova & Miao, 2024).

Empirical research on how educators use GenAI in storytelling remains limited, partly because the field is emerging and classroom practices are evolving rapidly. Understanding AI-related affordance uses and design patterns is important because it clarifies how creators coordinate human and AI initiatives across the stages of story design, what kinds of agency are retained or delegated, and what trade-offs are made between efficiency and ethical judgement. Such evidence can inform teacher education by identifying the capabilities and support teachers need to use GenAI responsibly.

We focus on pre-service early childhood educators in Hong Kong because storytelling is a common and curriculum-relevant pedagogical practice for supporting young children's oral language, emergent literacy, and social development (Nicolopoulou et al., 2015). It is also a high-stakes domain where issues of representation, language appropriateness, and child development are especially consequential (Gay, 2018). By examining how pre-service early childhood educators use GenAI tools to design mathematics storybooks for young learners, this study contributes an empirically grounded account of workflow patterns in AI-mediated storytelling, the GenAI-related tool affordances participants took up, and the pedagogical considerations that shaped their design decisions. It also offers implications for teacher education, including ways to scaffold storybook creation, strengthen pre-service teachers' pedagogical judgement and agency, and support critical evaluation of AI-generated materials before they are used as classroom-ready resources.

2. Digital Storytelling in the GenAI Era

Digital storytelling generally refers to the use of digital multimodal resources, such as images, audio, and video, to develop and communicate narratives (Robin, 2006). Its educational value lies not simply in the use of digital media, but in the design work it requires. Creators must select and coordinate modes, maintain narrative coherence, and anticipate how audiences will interpret the resulting story. Accordingly, research on digital storytelling and multimodal literacy emphasizes creator agency and the purposeful orchestration of resources in meaning-making (Petit, 2020).

In educational contexts, digital storytelling has been used to support multimodal and digital literacy as well as reflective learning. Studies have examined its potential to strengthen learners' digital literacy through multimedia production and information work (Chan et al., 2017; Churchill, 2020) and have highlighted its integration of narrative writing, design choices, and audiovisual composition (Wu & Chen, 2020). It is therefore often associated with student-centred learning designs that promote engagement, creativity, and reflective meaning-making (Bilici & Yilmaz, 2024; Staley & Freeman, 2017).

Wu and Chen's (2020) synthesis of 57 empirical studies further demonstrates that educational digital storytelling is not a single pedagogical form. They identified five orientations: appropriative, agentive, reflective, reconstructive, and reflexive. These orientations range from using story production to apply a learned concept to supporting learner autonomy, reflection, critical re-examination of assumptions, and identity work. Collectively, this literature positions digital storytelling as a designed and potentially value-laden meaning-making practice in which creators make consequential decisions about narrative, representation, and audience.

Recent advances in GenAI make this design- and agency-oriented understanding especially important. GenAI tools can generate textual, visual, and auditory materials, enabling creators to produce story elements with less specialized technical expertise (Colton & Wiggins, 2012; Nounanaki, 2025). In storybook creation, these tools may support brainstorming, script generation, image generation, revision, and multimodal assembly. However, they also introduce new concerns about the opacity of generated outputs, the distribution of authorial control, the accuracy and appropriateness of content, and the cultural assumptions embedded in generated narratives and images. The key issue is therefore not whether GenAI is used, but how creators utilize its possibilities and exercise judgement throughout the production process. Addressing this question requires distinguishing between the action possibilities that GenAI tools make available and the recurring ways creators organize those possibilities across the storybook-creation workflow.

3. Tool Affordances and Design Patterns for GenAI-Mediated Stories

In this study, tool affordances refer to the action possibilities made available through the relationship between a technology's functions and users' goals, knowledge, and capabilities (Nagy & Neff, 2015). For example, GenAI tools may enable users to specify constraints, generate alternatives, select or recombine outputs, revise content, and assemble textual and visual materials. Prompting, selecting, revising, and integrating are the enacted practices through which users actualize these affordances. Design patterns, by contrast, refer to recurring ways in which these practices are organized and sequenced over time across the stages of storybook design (Tsiakas & Murray-Rust, 2024). They not only capture what users can do with a tool, but also how they combine and coordinate available affordances throughout the storybook-creation workflow.

Individual interaction practices do not, by themselves, explain how users organize initiative and oversight across an extended creative task. Pattern-oriented human–AI interaction research offers a useful vocabulary for examining these processes. Luera et al. (2024) identified recurring user-guided practices, including prompting, selecting generated alternatives, manipulating system behavior, and transforming generated objects. Extending this interaction-level perspective, Lin et al. (2023) mixed-initiative co-creativity framework characterizes human–AI collaboration through configurations of control, including who initiates an action, whether interaction elaborates new possibilities or reflects on prior outputs, and whether attention is directed to the overall artefact or particular components. Evidence from AI-mediated picture-book creation suggests that these workflow-level processes are especially important in storytelling contexts. Li and Chuprina (2024) described iterative processes of stimulation, perception, resonance, and feedback, through which users generated, compared, selected, assembled, and refined story and image materials. They further showed that technically plausible audiovisual outputs may not produce meaningful narrative or emotional coherence without substantial human input. Together, these studies suggest that GenAI-supported creation involves patterned cycles of generation, evaluation, selection, and integration rather than isolated prompts or outputs.

While affordances and workflow patterns explain the functions of GenAI tools and how users organize their practices over time, they do not fully explain how users evaluate the technical, ethical, and cultural implications of their decisions. To address this issue, the present study draws on the Model of AI-Mediated Storytelling (AIMS) proposed by Shahamat (2023) in an unpublished preprint. AIMS conceptualizes AI-mediated storytelling through three interrelated dimensions: technical innovation, ethical accountability, and cultural impact. In this study, technical innovation refers to how participants used available GenAI functions to extend, accelerate, or reorganize their design work. Ethical accountability refers to how participants checked, justified, revised, or assumed responsi-

bility for the accuracy, developmental appropriateness, pedagogical suitability, and consequences of generated content. Cultural impact refers to how participants considered, retained, adapted, or revised cultural, linguistic, and representational elements in their storybooks. Importantly, these AIMS dimensions provide evaluative dimensions through which participants' uptake of tool affordances and workflow practices can be examined.

Building on these perspectives, this study examines how pre-service early childhood educators use GenAI to create mathematics storybooks. As this area remains underexplored, the study is guided by the following exploratory research questions rather than hypothesis-testing questions:

1. What GenAI-related affordances do pre-service teachers report taking up across the stages of planning, generation, curation, and integration?
2. What recurring process patterns characterize pre-service teachers' GenAI-mediated storybook creation across the stages of planning, generation, curation, and integration?
3. How do the identified process patterns and affordance choices surface and negotiate tensions among technical innovation, ethical accountability, and cultural impact across the stages of storybook creation?

4. Methodology

4.1. Research Design

This study adopted an exploratory qualitative design (Stebbins, 2025) to investigate the underexplored design processes through which pre-service early childhood educators use GenAI to create mathematics storybooks. This study examined the GenAI-related affordances participants reported, recurring workflow patterns across storybook creation, and how these practices related to technical innovation, ethical accountability, and cultural impact. To examine how these design processes unfolded over time, we used cross-case workflow analysis, which is a process-oriented analytic strategy that combines within-case reconstruction of each participant's design trajectory with cross-case comparison of recurring workflow configurations (Miles et al., 2014). Each participant's design process was treated as an analytical case, drawing on their interaction logs with GenAI tools, video-recorded process presentations, and storybook design artefacts. We traced decisions across planning, generation, curation, and integration within each case before comparing cases to identify the affordance of GenAI uptake, recurring workflow patterns, and points at which technical innovation, ethical accountability, and cultural impact were negotiated.

4.2. Participants

We used criterion-based purposive sampling within a convenience sample to recruit information-rich participants with direct experience of designing GenAI-mediated mathematics storybooks. The convenience sample comprised students enrolled in an accessible university-level early childhood mathematics course that included an individual GenAI-supported storybook-design assignment. Within this cohort, participants were purposively included because their completed design activities and associated materials could provide detailed evidence of the GenAI-mediated storybook-creation process. Eligibility criteria were: (a) enrolment in the Bachelor of Education (Early Childhood Education) programme; (b) participation in course workshops regarding designing storybooks using GenAI tools; (c) completion of the required individual storybook-design assignment; (d) submission of the required GenAI interaction logs and design materials; (e) completion of a process presentation; and (f) provision of informed consent for the use of their data in this study.

At the beginning of the course, students were informed about the study through an in-class announcement and were invited to participate voluntarily. Participation or non-

participation did not affect course assessment or grades. Ethical approval was obtained from the Human Research Ethics Committee of the corresponding author's university before data collection. No incentives were provided, and all data were anonymized for analysis and reporting. Twenty-seven students who met the eligibility criteria provided consent and contributed a complete set of design materials. All participants were female, reflecting the programme's gender composition. Eighteen participants were aged 18–20, while the remaining nine participants (33.33%) were aged 21–25. Table 1 presents the characteristics of participants' GenAI-mediated mathematics storybooks, including the intended learner age group and target mathematics concept.

Table 1. Characteristics of Participants' GenAI-Mediated Mathematics Storybooks.

Participant ID	Intended Learner Age Group	Target Mathematics Concept
P001	K3	Recognizing coins and values of each coin
P002	K1	Counting
P003	K1	Counting
P004	K3	Simple addition and subtraction
P005	K3	Recognizing coins and values of each coin
P006	K1	Comparing quantities using “more”, “less”, and “equal”
P007	K3	Recognising three-dimensional shapes
P008	K2	Counting from 1 to 20
P009	K2	Comparing object lengths
P010	K3	Telling time to the hour using an analogue clock
P011	K1	Counting
P012	K1	Recognizing shapes
P013	K2	Comparing quantities using “more”, “less”, and “equal”
P014	K3	Telling time to the hour using an analogue clock
P015	K3	Understanding and using positional language: above and below
P016	K3	Understanding and identifying the months of the year
P017	K1	Recognizing colors and shapes
P018	K3	Measuring and comparing the lengths of objects using non-standard units
P019	K1	Counting
P020	K3	Measuring and comparing the lengths of objects using non-standard units
P021	K1	Recognizing shapes
P022	K2	Recognizing colors and shapes
P023	K2	Recognizing coins and values of each coin
P024	K3	Comparing quantities using “more”, “less”, and “equal”
P025	K3	Ordering numbers from greatest to smallest
P026	K1	Comparing and describing the relative sizes of shapes
P027	K2	Comparing quantities using “more”, “less”, and “equal”

Note. K1, K2, and K3 refer to the first, second, and third years of kindergarten in Hong Kong, respectively. K1 serves children aged 3–4 years, K2 serves those aged 4–5 years, and K3 serves those aged 5–6 years.

Sample adequacy was related to the study's focused aim and the depth of multimodal data available for each participant. Consistent with the concept of information power (Malterud et al., 2016), each participant contributed a distinct design trajectory supported by GenAI interaction logs, a video-recorded process presentation, and storybook design artefacts. These materials enabled detailed within-case reconstruction of decisions across planning, generation, curation, and integration, followed by systematic cross-case comparison. The inclusion of 27 students was therefore considered adequate to identify recurring workflow configurations and GenAI-affordance use patterns, while also attending to variations and discrepant cases across the storybook-creation process (Malterud et al., 2016).

4.3. Procedures

Participants were enrolled in a university-level course on teaching mathematics in early childhood education. At the outset of the course, students received an individual take-home assignment to design a classroom-ready mathematics storybook for possible use in an early childhood teaching practicum with GenAI assistance. The assignment required students to select an age-appropriate mathematical concept, develop a coherent story plot and script, create corresponding visual elements, and assemble the final storybook using Book Creator. Students were also required to document their GenAI-supported design process by retaining records of one complete GenAI interaction for text-related brainstorming and generation and one for image-related brainstorming and generation. Assessment criteria addressed the appropriateness and accuracy of the mathematical content, narrative coherence, quality and consistency of visual design, integration of textual and visual elements, and the reflective use of GenAI tools.

Over four weekly workshops, the instructor guided participants through a structured storybook-design sequence. Participants used the same GenAI tools during this four-week period and while completing their storybooks following the workshops. In Workshop 1, the instructor introduced the pedagogical value of creating a mathematics storybook and elicited participants' prior knowledge and practicum-related teaching needs. Participants then began to define their design tasks by identifying relevant practicum contexts and selecting mathematical concepts as the thematic basis of their storybooks. During this process, participants used ChatGPT (GPT-5), a conversational GenAI tool that generates text in response to user prompts, to brainstorm and select mathematical concepts. Participants were instructed to use ChatGPT (GPT-5). Based on the model information available in the submitted interaction records, the documented ChatGPT interaction of each participant was conducted using GPT-5. In Workshop 2, participants explored GenAI tools for generating textual content, including story plots, narration, and dialogue. They used ChatGPT (GPT-5) to brainstorm and develop the storybook script. The instructor did not provide any fixed prompt template, and participants formulated and refined their own prompts. In Workshop 3, participants explored possible approaches to generating visual content with generative AI (GenAI) tools. They compared the benefits and limitations of different platforms and discussed the suitability of particular tools for different design purposes. Following this exploration, participants were required to use Krea AI, an AI-powered image-generation platform, to brainstorm and generate visual elements for their storybooks through text-based prompts. Although participants were introduced to multiple GenAI tools during the workshops, Krea AI and ChatGPT were specified for image and text generation, respectively, to ensure that comparable interaction records could be collected across participants. In Workshop 4, participants further developed and refined their storybook prototypes by confirming the storybook theme, visual style, illustrations, storyline, and mathematical content.

Following the workshops, each participant delivered an individual presentation of their process, explaining how they had developed the storybook. The presentations were conducted in Cantonese and video-recorded, capturing both participants' verbal explanations and on-screen materials. Each presentation lasted approximately 10 min. In addition to the presentation, each participant submitted: (a) the final storybook; (b) one complete ChatGPT interaction record documenting text-related brainstorming and script generation; and (c) one complete Krea AI interaction record documenting image-related brainstorming and visual generation. These interaction records included participants' prompts, AI-generated outputs, and iterative refinements.

4.4. Data Analysis

The design trajectory of each participant constituted an analytical case. Within each case, the primary unit of analysis was a participant-stage episode: a bounded segment of planning, generation, curation, or integration reconstructed through the participant's process-presentation, GenAI interaction records, and completed storybook. The Cantonese presentation recordings were transcribed and translated into English by a trained bilingual research assistant, with the translations checked against the original recordings for accuracy. The resulting corpus therefore integrated participants' narrated accounts in the presentation transcripts, prompts and AI-generated outputs documented in the interaction records, and the textual and visual content of the completed storybooks, enabling triangulation across participant reports, documented GenAI interactions, and artefact-based evidence. For analytic purposes, the planning stage involved identifying the intended learner group, the mathematical learning objective, the narrative direction, and initial design requirements. Generation involved using GenAI tools to produce textual or visual materials, including story ideas, scripts, and images. Curation referred to evaluating, selecting, rejecting, recombining, or revising generated materials in relation to pedagogical, mathematical, narrative, and visual considerations. Integration referred to assembling selected textual and visual materials into a coherent storybook through page sequencing, layout, multimodal coordination, and final revision.

The analysis combined inductive coding with deductive, theory-informed interpretation to identify affordances of GenAI use and recurring workflow patterns within and across cases. Using the stage framework of planning, generation, curation, and integration, we organized each participant's presentation transcript, interaction records, and submitted storybooks into case-by-stage matrices, with each row representing one participant case and columns representing planning, generation, curation, and integration. Each cell summarized relevant participant reports, documented GenAI interactions, and artefact-based evidence. This enabled us to examine reported and documented design decisions as sequences across the storybook-creation process, rather than as isolated actions.

We then conducted inductive coding to identify affordances and recurring workflow sequences among participants using GenAI. The GenAI-related tool affordance was coded as taken up when a participant actively used a GenAI capability or generated output to advance a discernible design purpose. This included specifying constraints in prompts, generating or expanding alternatives, selecting or recombining outputs, revising generated content, and integrating textual and visual materials into the storybook. We then conducted within-case analysis to identify recurring sequences of actions in each participant's workflow, followed by cross-case comparison to refine shared pattern bundles.

After the initial inductive coding of affordance use and workflow actions, we conducted a second-cycle deductive coding process using the AIMS framework. Each participant-stage episode was examined for evidence related to three AIMS dimensions: technical innovation, ethical accountability, and cultural impact. Episodes were coded for

technical innovation when participants used GenAI to increase efficiency, generate variation, or extend design possibilities. Episodes were coded for ethical accountability when participants checked, justified, revised, or assumed responsibility for the accuracy, appropriateness, or consequences of generated content. Episodes were coded for cultural impact when participants considered, retained, or revised cultural, linguistic, or representational elements. We then examined the co-occurrence and sequencing of these codes within and across cases to identify points at which participants negotiated tensions among the three dimensions.

To enhance trustworthiness, we maintained analytic memos and an audit trail documenting codebook development, coding decisions, and pattern refinement. Two trained research assistants independently coded a subset of seven participant cases (approximately 26% of the 27 cases) during codebook development. They compared their segmentation decisions and code applications, discussed discrepancies, and refined the code definitions through consensus. Following the calibration process, the two research assistants coded the remaining 20 participant cases using the refined codebook, with 10 cases assigned to each assistant. The first author reviewed the completed case-by-stage matrices, code applications, and preliminary pattern assignments. Ambiguous or discrepant cases were discussed among the research team until consensus was reached.

4.5. Findings

4.5.1. Affordance Uptake Across Workflow Stages

We first examined the specific GenAI-related tool affordances that pre-service teachers took up across planning, generation, curation, and integration. The affordance families were inductively developed by comparing coded actions and decisions across participant-stage episodes and grouping closely related practices into higher-order categories. The resulting families were refined through iterative within-case and cross-case comparison, with attention to discrepant cases and corroborating evidence across transcripts, interaction records, and completed storybooks. To ensure the empirical basis for the findings is visible, Table 2 presents the operational definition and case count for each affordance family.

Table 2. GenAI-affordance uptake across the storybook-creation workflow.

Affordance Family	Operational Definition	Case Count
Steering and constraint-setting	Using prompts to specify pedagogical or design requirements that guide GenAI generation, such as learner age, mathematical topic, story length, or narrative structure.	21 of 27 cases
Option expansion and variation generation	Using GenAI to generate multiple possible storylines, openings, characters, settings, explanations, or visual alternatives before selecting a preferred direction.	19 of 27 cases
Selection, rejection, and recombination	Evaluating, retaining, rejecting, combining, or manually revising GenAI-generated outputs to improve pedagogical fit, mathematical clarity, narrative coherence, or age appropriateness.	17 of 27 cases

Note. Case counts are non-exclusive because participants could enact more than one affordance family across different workflow stages.

4.5.2. Steering and Constraint Affordances

The affordance family most consistently taken up involved steering and constraint-setting (21 of 27 cases). Participants used prompts to translate pedagogical intentions into requirements that shaped generation, including the intended learner age or developmental

level, the mathematical topic, story length, narrative structure, and desired instructional opportunities. Across these 21 cases, participants' narrated accounts in the presentations were broadly consistent with their GenAI interaction records and completed storybooks.

For example, P007 explained in her presentation that she wanted to design a storybook to help children aged between five and six recognize three-dimensional shapes. Her corresponding ChatGPT interaction record included the prompt, "Please suggest storybooks for children aged five to six years to learn to recognize three-dimensional shapes." In her Krea AI interaction record, she also specified this requirement, "Please give me a picture with three objects: a sphere, a cylinder, and a rectangular prism." The completed storybook also retained this focus by presenting three-dimensional objects for children to identify. Together, these sources show how her early pedagogical intentions were carried into prompting and subsequently retained in the final artefact.

4.5.3. Option Expansion and Variation Generation

The second most frequently taken-up affordance family involved option expansion and variation generation (19 of 27 cases). Participants used GenAI to produce multiple story openings, characters, settings, mathematical explanations, visual styles, or alternative endings before selecting a preferred direction. This practice was often described as a way to accelerate ideation and compare narrative possibilities.

For example, P001 used ChatGPT to generate and compare alternative everyday contexts for teaching children to recognize Hong Kong coins and their values. Her initial prompt asked for "story plots that give a chance for children aged between 5 and 6 to recognize coins and values of each coin in Hong Kong." ChatGPT first suggested "The Coin Parade for Mooncake Day", in which coins were used to buy mooncakes. P001 considered this context less closely connected to children's daily lives and prompted ChatGPT again for "a more daily life scenario that involves recognizing coins and values of each coin in Hong Kong". She then selected a supermarket scenario, in which the main character used coins to purchase requested items throughout the story.

P001 also explored alternative character designs during image generation, including a boy and a girl, before selecting a bear as the main character because she considered it more interesting. After selecting the supermarket setting and bear character, she refined the mathematical task through a more detailed prompt, "Create a simple problem in which the character needs to pay HK\$18 but does not have eighteen \$1 coins." The final storybook incorporated this selected scenario, with the bear using the specified coin combination to make an HK\$18 payment. In one page of the final storybook, the character explains, "I do not have eighteen one-dollar coins. I can pay with four two-dollar coins and two five-dollar coins instead." The accompanying illustration displays the selected coins, enabling children to connect the values of different coins with a total payment of HK\$18. This example illustrates the potential to expand options as P001 used GenAI to explore multiple possible ways of embedding coin-value recognition in the narrative before selecting and refining one scenario for the final storybook.

4.5.4. Selection, Rejection, and Recombination

During curation, participants used affordances related to selection, rejection, and recombination (17 of 27 cases). They retained useful elements of AI-generated outputs while rejecting or revising content that did not fit the intended learner level, mathematical objective, narrative coherence, or visual design.

For example, P004 explained in her presentation that ChatGPT initially generated an engaging story about inviting animals to a birthday party. In the narrative, children were expected to add up the prices of various items seen in a supermarket while preparing for

the birthday party. However, she considered this task too challenging for young children because it required them to perform addition using relatively abstract numerical information. She therefore rejected the original mathematical activity and revised it to provide a more concrete counting-based task. As indicated in her final storybook, the revised activity presented the animals' faces and invited children to add by counting the visible animals. This selective revision retained the original birthday-party narrative while modifying the mathematical explanation to suit the intended learners.

4.5.5. Process Patterns in GenAI-Mediated Storybook Creation

The second objective of this study was to identify recurring process patterns in pre-service teachers' GenAI-mediated storybook creation across the four workflow stages of planning, generation, curation, and integration. Whereas the preceding affordance analysis identified specific tool-enabled action possibilities that participants utilized, this analysis examined how participants combined and sequenced these practices across their overall design trajectories. Analysis of the 27 participant cases identified two recurring process patterns. The first, constraint-guided planning-to-generation, involved establishing pedagogical requirements before or early on in the generation and using them to guide subsequent prompting, curation, and integration. The second, exploration-to-selection, involved generating and comparing alternatives before deciding on a preferred pedagogical and narrative direction through selection, recombination, and refinement. Table 3 presents the operational definitions and case counts for each pattern.

Table 3. Design patterns in GenAI-mediated storybook creation.

Process Pattern	Operational Definition	Case Count
Constraint-guided planning-to-generation	Participants articulated explicit pedagogical and design constraints before, or early in, GenAI prompting. These constraints included the intended learner group, mathematical objective, narrative structure, or visual requirements, and were subsequently used to guide prompting, curation, and integration decisions.	18 of 27 cases
Exploration-to-selection	Participants initially used GenAI to generate and compare multiple possible story directions, characters, scripts, settings, or visual options before settling on a more stable pedagogical or narrative direction. Selection, recombination, and refinement were used to develop a final design from multiple alternatives.	9 of 27 cases

4.6. Pattern 1: Constraint-Guided Planning-to-Generation

In the constraint-guided planning-to-generation pattern (18 of 27 cases), participants treated planning as a teacher-led goal-setting stage that established criteria for subsequent AI generation. Presentations in these cases typically began with an explicit specification of the intended learner group, mathematical objective, and narrative structure, often framed as a beginning-to-end storyline or a page-by-page outline. These early decisions were then translated into prompting constraints through steering and constraint-setting affordances and used as reference points during later curation and integration.

For example, P002 explained in her presentation that she intended to design a storybook for children aged four to five years to develop one-to-one correspondence. Before generating the story, she specified the intended age group, the required 12-page structure, simple language, and opportunities for children to count objects. Her ChatGPT prompt stated, "Please give me a twelve-page story using twelve simple sentences with explicit counting opportunities." The completed storybook retained this planned structure by pre-

senting animals in quantities that corresponded to the target numbers and inviting children to count them. This alignment across the presentation, interaction record, and final artefact illustrates how the participant's initial pedagogical requirements structured later design decisions.

In these cases, curation did not primarily function as a stage for exploring entirely new directions. Instead, it involved checking whether generated text and images remained consistent with the initial learning objective and intended learner level. In the same case, P002 initially received an engaging storyline that included the Chinese word for “owl” (貓頭鷹), which she considered too difficult for children aged four to five because the characters contained many strokes. Rather than abandoning the planned storyline, she retained its central narrative premise and used selection, rejection, and revision practices to modify the language for the intended learners. She revised the prompt to request simpler animal names in Chinese, “Please give me easier animal names in Chinese.” The final storybook therefore preserved the original mathematical objective while modifying the language to suit the intended learners.

During integration, P002 assembled the selected text and visuals into a 12-page storybook while maintaining the planned counting structure across pages. For example, each page presented a designated number of animals and invited children to count them one by one, linking the visual objects to the corresponding number. After replacing the more complex animal name with a simpler alternative, P002 also ensured that the language remained accessible to children aged four to five throughout the storybook. The final artefact therefore showed alignment between the initial one-to-one correspondence objective, the simplified vocabulary, the quantities represented in the illustrations, and the counting activities embedded across the pages.

4.7. Pattern 2: Exploration-to-Selection

In the exploration-to-selection pattern (9 of 27 cases), participants began using GenAI to generate possibilities rather than committing to a detailed narrative plan at the outset. They explored alternative story contexts, characters, settings, scripts, or visual styles; compared these alternatives; and gradually selected a direction for further development. In these cases, planning was not absent. Rather, it was often tentative, delayed, or carried out alongside generation.

For example, P020 began without a fixed learning objective or narrative plan. She first asked ChatGPT, “Please suggest some possible mathematics learning objectives for children aged five to six years.” ChatGPT proposed several topics, including basic addition and subtraction, comparing sizes and lengths, and positional concepts. P020 then asked ChatGPT to suggest possible story plots for each theme, “Please suggest interesting stories for each theme.” In her presentation, she explained that she evaluated these alternatives and chose to compare lengths because she considered it the most interesting option.

As the selected concept did not yet specify which objects would be compared, P020 subsequently used Krea AI to explore alternative visual representations. The Krea AI interaction record included generated images of pencils, crayons, and toy trains of different lengths. In her presentation, she explained that she selected the objects that she considered most visually appealing and most easily distinguishable for young children.

After selecting this visual direction, P020 developed a more specific story sequence and mathematical task. Her subsequent ChatGPT prompts requested simple dialogue using the terms “longer” and “shorter” as well as a storyline in which characters compared pencils of different lengths. During curation, she rejected images in which the differences in length were unclear or visually inconsistent and retained those that most clearly supported the intended comparison. The final storybook therefore emerged through the

progressive generation, comparison, selection, and refinement of alternatives, rather than from a fully specified pedagogical and narrative plan at the outset.

4.8. Negotiating AIMS Tensions Across the Lifecycle

We next examined how participants' workflow patterns and uptake of GenAI tool affordances surfaced tensions among the AIMS dimensions of technical innovation, ethical accountability, and cultural impact. These tensions were not usually articulated by participants as abstract ethical concerns. Rather, they were identified when participants reported or documented efforts to use GenAI efficiently, creating a need to check, revise, or justify the suitability of generated content for a classroom-ready storybook. Such tensions were most visible during curation and integration, when participants could evaluate the coherence, mathematical accuracy, developmental appropriateness, and representational fit of text and visuals for the completed artefact. Table 4 presents the operational definitions and case counts for the three AIMS tensions identified across the storybook-creation lifecycle.

Table 4. AIMS tensions across the GenAI-mediated storybook-creation lifecycle.

AIMS Tension	Operational Definition	Case Count
Efficiency and automation versus sustained oversight	Tension between using GenAI to accelerate drafting or production and maintaining human responsibility for mathematical accuracy, developmental appropriateness, and instructional clarity.	19 of 27 cases
Creative variation versus coherence and pedagogical fit	Tension between using GenAI to generate multiple creative alternatives and maintaining a coherent storyline, developmental progression, and clear mathematical focus across the completed book.	15 of 27 cases
Generic Outputs and Cultural Defaulting versus Local Contextual Fit	Tension between accepting model-generated names, settings, examples, or visual representations and adapting these elements to better reflect the intended local classroom context and avoid inappropriate or stereotypical representation.	8 of 27 cases

Note. Tensions were not mutually exclusive. Individual cases could demonstrate more than one tension across different workflow stages.

4.9. Efficiency and Automation Versus Sustained Oversight

The tension between efficiency and sustained oversight was evident in 19 of the 27 cases. Participants frequently described GenAI as useful for generating story ideas, scripts, and visuals efficiently. However, interaction records and completed storybooks showed that rapid generation did not necessarily produce content that was immediately suitable for young children. In some cases, participants retained generated content with only minor modification. In others, they slowed the workflow and introduced additional checking or revision after identifying a mismatch between the generated output and the intended learning objective, learner level, or classroom context. Across these 19 cases, participants described GenAI as improving efficiency while also recognizing that its outputs required human checking and revision.

One common response involved using professional judgement to refine prompts after an initial output revealed ambiguity or an inappropriate level of complexity. For example, P021 initially used ChatGPT to develop a story for K1 children about recognizing shapes. She then used Krea AI to generate images of the target shapes; however, the shapes were displayed in different colors. To reduce unnecessary visual variation, she re-prompted Krea AI, "Please give me one blue square, one blue triangle, and one blue rectangle." In her presentation, P021 explained that she wanted the shapes to be the same color so that children could focus on shape as the only varying feature, rather than having to distinguish

among colors. This episode illustrates how the efficiency of initial image generation was followed by an accountability-oriented revision when the participant identified a potential developmental mismatch in the instructional materials.

4.10. Creative Variation Versus Coherence and Pedagogical Fit

The tension between creative variation and pedagogical coherence was evident in 15 of the 27 cases. Participants often used GenAI to generate multiple story ideas, characters, visual styles, or page-level activities to make their storybooks more engaging. However, the interaction records and completed artefacts showed that variation could also create difficulties when elements from different outputs were combined. These difficulties included uneven narrative pacing, inconsistent visual styles, and discontinuities in the progression of the target mathematical concept.

For example, P013 developed a storybook on comparing quantities using the concepts of “more”, “less”, and “equal” for children aged 4 to 5 years. During generation, she used ChatGPT and Krea AI to explore alternative story contexts and visual representations of quantity, including animals collecting food, children sharing toys, and baskets containing different numbers of fruit. In her presentation, she explained that she selected the fruit-basket scenario because she considered it more engaging and concrete for young children. However, once she assembled the initial pages, she found that the generated scenes did not consistently support the intended mathematical progression. In particular, one illustration depicted several fruits that did not relate to the accompanying description, making it unclear which basket contained more or fewer fruits.

This mismatch prompted further curation and integration. P013 retained the fruit-basket context but revised the problematic pages to ensure that the quantities shown in the illustrations corresponded to the accompanying text and comparison task. She also refined her prompts to Krea AI to generate illustrations that more closely matched the dialogue and intended quantities. These revisions enabled children to compare unequal sets of fruit using terms such as “more” and “less” before encountering examples of equal quantities. The final storybook, therefore, reflected a more coherent mathematical progression than the initial set of generated materials.

This case illustrates that creative variation was not inherently problematic. Rather, it increased the need for later evaluative work. GenAI enabled participants to explore multiple engaging possibilities, but pedagogical fit depended on their ability to identify inconsistencies, select appropriate elements, and restore coherence at the whole-book level. In this sense, technical innovation shifted part of the accountability work downstream, making the quality of the final instructional product contingent on careful curation and integration rather than on generation alone.

4.11. Generic Outputs and Cultural Defaulting Versus Local Contextual Fit

The tension between generic AI-generated outputs and local contextual fit was evident in 8 of the 27 cases. Some participants retained AI-generated names, settings, and visual styles with limited discussion of their cultural or contextual implications. In these cases, model-generated defaults are included in the final artefact without being explicitly evaluated for the intended classroom audience. Other participants revised names, settings, examples, or visual details to make the storybook more locally familiar or contextually appropriate for young children in Hong Kong.

Such considerations were most evident during curation and integration, when participants could evaluate how textual and visual elements worked together for the completed storybook. For example, P026 initially retained the AI-generated character name and visual setting during generation. However, while assembling the final storybook in Book

Creator, she reconsidered the name “Anika”, judging it to be less familiar or suitable for her intended Hong Kong kindergarten audience. She subsequently changed the character’s name to “Siu Ming (小明)”, and this revision was reflected in the final storybook. Importantly, the example does not imply that the original name was inherently inappropriate. Rather, it shows how P026 evaluated the name in relation to the particular audience she imagined for the storybook. Overall, these cases indicate that considerations of cultural impact were frequently reactive and triggered by visibility. GenAI-enabled production made it easy to generate plausible names, settings, and images; however, local contextual fit depended on participants’ willingness and capacity to notice, question, and revise those defaults during later stages of the workflow.

5. Discussion

This study examined how 27 pre-service early childhood educators in Hong Kong designed GenAI-mediated mathematics storybooks across four workflow stages: planning, generation, curation, and integration. It investigated the GenAI-related tool affordances that participants used, the process patterns applied to sequence these practices, and the tensions among technical innovation, ethical accountability, and cultural impact interpreted through the AI-mediated storytelling (AIMS) framework (Shahamat, 2023). The findings showed that participants commonly used GenAI to steer generation through pedagogical constraints, expand story and visual alternatives, and select, reject, or revise generated materials. These practices were organized into two recurring workflow patterns: constraint-guided planning-to-generation, in which pedagogical requirements were established early and used to guide later decisions, and exploration-to-selection, in which participants generated and compared alternatives before gradually stabilizing a preferred direction. Across both patterns, tensions related to efficiency and oversight, creative variation and pedagogical coherence, and generic outputs and local contextual fit became particularly visible during curation and integration.

5.1. Process Patterns and the Timing of Pedagogical Intent

Two recurring process patterns were identified. In the constraint-guided planning-to-generation pattern, participants established pedagogical requirements before or early on in storybook generation, including the intended learner group, mathematical objective, narrative structure, language level, and page-level learning opportunities. These requirements subsequently guided prompting, selection, revision, and assembly. Planning, therefore, functioned as an early form of pedagogical governance. It provided participants with criteria for evaluating whether generated outputs remained aligned with the storybook’s intended mathematical and developmental purposes. This pattern is consistent with accounts of digital storytelling as purposeful design work, in which creators coordinate narrative, visual, and linguistic resources in relation to intended meanings and audiences (Pettit, 2020). In a mathematics storybook context, however, such coordination also requires attention to the progression of the target concept. A storybook may be visually attractive and narratively engaging while remaining pedagogically weak if its illustrations, dialogue, and activities do not accurately represent the mathematical idea or support an appropriate learning sequence. Research on mathematics picture books similarly emphasizes that their educational value depends on the relationship between mathematical content, representations, narrative integration, and the developmental suitability of the text and illustrations (Marston, 2010; van den Heuvel-Panhuizen & Elia, 2012).

By contrast, the exploration-to-selection pattern involved generating and comparing possible learning objectives, story contexts, characters, visual representations, and narrative directions before stabilizing a preferred approach. In these cases, planning was not absent. Rather, planning was tentative, delayed, and occurred alongside generation. Participants used GenAI to efficiently identify a broad range of possible directions before evaluating these alternatives in relation to their emerging ideas about what might be engaging, feasible, and appropriate for young children. This pattern aligns with human–AI co-creation research that distinguishes between elaborating new possibilities and reflecting on previously generated material (Lin et al., 2023). It also illustrates that GenAI can support early pedagogical ideation, particularly when participants have not yet settled on a specific mathematical focus, story context, or visual representation.

However, the two patterns involved different distributions of pedagogical risk. Constraint-guided planning-to-generation is consistent with digital storytelling guidance that emphasizes clarifying learning goals, audience, and narrative structure before production so that these criteria can guide later development and evaluation (Robin & McNeil, 2012). This may be particularly important in mathematics storybook design, where mathematical representations, illustrations, narrative events, and conceptual progression need to work together coherently (van den Heuvel-Panhuizen & Elia, 2012). By contrast, exploration-to-selection aligns with mixed-initiative co-creation research, using generated alternatives to support elaboration and reflection (Lin et al., 2023). However, the findings also show that this wider option space placed greater demands on later selection, recombination, and repair. Participants needed to determine whether engaging narrative and visual options also supported conceptual accuracy, developmental appropriateness, and a coherent progression of mathematical ideas. Thus, the critical issue was not whether participants began with constraints or exploration, but whether they could make pedagogical intentions sufficiently explicit to guide and evaluate later decisions.

5.2. Tool Affordances, Pedagogical Judgement, and Professional Agency

The affordance findings further clarify how participants enacted these workflow patterns. Steering and constraint-setting; option expansion and variation generation; and selection, rejection, and recombination were frequently taken up across cases. This is consistent with relational accounts of technological affordances, which emphasize that action possibilities emerge through the relationship between a technology's functions and users' purposes, knowledge, and situated goals (Nagy & Neff, 2015). It also relates to recent human–AI interaction research identifying prompting, selection, and manipulation as common user-guided ways of directing generative systems (Luera et al., 2024; Tsiakas & Murray-Rust, 2024). However, the present findings show that, in mathematics storybook design, these actions were not merely technical operations. Participants used prompts to translate instructional intentions into generation requirements, used alternative outputs to explore possible narrative and visual representations, and used selection and revision to judge whether generated materials fit the intended learner level and mathematical objective. Such judgement is especially consequential in early mathematics storybooks, where the educational value of a book depends on the alignment among mathematical ideas, illustrations, language, narrative context, and opportunities for children to engage with the target concept (van den Heuvel-Panhuizen & Elia, 2012).

The findings, therefore, complicate a simple claim that GenAI either enhances or diminishes professional agency. GenAI expanded the range and speed of available design options, but it also shaped the set of alternatives from which participants selected. This is consistent with ecological accounts of teacher agency, which conceptualize agency not as an individual possession but as an achievement arising through the interac-

tion of professional purposes, available resources, knowledge, and contextual conditions (Priestley et al., 2026). In the present study, participants exercised pedagogical agency most clearly when they articulated criteria, rejected unsuitable outputs, or revised generated materials in response to observed mismatches. When alternatives were retained with limited evaluation, however, participants appeared to be working primarily as curators of model-produced content rather than as designers actively governing the learning pathway. The findings therefore suggest that GenAI-supported design does not automatically strengthen professional agency. Its contribution depends on whether pre-service teachers can critically evaluate and reshape outputs based on independent pedagogical criteria. This capacity is central to AI literacy, which includes evaluating AI-generated content and making informed decisions about its use rather than simply operating AI tools (OECD/European Commission, 2026).

5.3. AIMS Tensions as Bounded Ethical and Pedagogical Decisions

The AIMS analysis showed that tensions among technical innovation, ethical accountability, and cultural impact were generally enacted through concrete design decisions rather than articulated as abstract ethical beliefs. Accountability and cultural considerations became clear when generated material appeared unsuitable for inclusion in a classroom-ready storybook. These findings can be situated cautiously alongside broader principlist frameworks for responsible AI, which identify beneficence, non-maleficence, autonomy, justice, and explicability as core ethical principles (Floridi et al., 2018). However, the evidence in the present study relates only to a narrower set of concerns. It primarily reflects participants' professional responsibility to avoid mathematically inaccurate, developmentally unsuitable, visually unclear, or contextually inappropriate materials, together with limited attention to local contextual fit. The findings do not provide sufficient evidence to draw conclusions about other responsible AI concerns, such as autonomy, justice, or fairness.

Within this bounded pedagogical scope, the first observed tension concerned the relationship between efficiency and sustained oversight. This tension was evident when participants valued GenAI for quickly generating stories, scripts, and images but later identified problems in learner level, mathematical accuracy, visual clarity, or instructional suitability. Participants responded by tightening prompts, regenerating content, or manually revising materials. This finding supports the need for wider research showing that pre-service teachers may find GenAI useful for instructional planning while still requiring explicit support to evaluate the quality and correctness of generated outputs (Gurl et al., 2025; Karataş & Yüce, 2024). In the present study, sustained oversight was not limited to checking for generic factual errors. It involved evaluating whether materials supported the intended purpose of mathematical learning. For instance, participants considered whether visual sets corresponded to accompanying descriptions, whether wording was appropriate for the intended age group, whether illustrations directed attention to the relevant mathematical feature, and whether the sequence of pages supported a coherent conceptual progression. Ethical accountability was also considered, referred to as pedagogical responsibility for the suitability of materials intended for young children. However, evaluating the suitability of individual outputs did not necessarily ensure coherence once multiple generated elements were assembled into a complete storybook. This broader, whole-book challenge was reflected by the second tension: creative variation versus coherence and pedagogical fit.

The second tension was particularly evident when participants combined elements from different generated alternatives. Variation expanded creative possibilities, but it could also produce uneven visual styles, disrupted narrative pacing, or discontinuities in

mathematical progression. The repair work required during later-stage integration demonstrates that generated outputs cannot be evaluated in isolation. A page may be visually appealing or mathematically plausible on its own while having a weak instructional trajectory when combined with other pages in a complete storybook. This has specific implications for mathematics education. The mathematical quality of a storybook depends not only on whether individual statements or quantities are correct, but also on whether the story provides meaningful opportunities to engage with the target concept and whether its illustrations, language, and narrative sequence support rather than obscure that concept (Marston, 2010; van den Heuvel-Panhuizen & Elia, 2012). Participants' attention to quantity matching, age-appropriate vocabulary, visual distinctions, and sequencing, therefore, represented important forms of engaging in pedagogical content knowledge. At the same time, the need for late-stage repair indicates that such knowledge was not always applied as an explicit design criterion before generation began. Beyond whether the completed storybook was mathematically and narratively coherent, participants also needed to consider whether its names, settings, and representations were appropriate for the particular children for whom it was intended. This concern was reflected in the third tension: generic outputs versus local contextual fit.

The third tension, between generic outputs and local contextual fit, was less common but analytically important. A small number of participants revised names, settings, examples, or visual details after recognizing they did not fit the imagined Hong Kong kindergarten audience. These findings suggest that contextual considerations were often triggered by visibility and became more salient when participants viewed the completed text and images as a unified storybook rather than as isolated outputs. Nevertheless, localizing a character name or setting should not automatically be equated with culturally responsive pedagogy. Culturally responsive teaching requires more sustained attention to learners' cultural knowledge, experiences, identities, and participation in meaning-making (Gay, 2018). The present findings suggest that some participants noticed and addressed contextual mismatches in final artefacts, whereas others retained generic outputs with limited discussion. Cultural impact was, therefore, rarely established as an explicit requirement during the planning or generation stages. This suggests the need for teacher education to help move students beyond reactive localization towards more careful consideration of audience, representation, and local relevance.

5.4. Implications for Teacher Education and Scalable Implementation

The findings have several implications for teacher education. First, teacher education should support both constraint-guided and exploration-based workflows, rather than privileging one pattern unconditionally. This recommendation is consistent with mixed-initiative co-creation research, which distinguishes between elaborating on new possibilities and reflecting on previously generated material, suggesting that effective human–AI collaboration may involve movement between these modes (Lin et al., 2023). Constraint-guided planning may be appropriate when the learner group, mathematical objective, and instructional sequence are already clearly defined. Exploration-to-selection may be useful when pre-service teachers are developing story contexts, characters, or visual representations. However, exploration-first workflows require explicit convergence points, including structured review of mathematical accuracy, developmental appropriateness, language complexity, text–image alignment, conceptual sequencing, narrative coherence, and local contextual fit. These review criteria are particularly important for mathematics picture books, as their instructional value depends on the alignment of mathematical content, illustrations, narrative structure, and opportunities for children to engage with the target concept (van den Heuvel-Panhuizen & Elia, 2012). They also align with human–AI inter-

action guidance that emphasizes users' ability to correct, refine, and remain appropriately informed about system outputs throughout an interaction, rather than treating AI output as the final product (Amershi et al., 2019).

Second, the findings suggest that AI literacy should be integrated with digital literacy and subject-specific pedagogical content knowledge, rather than taught as a stand-alone set of technical skills. Students need practical competence in using text-generation, image-generation, and digital-authoring tools; however, technical fluency alone is insufficient for effective educational design. Consistent with the TPACK framework, meaningful technology integration depends on the coordinated application of technological, pedagogical, and content knowledge (Koehler & Mishra, 2009). In the present context, this includes evaluating whether AI-generated text, images, and story activities accurately represent mathematical ideas, direct children's attention to the intended feature, and remain developmentally appropriate. This emphasis also aligns with AI literacy frameworks that foreground critical evaluation of AI outputs and informed decision-making about their use (OECD/European Commission, 2026). Similarly, recent research with pre-service secondary mathematics teachers found that participants benefited from reflective engagement with ChatGPT but were sometimes less accurate in recognizing problematic mathematical output, suggesting a need for explicit frameworks to critically examine AI-generated mathematics materials (Gurl et al., 2025). Teacher education should therefore integrate opportunities for pre-service teachers to use, evaluate, revise, and justify GenAI-generated content in relation to specific subject matter and learner needs, rather than treating prompt construction or tool operation as sufficient evidence of AI competence.

At a larger scale, these insights could be incorporated into teacher-education modules through reusable design and formative-assessment routines rather than intensive individual supervision alone. Research on formative rubrics suggests that making quality criteria explicit can support learners' self-regulation and reflection, particularly when rubrics are used to guide revision rather than assign marks (Panadero & Jonsson, 2013). Peer assessment may also provide a scalable means of expanding feedback opportunities while engaging students in evaluating the quality of alternative designs (Double et al., 2020). In this context, programmes could use common task briefs, staged design templates, prompt-and-revision logs, mathematics-storybook evaluation rubrics, calibrated peer-feedback protocols, and sample-based instructor audits. Such structures could make GenAI-mediated design tasks more manageable in larger cohorts while preserving attention to pedagogical reasoning.

5.5. Limitations and Future Research

Several limitations should inform the interpretation of the findings. First, the study involved a small and demographically homogeneous sample of 27 female pre-service early childhood educators from a single university in Hong Kong. Although this context enabled an in-depth exploratory analysis of detailed, multimodal design trajectories, it limits the transferability of the findings. The observed patterns should not be interpreted as estimates of how frequently similar practices occur among pre-service teachers more broadly, nor as evidence that the same patterns necessarily emerge in other cultural, disciplinary, institutional, or gender-diverse settings. The findings may nevertheless offer analytically useful insights for contexts in which teacher candidates design AI-supported materials for young children. Future research should involve multiple institutions; more culturally, linguistically, and gender-diverse cohorts; in-service teachers; and teacher candidates working across different subject areas and educational systems. In addition, this study examined the design process and quality of completed storybooks rather than their implementation with children. Therefore, it cannot be established whether the final arte-

facts improved children's mathematical understanding, engagement, language development, or sense of contextual relevance. Future classroom-based research could examine how children, teachers, and parents respond to GenAI-mediated mathematics storybooks and whether the process patterns identified here are associated with differences in instructional quality, classroom enactment, or student learning outcomes.

6. Conclusions

The GenAI-mediated creation of mathematics storybooks should not be understood simply as an accelerated form of text or image production. Rather, it is a pedagogical design process in which pre-service teachers must determine when to articulate instructional constraints, when to use generated alternatives to explore certain topics further, and how to evaluate, recombine, and revise outputs in relation to children's learning needs. The findings show that GenAI expanded the narrative range and visual possibilities available but did not resolve the professional work required to ensure mathematical accuracy, developmental appropriateness, conceptual coherence, and local contextual fit. Instead, it reconfigured pedagogical agency by making curation and integration consequences of decision-making. Whether GenAI-supported storybooks become meaningful educational resources therefore depends not primarily on the fluency of generated content, but on teachers' capacity to apply their independent pedagogical judgement throughout the design lifecycle.

Author Contributions: Conceptualization, Y.Y.C. & K.M.K.H.; methodology, Y.Y.C.; software, Y.Y.C.; validation, Y.Y.C.; formal analysis, Y.Y.C.; investigation, Y.Y.C. & K.M.K.H.; resources, Y.Y.C. & K.M.K.H.; data curation, Y.Y.C.; writing—original draft preparation, Y.Y.C.; writing—review and editing, Y.Y.C. & K.M.K.H.; visualization, Y.Y.C.; project administration, K.M.K.H.; funding acquisition, K.M.K.H. All authors have read and agreed to the published version of the manuscript.

Funding: The work was supported by the Hong Kong Metropolitan University R&D Fund of the Hong Kong Special Administrative Government (Project Number: RD/2025/1.6).

Institutional Review Board Statement: Ethical approval for this study was obtained by the corresponding author from the Research Ethics Committee (REC) of Hong Kong Metropolitan University (REC reference number: HE-RD/2025/1.6, date of approval: 15 September 2025). The study was conducted in accordance with the guidelines and regulations of the REC. All research protocols and consent forms were reviewed and approved by the REC.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study. All participants provided signed informed consent for the publication of anonymized data prior to participation in this study.

Data Availability Statement: The supporting data for this study are available from the corresponding author upon reasonable request. The data are not publicly available due to privacy and ethical restrictions.

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Amershi, S., Weld, D., Vorvoreanu, M., Fournery, A., Nushi, B., Collisson, P., Suh, J., Iqbal, S., Bennett, P. N., Inkpen, K., Teevan, J., Kikin-Gil, R., & Horvitz, E. (2019). Guidelines for human-AI interaction. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1–13). Association for Computing Machinery.
- Bilici, S., & Yilmaz, R. M. (2024). The effects of using collaborative digital storytelling on academic achievement and skill development in biology education. *Education and Information Technologies*, 29(15), 20243–20266. [CrossRef]
- Chan, B. S., Churchill, D., & Chiu, K. F. (2017). Digital literacy learning in higher education through digital storytelling approach. *Journal of International Education Research*, 13(1), 1–16. [CrossRef]
- Churchill, N. (2020). Development of students' digital literacy skills through digital storytelling with mobile devices. *Educational Media International*, 57(3), 271–284. [CrossRef]

- Colton, S., & Wiggins, G. A. (2012). Computational creativity: The final frontier? In L. De Raedt, C. Bessiere, D. Dubois, P. Doherty, P. Frasconi, F. Heintz, & P. J. F. Lucas (Eds.), *ECAI 2012: 20th European conference on artificial intelligence* (Vol. 242, pp. 21–26). Frontiers in Artificial Intelligence and Applications. IOS Press.
- Cukurova, M., & Miao, F. (2024). *AI competency framework for teachers*. UNESCO Publishing.
- Double, K. S., McGrane, J. A., & Hopfenbeck, T. N. (2020). The impact of peer assessment on academic performance: A meta-analysis of control group studies. *Educational Psychology Review*, 32(2), 481–509. [\[CrossRef\]](#)
- Floridi, L., Cows, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689–707. [\[CrossRef\]](#) [\[PubMed\]](#)
- Gay, G. (2018). *Culturally responsive teaching: Theory, research, and practice*. Teachers College Press.
- Gurl, T. J., Markinson, M. P., & Artzt, A. F. (2025). Using ChatGPT as a lesson planning assistant with preservice secondary mathematics teachers. *Digital Experiences in Mathematics Education*, 11(1), 114–139. [\[CrossRef\]](#)
- Karataş, F., & Yüce, E. (2024). AI and the future of teaching: Preservice teachers' reflections on the use of artificial intelligence in open and distributed learning. *International Review of Research in Open and Distributed Learning*, 25(3), 304–325. [\[CrossRef\]](#)
- Koehler, M., & Mishra, P. (2009). What is technological pedagogical content knowledge (TPACK)? *Contemporary Issues in Technology and Teacher Education*, 9(1), 60–70.
- Li, Y., & Chuprina, N. V. (2024). Perception, and resonance: Research on generative ai in digital picture book design. *Theory and Practice of Design*, 34, 488–497. [\[CrossRef\]](#)
- Lin, Z., Ehsan, U., Agarwal, R., Dani, S., Vashishth, V., & Riedl, M. (2023). Beyond prompts: Exploring the design space of mixed-initiative co-creativity systems. *arXiv*, arXiv:2305.07465.
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems* (pp. 1–16). Association for Computing Machinery.
- Luera, R., Rossi, R. A., Siu, A., Dernoncourt, F., Yu, T., Kim, S., Zhang, R., Chen, X., Salehy, H., Zhao, J., Basu, S., Mathur, P., & Lipka, N. (2024). Survey of user interface design and interaction techniques in generative AI applications. *arXiv*, arXiv:2410.22370.
- Malterud, K., Siersma, V. D., & Guassora, A. D. (2016). Sample size in qualitative interview studies: Guided by information power. *Qualitative Health Research*, 26(13), 1753–1760. [\[CrossRef\]](#) [\[PubMed\]](#)
- Marston, J. (2010). Developing a framework for the selection of picture books to promote early mathematical development. In L. Sparrow, B. Kissane, & C. Hurst (Eds.), *Proceedings of the 33rd annual conference of the mathematics education research group of Australasia* (pp. 383–390). Mathematics Education Research Group of Australasia.
- Miles, M. B., Huberman, A. M., & Saldana, J. (2014). *Qualitative data analysis*. Sage Publications.
- Nagy, P., & Neff, G. (2015). Imagined affordance: Reconstructing a keyword for communication theory. *Social Media + Society*, 1(2), 2056305115603385. [\[CrossRef\]](#)
- Nicolopoulou, A., Cortina, K. S., Ilgaz, H., Cates, C. B., & de Sá, A. B. (2015). Using a narrative-and play-based activity to promote low-income preschoolers' oral language, emergent literacy, and social competence. *Early Childhood Research Quarterly*, 31, 147–162. [\[CrossRef\]](#) [\[PubMed\]](#)
- Nounanaki, A. L. (2025). AI augmented legend narratives on TikTok: When the acting parts become the tellers of their own stories. Planetary. *Journal of Social Sciences and Humanities Research*, 2(3), 1–8. [\[CrossRef\]](#)
- OECD/European Commission. (2026). *Empowering learners for the age of AI: An AI literacy framework for primary and secondary education*. OECD Publishing.
- Panadero, E., & Jonsson, A. (2013). The use of scoring rubrics for formative assessment purposes revisited: A review. *Educational Research Review*, 9, 129–144. [\[CrossRef\]](#)
- Petit, E. (2020). *Digital storytelling for developing students' agency through the process of design: A case study*. Research-Publishing.
- Priestley, M., Biesta, G., & Robinson, S. (2026). *Teacher agency: An ecological approach*. Bloomsbury Publishing.
- Robin, B. R. (2006). The educational uses of digital storytelling. In *Society for information technology & teacher education international conference* (pp. 709–716). Association for the Advancement of Computing in Education (AACE).
- Robin, B. R., & McNeil, S. G. (2012). What educators should know about teaching digital storytelling. *Digital Education Review*, 22, 37–51.
- Shahamat, R. (2023). *Media intelligence: Ethical, technical, and cultural dimensions of AI-augmented storytelling [preprint]*. SSRN.
- Staley, B., & Freeman, L. A. (2017). Digital storytelling as student-centred pedagogy: Empowering high school students to frame their futures. *Research and Practice in Technology Enhanced Learning*, 12(1), 21. [\[CrossRef\]](#) [\[PubMed\]](#)
- Stebbins, R. A. (2025). *Exploratory research in the social sciences*. Sage Publications.
- Tsiakas, K., & Murray-Rust, D. (2024). Unpacking human-AI interactions: From interaction primitives to a design space. *ACM Transactions on Interactive Intelligent Systems*, 14(3), 1–51. [\[CrossRef\]](#)

- van den Heuvel-Panhuizen, M., & Elia, I. (2012). Developing a framework for the evaluation of picturebooks that support kindergartners' learning of mathematics. *Research in Mathematics Education*, 14(1), 17–47. [[CrossRef](#)]
- Wu, J., & Chen, D. T. V. (2020). A systematic review of educational digital storytelling. *Computers & Education*, 147, 103786. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.