



Design Implications for Teacher-Configurable Generative AI Writing Tools: A Participatory Design Study with StoryBot

Ariel Han, Shenshen Han, Zhenyao Cai, Joey Huang, Franceli Cibrian & Kylie Peppler

To cite this article: Ariel Han, Shenshen Han, Zhenyao Cai, Joey Huang, Franceli Cibrian & Kylie Peppler (13 Oct 2026): Design Implications for Teacher-Configurable Generative AI Writing Tools: A Participatory Design Study with StoryBot, International Journal of Human-Computer Interaction, DOI: [10.1080/10447318.2026.2739766](https://doi.org/10.1080/10447318.2026.2739766)

To link to this article: <https://doi.org/10.1080/10447318.2026.2739766>



© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC



Published online: 13 Oct 2026.



Submit your article to this journal [↗](#)



Article views: 256



View related articles [↗](#)



View Crossmark data [↗](#)

Design Implications for Teacher-Configurable Generative AI Writing Tools: A Participatory Design Study with StoryBot

Ariel Han^a , Shenshen Han^b , Zhenyao Cai^c , Joey Huang^d , Franceli Cibrian^e  and Kylie Pepler^f 

^aDepartment of Instructional Technology & Learning Sciences, Utah State University, Logan, Utah, USA; ^bDepartment of Informatics, University of California, Irvine, Irvine, California, USA; ^cSchool of Education, University of California, Irvine, Irvine, California, USA; ^dCollege of Education, North Carolina State University, Raleigh, North Carolina, USA; ^eSchool of Engineering, Chapman University, Orange, California, USA; ^fDepartment of Informatics & School of Education, University of California, Irvine, Irvine, California, USA

ABSTRACT

As generative AI (GenAI) tools become accessible, their integration into K–12 writing presents opportunities and challenges. This study partnered with 21 primary school teachers in a participatory co-design process to develop StoryBot, a teacher-configurable GenAI story-authoring platform. Teachers identified instructional needs and learning outcomes that informed system features, including genre-specific writing scaffolds, configurable prompts, and constraints on full-text generation to preserve student voice. The platform positions educators as co-designers of AI behavior rather than passive users. We conducted a pilot writing workshop with 13 youth (ages 8–12) and a focus group with six educators to examine how the platform influenced students' writing behaviors and classroom integration. Findings demonstrate the value of teacher co-design and configurable AI systems in aligning GenAI tools with pedagogical priorities and practices. We present design implications for teacher-configurable GenAI writing systems that support student agency, ethical safeguards, and instructional alignment in human–AI coauthoring environments.

KEYWORDS

Generative AI; artificial intelligence; K-12 education; co-design

CCS CONCEPTS

• Human-centered computing → empirical studies in HCI

1. Introduction

Writing proficiency is widely recognized as foundational to academic success yet develops slowly through sustained practice, making upper elementary years a particularly important period for supporting emergent writers (Bazerman, 2016; Graham & Perin, 2007a). From Quill¹ and StoryJumper² to emerging platforms like Writable³ and Khanmigo,⁴ upper elementary classrooms (ages 8–12; often grades 3–6) already weave digital and traditional tools to support writing. These platforms help emergent writers transition from brainstorming to structured drafts, offering prompts, vocabulary support, and teacher feedback along the way (Jeon, 2024; Jeon & Lee, 2024). At the same time, reports indicate that almost 26% of US students now use GenAI for schoolwork, often experimenting with tools such as ChatGPT⁵ (Holtermann, 2025). This dual reality highlights a pressing dilemma: while scaffolding is essential for building student writing skills and fluency, rapid uptake of Generative Artificial Intelligence (GenAI) in classrooms risks automating writing in ways that bypass student agency (Perkins, 2023; Werdiningsih et al., 2024). In this paper, agency is grounded in Bandura (1982) conceptualization as the individual's ability to influence their own actions and circumstances through intentional effort—often described as a sense of control over their behavior (Bennett et al., 2023), which in turn improves the engagement and autonomy of learners (Brod et al., 2023). In the context of writing,

CONTACT Ariel Han  ariel.han@usu.edu  Department of Instructional Technology & Learning Sciences, Utah State University, Logan, UT, USA

© 2026 The Author(s). Published with license by Taylor & Francis Group, LLC

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. The terms on which this article has been published allow the posting of the Accepted Manuscript in a repository by the author(s) or with their consent.

the preservation of agency means that students retain meaningful control over ideas, narrative decisions, and language choices, rather than delegating those decisions to the system (Bennett et al., 2023).

While Human-Computer Interaction (HCI) research has explored the implications of GenAI in writing, focusing on writers' experiences (Bhat et al., 2023; Gero et al., 2025), creativity and output quality (Kim et al., 2025a; Liu et al., 2024), and tool design (Chung et al., 2025; Fan et al., 2025b), work at the intersection of GenAI, formal education, and classroom practices remains underexplored. Recent studies have begun to explore GenAI in learning contexts, often in informal or home-based settings (Chen et al., 2025b; Xu et al., 2023; Zhang et al., 2022). Yet, a critical gap persists: how classroom teachers—who are central to shaping pedagogy, assessment, and ethical norms—can directly inform the design of GenAI writing tools in ways that integrate meaningfully into school-based instruction (Han et al., 2024).

We draw on participatory design (Muller & Kuhn, 1993) and co-design (Lin & Van Brummelen, 2021; Penuel et al., 2007) to support knowledge co-construction, engaging teachers as design partners who bring valuable professional expertise (Chang & Philip, 2025; Kali et al., 2015; Sadek et al., 2024; Stefaniak et al., 2025). In doing so, it contributes new design implications about how to translate instructional priorities into concrete GenAI features that balance student agency,⁶ scaffolding, and ethical safeguards (Kim et al., 2025b). This study was led by the following research questions:

1. How can participatory co-design with K–12 teachers inform the design of teacher-configurable GenAI story-authoring platforms for classroom instruction?
2. In what ways can teachers' instructional goals and classroom practices be translated into features and scaffolding strategies within a GenAI-based writing platform?
3. How do students engage with and make use of a teacher-co-designed GenAI story-authoring platform during writing activities?
4. How do teachers perceive the pedagogical value and classroom applicability of a co-designed GenAI writing tool?

These research questions are addressed across four phases of the study. RQ1 is examined through Phase 1: Co-Designing a Logic Model (Section 3.2). RQ2 is addressed in Phase 2: Design and Implementation (Section 4). RQ3 is explored through Phase 3: Pilot Study with Students (Section 5). RQ4 is examined in Phase 4: Teachers' Focus Group Workshop on the Classroom Integration of StoryBot (Section 5.3). This study contributes to research on GenAI in education by examining how story-authoring tools can be meaningfully integrated into classroom practice through co-design with educators. It emphasizes the importance of viewing GenAI as a sociotechnical system embedded within complex instructional ecologies (Lin & Van Brummelen, 2021; Nicholson et al., 2022; Ravi et al., 2025). Specifically, this study offers three key contributions:

1. A Teacher-Co-Designed Logic Model for GenAI Integration: We engaged 21 classroom teachers in developing a logic model that articulated contextual needs, ethical concerns, and desired outcomes. This process surfaced the kinds of scaffolds, controls, and affordances educators deem essential for classroom alignment.
2. The Design and Development of StoryBot: Guided by the logic model, we prototyped StoryBot, a GenAI writing educational platform with a dual design: (a) a teacher-facing view for customizing genre-specific tasks, prompts, and scaffolds, and (b) a student-facing interface for interactive, genre-aligned writing support.
3. Pilot Evaluation: Through an exploratory writing workshop with 13 students and a follow-up focus group with six educators, we conducted a preliminary evaluation of StoryBot's use in classroom contexts. These findings offer early insights into how the platform may support writing engagement and practice, highlighting the promise of co-designed GenAI tools while underscoring the need for further empirical investigation in formal educational settings.

2. Related work

2.1. AI-assisted Co-writing systems in HCI

HCI research on AI-assisted co-writing has examined writers' experiences and interaction techniques across different stages of the writing process (Gero et al., 2023; Yuan et al., 2022). Studies of ideation and planning show that the timing and level of AI intervention can substantially shape writing outcomes and perceptions (Dhillon et al., 2024; Qin et al., 2025). Qin et al. (2025) found that introducing LLM assistance after independent ideation increased idea generation and reduced similarity to AI suggestions, whereas introducing it from the outset increased idea fixation and reduced creative self-efficacy and self-credit (Qin et al., 2025). Similarly, Dhillon et al. (2024) found that paragraph-level AI scaffolding improved writing quality and productivity, particularly for less frequent writers and less technology-savvy users, but reduced perceived text ownership and satisfaction (Dhillon et al., 2024). Research on later stages has likewise explored AI support for reviewing and revising text (Lu et al., 2026; Zhang et al., 2025).

More broadly, AI-assisted writing systems can help writers translate ideas into written language (Bhat et al., 2023; Booten & Gero, 2021), generate inspiration (Gero et al., 2022; Yuan et al., 2022), and consider alternative perspectives (Göldi et al., 2024; Wan et al., 2024). Yet these benefits can introduce tradeoffs for writers' agency and sense of authorship. For instance, Qin et al. (2025) found that greater autonomy and ownership were associated with the benefits of delaying AI assistance (Qin et al., 2025), while Reza et al. (2025) found that writers' preferences for AI involvement vary across writing processes and contexts depending on what they seek to retain control over (Reza et al., 2025). So, this work suggests that AI's effects on agency and authorship depend not simply on whether AI is used, but on when, where, and how it participates in the writing process.

Existing AI writing systems can be broadly considered across two complementary design contexts. *General-purpose AI writing assistants*, such as ChatGPT, Claude, and Gemini, are designed for broad individual use and can support writers with activities such as ideation, drafting, revision, and creativity (Bhat et al., 2023; Li et al., 2024; Zhang et al., 2026).

In contrast, *Classroom-embedded AI writing systems*, such as Khanmigo and MagicSchool.ai, operate within pedagogical contexts in which AI assistance must also align with learning objectives, curricular expectations, assessment practices, and students' developmental needs. Yet existing educational GenAI platforms offer varying degrees of teacher control over how AI participates in students' writing, and generative features that produce substantial portions of text can risk shifting the system from scaffolding students' writing toward performing the writing for them. These design tensions shift the design problem from determining how AI can assist writing to determining how that assistance should be configured within instruction. In writing instruction, these decisions include what forms of assistance AI should provide, when assistance should occur, and how much support should be available to students. Relatively little HCI research has examined how teachers themselves can participate in designing and configuring these forms of AI assistance for classroom writing. We address this gap by investigating how participatory co-design with teachers can inform teacher-configurable AI writing tools that align AI assistance with pedagogical goals and classroom practices.

2.2. Affordances and tensions in LLM-mediated literacy learning

AI-powered educational tools have made notable advances in literacy education (e.g., reading comprehension, vocabulary acquisition) (Fan et al., 2025a; Leong et al., 2024; Shao et al., 2026), especially GenAI language models (LLM). LLMs contribute to creating dynamic learning environments that can simulate challenges and provide instant feedback, fostering deeper concept understanding (Kang et al., 2025). In language learning, LLM-based systems support the development of literacy skills through personalized experiences tailored to individual proficiency and pace (Leong et al., 2024). For example, research by Xu et al. (2023) introduced a conversational agent coupled with an e-book to facilitate parent-child interactions during bilingual reading sessions, resulting in improved language use and engagement. Bruner (1985) further underscores the importance of conversational exposure as a crucial component in language acquisition, highlighting that learning to use language effectively extends

beyond the simple acquisition of vocabulary to encompass social, contextual and pragmatic aspects of communication. Dialogic interactions, in which adults encourage children to elaborate on their thoughts, not only facilitate richer language experiences, but also optimize the use of feedback and clarity during exchanges (Paatsch et al., 2019; Scull et al., 2013).

Building on the inherently dialogic nature of classroom instruction, where educators and students engage in active, reciprocal communication, LLM-based story-authoring leverages LLM to replicate this critical pedagogical exchange. Studies such as those by Zhang et al. (2022), Leong et al. (2024), and Xu et al. (2023) illustrate how LLM-based language learning platforms can boost vocabulary, speaking ability, and overall language proficiency. In addition, researchers such as Wang et al. (2023b), Cai et al. (2025), and Xu and Warschauer (2020) emphasize the need to design LLM systems that adapt to the individual pace and learning style of each student, noting that high-quality child-AI interactions are crucial for the successful integration of LLM into educational settings.

Despite the promise of LLMs, their use in educational settings raises significant ethical and pedagogical concerns (Adnin et al., 2025; Harvey et al., 2025). Park and Ahn (2024) identified challenges including hallucinations and algorithmic bias, alongside concerns about diminished learning and academic integrity (Park & Ahn, 2024). Similarly, Harvey et al. (2025) identified risks ranging from biased or inaccurate content and privacy violations to broader harms such as inhibited student learning, reduced educator autonomy, and educational inequities (Harvey et al., 2025). Additionally, the ability of LLMs to generate content presents both opportunities and challenges in redefining plagiarism and co-authorship (McGuire, 2023; Tajik, 2025). The systematic review by Yan et al. (2024) further emphasizes the necessity of careful evaluation of the ethical implications associated with LLM use to prevent misuse or negative outcomes, which may hinder their acceptance within academic circles. This dynamic calls for urgent discussions around responsible usage and the establishment of standards to ensure that the educational benefits of these technologies do not come at the cost of ethical standards (Cotton et al., 2024; Tanksley et al., 2025). This paper builds on these tensions by examining how teachers articulate acceptable assistance and how those commitments can be translated into concrete design constraints and scaffolds for classroom writing.

2.3. Research co-design in AI-mediated learning technology

Participatory design (PD) methodologies emphasize meaningful collaboration between designers and end users in the creation of tools, resources, and experiences that reflect the values and goals of the intended audience (Muller & Kuhn, 1993; Scaife et al., 1997). PD provides concrete strategies, such as co-design workshops and collaborative prototyping, that operationalize these values through power-sharing and situated practice (Nicholson et al., 2022; Yip et al., 2013). In the context of AI research in education in the HCI community, there is a growing emphasis on leveraging PD approaches that bring together educators, students, and researchers to collaboratively design and implement educational innovations (He et al., 2025; Xie et al., 2024). Involving educational stakeholders (e.g., students and teachers) in the design process helps address the diverse and nuanced needs of real-world learning environments, aligning technological advancements with pedagogical priorities to ensure that solutions are both theoretically grounded and practically relevant (Kali et al., 2015; Lawrence et al., 2022).

We employ participatory methods, specifically co-design, to collaboratively identify which features and functions of AI-based authoring tools are the most meaningful, usable, and equitable in writing instruction contexts. Prior research has shown how involving educators in the design process leads to tools that better align with instructional workflows, classroom culture, and diverse student needs (Hutchins & Biswas, 2024; Stefaniak et al., 2025). Similarly, co-design processes have been employed to create AI-driven tools for classroom orchestration, allowing educators to have a voice in the iterative design of classroom technologies (Chang & Philip, 2025; Tatar et al., 2025). While prior work demonstrates the value of participatory design for developing educational AI systems (Hutchins & Biswas, 2024; Ravi et al., 2025), less attention has been given to how teacher participation can be translated into configurable AI scaffolds grounded in learning theory. We address this gap through the theoretical framework described below.

2.4. Guiding theoretical framework

This study is guided by Participatory Design (PD) (Scaife et al., 1997), Vygotsky's Zone of Proximal Development (ZPD), and scaffolding theory (e.g., Gibbons (2002)). These perspectives frame teachers as active contributors to AI design and explain how AI support can help students engage in writing tasks beyond what they can accomplish independently. We integrate these perspectives through a theory of change centered on teacher configurability, connecting teachers' pedagogical expertise to the design and enactment of AI assistance in classroom writing activities.

Therefore, PD serves as both a theoretical framework and a methodological approach. Theoretically, PD is grounded in sociocultural and critical design traditions, emphasizing equity, inclusivity, and the recognition of teachers as co-creators rather than passive recipients of technology (Scaife et al., 1997). Our design processes of AI-powered story-authoring tools are grounded in sociocultural learning theory, particularly Vygotsky's (Semenovich Vygotsky & Cole, 1978) concept of the Zone of Proximal Development (ZPD) and scaffolding theory (Gibbons, 2002), to analyze how AI supports students in performing tasks that exceed their current independent capabilities (Sanders & Welk, 2005). Pea and Kurland (1987) argued that the design of cognitive technology for writing must be informed by the pedagogical practices of expert writing instructors, highlighting the need to translate effective teaching strategies into system design principles, enabling computer-assisted writing environments to emulate high-quality human feedback and scaffold writers' development over time.

Building on these perspectives, we conceptualize teacher configurability (i.e., teachers' ability to configure the type, timing, and level of AI assistance) as the mechanism that connects participatory design with scaffolding theory (Gibbons, 2002). Through participatory design, teachers translate their pedagogical expertise into configurable AI-supported writing activities, including prompts, genre-specific scaffolds, and constraints on AI assistance (Kang et al., 2025). Rather than treating scaffolding as a fixed capability of the AI system, a teacher-configurable AI writing system enables teachers to calibrate the type, timing, and intensity of AI support according to instructional goals and teachers' understanding of students' ZPD. In this way, participatory design provides the process through which pedagogical knowledge is translated into configurable AI behaviors, while scaffolding theory explains how these teacher-configured supports position learners within their ZPD across different classroom contexts.

In literacy education, scaffolds can include modeling strategies, providing prompts or offering feedback to support students in articulating ideas, organizing text, and refining language use, offering real-time responses that guide students in tackling writing tasks that would be difficult to accomplish independently (Nordlof, 2014). Previous research has shown that such scaffolds improve critical writing skills, including coherence and vocabulary development (Graham et al., 2012; Traga Philippakos et al., 2018). GenAI chatbots extend this tradition by enabling real-time, adaptive scaffolding, facilitating brainstorming, encouraging revision, and structuring the writing process (Ait Baha et al., 2024; Jeon & Lee, 2024).

Accordingly, our theory of change positions scaffolding as a participatory and teacher-configurable process that evolves through iterative cycles of design and educational engagement. This lens guided our research design across four phases (described in Section 3). These phases respond to our research questions by demonstrating how teacher input informs platform design, how instructional goals materialize as scaffolding strategies, and how student-AI interactions reveal both opportunities and challenges for integrating GenAI into classroom literacy practice (Park & Ahn, 2024). We aim to extend scaffolding theory into the context of human-AI collaboration for both the pedagogical and design implications of AI-mediated writing support (Lee et al., 2024).

3. Overarching study design & research methods

The iterative co-design and development of StoryBot unfolded across four phases:

1. **Phase 1: Co-Design a Logic Model with Teachers.** We conducted 90-minute individual co-design sessions with 21 K–12 classroom teachers from across the United States (U.S.). These sessions engaged teachers in a participatory process to collaboratively develop a logic model, aiming to

understand the usability of current GenAI platforms and youths' strategies and experiences with AI-supported writing.

2. **Phase 2: Designing StoryBot.** This phase involved iterative design and development of StoryBot through collaboration with a multidisciplinary team of PhD students ($n=7$), including HCI and learning sciences researchers ($n=3$), interaction designers ($n=2$), and software engineers ($n=2$). All team members had formal training and research experience in educational technology, child-computer interaction, and AI-enabled system design. The goal was to design a system that supports young learners in developing narrative writing and storytelling abilities.
3. **Phase 3: Pilot Study with Students.** Once the platform was developed, we conducted a 2-day writing workshop with 13 students at a community center on the West Coast, USA. This phase investigated the applicability and efficacy of StoryBot, examining its implications for writing education among young learners.
4. **Phase 4: Teachers' Focus Group Workshop, Feedback on StoryBot.** We conducted a three-hour remote focus group workshop with six classroom teachers to explore their perspectives on the pedagogical role of StoryBot and its potential impact on teaching and learning. Our goal was to elicit feedback on the platform's design, functionality, and instructional relevance to inform future iterations.

Each phase is described in detail below, including sample characteristics, procedures, and key findings.

3.1. Ethical approval and informed consent

This study was conducted in accordance with the ethical principles of the Declaration of Helsinki and was approved by the Institutional Review Board (IRB) at the University of California, Irvine (UCI; Protocol No. 2356) prior to data collection. All adult participants provided verbal informed consent after reviewing the study procedures and prior to participation and recording. For the student study, written informed consent was obtained from a parent or legal guardian prior to participation, and children provided oral assent at the beginning of the workshop after the study procedures were explained to them. Caregivers were informed that students would interact with an LLM-based writing tool during the study. To protect student privacy, the research platform did not require user accounts or collect personal identifying information, and student interaction data were stored locally and securely deleted after analysis.

3.2. Phase 1 co-design a logic model

The co-design phase aimed to meaningfully involve teachers in shaping GenAI technology around their instructional needs and classroom contexts (Hutchins & Biswas, 2024; Stefaniak et al., 2025). Rather than focusing primarily on interface design, this phase focused on co-designing a theory of change and identifying the pedagogical and equity conditions under which GenAI could support K-12 writing instruction.

We therefore used a logic model, an established approach in educational program planning and evaluation for connecting resources and activities with intended outputs and learning outcomes (Kekahio et al., 2014; Knowlton & Phillips, 2012). Logic models have also been applied in educational technology to articulate how users, platform features, and activities contribute to intended educational outcomes (Kao et al., 2018). This structure was particularly appropriate for our research because our goal was not only to identify desired system features, but to understand how teachers connected those features to instructional practices and learning goals. Teachers therefore co-constructed the logic model as a shared framework for identifying contextual needs, prioritizing features, and aligning subsequent design decisions with intended literacy outcomes (Nicholson et al., 2022; Ravi et al., 2025).

3.3. Participants

We conducted 21 individual co-design sessions with K–8⁷ teachers between Fall 2023 and March 2024. Each session lasted approximately 90 min and was conducted via Zoom. In total, 21 teachers participated (16 female, 5 male; see Table 1). Nine participants were affiliated with the National Writing Project (NWP), a professional development network focused on strengthening writing instruction (Kaplan, 2008). All instruction and writing in this context are in English, and, given the regional context, classrooms tend to be linguistically diverse, including students formally identified as English-language learners (ELLs), i.e., students who speak a home language in addition to English.

Participants were recruited via snowball sampling through professional and personal networks. Eligibility criteria required that participants be current or former K–12 teachers. The teachers represented grades 1–8 (ages 6–14), with an average of 14.8 years of experience (range: 2–31 years). While most reported little to no prior experience with GenAI platforms such as ChatGPT, Gemini, or Claude, all reported frequent use of digital tools for reading and writing instruction (e.g., Epic,⁸ StoryBird,⁹ Canva,¹⁰ MagicSchool.ai,¹¹ and Google Docs¹²). The teacher sample represented multiple U.S. locations and included variation in teaching experience, grade levels, subject areas, and familiarity with GenAI, providing a range of perspectives within the study context.

3.4. Procedures

Each 90-minute co-design session followed a consistent structure (Table 2):

Teachers were first introduced to the capabilities of GenAI (e.g., text generation, synthesis, summarization) and basic prompt engineering strategies (e.g., role definition, audience specification, tone/style parameters). Educational materials were adapted from Google's GenAI teaching resources (i.e., Grow with Google¹³). Next, teachers engaged in a hands-on activity with Gemini, exploring classroom applications such as text-to-speech, speech-to-text, translation, genre-based text generation, and image

Table 1. Educator co-design participants, ELA (English Language Arts, U.S. subject area for reading and writing).

Teacher	City, State	Specialty area	Grade(s) taught	Years of teaching	Familiarity with GenAI
Ms. Kim	Los Angeles, CA	ELA	3-6	10	Limited
Ms. Bennett	San Diego, CA	ELA	3-6	14	Limited
Ms. James	Oceanside, CA	General	3-6	13	Limited
Ms. Garcia	San Diego, CA	ELA	1-6	31	Limited
Mr. Ethan	Los Angeles, CA	ELA	1-6	7	Moderate
Ms. Rachel	Los Angeles, CA	ELA	1-6	11	Moderate
Mr. Lee	Encinitas, CA	General	3-8	12	Limited
Mr. Brian	San Diego, CA	General	3-8	24	Limited
Ms. Jina	La Habra, CA	Math	3-6	18	Limited
Mr. Daniel	Hamlet, NC	Social Studies	3-6	21	Limited
Ms. Laura	Apex, NC	ELA	3-6	6	Limited
Ms. Megan	Raleigh, NC	ELA	3-8	5	Limited
Ms. Grace	Oceanside, CA	ELA	3-8	30	Limited
Mr. Sean	Issaquah, WA	General	3-8	12	Moderate
Ms. Elise	Oceanside, CA	General	3-6	13	Fluent
Ms. Lily	Oceanside, CA	General	3-6	2	Fluent
Ms. Olivia	San Diego, CA	ELA	3-6	24	Limited
Ms. Lin	Orange, CA	Math	3-8	8	Moderate
Ms. Molly	Raleigh, NC	General	1-6	14	Moderate
Ms. Sophia	La Habra, CA	General	1-6	20	Limited
Ms. Ingrid	Orange, CA	General	1-6	15	Moderate

Table 2. An overview of the co-design schedule.

Duration	Activities
10 mins	Welcome and Introduction
15 mins	Lesson about Generative AI and Logic Model
15 mins	Prompt Engineering Activity with Gemini
15 mins	Semi-structured Interview (Open-discussion)
25 mins	Create a Logic Model
10 mins	Q & A and Open Discussion

Co-Designing Literacy Tools: Crafting a Logic Model for a GenAI Story-Authoring Platform

Please list your ideas in the tables provided and think aloud as you work through each section. Your verbal reflections will help us better understand your design decisions and priorities.

Name:

Context & Users	Inputs & Features	Outputs & Outcomes	Ultimate Goals
Who are the primary user groups this product is designed for? - Idea 1 - Idea 2 - Idea 3	What features and functionalities does the system offer? - Idea 1 - Idea 2 - Idea 3	What kinds of user-generated data will be collected or produced during interaction? - Idea 1 - Idea 2 - Idea 3	What overarching goal does the system aim to achieve? - Idea 1 - Idea 2 - Idea 3

Figure 1. Example of a logic model blank worksheet distributed to participants outlining processes in the logic model.

generation. This activity encouraged open-ended ideation, allowing teachers to imagine features beyond the provided examples.

We then introduced the logic model framework to map relationships among inputs, activities, outputs, and outcomes. Using a shared Google Slides template (Figure 1), teachers populated components of the model by identifying intended users, classroom contexts, desired AI functionalities, and anticipated learning outcomes. Throughout this process, participants engaged in a think-aloud protocol, articulating their reasoning as they completed the model. Semi-structured interviews were conducted in parallel to probe rationales for design choices (e.g., prioritizing “story maps” as features or emphasizing “lexical diversity” as an outcome). All sessions were recorded, transcribed, and supplemented by the completed logic model worksheets (Figure 1).

3.5. Data sources and analytical techniques

To contextualize findings, 21 participants completed a pre-survey capturing demographic information, teaching background, and prior experience with GenAI. To address RQ1, we analyzed three primary data sources: (1) completed logic model worksheets; (2) think-aloud transcripts; and (3) semi-structured interview transcripts.

We employed a hybrid coding approach (Azungah, 2018; Saldaña, 2013). Deductive codes were derived from the logic model framework (e.g., instructional context, inputs, outputs, outcomes), while inductive coding captured emergent themes (Low, 2019). Two researchers independently coded transcripts, iteratively refining a shared codebook over eight weeks. Inter-coder reliability reached 0.87 Cohen’s kappa.

Following initial coding, we conducted a comparative analysis by mapping our categories against established models of writing education (e.g., Bradford et al. (2016), Coker et al. (2016), Graham et al. (2012)) and foundational frameworks on writing processes (Flower & Hayes, 1981). Although the categories emerged inductively, this alignment helped situate our findings within broader theoretical and pedagogical contexts (Table A.1). Finally, based on teacher-identified goals and rationales, the research team synthesized results into a systematic logic model.

3.6. Phase 1 findings

Our findings were organized into four thematic categories aligning with parts of the logic model: (1) context and user identification, (2) inputs and feature prioritization, (3) outputs and outcome

Table 3. Teacher-identified context and user identification for AI writing support.

User group	Key challenges	Illustrative teacher quote
Emergent writers	Difficulty with planning, organizing, and revising texts	<i>"They often get stuck before they can organize their ideas. Having AI guide them through planning could help."</i> – Ms. Garcia, Grade 1-3, 31 years of experience
Struggling readers	Limited literacy fluency, difficulty generating coherent text	<i>"Some students can't get past the first draft; AI could help them produce a first version to revise."</i> – Mr. Ethan, Grade 4-6, 7 years of experience
English Language Learners (ELLs)	Limited vocabulary, language barriers	<i>"Our school has many multilingual students and ELLs. I think tools like this can be especially helpful for them to learn new vocabulary and sentence structures as they write."</i> – Ms. Kim, Grade 3-6, 10 years of experience
Teachers	Time constraints, need for differentiated instruction, personalized feedback	<i>"It's not just for students—AI could help us plan lessons and provide individualized feedback more efficiently."</i> – Ms. Olivia, Grade 3-6, 24 years of experience

alignment, and (4) desired goals derived through a deductive approach. Throughout, we tracked ethical considerations and design strategies generated inductively from teacher interviews and think-aloud protocols.

3.6.1. Context and user identification

Teachers highlighted emergent writers, struggling readers, ELLs, and teachers themselves as the primary users of AI writing support. Table 3 summarizes these groups and the key challenges they face in the classroom.

Across sessions, teachers consistently emphasized that AI support should be targeted and scaffolded, focusing on students who face the greatest literacy challenges while also enhancing teacher capacity. For emergent writers and struggling readers, teachers hypothesized ways that AI could help structure ideas, prompt revisions, and model coherent writing. For ELLs, in particular, language scaffolds and translation tools were viewed as essential to enabling meaningful participation in writing tasks.

Teachers also positioned themselves as active users: the platform should allow customization of prompts, lesson plans, evaluation criteria, and sample responses. Importantly, educators described the need for AI to be a collaborative partner, not a replacement, extending their ability to provide equitable, personalized instruction across diverse classrooms.

3.6.2. Inputs and feature prioritization

Teachers identified six interconnected design inputs for GenAI-supported writing, spanning scaffolding, planning, genre-specific instruction, inclusion, engagement, and teacher configurability (Table 4). These priorities reflect teachers' broader emphasis on supporting writing development while maintaining equitable access and instructional control.

Scaffolded writing support was the most frequently identified priority (16/21 teachers), particularly for helping emergent writers elaborate and structure their ideas. Personalization was closely related to this priority, with 15/21 teachers emphasizing student choice in writing topics as a means of sustaining engagement. Teachers also prioritized planning tools such as story maps and genre-specific scaffolds to support idea organization and students' understanding of structure, purpose, voice, and audience.

Equity and inclusion were also prominent. Twelve of 21 teachers prioritized text-to-speech, speech-to-text, and multilingual support, while 13/21 emphasized developmentally appropriate, genre-specific mentor texts. Similarly, 13/21 identified gamification and culturally responsive imagery as potential supports for engagement and representation.

Finally, teacher configurability emerged as a central requirement, with 14/21 teachers emphasizing control over assignments, lesson settings, and students' use of AI assistance. Across these priorities, teachers envisioned GenAI as a configurable scaffold that could support diverse learners while preserving teacher oversight of instructional decisions.

3.6.3. Outputs and outcome alignment

Twelve out of 21 teachers provided detailed insights about the types of data and learning outcomes that would make an AI writing platform pedagogically meaningful. They emphasized that outputs should align with measurable writing development goals and support their instructional practices (Table 5).

Table 4. Teacher-identified logic model inputs, rationales, and representative quotes.

Input	Rationale	Representative teacher quote
Scaffolded supports for writing development	AI as a writing partner that encourages elaboration, structure, and student choice.	<i>"My students, especially those just beginning to write, often stop after a few sentences and say they're done. I think AI could help engage them to keep going, to add more detail to their writing. I want them to understand that writing has structure; there's a beginning, and there's an ending."</i> – Ms. Garcia, Grade 1-3, 31 years of experience
Visual scaffolding: Story mapping tools	Visual scaffolds to support idea generation and narrative structure planning.	<i>"I think AI could be especially helpful as students build out their story ideas, brainstorming together with the AI. A story map is useful because it helps them lay out the storyline before they begin writing."</i> – Ms. Olivia, Grade 3-6, 24 years of experience
Genre-specific writing as a developmental trajectory	Writing instruction progresses through genre stages, fostering skills in voice, perspective, and purpose.	<i>"In my school district, younger students start with creative, narrative writing, both fiction and nonfiction. By around fourth grade (9–10 years old), they begin developing skills for opinion writing. It's important for them to think about the purpose of their writing, their voice, viewpoint, and the intended audience. These are all essential elements of writing education."</i> – Ms. Grace, Grade 3-6, 30 years of experience
Inclusive features for ELL/multilingual students	Text-to-speech, speech-to-text, and customized mentor texts to reduce barriers and provide differentiated support.	<i>"Finding strong mentor texts that fit what I'm teaching and hit the exact skills I want students to practice is really time-consuming. That's a big part of our work as teachers. I think AI could really help by generating leveled writing models that meet students where they are."</i> – Ms. James, Grade 3-6, 13 years of experience
Culturally responsive image generation	Gamified, representative imagery to motivate reluctant writers and support diverse identities.	<i>"Who's your character? Are they white? Are they Black? You know, so that now they can have something that represents them. Right? If they're a Black kid, they want to see a story about a Black kid, they don't want to see a white kid with blond hair all the time."</i> – Ms. Bennett, Grade 3-6, 14 years of experience
Teacher configuration	Dashboards and controls for class creation, assignments, and monitoring AI use.	<i>"It would be helpful to have the ability to create classes and assign specific writing tasks to students, including setting topics and guidelines. Additionally, I'd like to track how frequently students use AI assistance throughout the writing process."</i> – Mr. Lee, Grade 3-8, 12 years of experience

Teachers emphasized that AI-generated outputs should directly connect to learning goals. Metrics like word count, essay frequency, child-AI conversation, time on task, and lexical diversity were valued because they provide actionable insights into fluency, engagement, and vocabulary development, while alignment with syntax, spelling, and verbal reasoning ensures that outputs remain instructionally meaningful. Ms. Garcia (Grade 1-3) summarized the overarching instructional goal:

"I think what you want your writers to do is to be able to improve their writing, by improving their workflow choices, their structure, sentence skills, organization, and paragraphing."

By centering outputs on measurable, goal-aligned indicators, the platform can provide teachers with data that supports formative assessment, informs differentiated instruction, and ensures that AI interventions strengthen existing literacy practices. These insights directly shaped the logic model, linking student data outputs to short- and long-term learning outcomes.

3.6.4. Desired goals

Teachers identified both short- and long-term goals for GenAI-supported writing. Short-term goals focused on writing skills, including spelling, grammar, and lexical diversity, while long-term goals emphasized self-efficacy, motivation, and engagement. Teachers particularly noted that emergent writers

Table 5. Output/metric and instructional goal.

Output/metric	Purpose/instructional goal	Representative teacher quote
Word count	Track fluency and writing volume	<i>"How much students write really shows me their fluency. A lot of times, my kids will write one sentence and think they're done. That's when I step in and ask, 'Can you add more detail?' It encourages them to stretch their thinking and expand their writing."</i> - Ms. Garcia, Grade 1-3, 31 years of experience
Number of essays completed	Monitor engagement and practice frequency	<i>"I'd love to see students come back and write more essays over time. That consistency not only shows their engagement but also helps nurture a genuine love of writing."</i> - Ms. Laura, Grade 3-6, 6 years of experience
Time on task	Evaluate student focus and effort	<i>"What I really look for is whether they're staying with their writing. It's about seeing how much focus and effort they're putting into it."</i> - Mr. Daniel, Grade 3-6, 21 years of experience
Lexical diversity & Verbal reasoning	Support vocabulary development & Assess higher-order writing skills	<i>"Vocabulary is such a big part of learning to read and write well. For all kids, especially ELLs and emergent writers, it's important that they not only learn new words but also practice using them in different ways."</i> - Ms. Megan, Grade 1-6, 14 years of experience
Syntax, spelling, paragraphing	Align feedback with specific writing skills	<i>"The basics, like spelling, grammar, and paragraphing take a lot of practice. Kids really learn through making mistakes and then correcting them."</i> - Ms. Olivia, Grade 3-6, 24 years of experience

Table 6. Teacher concerns & design suggestions for ethical GenAI writing support.

Theme	Concern (with prevalence)	Representative teacher quote
Agency & active engagement	12/21 teachers worried students might passively accept AI text instead of engaging in writing.	<i>"I think the most important part is ensuring that kids don't just copy from AI, but instead think independently and write in their own words."</i> - Ms. Rachel, Grade 4-6, 11 years of experience
Authenticity & student voice	15/21 teachers worried AI text may suppress students' unique voices and limit creativity.	<i>"It would be really helpful to have a teacher-facing interface that allows us to customize the level of AI assistance provided to students."</i> - Mr. Brian, Grade 3-8, 24 years of experience
Dependency & skill development	12/21 teachers concerned students may become over-reliant on AI, hindering independence.	<i>"My biggest concern is that my students depend on and rely on AI most of the time, so they cannot write a story without an AI assistant."</i> - Mr. Lee, Grade 3-8, 12 years of experience <i>"For example, in a classroom setting, I might allow AI use only during ideation, whereas in one-on-one tutoring, I could monitor how students write asynchronously."</i> - Mr. Sean, Grade 3-8, 12 years of experience

often struggle not only with writing skills but also with initiating, sustaining, and revising their work. As Ms. James (Grade 3–6) explained:

"I think that as a writing teacher, a lot of us want the students to feel curious, feel confident, feel motivated, feel activated as writers are encouraged. I think a lot of writing teachers will say 'I can work with the student whose writing is weak', but getting them to start, getting them to continue and finish, and then getting them to revise. Those are the psychological hurdles, right?"

This insight captured a sentiment expressed across multiple interviews, that while educators feel equipped to address technical writing skills through traditional instruction, the motivational and emotional barriers often prove more challenging to overcome. Teachers explained that their students' reluctance to begin writing, difficulty sustaining effort, and resistance to revision created significant instructional obstacles that an AI platform could uniquely address by providing personalized, non-judgmental support for both skill development and confidence building.

3.6.5. Ethical considerations

Throughout the co-design process, teachers raised ethical and pedagogical concerns about how GenAI writing platforms might affect student learning. Three interconnected themes emerged: *agency*, *authenticity*, and *dependency* (Table 6).

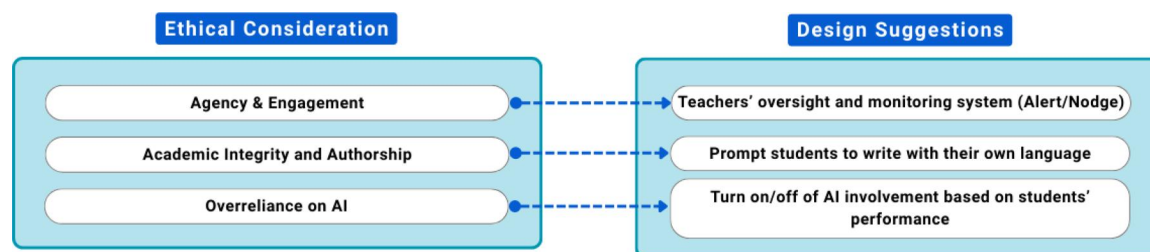


Figure 2. Ethical consideration and design insight from the logic model development.

Twelve of 21 teachers expressed concerns that students might passively accept AI-generated text rather than actively engage in writing. In response, they recommended monitoring and transparency features that distinguish student and AI contributions. Authenticity and student voice were the most prevalent concerns (15/21), leading teachers to advocate for configurable levels of AI assistance that provide scaffolding without replacing students' own expression. Dependency was similarly raised by 12/21 teachers, who recommended constraints on when and how AI could be used, such as supporting ideation while limiting assistance during drafting.

All in all, these findings suggest that classroom GenAI tools should preserve *agency*, protect *authenticity*, and mitigate *dependency*. Teachers translated these concerns into concrete design recommendations, including monitoring, adaptive assistance, and configurable AI settings, which informed the logic model (Figure 2).

4. Phase 2: Design & implementation phase

Insights from the development of the logic model in Phase 1 provided a foundation for the StoryBot platform as a design probe, an exploratory research instrument to learn about how students and teachers engage with writing and AI in authentic ways (Hutchinson et al., 2003). StoryBot is a custom web-based story-authoring environment designed to scaffold students through the writing process, from brainstorming and planning to drafting. The platform functions simultaneously as a learning tool and a research instrument, allowing for systematic comparison of AI-driven scaffolds while preserving ecological validity in authentic classroom writing tasks.

4.1. Design strategies & platform overview

Phase 1 identified recurring instructional challenges across learner groups and contexts. As summarized in Table 7, we translated teachers' identified user needs, prioritized design inputs, desired learning outcomes, and ethical considerations (Tables 3–6) into instructional challenges that informed StoryBot's design strategies and platform features.

StoryBot combines an interactive AI chatbot with visual story-mapping and writing tools (Figure 3). Figure 3 illustrates the end-to-end workflow. Teachers first configure lessons by selecting writing topics and providing multimedia resources and mentor texts aligned with instructional objectives (Figure 3a). Students then select a topic and receive guided prompts from the chatbot before transitioning to the writing editor (Figure 3b). The editor organizes writing into genre-specific sections and provides on-demand access to teacher-provided mentor texts, images, and story maps. Mentor texts may include exemplar essays, prompts, or instructional guidance and serve as reference materials rather than sources for generating complete essays.

To support students in initiating and developing their writing, StoryBot uses guided brainstorming and connects topics to students' interests. For opinion writing, prompts scaffold the development of an opinion, supporting reasons, and relevant evidence (Table 7, Rows 1–3). Importantly, StoryBot does not generate sentence-level text for students; instead, it uses prompt-based guidance to encourage students to compose in their own words (Table 7, Row 2). Table 8 illustrates students' responses to these prompts, while Table A.2 summarizes the prompting strategies used to support writing without replacing student authorship.

Table 7. Design strategies and platform features synthesized from the teacher-identified logic model (Tables 3–6).

Logic model findings	Derived design challenge	Design strategies	Platform features
1 Genre-specific writing progression; purpose, voice, and audience (Table 4)	Learners struggle to understand rhetorical purpose and audience	Genre-specific writing as a developmental trajectory; modular integration into narrative, opinion, and AI literacy lessons	• Age- and grade-appropriate writing prompts • Scaffolds for voice, audience awareness, and purpose
2 Scaffolded writing development; vocabulary and language support (Tables 4 and 5)	Learners use limited vocabulary and language	Scaffolded writing supports with guided language suggestions	• Sentence- and paragraph-level scaffolds • Guided elaboration suggestions • Optional narrative pathways for idea expansion
3 Story mapping and planning supports for writing (Tables 3 and 4)	Learners struggle to initiate, organize, or complete writing tasks	Story mapping tools for planning and organizing ideas	• Interactive planning for characters, plot events, and story arcs
4 Differentiated instruction, ELL supports, and inclusive learning (Tables 3, 4, and 6)	Learners have diverse language and literacy support needs	Inclusive supports for multilingual and emerging writers	• Text-to-speech and speech-to-text • AI-generated leveled mentor texts • Choice-driven prompts (genre, topic, images)
5 Learning outcomes and formative assessment metrics (Table 5)	Teachers need to monitor student progress and writing behaviors	Learning analytics and teacher dashboards	• Word count • Time on task • Lexical diversity • Number of essays
6 Ethical considerations: student agency, teacher oversight, and responsible AI use (Table 6)	Teachers require oversight of AI use while preserving student agency	Teacher-configurable AI with monitoring and instructional controls	• Assignment management • Prompt customization • Student progress dashboard • AI usage analytics • Text-similarity analysis

StoryBot also incorporates a text-similarity feature to support teacher awareness of AI-student co-writing patterns (Wang & Dong, 2020). The feature compares AI-generated text presented during an interaction with the student's subsequent writing. Both are represented as 768-dimensional embedding vectors, and cosine similarity is used to estimate semantic overlap. Higher scores indicate greater similarity and provide teachers with an additional signal of potential AI influence; the measure is not intended as a definitive indicator of plagiarism or authorship.

Finally, StoryBot's features reflect teachers' complementary short- and long-term goals. Short-term goals include writing competencies such as spelling, grammar, and lexical diversity, supported through writing feedback, vocabulary prompts, and writing metrics (Table 7, Rows 2 and 5). Long-term goals include self-efficacy, engagement, revision, and persistence. These goals are supported through personalized scaffolding, multimodal supports such as text-to-speech and speech-to-text, and choice-driven interactions with prompts, images, and story elements (Table 7, Rows 2 and 4).

4.1.1. Teacher configuration

Teachers emphasized the importance of controlling lesson design, particularly the selection of topics aligned with their instructional goals. To support this need, we developed a teacher configuration interface (Figure 3) that enables teachers to design writing activities before they are delivered to students. Teachers can configure writing topics, genre-specific activities, multimedia resources (e.g., videos and images), vocabulary lists, and mentor texts aligned with their instructional objectives. The interface provides teachers with meaningful control over AI-supported instruction while supporting differentiated teaching and preserving student agency (Table 7, Row 6).

Each lesson functions as a configurable template that preserves its core pedagogical sequence while allowing teachers to adapt instructional content across contexts. For example, in the Narrative Writing module, teachers can specify available genres (e.g., biography, fantasy, or fable). If a student selects "Biography," the system dynamically adapts the instructional prompts: the LLM provides a concise explanation of the genre, generates genre-specific scaffolded ideas, and tailors subsequent writing activities accordingly (Figure 4). Thus, teachers can modify elements such as genres, themes, prompts, and examples without altering the underlying instructional scaffold. Drawing on template-based instructional design and learning-object principles, this approach enables multiple learning objects to be

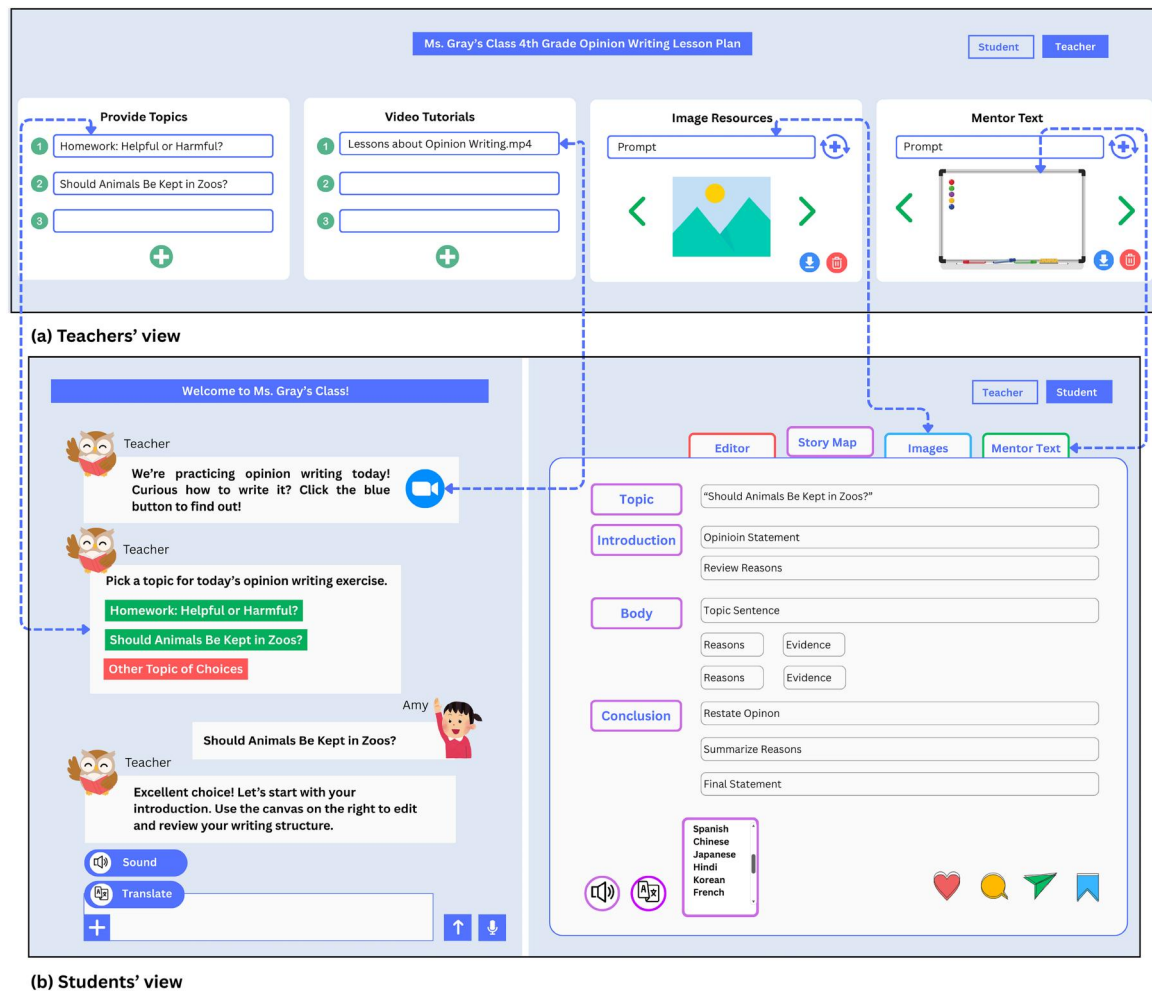


Figure 3. (a) Teacher customization panel for lesson configuration; (b) students' view *left*: student–AI chatbot interface with option buttons and input field with multimodal interaction (sound, translation); *right*: workspace panels including editor, story map, image resources, and mentor text repository.

generated from a reusable lesson template while maintaining consistent educational goals (Hedderich et al., 2024; Kang et al., 2025).

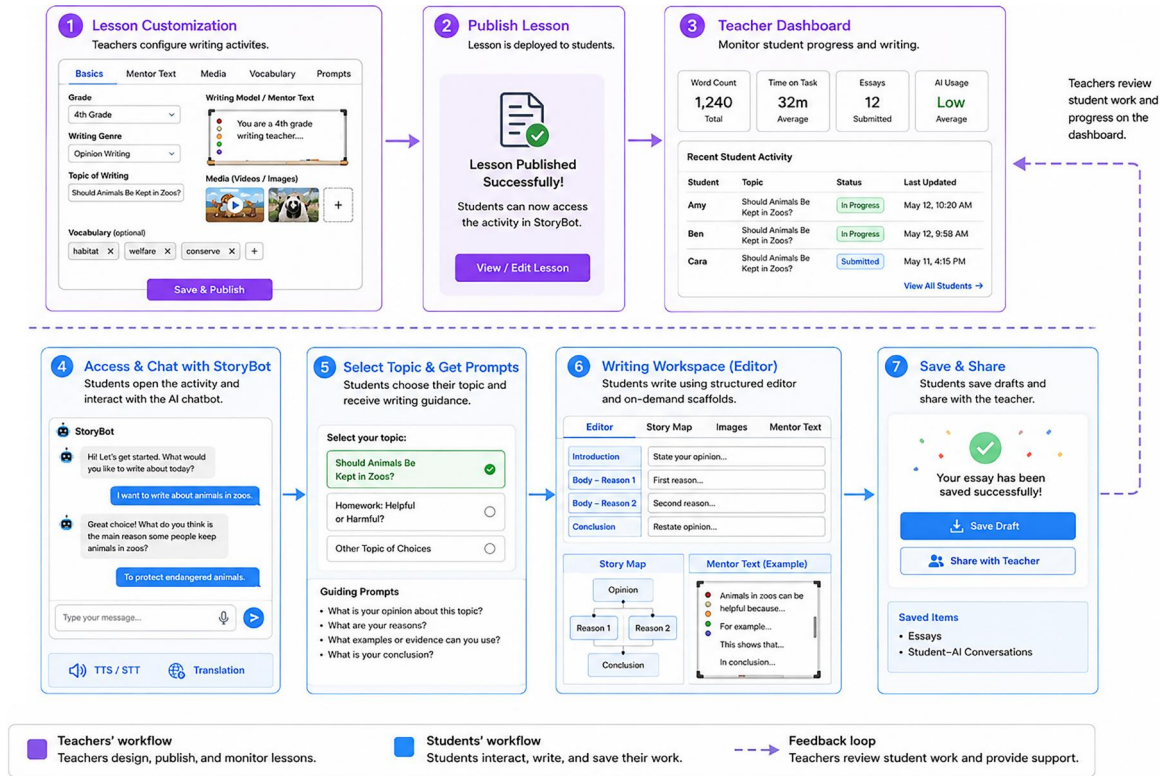
To protect students' privacy when using LLMs, we adopted four safeguards: (1) StoryBot uses open-source LLM models (e.g., Llama 3.3, gpt-4o-mini) and localized LLM instances (e.g., gpt-3.5-turbo), with usage data stored locally and securely discarded after analysis; (2) the research platform does not require sign-in or collect personal information; (3) caregivers were informed during consent that students would interact with an LLM-based chatbot during writing; and (4) during workshops, two staff members or researchers were assigned to each table to monitor students' LLM interactions, maintaining a 3:1 student-to-staff ratio (Table 7, Row 6).

5. Phase 3: Pilot study with students

Following co-design sessions with teachers and iterative prototyping, we conducted a pilot study to explore how students engaged with StoryBot during authentic writing tasks. This phase of the research examined the affordances and limitations of the platform's scaffolding in shaping students' learning behaviors. Specifically, this study sought to capture both quantitative indicators of writing fluency, volume, and genre-specific strategy use, as well as qualitative insights into students' cognitive and affective engagement.

Table 8. Examples of StoryBot prompts supporting student elaboration.

Writing topic	StoryBot prompt	Student response
Should animals be kept in zoos?	"That's a great reason! Can you provide an example?"	"For example, pandas are kept safe from extinction in zoos because they are protected and fed."
Homework: Helpful or harmful?	"Let's add another paragraph to explain your opinion."	"I think homework is helpful because it helps me remember things I learned in school."

**Figure 4.** StoryBot workflow.

5.1. Procedure

The writing workshop consisted of two consecutive 90-minute sessions and was designed to evaluate how StoryBot translated into students' writing experiences. We focused on assessing both the outputs identified in our logic model (i.e., increased writing time, word count, and use of genre-specific strategies) and the outcomes specific to the opinion writing module. The selected outputs were chosen deliberately to capture both short-term, measurable writing skills and indicators of engagement, reflecting the dual emphasis articulated by teachers during co-design. For example, metrics such as word count and writing time serve as proxies for fluency and persistence, while lexical diversity and syntactic markers capture foundational language growth. Similarly, genre-specific strategies were emphasized because opinion writing requires the integration of claims, reasons, and evidence—skills teachers flagged as developmentally important yet challenging for students at this age. Ultimately, we aim to measure a balanced view of technical writing progress and broader motivational engagement.

Data collection employed multiple complementary methods, including system log analysis (e.g., writing time, word count, number of essays completed), video recordings, and think-aloud protocols. This multi-method approach enabled triangulation of both quantitative measures of writing outputs and qualitative insights into the cognitive and emotional aspects of students' engagement with StoryBot.

5.1.1. Participants

Thirteen youth participants, aged 8 to 12, were recruited through snowball sampling from our institution via advertisements on mailing lists for an in-person writing workshop featuring the first iteration

of StoryBot. This age range was intentionally selected to reflect the target user group identified during co-design sessions with teachers: emergent and developing writers in upper elementary school. This developmental window is widely recognized as critical for establishing foundational reading and writing competencies that shape long-term academic success (Kara-Soteriou & Kaufman, 2002; Wilcox et al., 2016).

The final participant group consisted of 13 students (average age 10.2 years). Eight of the participants were girls. Approximately one-third spoke English as their primary language, while the remaining two-thirds spoke additional languages, predominantly Chinese and Korean. The majority of the participants ($n = 10$) identified as Asian, with two identifying as White, and one as multiracial (Table 9).

In terms of prior AI experience, all participants reported familiarity with voice assistants such as Alexa or Google Home. Two used these tools monthly, while six reported more frequent use (weekly or daily).

5.1.2. Study procedure

The StoryBot writing workshop was conducted at the institute's facility, located in a suburban area on the West Coast of the United States. Two researchers facilitated the workshop with the support of one assistant. One researcher conducted the semi-structure interviews (e.g., what are you writing about?, what do you like about StoryBot, do you think it helps or distract your writing?), while the other, together with the assistant, led the workshop activities and provided technical support, such as helping students set up their laptops and connect to the internet. We employed a within-subjects design in which participants completed writing activities in both the StoryBot and Google Docs conditions. To control for potential order effects, participants were assigned to one of two counterbalanced condition sequences (Table 10). In both conditions, students were not restricted by time limits but instead allowed to write freely, enabling us to observe their writing behaviors in a more natural setting.

We obtained parental consent and children's assent after explaining the purpose of the study and what students would be working on. During the recruitment process, parents were invited to try out the StoryBot tool as part of the consent survey. The survey collected demographic information (e.g., age, gender, location, parental education level) and included an interactive demo of StoryBot to help parents make an informed decision about whether they consented to their child using the tool for writing activities.

Before the workshop began, students participated in short semi-structured pre-interviews focused on their writing efficacy and prior experience with AI. The interviews explored students' self-reported confidence and attitudes toward writing (e.g., "Do you enjoy writing at school or at home?") as well as

Table 9. Youth participants in StoryBot pilot workshop.

Name	Gender	Age	Grade	Primary language	ELL (Y/N)	Parent education
Alex	Boy	8	3	Chinese	Y	Bachelor
Erin	Girl	9	4	Korean/English	N	Master
Zoe	Girl	9	4	English	N	Master
Aiden	Boy	9	4	English	N	Bachelor
Zara	Girl	10	4	English	N	Bachelor
Julia	Girl	10	5	Korean/English	N	Doctorate
Riley	Girl	10	5	English	N	Bachelor
Kai	Boy	10	5	Chinese	Y	Bachelor
Noah	Boy	11	5	Chinese/English	N	Bachelor
Owen	Boy	11	6	Korean/English	N	Bachelor
Layla	Girl	11	6	Korean	Y	Master
Ava	Girl	12	6	Korean	Y	Master
Hana	Girl	12	6	Chinese	Y	Bachelor

Table 10. An overview of the daily workshop schedule by group.

	Day 1	Day 2
Topic	Should Animals Be Kept in Zoos?	Homework: Helpful or Harmful
Group A	Writing with StoryBot ($n = 7$)	Writing without StoryBot ($n = 6$)
Group B	Writing without StoryBot ($n = 6$)	Writing with StoryBot ($n = 7$)

their familiarity with AI concepts (e.g., “Can you explain what AI is?” and “Have you used any AI systems before?”).

The students were then divided into two groups and randomly assigned to write an opinion essay on either “Should Animals Be Kept in Zoos?” or, “Homework: Helpful or Harmful?” with or without AI-assistance. Although the StoryBot platform allows educators to customize writing prompts, we chose to standardize topics for the purposes of this pilot study. The selected prompts were intentionally chosen because they are age appropriate, commonly used in classroom settings, and likely to elicit diverse viewpoints and reasoning strategies, making them suitable for evaluating opinion writing performance.

On the second day, the groups switched tasks, writing essays on the opposite topic using the alternative method. At the end of the two sessions of the workshop, we conducted post-interviews to assess the students’ experiences with the StoryBot platform. The interviews aimed to gather their general impressions and identify both the strengths and the challenges they encountered while using the tool. Sample questions included: “Tell me about your overall experience using StoryBot for your writing”, “What did you like the most about StoryBot?” “What did you find most difficult?” and “If you could change something about StoryBot, what would it be?” These questions were designed to surface what aspects of the platform were effective, which were less successful, and to elicit suggestions for improvement.

5.1.3. Data sources and analytical techniques

We collected data through screen recordings of students’ writing activities and saved their written outputs as text files ($n = 26$). Students were asked to verbalize their thought processes as they wrote, following a think-aloud protocol.

To assess the direct outcomes of the system experience, as defined in the logic model, we measured: 1) students’ time spent in writing, 2) total word count, and 3) lexical diversity to evaluate the platform’s impact on writing efficacy (McNamara et al., 2010). Additionally, we recorded students’ utterances using voice recorders and logged interactions between students and the AI writing platform to evaluate students’ engagement. To analyze students’ written outputs, we used Text Inspector,¹⁴ a web-based text analysis tool to measure lexical diversity. This tool provides the Measure of Textual Lexical Diversity (MTLD), which calculates the variety of vocabulary used within a text. A higher MTLD indicates a broader vocabulary, reflecting linguistic proficiency and complexity in writing (McCarthy & Jarvis, 2010). Lexical diversity is particularly important in enhancing the richness and depth of communication, reducing repetitiveness, and demonstrating the writer’s competence (McNamara et al., 2010). This metric is especially relevant for contexts such as creative and academic writing, where a rich vocabulary is advantageous.

We analyzed each student’s writing performance across the two conditions (with StoryBot and without StoryBot) to examine our research question: how pedagogically co-designed GenAI scaffolding influences student engagement and elaboration during writing. Given the small sample size, we focused on within-student comparisons and emphasized effect size estimates rather than statistical significance testing, enabling us to characterize the practical relevance of observed differences (Raudys & Jain, 1991; Slavin & Smith, 2009).

For the quantitative analysis, we computed descriptive statistics and Cohen’s d (Cohen, 2013) to compare writing outcomes across conditions (Fritz et al., 2012). Students using StoryBot demonstrated higher lexical diversity (MTLD; $d = 1.33$, large effect) and produced longer texts (total word count; $d = 0.68$, moderate effect). They also spent more time engaged in writing ($d = 0.72$, moderate effect). Notably, increased writing time may reflect either productive persistence or additional cognitive effort; therefore, we interpret this measure conservatively as sustained writing time rather than writing efficiency.

All students were able to successfully engage with StoryBot and complete their essays under the StoryBot condition (see Table 11).

For qualitative analysis, we audio-recorded think-aloud protocols and generated transcripts using automatic speech recognition, followed by manual correction to ensure accuracy. We then conducted an inductive thematic analysis, which involved generating open codes and iteratively comparing them to identify underlying themes (Braun & Clarke, 2019). The first author led the open coding of

transcripts and refined patterns and themes through multiple rounds of peer debriefing. Importantly, we leveraged this qualitative analysis to complement our understanding of quantitative writing processes and outcomes, moving beyond writing output and behavioral measures to also account for students' perceptions and lived experiences. This process surfaced a central theme around students' engagement, perceptions, and experiences.

5.2. Phase 3 findings

Students using StoryBot showed higher lexical diversity (MTLD; $d = 1.33$, large), produced longer texts (TWC; $d = 0.68$, moderate), and spent more time writing ($d = 0.72$, moderate) than when writing without StoryBot. We interpret the latter conservatively as sustained writing time rather than direct evidence of increased engagement. While additional time may reflect productive persistence, it may also reflect additional AI interactions, increased cognitive effort, or interruptions associated with the scaffolded writing process. All students successfully used StoryBot to complete their essays (see Table 11).

To better understand how students engaged with StoryBot alongside these quantitative patterns, we conducted an inductive thematic analysis of students' think-aloud and post-interview data. The analysis identified three interrelated themes: (1) *scaffolding expanded students' ideas and linguistic choices*, (2) *conversational interaction sustained engagement*, and (3) *students negotiated AI assistance and agency*. As summarized in Table 12, these themes capture recurring patterns in how students experienced and responded to StoryBot's scaffolding.

Prompted idea development and vocabulary exploration provide qualitative context for the observed differences in writing volume and lexical diversity, while perceived attentiveness and conversational partnership help contextualize the longer sustained writing time. The third theme cuts across these outcomes by illustrating how students evaluated and selectively incorporated StoryBot's assistance while maintaining control over their writing. These qualitative findings are therefore used to contextualize, rather than causally explain, the quantitative differences observed between conditions. Representative excerpts below illustrate the recurring patterns identified through the thematic analysis.

Table 11. Writing outcomes (M, SD) for lexical diversity (MTLD), total word count (TWC), and time spent (minutes) with and without StoryBot. Cohen's d is reported for the between-group difference (with – without), using a pooled SD.

	MTLD ($p = 0.004^{**}$)			TWC ($p = 0.03^{*}$)			Time spent (min) ($p = 0.001^{**}$)		
	M	SD	d	M	SD	d	M	SD	d
Without StoryBot	64.90	12.65		162.62	60.59		26.15	13.31	
With StoryBot	79.83	9.57		210.77	79.79		36.77	16.05	
Effect size (Cohen's d)			1.33			0.68			0.72

Note. $^{*}p < .05$; $^{**}p < .01$.

Table 12. Themes and Sub-themes identified through the Phase 3 thematic analysis.

Theme	Sub-theme	Interpretation
Scaffolding expanded students' ideas and linguistic choices	Prompted idea development and elaboration	Step-by-step follow-up questions supported students in developing initial ideas through additional reasons, examples, and details.
	Vocabulary exploration	Guided linguistic alternatives exposed students to more precise and descriptive vocabulary while preserving choice over whether to use suggestions.
Conversational interaction sustained engagement	Perceived attentiveness and encouragement	Responsive prompts and encouragement contributed to students' perceptions that StoryBot was attending to their ideas and encouraging continued participation.
	Conversational writing partnership	Turn-taking interaction provided a sounding board through which students could articulate, develop, and reconsider their ideas.
Students negotiated AI assistance and agency	Selective uptake and creative control	Students evaluated StoryBot's suggestions against their own intentions, selectively incorporating assistance while maintaining responsibility for their writing.
	Awareness of AI reliance	Students recognized a tension between benefiting from immediate AI scaffolding and developing independent writing capacity.

5.2.1. Scaffolding expanded students' ideas and linguistic choices

The first theme captured how StoryBot's scaffolding supported students in developing their ideas and expanding their linguistic choices. Two related sub-themes emerged from the qualitative analysis: *prompted idea development and elaboration* and *vocabulary exploration*. These patterns provide qualitative context for the greater writing volume and lexical diversity observed in the StoryBot condition.

5.2.1.1. Prompted idea development and elaboration. Students described StoryBot's step-by-step prompts as helping them develop initial ideas through additional reasons, examples, and details. Rather than generating complete responses, StoryBot used follow-up questions such as "Can you provide an example?" and "Why do you think that?" to encourage students to elaborate on their own ideas. Noah (age 11) described this process:

"It helped me develop the story step by step. I really like how it explains to you and it gives you ideas and it makes you think. And then once you're done, you're like, okay, let's move on to the next thing."

This pattern was also visible in students' interactions with StoryBot. For example, when responding to the prompt "Should animals be kept in zoos?", Julia (age 10) initially wrote, "I disagree with animals being kept in the zoos." StoryBot followed with, "Great start! Can you explain why you disagree with animals being kept in zoos? Try to give a reason or example." Julia subsequently expanded her response:

"The reason why I don't agree with Zoos is that Zoos are meant to entertain humans. All people hear is almost nothing. This shows that animal keepers are basically keeping animals in one 'tiny' space and torturing the animal's quality of life and stopping them from running wild in the woods."

The interaction illustrates the recurring pattern identified in the thematic analysis: StoryBot's follow-up questions created opportunities for students to move beyond an initial position by articulating reasons, examples, and further detail. This pattern provides qualitative context for the difference in writing volume. Students produced an average of 210.77 words ($SD = 79.79$) with StoryBot compared with 162.62 words ($SD = 60.59$) without StoryBot ($d = 0.68$, moderate; Table 11). Rather than attributing this difference to a single feature, the qualitative findings suggest that iterative elaboration prompts were one way students experienced support for extending their writing.

5.2.1.2. Vocabulary exploration. A second recurring pattern concerned students' engagement with alternative ways of expressing their ideas. StoryBot provided targeted vocabulary prompts when students used repetitive or general language, encouraging them to consider more precise or descriptive alternatives. Importantly, these suggestions were presented as options rather than replacements, allowing students to decide whether they fit their intended meaning.

Julia (age 10) described this process:

"I liked that StoryBot sometimes gave me new words or details I wouldn't think of. Like, I wrote 'daisies' in my sentence, and then it said I could write 'the vibrant colors of the daisies.' That was cool. If I like the idea, I use it, but if I don't like it, I just skip it."

Her account illustrates both vocabulary exploration and selective uptake: StoryBot introduced linguistic alternatives, while the student retained control over whether to incorporate them. This pattern complements the quantitative finding that students' lexical diversity was higher with StoryBot (MTLD: $M = 79.83$, $SD = 9.57$) than without StoryBot ($M = 64.90$, $SD = 12.65$), with a large effect size ($d = 1.33$; Table 11). The qualitative findings therefore provide context for the MTLD difference by showing how students encountered, evaluated, and selectively incorporated more varied linguistic choices during their interactions with StoryBot.

5.2.2. Conversational interaction sustained engagement

The second theme captured how students experienced the conversational qualities of StoryBot during writing. Two related sub-themes emerged: *perceived attentiveness and encouragement* and *conversational writing partnership*. Students described StoryBot not only as providing prompts, but also as responding to their contributions in ways that encouraged continued interaction with the writing task. These

patterns provide qualitative context for the longer sustained writing time observed in the StoryBot condition.

5.2.2.1. Perceived attentiveness and encouragement. Students described StoryBot's acknowledgements, encouragement, and follow-up questions as indications that the system was responding to their ideas. Layla (age 11) explained:

"AI engaged me with compliments and showed me like the AI is actually paying attention to what I'm saying."

This perception of attentiveness was reinforced by StoryBot's turn-taking structure. Rather than presenting a writing task and waiting for a completed response, StoryBot acknowledged students' contributions and followed them with questions that moved the writing forward. For example, after students stated an opinion, StoryBot prompted them to explain their reasoning, provide an example, or elaborate on a particular point. Students therefore experienced the interaction as a sequence of responsive exchanges rather than a single prompt-response interaction.

5.2.2.2. Conversational writing partnership. Students also described StoryBot as functioning as a writing partner or sounding board through which they could articulate and reconsider their ideas. Owen described this experience by comparing it to the "rubber ducky"¹⁵ technique used in programming:

"I never knew what a difference a writing partner or bot can make. It's like the rubber ducky technique that programmers use, you explain your work step-by-step, as if to someone who knows nothing about it, and in doing so, you find errors or new ideas."

This account suggests that the conversational value of StoryBot did not depend solely on receiving suggestions. Articulating ideas in response to successive questions also provided students with opportunities to reflect on and further develop their own reasoning.

These patterns complement the quantitative writing-time results. Students spent an average of 36.77 min (SD = 16.05) writing with StoryBot compared with 26.15 min (SD = 13.31) without StoryBot, a difference of 10.62 min ($d = 0.72$, moderate; Table 11). We interpret this difference as *sustained writing time* rather than greater writing efficiency, as additional time may reflect productive persistence as well as the interactional effort required by the chatbot. The thematic findings help characterize this additional engagement: students described repeated cycles of responding, explaining, and reconsidering ideas within an interaction they perceived as attentive and conversational.

5.2.3. Students negotiated AI assistance and agency

The third theme captured how students negotiated the role of AI assistance within their writing process. Unlike the first two themes, this theme was not associated with a single quantitative outcome; instead, it represented a cross-cutting pattern in students' interactions with StoryBot. Two related sub-themes emerged: *selective uptake and creative control* and *awareness of AI reliance*.

5.2.3.1. Selective uptake and creative control. Students' accounts indicated that receiving a suggestion did not necessarily mean accepting it. Instead, students evaluated StoryBot's suggestions in relation to their own intentions and decided whether the assistance was useful. Julia's description of vocabulary support illustrates this process:

"If I like the idea, I use it, but if I don't like it, I just skip it."

Selective uptake was supported by StoryBot's constrained scaffolding approach. Rather than generating complete sentences or paragraphs for students, StoryBot primarily used short prompts and word-level suggestions, such as "Can you give an example?", "Why do you think that?", and "Is there a more precise word you could use?" These prompts created opportunities for assistance while leaving substantive decisions about content and expression with the student. In this sense, students' agency was reflected not in writing without AI support, but in deciding how AI support would contribute to their writing.

5.2.3.2. Awareness of AI reliance. Students also reflected on the boundary between receiving useful support and becoming dependent on AI assistance. Some students emphasized the importance of maintaining the ability to write independently, particularly in contexts where AI might not be available. These reflections suggest that students were aware of potential tensions between the immediate benefits of AI scaffolding and longer-term independent writing development.

Ultimately, these findings qualify the quantitative differences observed in Phase 3. Greater writing volume, lexical diversity, and sustained writing time capture changes in students' writing behaviors and outputs, but do not indicate how students understood their own role in AI-supported writing. The qualitative analysis adds this dimension by showing that students actively evaluated StoryBot's assistance and sought to maintain control over their ideas and writing. Thus, the findings suggest that the value of StoryBot's scaffolding lies not simply in providing additional assistance, but in structuring that assistance so that students remain active participants in generating, evaluating, and expressing their own ideas.

5.3. Phase 4: Teachers' focus group workshop educator reflections on the classroom potential of StoryBot

We conducted a three-hour remote focus group workshop with six classroom teachers (different from the co-design teachers) via Zoom to explore their perspectives on the pedagogical role of StoryBot and its potential impact on teaching and learning. The workshop served as both an introduction to the StoryBot platform and a collaborative opportunity for educators to participate in its user evaluation. Our goal was to elicit feedback on the platform's design, functionality, and instructional relevance to inform future iterations. The session began with a pre-survey assessing participants' prior experiences with AI in education, followed by an overview of the workshop agenda. The research team then presented the StoryBot platform and its underlying design rationale, setting the stage for hands-on exploration and group discussion.

Participants engaged in hands-on activities with the StoryBot prototype, exploring different writing genres (e.g., opinion and narrative writing) (see [Figure 1.2](#) and [1.3](#)) and sharing their work in a shared Google Doc. They identified strengths and areas for improvement. The workshop concluded with discussions on integrating StoryBot into teaching contexts and identifying challenges. A post-survey gathered feedback on the tool's usability, potential for skill development, and implementation challenges, along with suggestions for improvement (See the agenda in [Table 13](#)).

5.3.1. Participants

The focus group workshop included six K-12 teachers from elementary and middle schools interested in integrating AI into their teaching practices. Participants ranged from an 8th-grade ELA teacher using OpenAI for lesson planning to a K-5 school librarian incorporating AI into interactive lessons. Other participants included teachers utilizing AI for proofreading, parent communication, and lesson planning across various subjects, including math and social studies. Detailed information about the participating teachers is provided in [Table 14](#).

5.3.2. Data collection and analysis

Pre- and post-surveys assessed participants' prior AI experiences and their perceptions of StoryBot following the workshop. The three-hour workshop was recorded and transcribed, and chat logs and open-ended survey responses provided additional qualitative data.

Table 13. Teacher focus group session schedule.

Duration	Activities
15 mins	Welcome and Introduction
20 mins	Introduction on StoryBot
45 mins	Interactive Session: Exploring StoryBot - Hands-on Activity - Group Discussion
10 mins	Break
40 mins	Integrating StoryBot into teaching
10 mins	Q & A and Open Discussion

Table 14. Teacher participants and their AI usage practices.

Name	Grade level	Subject area	Location	Years of experience	AI usage
Sara	8th grade	ELA	Raleigh, NC	19	Utilizes OpenAI for lesson planning and assisting students with writing tasks.
Rachel	8th grade	General	Apex, NC	20	Uses AI for proofreading and generating writing ideas.
Jamie	6th grade	ELA	Encinitas, CA	14	Has employed AI for parent communication and plans to integrate it more with student activities.
Ken	Grades 3–8	Math	Apex, NC	18	Programs in ChatGPT and is keen on exploring new AI applications in the classroom.
Nina	7th grade	Social Studies	Raleigh, NC	23	Uses AI for lesson and project planning, aiming to enhance proficiency.
Carla	K–5	School Librarian	Raleigh, NC	16	Uses MagicSchool.ai for interactive lessons.

Table 15. Themes and Sub-themes identified through the Phase 4 thematic analysis.

Theme	Sub-themes
Affordances and tradeoffs of a GenAI-based story-authoring platform	Interactive scaffolding for writing engagement; AI as a complementary instructional partner; tensions between AI assistance and independent writing
Teacher-identified design needs for instructional alignment	Curricular and genre alignment; developmental differentiation and adaptive scaffolding; teacher configurability and oversight

To address RQ4 (i.e., *How do teachers perceive the pedagogical value and classroom applicability of a co-designed GenAI writing tool?*) we conducted an inductive thematic analysis of workshop transcripts, chat logs, and open-ended survey responses. Initial codes captured recurring evaluations, concerns, and design recommendations expressed by teachers. Related codes were iteratively compared and grouped into sub-themes, which were then organized into higher-order themes representing teachers' perceptions of StoryBot's pedagogical value and requirements for classroom integration. Representative quotations are used in the findings to illustrate these recurring patterns rather than as standalone evidence.

5.4. Phase 4 findings

The thematic analysis identified two higher-order themes concerning teachers' perceptions of StoryBot's pedagogical value and classroom applicability: (1) *StoryBot offered pedagogical affordances while introducing tradeoffs in AI-supported writing*, and (2) *classroom integration required teacher-configurable and developmentally responsive scaffolding*. Each theme comprised three related sub-themes (Table 15).

5.4.1. Affordances and tradeoffs of a GenAI-based story-authoring platform

The first theme captured teachers' perceptions of StoryBot's pedagogical affordances alongside concerns about how AI assistance might affect students' independent writing. Three related sub-themes emerged: *interactive scaffolding for writing engagement*, *AI as a complementary instructional partner*, and *tensions between AI assistance and independent writing*.

5.4.1.1. Interactive scaffolding for writing engagement. Teachers viewed StoryBot's interactive prompts as potentially useful for helping students initiate and sustain writing. Three of the six teachers described the platform as making writing more engaging and less intimidating, particularly for students who may struggle to get started or continue developing their ideas. Jamie explained:

"I liked the idea of this platform design. I teach writing, and it could be a useful tool to help my students get started on writing a story based on a particular genre."

Similarly, Rachel described StoryBot as "easy and fun to use" and noted that "the prompts keep them going." These responses suggest that teachers perceived conversational prompting not simply as an interface feature, but as a scaffold for sustaining students' participation in the writing process.

5.4.1.2. AI as a complementary instructional partner. Teachers also described StoryBot as a supplementary instructional resource rather than a replacement for teacher guidance. Ken, for example, characterized AI as a "third party" in the classroom that could probe students' thinking while the teacher

retained primary responsibility for lesson planning and instruction. This perspective positions StoryBot as an additional layer of support within teacher-directed learning rather than as an autonomous tutor.

5.4.1.3. Tensions between AI assistance and independent writing. Alongside these perceived affordances, four of the six teachers expressed concern that students could become overly reliant on AI-generated suggestions, potentially reducing opportunities for independent thinking and writing. Teachers therefore emphasized the importance of calibrating the amount of AI assistance available to students. Sara explained:

“Since it is hard to tell whether students passively accept the AI’s help or are actively engaged in their writing, perhaps giving students or teachers a choice of how much AI can help would be beneficial.”

Rather than rejecting AI assistance, teachers emphasized the need to balance scaffolding with opportunities for students to remain active decision-makers in the writing process. This tradeoff between support and independence emerged as a central condition shaping teachers’ perceptions of StoryBot’s pedagogical value.

5.4.2. Teacher-identified design needs for instructional alignment

The second theme concerned the conditions teachers considered necessary for StoryBot to align with classroom instruction. Three related sub-themes emerged: *curricular and genre alignment*, *developmental differentiation and adaptive scaffolding*, and *teacher configurability and oversight*.

5.4.2.1. Curricular and genre alignment. Teachers emphasized that StoryBot’s scaffolds should align more closely with grade-level expectations, writing genres, and curricular standards. Rachel, for example, noted limitations in the platform’s support for opinion writing:

“It is very basic while evaluating responses to an opinion-based topic. It appears to be a good tool for planning an opinion piece but would require teacher guidance to effectively present an opinion based on the requirements of [the state] 8th grade standards.”

This feedback suggests that teachers viewed StoryBot as useful for planning and early-stage writing support, while also identifying a need for stronger alignment with genre-specific criteria and classroom standards.

5.4.2.2. Developmental differentiation and adaptive scaffolding. Five of the six teachers emphasized the need to adjust the difficulty and level of support according to students’ grade levels and language proficiency. Teachers described the importance of lowering vocabulary demands for emergent writers while increasing complexity for more advanced students. They also suggested support mechanisms, such as an “I need help” option, that would allow students to request assistance when necessary without requiring constant AI intervention.

Teachers generally viewed the current version of StoryBot as most appropriate for students in grades 4-6, while recommending a stronger “low floor, high ceiling” approach so that the platform could support both beginning and more advanced writers. Three teachers also suggested using more open-ended prompts to encourage critical thinking and original idea generation.

5.4.2.3. Teacher configurability and oversight. Four of the six teachers emphasized that successful classroom integration would require teachers to control how StoryBot is configured and used. They expressed a preference for teacher-in-the-loop features that would allow educators to curate lessons, adjust scaffolding levels, and tailor writing activities to specific instructional goals. Jamie described this need:

“I would love StoryBot to scaffold vocabulary levels according to the language proficiency and grade levels of each student; it would be nice to have custom features for us to accommodate students’ learning goals according to their level.”

This emphasis on configurability reflects teachers’ broader expectation that AI support should be adaptable to diverse classroom contexts while preserving educator oversight over instructional decisions.

Overall, these findings indicate that teachers viewed StoryBot's pedagogical value as conditional on how its scaffolds are configured and integrated into instruction. While participants recognized the platform's potential to support writing engagement and development, teachers consistently emphasized that classroom use should remain developmentally appropriate, curricularly aligned, and under teacher control. Across both themes, a recurring tension emerged between providing sufficient AI support to help students progress and preserving opportunities for independent writing and teacher-directed instruction.

6. Discussion

6.1. Design tensions and recommendations

Across these findings, we identify two broad design tensions for teacher-configurable GenAI writing platforms: teachers' need for oversight and control and the need to prioritize augmentation over automation in student writing. These tensions point to broader recommendations for designing classroom GenAI systems that enable teacher-in-the-loop orchestration while positioning AI as a scaffold for, rather than a replacement of, students' writing processes. Table 16 summarizes these recommendations and maps them to concrete platform-level features.

6.1.1. Tension 1: Needs for teachers' oversight and control

Striking an appropriate balance between the generative capacity of GenAI systems and teachers' need for instructional oversight remains an ongoing design challenge in AI-supported learning environments (Chen et al., 2025a). Across our participatory study, teachers expressed concerns about unrestricted text generation and the potential loss of instructional control. These concerns highlighted three interrelated considerations for classroom-ready GenAI writing tools: (1) how the level and timing of AI assistance should vary across learners with different writing proficiency and language backgrounds; (2) how AI-generated contributions can function as formative scaffolds rather than substitutes for students' learning or indicators of writing ability; and (3) how student agency, authorship, and cognitive ownership can be preserved throughout AI-supported writing (Chakrabarty et al., 2025; Chen et al., 2025b; Tatar et al., 2025). These considerations reflect broader debates in AIED and HCI concerning authorship, responsibility, and the sociotechnical nature of AI-mediated learning environments (Ahn & Lim, 2025; Aleman et al., 2024; Min et al., 2026).

Our co-design process foregrounded teacher participation as one way of negotiating these tensions. Rather than treating teachers as recipients of predefined AI functionality, the process engaged them in determining how AI assistance should be configured around classroom practices, pedagogical goals, and ethical concerns. This approach aligns with prior work showing that participatory and co-design processes can strengthen users' agency by involving them in design and decision-making (Iivari et al., 2024). In StoryBot, these priorities were reflected in teacher-facing configuration of writing activities and AI scaffolds, positioning teachers as active mediators of how GenAI enters the writing process.

Table 16. Summary of design tensions, recommendations, design implications, and platform-level features for teacher-configurable GenAI writing platforms.

Design tension → recommendation	Design implications	Platform-Level features
Teacher Oversight and Control → <i>Enable adaptive teacher-in-the-loop orchestration.</i>	Enable teachers to regulate the level, timing, and type of AI assistance according to students' writing proficiency, language backgrounds, and instructional goals.	Live teacher dashboards; per-student AI controls; phase-based controls across ideation, drafting, and revision; teacher-configurable activities and instructional content.
Augmentation over Automation → <i>Position AI as a cognitive partner and scaffold.</i>	Prioritize AI assistance during formative writing processes while avoiding the replacement of students' critical thinking, composition, and independent writing.	Guided and scaffolded prompting; scripted/rule-based prompts with limited generative responses; phase-sensitive assistance; teacher-curated mentor texts.
Augmentation over Automation → <i>Make AI contributions transparent and authorship boundaries legible.</i>	Make the extent and role of AI assistance visible so teachers and students can distinguish student-authored writing from AI-mediated scaffolding.	Visual differentiation of AI contributions; color-coded highlighting; text-similarity or AI-influence visualization; student-AI interaction records.

However, the implementation also revealed an important distinction between *configuring* AI assistance and *orchestrating* it during classroom activity. StoryBot supported task-level configuration and post-hoc monitoring, but provided limited support for teachers to respond dynamically to students' interactions with AI. Teachers could not, for example, adjust scaffolding intensity for an individual student during an activity, selectively enable or restrict AI assistance across different phases of writing, or intervene directly in an ongoing student–AI interaction. These limitations reflect both the technical complexity of fine-grained control over LLM behavior and a broader design challenge of providing meaningful teacher control without introducing excessive cognitive or orchestration demands during instruction (Kang et al., 2025).

6.1.2. Design recommendation 1: Adaptive teacher-in-the-loop orchestration

Addressing this tension requires moving beyond static teacher configuration toward adaptive, teacher-in-the-loop orchestration of AI assistance. We suggest that classroom GenAI platforms enable teachers to regulate the *level, timing, and type* of AI support in response to learners' needs and instructional goals. Such orchestration can operate at multiple levels: *live dashboards* can provide awareness of student–AI interaction patterns; *per-student controls* can enable differentiated levels of assistance based on students' writing proficiency and language backgrounds; and *phase-based controls* can allow teachers to enable, restrict, or gradually fade AI support across ideation, drafting, and revision. Teacher-facing interfaces should also retain control over instructional content and writing activities, including classroom topics, lesson goals, images, and worksheets (Kang et al., 2025; Min et al., 2026).

Importantly, teacher-in-the-loop orchestration should not be understood simply as providing educators with more settings. Rather, the design goal is to make AI assistance responsive to the pedagogical judgments teachers already make about when learners need support and when they should work more independently. For example, an emergent writer may benefit from more intensive ideation scaffolds, whereas a developing writer may receive reduced generative assistance and more targeted feedback during drafting. Similarly, AI support might be stronger during brainstorming and intentionally faded as students move toward independent composition. In this sense, configurability becomes a mechanism through which teachers can calibrate AI assistance to learners' proficiency, language backgrounds, and instructional contexts (Borge et al., 2024; Harvey et al., 2025; Wang & Cornely, 2024).

All in all, these findings suggest that teacher-configurable GenAI writing platforms should be designed not only as student-facing learning tools, but also as *classroom orchestration systems*. Such systems should support teacher awareness, differentiated control, and timely intervention while allowing AI to shift between roles—such as cognitive partner, targeted scaffold, or background support—according to the pedagogical context. This framing preserves teachers' instructional authority while creating conditions for AI assistance to remain responsive to learners without becoming an autonomous substitute for teaching or writing.

6.1.3. Tension 2: Augmentation over automation

Across the co-design sessions, teachers consistently highlighted students' difficulties in developing well-elaborated texts and expressed interest in how GenAI might provide additional support during the writing process. At the same time, they raised concerns that students might engage passively with AI-generated content, potentially bypassing critical thinking and authentic composition processes (Borge et al., 2024). This tension raises an important design question for AI-supported writing: how can GenAI provide meaningful assistance without shifting from supporting students' writing processes to producing writing on their behalf? Questions of authorship in AI-assisted writing remain particularly unresolved and evolving (Gero et al., 2025; Harvey et al., 2025).

These concerns point to the importance of designing AI systems that promote active engagement and transparency while maintaining student-centered learning (Chen et al., 2025b; Weisz et al., 2024). In StoryBot, we responded to this tension by deliberately constraining aspects of the system's generative autonomy. For example, rather than allowing students to freely generate mentor texts, teachers could curate and assign specific examples as instructional anchors. Although this approach limits the range of stylistic and rhetorical examples that students can generate independently, it reduces the risk of

exposing students to low-quality or instructionally misaligned content. More importantly, this design choice positions generative capability within a teacher-mediated instructional structure, such that AI augments rather than replaces pedagogical intent (Fui-Hoon Nah et al., 2023; Mosaiyebzadeh et al., 2023).

6.1.4. Design recommendation 2: AI as a cognitive partner and scaffold

Our student workshops and teacher feedback suggest that AI-supported writing systems should position AI as a cognitive partner and scaffold rather than a shortcut for producing final text (Borge et al., 2024; Kim & Baylor, 2006; Tankelevitch et al., 2024). Teachers particularly valued AI support during formative stages such as brainstorming, idea development, and revision. This extends traditional scaffolding toward learner–AI co-regulation, in which adaptive and dialogic support can be provided across phases of the writing process (Chung et al., 2025; Fan et al., 2025a).

Students' interactions with StoryBot further suggest that AI can serve different cognitive functions depending on learners' writing needs. Some students used AI to elaborate ideas, while others used it to refine expression; students also described StoryBot as a writing partner, attentive listener, or step-by-step guide (Cress & Kimmerle, 2023; Ojeda-Ramirez et al., 2025; Wang et al., 2023a). These patterns align with perspectives on shared agency in AI-supported learning and sociocultural accounts of tool-mediated learning (Hayashi et al., 2024; Kim et al., 2025b; Nordlof, 2014).

Designing for augmentation therefore requires calibrating *what*, *when*, and *how* AI provides assistance. Writing platforms should prioritize guided prompts and scaffolded interactions over direct text generation, with support adapted to students' learning goals, language proficiency, interests, and stages of writing. For example, stronger scaffolding may support brainstorming, while reduced assistance during drafting can create greater space for independent composition.

Platforms should also make AI involvement transparent by distinguishing student-authored writing from AI-mediated support. Features such as color-coded highlighting or text-similarity analysis could help teachers and students understand the extent of AI influence, supporting accountability, authorship, and student agency.

Overall, AI assistance should be *bounded*, *scaffolded*, and *transparent*: bounded to avoid substituting for student composition, scaffolded to provide appropriate support throughout the writing process, and transparent to preserve visibility into student and AI contributions. Future research should examine whether students internalize these scaffolds over time and develop transferable writing strategies and greater independence (Borge et al., 2024).

6.2. Implications for co-designing GenAI learning technologies with educators

Beyond the design implications for StoryBot, our study offers methodological insights for HCI researchers and educators engaging in the co-design of GenAI learning technologies. First, our findings suggest the value of beginning co-design with *pedagogical goals and conditions of use* rather than predefined AI capabilities or interface features. Prior work in participatory design and educational technology similarly emphasizes involving educators as design partners whose situated knowledge can guide the alignment of technology with instructional practices and classroom needs (Nicholson et al., 2022; Penuel et al., 2007; Ravi et al., 2025). In Phase 1, teachers did not initially design a specific interface; instead, the logic-model activity invited them to articulate intended users, instructional challenges, desired forms of support, learning outcomes, and ethical concerns. For co-design with rapidly evolving AI, prioritizing pedagogical goals over technical capabilities can ensure that design decisions are guided by educators' instructional priorities, classroom practices, and learners' needs rather than by current AI capabilities (Ravi et al., 2025; Xie et al., 2024).

Second, our findings suggest that educator co-design for GenAI should address not only interface features, but also how AI behaves and where its role should be limited. Teachers helped shape decisions about when AI should intervene, how much support it should provide, and what forms of generation should be restricted to preserve student agency and authorship. These considerations align with prior work showing that the level and form of AI assistance can shape users' agency, ownership, and engagement in AI-supported writing (Kim et al., 2025b; Reza et al., 2025). Translating these pedagogical

judgments into StoryBot required treating teachers' expertise as input into the configuration of AI behavior rather than positioning teachers solely as evaluators of a completed system (He et al., 2025; Nicholson et al., 2022). For HCI researchers, this suggests that co-design activities for GenAI can explicitly elicit decisions about the type, timing, level, and visibility of AI assistance. Such an approach may be particularly important in educational contexts, where the appropriateness of AI behavior depends on learners' developmental needs, instructional objectives, and classroom norms (Jeon & Lee, 2023; Ravi et al., 2025).

Finally, our four-phase process shows the value of treating co-design as an iterative cycle. Teacher priorities informed StoryBot's design, student use revealed how those decisions worked in practice, and returning the platform to educators surfaced additional needs around curriculum, developmental differences, and AI assistance. Rather than ending educator involvement after initial ideation, researchers can repeatedly bring design decisions back to educators for reflection and refinement, consistent with participatory and iterative approaches that engage stakeholders throughout the design and integration of emerging technologies (Krol et al., 2025; Nicholson et al., 2022). Overall, our experience suggests that co-designing educational GenAI involves not only asking educators what an AI system should contain, but repeatedly negotiating what the AI should do, when it should act, and where its role should remain limited.

6.3. Limitations

While this study sheds light on the co-design and development of a GenAI-based story-authoring platform for K–12 writing instruction, it's also important to acknowledge its limitations.

6.3.1. Sample characteristics and generalizability

Participants were recruited through professional and personal networks using snowball sampling, which may have introduced selection bias toward educators and families with greater interest in educational technology and GenAI. Although the teacher sample represented multiple U.S. locations and included variation in teaching experience, grade levels, subject areas, and familiarity with GenAI, all participants were situated within U.S. educational contexts, and the study focused on English-language writing instruction. The findings may therefore not fully reflect the perspectives of educators working in other cultural, linguistic, or pedagogical contexts. Given the exploratory and participatory nature of the study, we also did not systematically compare perspectives across teacher experience levels or subject domains; such differences warrant investigation with larger and more targeted samples.

The student sample was similarly limited in size and included learners across ages 8–12 and different language backgrounds. Although both multilingual learners and English language learners (ELLs) participated, the small subgroup sizes did not support meaningful comparisons by language status. Thus, our findings should not be interpreted as demonstrating differential effects for ELLs or multilingual learners. Future research should recruit larger and more diverse samples to examine variation across geographic, linguistic, developmental, and instructional contexts.

6.3.2. Pilot study

This pilot study should be interpreted as exploratory rather than evaluative. Rather than assessing the effectiveness of StoryBot or its impact on students' writing outcomes, the study aimed to provide an exploratory case study of how students engaged with and used StoryBot during writing activities. The findings therefore offer preliminary insights into students' interactions with the platform and how its AI-supported writing features were used in practice. StoryBot was compared with Google Docs as a conventional writing baseline; therefore, the findings do not establish StoryBot's advantages over commercial GenAI writing tools. Future evaluations should include general-purpose or commercial GenAI systems as benchmark conditions to examine the contributions of StoryBot's teacher-configurable scaffolding relative to other forms of AI assistance.

The pilot was intentionally restricted to opinion writing to maintain a consistent task across conditions, and the findings may not generalize to other writing genres. Although the within-subjects, counterbalanced design helped reduce between-participant variability and potential order effects, we did not

systematically control for prior writing proficiency or previous experience with AI-assisted writing. These individual differences may have influenced students' interactions with the platform and therefore limit causal interpretation of the quantitative findings.

Although the logic model identified multilingual learners and ELLs as populations that could benefit from targeted support, features such as translanguaging, text-to-speech, and speech-to-text were not implemented or evaluated in this pilot. Other proposed features, including gamification and culturally responsive imagery, were also outside the scope of the prototype. Moreover, behavioral measures such as word count and time-on-task capture observable writing behaviors but do not fully explain writing quality, engagement, or the mechanisms underlying students' interactions with AI. Future studies should incorporate larger samples, baseline measures of writing proficiency and prior GenAI experience, multiple writing genres, AI benchmark conditions, and interaction-level analyses to better understand how different forms of AI scaffolding influence students' writing processes and outcomes.

7. Conclusion

By integrating qualitative insights from teacher co-design sessions with quantitative analyses of student writing outcomes, we demonstrate how such systems can meaningfully bridge design and classroom practice (Hutchins & Biswas, 2024; Stefaniak et al., 2025). Our findings show that when such systems are intentionally designed, they can function as responsive scaffolds that support students' writing development and engagement, while teacher-informed design choices address long-standing instructional challenges (Chang & Philip, 2025; Han et al., 2024; Han & Cai, 2023). Teachers emphasized the importance of maintaining a teacher-in-the-loop approach, enabling them to configure tasks, calibrate levels of AI assistance, and sustain active learning experiences tailored to their students. Our findings highlight the dual promise of human-centered AI in education: fostering student growth while empowering teachers to deliver inclusive, responsive instruction. The study's contribution lies not only in providing a pathway for the design of GenAI-based story-authoring platform, but also in emphasizing the value of a collaborative approach in the development procedure with stakeholders in education (Van Mechelen et al., 2023).

Notes

1. <https://www.quill.org/>
2. <https://www.storyjumper.com/>
3. <https://www.writable.com/>
4. <https://www.khanmigo.ai/>
5. <https://chatgpt.com/>
6. In technology-mediated interactions, this sense of agency supports users' perception that outcomes result from their own actions rather than from the system itself (Bennett et al., 2023; Cornelio et al., 2022).
7. In U.S. education, "K–12" denotes kindergarten through 12th grade (primary–secondary; ages 5–18). "K–8" indicates kindergarten through 8th grade (ages 5–14).
8. <https://www.getepic.com/>
9. <https://storybird.com/>
10. <https://www.canva.com/>
11. <https://www.magicschool.ai/>
12. <https://docs.google.com/>
13. <https://grow.google/ai-for-educators/>
14. <https://textinspector.com/>
15. The "rubber ducky" technique refers to explaining a problem step-by-step to an inanimate or non-expert listener as a way to clarify one's reasoning and identify errors or new ideas.

Author contributions

CRedit: **Ariel Han**: Conceptualization, Data curation, Formal analysis, Methodology, Project administration, Writing – original draft, Writing – review & editing; **Shenshen Han**: Conceptualization, Methodology, Project administration, Software, Writing – original draft; **Zhenyao Cai**: Conceptualization, Data curation, Formal analysis, Methodology, Writing – review & editing; **Joey Huang**: Data curation, Formal analysis, Methodology, Project administration, Writing – review & editing; **Franceli Cibrian**: Writing – review & editing; **Kylie Peppler**:

Conceptualization, Investigation, Project administration, Supervision, Writing – original draft, Writing – review & editing.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Ariel Han  <http://orcid.org/0000-0002-0047-0817>
 Shenshen Han  <http://orcid.org/0009-0008-8727-7585>
 Zhenyao Cai  <http://orcid.org/0009-0009-6933-7000>
 Joey Huang  <http://orcid.org/0000-0003-4362-326X>
 Franceli Cibrian  <http://orcid.org/0000-0002-7084-6904>
 Kylie Pepler  <http://orcid.org/0000-0002-5472-4974>

Data availability statement

Research materials and de-identified or aggregate data supporting the findings of this study are publicly available through the Open Science Framework (OSF) at <https://doi.org/10.17605/OSF.IO/NVX2Y>. Raw participant-level data, including interview and focus-group transcripts, recordings, student interaction data, and other potentially identifiable materials, are not publicly available due to participant privacy, confidentiality, and ethical considerations. Where ethically and institutionally permissible, appropriately de-identified materials may be made available by the corresponding author upon reasonable request.

References

- Adnin, R., Pandkar, A., Yao, B., Wang, D., & Das, M. (2025). Examining student and teacher perspectives on undisclosed use of generative AI in academic work. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1–17). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713393>
- Ahn, D., & Lim, H. (2025). Exploring K-12 physical education teachers' perspectives on opportunities and challenges of AI integration through ideation workshops. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1–16). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713646>
- Ait Baha, T., El Hajji, M., Es-Saady, Y., & Fadili, H. (2024). The impact of educational chatbot on student learning experience. *Education and Information Technologies*, 29(8), 10153–10176. <https://doi.org/10.1007/s10639-023-12166-w>
- Aleman, E., Dilek, M., & Baran, E. (2024). Generative AI as a mirror: Educators exploring and challenging dominant discourses about the future of AI in education. In *Proceedings of the 18th International Conference of the Learning Sciences-ICLS 2024* (pp. 857–864). International Society of the Learning Sciences. <https://doi.org/10.22318/icls2024.157422>
- Azungah, T. (2018). Qualitative research: Deductive and inductive approaches to data analysis. *Qualitative Research Journal*, 18(4), 383–400. <https://doi.org/10.1108/QRJ-D-18-00035>
- Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, 37(2), 122–147. <https://doi.org/10.1037/0003-066X.37.2.122>
- Bazerman, C. (2016). What do sociocultural studies of writing tell us about learning to write. In *Handbook of writing research* (2nd ed., pp. 11–23). The Guilford Press.
- Bennett, D., Metatla, O., Roudaut, A., & Mekler, E. D. (2023). How does HCI understand human agency and autonomy? In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI '23)* (p. 18). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3544548.3580651>
- Bhat, A., Agashe, S., Oberoi, P., Mohile, N., Jangir, R., & Joshi, A. (2023). Interacting with next-phrase suggestions: How suggestion systems aid and influence the cognitive processes of writing. In *Proceedings of the 28th International Conference on Intelligent User Interfaces (IUI '23)* (pp. 436–452). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3581641.3584060>
- Booten, K., & Gero, K. I. (2021). Poetry machines: Eliciting designs for interactive writing tools from poets. In *Proceedings of the 13th Conference on Creativity and Cognition (Virtual Event, Italy) (C&C '21)* (p. 5). New York, NY, Association for Computing Machinery. <https://doi.org/10.1145/3450741.3466813>

- Borge, M., Smith, B. K., & Aldemir, T. (2024). Using generative ai as a simulation to support higher-order thinking. *International Journal of Computer-Supported Collaborative Learning*, 19(4), 479–532. <https://doi.org/10.1007/s11412-024-09437-0>
- Bradford, K. L., Newland, A. C., Rule, A. C., & Montgomery, S. E. (2016). Rubrics as a tool in writing instruction: Effects on the opinion essays of first and second graders. *Early Childhood Education Journal*, 44(5), 463–472. <https://doi.org/10.1007/s10643-015-0727-0>
- Braun, V., & Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative Research in Sport, Exercise and Health*, 11(4), 589–597. <https://doi.org/10.1080/2159676X.2019.1628806>
- Brod, G., Kucirkova, N., Shepherd, J., Jolles, D., & Molenaar, I. (2023). Agency in educational technology: Interdisciplinary perspectives and implications for learning design. *Educational Psychology Review*, 35(1), 25. <https://doi.org/10.1007/s10648-023-09749-x>
- Bruner, J. (1985). Child's talk: Learning to use language. *Child Language Teaching and Therapy*, 1(1), 111–114. <https://doi.org/10.1177/026565908500100113>
- Cai, Z., Han, A., Zhou, X., Gazulla, E. D., & Peppler, K. (2025). Child-AI co-creation: A review of the current research landscape and a proposal for six design considerations. In *Proceedings of the 24th Interaction Design and Children (IDC '25)* (pp. 916–922). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3713043.3731506>
- Chakrabarty, T., Laban, P., & Wu, C.-S. (2025). Can AI writing be salvaged? Mitigating idiosyncrasies and improving human-AI alignment in the writing process through edits. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1–33). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713559>
- Chang, M. A., & Philip, T. M. (2025). Caught between expansive world-building and the status quo: Using figured worlds to understand world-building in AI for educational co-design contexts. *Journal of the Learning Sciences*, 34(5), 651–699. <https://doi.org/10.1080/10508406.2025.2490514>
- Chen, J., Tang, M., Lu, Y., Yao, B., Fan, E., Ma, X., Xu, Y., Wang, D., Sun, Y., & He, L. (2025b). Characterizing LLM-empowered personalized story reading and interaction for children: Insights from multi-stakeholder perspectives. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1–24). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713275>
- Chen, R., Lee, V. R., & G Lee, M. (2025a). A cross-sectional look at teacher reactions, worries, and professional development needs related to generative AI in an urban school district. *Education and Information Technologies*, 30(11), 16045–16082. <https://doi.org/10.1007/s10639-025-13350-w>
- Chung, J. J. Y., Roemmele, M., & Kreminski, M. (2025). Toyteller: AI-powered visual storytelling through toy-playing with character symbols. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1–23). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713435>
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.
- Coker, D. L., Jr, Farley-Ripple, E., Jackson, A. F., Wen, H., MacArthur, C. A., & Jennings, A. S. (2016). Writing instruction in first grade: An observational study. *Reading and Writing*, 29(5), 793–832. <https://doi.org/10.1007/s11145-015-9596-6>
- Cornelio, P., Haggard, P., Hornbaek, K., Georgiou, O., Bergström, J., Subramanian, S., & Obrist, M. (2022). The sense of agency in emerging technologies for human–computer integration: A review. *Frontiers in Neuroscience*, 16(2022), 949138. <https://doi.org/10.3389/fnins.2022.949138>
- Cotton, D. R. E., Cotton, P. A., & Shipway, J. R. (2024). Chatting and cheating: Ensuring academic integrity in the era of ChatGPT. *Innovations in Education and Teaching International*, 61(2), 228–239. <https://doi.org/10.1080/14703297.2023.2190148>
- Cress, U., & Kimmerle, J. (2023). Co-constructing knowledge with generative AI tools: Reflections from a CSDL perspective. *International Journal of Computer-Supported Collaborative Learning*, 18(4), 607–614. <https://doi.org/10.1007/s11412-023-09409-w>
- Crowhurst, M. (1990). Teaching and learning the writing of persuasive/argumentative discourse. *Canadian Journal of Education / Revue canadienne de l'éducation*, 15(4), 348–359. <https://doi.org/10.2307/1495109>
- Dhillon, P. S., Molaei, S., Li, J., Golub, M., Zheng, S., & Robert, L. P. (2024). Shaping human-AI collaboration: Varied scaffolding levels in co-writing with language models. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)* (p. 18). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642134>
- Fan, H., Zhou, C., Yu, H., Wu, X., Gu, J., & Peng, Z. (2025b). LitLinker: Supporting the ideation of interdisciplinary contexts with large language models for teaching literature in elementary schools. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (p. 19). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3706598.3714111>
- Fan, M., Cui, X., Ma, W., Li, H., Tong, X., Yang, L., & Wang, Y. (2025a). From words to wonder: Designing and evaluating an AI-empowered creative storytelling system for elementary children. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1–15). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713478>

- Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition & Communication*, 32(4), 365–387. <https://doi.org/10.2307/356600>
- Fritz, C. O., Morris, P. E., & Richler, J. J. (2012). Effect size estimates: Current use, calculations, and interpretation. *Journal of Experimental Psychology: General*, 141(1), 2–18. <https://doi.org/10.1037/a0024338>
- Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *Journal of Information Technology Case and Application Research*, 25(3), 277–304. <https://doi.org/10.1080/15228053.2023.2233814>
- Gero, K. I., Desai, M., Schnitzler, C., Eom, N., Cushman, J., & Glassman, E. L. (2025). Creative writers' attitudes on writing as training data for large language models. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1–16). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713287>
- Gero, K. I., Liu, V., & Chilton, L. (2022). Sparks: Inspiration for science writing using language models. In *Proceedings of the 2022 ACM Designing Interactive Systems Conference (Virtual Event, Australia) (DIS '22)* (pp. 1002–1019). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3532106.3533533>
- Gero, K. I., Long, T., & Chilton, L. B. (2023). Social dynamics of AI support in creative writing. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (Hamburg, Germany) (CHI '23)* (p. 15). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3544548.3580782>
- Gibbons, P. (2002). *Scaffolding language, scaffolding learning*. Heinemann.
- Göldi, A., Wambsganss, T., Neshaei, S. P., & Rietsche, R. (2024). Intelligent support engages writers through relevant cognitive processes. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)* (p. 12). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642549>
- Graham, S., & Perin, D. (2007a). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99(3), 445–476. <https://doi.org/10.1037/0022-0663.99.3.445>
- Graham, S., & Perin, D. (2007b). *Writing next: Effective strategies to improve writing of adolescents in middle and high schools: A report to Carnegie Corporation of New York*. Alliance for Excellent Education.
- Graham, S., Bollinger, A., Booth Olson, C., Aoust, C. D., MacArthur, C., McCutchen, D., & Olinghouse, N. (2012). *Teaching elementary school students to be effective writers: A practice guide (NCEE 2012-4058)*. National Center for Education Evaluation and Regional Assistance, Institute of Education Sciences, U.S. Department of Education.
- Graham, S., Harris, K. R., Fishman, E., Houston, J., Wijekumar, K., Lei, P.-W., & Ray, A. B. (2019). Writing skills, knowledge, motivation, and strategic behavior predict students' persuasive writing performance in the context of robust writing instruction. *The Elementary School Journal*, 119(3), 487–510. <https://doi.org/10.1086/701720>
- Graham, S., Kiuhara, S. A., Harris, K. R., & Fishman, E. J. (2017). The relationship among strategic writing behavior, writing motivation, and writing performance with young, developing writers. *The Elementary School Journal*, 118(1), 82–104. <https://doi.org/10.1086/693009>
- Han, A., & Cai, Z. (2023). Design implications of generative AI systems for visual storytelling for young learners. In *Proceedings of the 22nd Annual ACM Interaction Design and Children Conference (IDC '23)* (pp. 470–474). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3585088.3593867>
- Han, A., Zhou, X., Cai, Z., Han, S., Ko, R., Corrigan, S., & Pepler, K. A. (2024). Teachers, parents, and students' perspectives on integrating generative AI into elementary literacy education. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)* (p. 17). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642438>
- Harris, K. R., & Graham, S. (2009). Self-regulated strategy development in writing: Premises, evolution, and the future. *British Journal of Educational Psychology*, 2(6), 113–135. <https://doi.org/10.1177/2372732215624216>
- Harvey, E., Koenecke, A., & Kizilcec, R. F. (2025). “Don't Forget the Teachers”: Towards an educator-centered understanding of harms from large language models in education. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1–19). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713210>
- Hayashi, Y., Shimojo, S., & Kawamura, T. (2024). Learning with a scripted AI tutor or a virtual AI learner: Experimental investigation of collaborative learning using CoCot. In *Proceedings of the 17th International Conference on Computer-Supported Collaborative Learning-CSCL 2024* (pp. 11–18). International Society of the Learning Sciences. <https://doi.org/10.22318/csl2024.360863>
- He, K., Liu, X., Xu, Y., Bustamante, A. S., & Warschauer, M. (2025). Carlitos the curious caterpillar”: Exploring teacher-AI co-creation of culturally responsive educational materials for young learners. In *Proceedings of the 24th Interaction Design and Children* (pp. 236–254). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3713043.3727056>
- Hedderich, M. A., Bazarova, N. N., Zou, W., Shim, R., Ma, X., & Yang, Q. (2024). A piece of theatre: Investigating how teachers design LLM chatbots to assist adolescent cyberbullying education. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)* (p. 17). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642379>

- Holtermann, C. (2025, May 17). How students are fending off accusations that they used A.I. to cheat. *The New York Times*. <https://www.nytimes.com/2025/05/17/style/ai-chatgpt-turnitin-students-cheating.html>
- Hutchins, N. M., & Biswas, G. (2024). Co-designing teacher support technology for problem-based learning in middle school science. *British Journal of Educational Technology*, 55(3), 802–822. <https://doi.org/10.1111/bjet.13363>
- Hutchinson, H., Mackay, W., Westerlund, B., Bederson, B. B., Druin, A., Plaisant, C., Beaudouin-Lafon, M., Conversy, S., Evans, H., Hansen, H., Roussel, N., & Eiderbäck, B. (2003). Technology probes: inspiring design for and with families. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)* (pp. 17–24). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/642611.642616>
- Iivari, N., Iversen, O. S., Smith, R. C., Schaper, M.-M., Ventä-Olkkonen, L., Hartikainen, H., Sharma, S., Kinnula, M., Lehto, E., Holappa, J., & Molin-Juustila, T. (2024). Transformative agency – the next step towards children's computational empowerment. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference (Delft, Netherlands) (IDC '24)* (pp. 322–337). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3628516.3655806>
- Jeon, J. (2024). Exploring AI chatbot affordances in the EFL classroom: Young learners' experiences and perspectives. *Computer Assisted Language Learning*, 37(1–2), 1–26. <https://doi.org/10.1080/09588221.2021.2021241>
- Jeon, J., & Lee, S. (2023). Large language models in education: A focus on the complementary relationship between human teachers and ChatGPT. *Education and Information Technologies*, 28(12), 15873–15892. <https://doi.org/10.1007/s10639-023-11834-1>
- Jeon, J., & Lee, S. (2024). Can learners benefit from chatbots instead of humans? A systematic review of human-chatbot comparison research in language education. *Education and Information Technologies*, 29(17), 23329–23360. <https://doi.org/10.1007/s10639-024-12725-9>
- Kali, Y., McKenney, S., & Sagy, O. (2015). Teachers as designers of technology enhanced learning. *Instructional Science*, 43(2), 173–179. <https://doi.org/10.1007/s11251-014-9343-4>
- Kang, W., Zhang, L., Peng, X., Zhang, H., Li, A., Wang, M., Huang, J., Tian, F., & Dai, G. (2025). TutorCraftEase: Enhancing pedagogical question creation with large language models. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1–22). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713731>
- Kao, Y. S., Matlen, B. J., Tiu, M., & Li, L. (2018). Logic models as a framework for iterative user research in educational technology: Illustrative cases. In *End-user considerations in educational technology design* (pp. 52–75). IGI Global Scientific Publishing. <https://doi.org/10.4018/978-1-5225-2639-1.ch003>
- Kaplan, J. S. (2008). The national writing project: Creating a professional learning community that supports the teaching of writing. *Theory Into Practice*, 47(4), 336–344. <https://doi.org/10.1080/00405840802329391>
- Kara-Soteriou, J., & Kaufman, D. (2002). Writing in the elementary school: The missing pieces. *New England Reading Association Journal*, 38(3), 25.
- Kekahio, W., Lawton, B., Cicchinelli, L., & Brandon, P. R. (2014). *Logic models: A tool for effective program planning, collaboration, and monitoring* (REL 2014–025). U.S. Department of Education, Institute of Education Sciences, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Pacific.
- Kim, C. Y., Sato, A. J., White, N. T., Ho, H.-R., Lee, C. P., Hwang, Y., & Mutlu, B. (2025a). Bridging generations using AI-supported co-creative activities. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1–15). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713718>
- Kim, S., So, H.-J., & Park, K. (2025b). Supporting learner agency in collaborative writing with generative AI. *British Journal of Educational Technology*, 57(4), 984–1008. <https://doi.org/10.1111/bjet.70015>
- Kim, Y., & Baylor, A. L. (2006). A social-cognitive framework for pedagogical agents as learning companions. *Educational Technology Research and Development*, 54(6), 569–596. <https://doi.org/10.1007/s11423-006-0637-3>
- Knowlton, L. W., & Phillips, C. C. (2012). *The logic model guidebook: Better strategies for great results*. Sage.
- Krol, S. J., Llano Rodriguez, M. T., & Llor Paredes, M. J. (2025). Exploring the needs of practising musicians in co-creative AI through co-design. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1–13). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713894>
- Lawrence, L. E. M., Guo, B., Yang, K., Echeverria, V., Kang, Z., Bathala, V., Li, C., Huang, W., Alevan, V., & Rummel, N. (2022). Co-designing ai-based orchestration tools to support dynamic transitions: Design narratives through conjecture mapping. In *Computer-Supported Collaborative Learning 2022* (pp. 139–146). International Society of the Learning Sciences. <https://doi.org/10.22318/csc2022.139>
- Lee, M., Gero, K. I., Chung, J. J. Y., Shum, S. B., Raheja, V., Shen, H., Venugopalan, S., Wambsganss, T., Zhou, D., Alghamdi, E. A., August, T., Bhat, A., Choksi, M. Z., Dutta, S., Guo, J. L. C., Hoque, M. N., Kim, Y., Knight, S., Neshaei, S. P., ... Siangliulue, P. (2024). A design space for intelligent and interactive writing assistants. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)* (p. 35). New York, NY: Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642697>
- Leong, J., Pataranutaporn, P., Danry, V., Perteneder, F., Mao, Y., & Maes, P. (2024). Putting things into context: Generative AI-enabled context personalization for vocabulary learning improves learning motivation. In

- Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)* (p. 15). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642393>
- Li, Z., Liang, C., Peng, J., & Yin, M. (2024). The value, benefits, and concerns of generative AI-powered assistance in writing. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)* (p. 25). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642625>
- Lin, P., & Van Brummelen, J. (2021). Engaging teachers to co-design integrated AI curriculum for K-12 classrooms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)* (pp. 1–12). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3411764.3445377>
- Liu, Y., Chen, S., Cheng, H., Yu, M., Ran, X., Mo, A., Tang, Y., & Huang, Y. (2024). How AI processing delays foster creativity: Exploring research question co-creation with an LLM-based agent. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)* (p. 25). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642698>
- Low, J. (2019). A pragmatic definition of the concept of theoretical saturation. *Sociological Focus*, 52(2), 131–139. <https://doi.org/10.1080/00380237.2018.1544514>
- Lu, X., Phyllis Ju, K., Dudley, M., Sano, L., & Wang, X. (2026). AI-mediated feedback improves student revisions: A randomized trial with feedbackwriter in a large undergraduate course. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)* (p. 26). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3772318.3791121>
- McCarthy, P. M., & Jarvis, S. (2010). MTLT, vocd-D, and HD-D: A validation study of sophisticated approaches to lexical diversity assessment. *Behavior Research Methods*, 42(2), 381–392. <https://doi.org/10.3758/BRM.42.2.381>
- McGuire, A. (2023). Leveraging ChatGPT for rethinking plagiarism, digital literacy, and the ethics of Co-authorship in higher education: A position paper and comparative critical reflection of composing processes. *Irish Journal of Technology Enhanced Learning*, 7(2), 21–31. <https://doi.org/10.22554/ijtel.v7i2.131>
- McNamara, D. S., Crossley, S. A., & McCarthy, P. M. (2010). Linguistic features of writing quality. *Written Communication*, 27(1), 57–86. <https://doi.org/10.1177/0741088309351547>
- Min, A., Dickerson, K., Park, S., Hicks, E., Han, A., Rubin, J. D., Lombard, E., Chen, K., Divanji, R., Odgers, C., & Hayes, G. R. (2026). Perceptions of AI-Driven EdTech: Nationwide survey and focus group insights from key end users. *ACM Transactions on Computer-Human Interaction*, 33(3), 1–40. <https://doi.org/10.1145/3798278>
- Mosaiyebzadeh, F., Pouriyeh, S., Parizi, R., Dehbozorgi, N., Dorodchi, M., & Macêdo Batista, D. (2023). Exploring the role of ChatGPT in education: Applications and challenges. In *Proceedings of the 24th Annual Conference on Information Technology Education (Marietta, GA, USA) (SIGITE '23)* (pp. 84–89). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3585059.3611445>
- Muller, M. J., & Kuhn, S. (1993). Participatory design. *Communications of the ACM*, 36(6), 24–28. <https://doi.org/10.1145/153571.255960>
- Nicholson, R., Bartindale, T., Kharrufa, A., Kirk, D., & Walker-Gleaves, C. (2022). Participatory design goes to school: Co-teaching as a form of co-design for educational technology. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)* (pp. 1–17). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3491102.3517667>
- Nordlof, J. (2014). Vygotsky, scaffolding, and the role of theory in writing center work. *The Writing Center Journal*, 34(1), 45–64. <https://www.jstor.org/stable/43444147>
- Ojeda-Ramirez, S., Han, A., Rich, A., Han, S., & Peppler, K. (2025). Reflective AI-partnerships: How middle schoolers balance creativity and AI collaboration. In *Proceedings of the 18th International Conference on Computer-Supported Collaborative Learning-CSCL 2025* (pp. 404–408). International Society of the Learning Sciences. <https://doi.org/10.22318/cscl2025.485037>
- Paatsch, L., Scull, J., & Nolan, A. (2019). Patterns of teacher talk and children's responses: The influence on young children's oral language. *The Australian Journal of Language and Literacy*, 42(2), 73–86. <https://doi.org/10.1007/BF03652028>
- Park, H., & Ahn, D. (2024). The promise and peril of ChatGPT in higher education: Opportunities, challenges, and design implications. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)* (pp. 1–21). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642785>
- Pea, R. D., & Kurland, D. M. (1987). Cognitive technologies for writing. *Review of Research in Education*, 14, 277–326. <https://doi.org/10.2307/1167314>
- Penuel, W. R., Roschelle, J., & Shechtman, N. (2007). Designing formative assessment software with teachers: An analysis of the co-design process. *Research and Practice in Technology Enhanced Learning*, 2(1), 51–74. <https://doi.org/10.1142/S1793206807000300>
- Perkins, M. (2023). Academic integrity considerations of AI large language models in the post-pandemic era: ChatGPT and beyond. *Journal of University Teaching and Learning Practice*, 20(2), 1–24. <https://doi.org/10.53761/1.20.02.07>
- Qin, P., Yang, C.-L., Li, J., Wen, J., & Lee, Y.-C. (2025). Timing matters: How using LLMs at different timings influences writers' perceptions and ideation outcomes in AI-assisted ideation. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (pp. 1–16). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713146>

- Raudys, S. J., & Jain, A. K. (1991). Small sample size effects in statistical pattern recognition: Recommendations for practitioners. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(3), 252–264. <https://doi.org/10.1109/34.75512>
- Ravi, P., Masla, J., Kakoti, G., Lin, G. C., Anderson, E., Taylor, M., Ostrowski, A. K., Breazeal, C., Klopfer, E., & Abelson, H. (2025). Co-designing large language model tools for project-based learning with K12 educators. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems* (pp. 1–25). Yokohama, Japan. ACM. <https://doi.org/10.1145/3706598.3713971>
- Reza, M., Thomas-Mitchell, J., Dushniku, P., Laundry, N., Williams, J. J., & Kuzminykh, A. (2025). Co-writing with AI, on human terms: Aligning research with user demands across the writing process. *Proceedings of the ACM on Human-Computer Interaction*, 9(7), 1–37. <https://doi.org/10.1145/3757566>
- Sadek, M., Constantinides, M., Quercia, D., & Mougenot, C. (2024). Guidelines for integrating value sensitive design in responsible AI toolkits. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)* (p. 20). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642810>
- Saldaña, J. (2013). *The coding manual for qualitative researchers* (2nd ed.). SAGE. OCLC: ocn796279115.
- Sanders, D., & Welk, D. S. (2005). Strategies to scaffold student learning: Applying Vygotsky's zone of proximal development. *Nurse Educator*, 30(5), 203–207. <https://doi.org/10.1097/00006223-200509000-00007>
- Scaife, M., Rogers, Y., Aldrich, F., & Davies, M. (1997). Designing for or designing with? Informant design for interactive learning environments. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (Atlanta, Georgia, USA) (CHI '97)* (pp. 343–350). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/258549.258789>
- Scull, J., Paatsch, L., & Raban, B. (2013). Young learners: Teachers' questions and prompts as opportunities for children's language development. *Asia-Pacific Journal of Research in Early Childhood Education*, 7(1), 69–91.
- Semenovich Vygotsky, L., & Cole, M. (1978). *Mind in society: Development of higher psychological processes*. Harvard University Press.
- Shao, Y., Xiong, J., Wu, C., Wang, X., Zhou, Z., Ouyang, Y., Tao, Q., & Li, Q. (2026). WordCraft: Scaffolding the keyword method for L2 vocabulary learning with multimodal LLMs. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)* (p. 32). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3772318.3790668>
- Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis*, 31(4), 500–506. <https://doi.org/10.3102/0162373709352369>
- Stefaniak, J. E., Dousay, T., Asino, T. I., & Bagdy, L. M. (2025). Promoting co-design in learning design research: Integrating design-based research, decision-making, and ethnographic approaches. *TechTrends*, 69(3), 645–655. <https://doi.org/10.1007/s11528-025-01067-z>
- Tajik, A. (2025). Exploring the role of AI-driven dynamic writing platforms in improving EFL learners' writing skills and fostering their motivation. *Research Square*, 8(4), 69–109. <https://doi.org/10.21203/rs.3.rs-5788599/v1>
- Tankelevitch, L., Kewenig, V., Simkute, A., Scott, A. E., Sarkar, A., Sellen, A., & Rintel, S. (2024). The metacognitive demands and opportunities of generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)* (pp. 1–24). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642902>
- Tanksley, T., Smith, A. D. R., Sharma, S., Jr., & Huff, E. W. (2025). “Ethics is not neutral”: Understanding ethical and responsible AI design from the lenses of black youth. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)* (p. 20). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3706598.3713510>
- Tatar, C., Jiang, S., Rosé, C. P., & Chao, J. (2025). Exploring teachers' views and confidence in the integration of an artificial intelligence curriculum into their classrooms: A case study of curricular co-design program. *International Journal of Artificial Intelligence in Education*, 35(2), 702–735. <https://doi.org/10.1007/s40593-024-00404-2>
- Traga Philippakos, Z. A., MacArthur, C. A., & Munsell, S. (2018). Collaborative reasoning with strategy instruction for opinion writing in primary grades: Two cycles of design research. *Reading & Writing Quarterly*, 34(6), 485–504. <https://doi.org/10.1080/10573569.2018.1480438>
- Van Mechelen, M., Smith, R. C., Schaper, M.-M., Tamashiro, M., Bilstrup, K.-E., Lunding, M., Graves Petersen, M., & Sejer Iversen, O. (2023). Emerging technologies in K–12 education: A future HCI research agenda. *ACM Transactions on Computer-Human Interaction*, 30(3), 1–40. <https://doi.org/10.1145/3569897>
- Wan, Q., Hu, S., Zhang, Y., Wang, P., Wen, B., & Lu, Z. (2024). “It felt like having a second mind”: Investigating human-AI co-creativity in prewriting with large language models. *Proceedings of the ACM on Human-Computer Interaction*, 8(CSCW1), 1–26. <https://doi.org/10.1145/3637361>
- Wang, G., Sun, K., Atabey, A., Pothong, K., Lin, G. C., Zhao, J., & Yip, J. (2023b). Child-centred AI design: Definition, operation, and considerations. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems (CHI EA '23)* (pp. 1–6). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3544549.3573821>

- Wang, J., & Cornely, P.-R. (2024). Addressing academic misconduct in the age of ChatGPT: Strategies and solutions. In *Proceedings of the 2023 7th International Conference on Education and E-Learning (ICEEL '23)* (pp. 19–25). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3637989.3638014>
- Wang, J., & Dong, Y. (2020). Measurement of text similarity: A survey. *Information*, 11(9), 421. <https://doi.org/10.3390/info11090421>
- Wang, X., Liu, Q., Pang, H., Tan, S. C., Lei, J., Wallace, M. P., & Li, L. (2023a). What matters in AI-supported learning: A study of human-AI interactions in language learning using cluster analysis and epistemic network analysis. *Computers & Education*, 194(2023), 104703. <https://doi.org/10.1016/j.compedu.2022.104703>
- Weisz, J. D., He, J., Muller, M., Hoefer, G., Miles, R., & Geyer, W. (2024). Design principles for generative AI applications. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (Honolulu, HI, USA) (CHI '24)* (p. 22). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3613904.3642466>
- Werdiningsih, I., Marzuki, & Rusdin, D. (2024). Balancing AI and authenticity: EFL students' experiences with ChatGPT in academic writing. *Cogent Arts & Humanities*, 11(1), 2392388. <https://doi.org/10.1080/23311983.2024.2392388>
- Wilcox, K. C., Jeffery, J. V., & Gardner-Bixler, A. (2016). Writing to the common core: Teachers' responses to changes in standards and assessments for writing in elementary schools. *Reading and Writing*, 29(5), 903–928. <https://doi.org/10.1007/s11145-015-9588-6>
- Xie, B., Sarin, P., Wolf, J., Garcia, R. C. C., Delaney, V., Sieh, I., Fuloria, A., Varuvel Dennison, D., Bywater, C., & Lee, V. R. (2024). Co-designing AI education curriculum with cross-disciplinary high school teachers. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(21), 23146–23154. <https://doi.org/10.1609/aaai.v38i21.30360>
- Xu, Y., & Warschauer, M. (2020). A content analysis of voice-based apps on the market for early literacy development. In *Proceedings of the Interaction Design and Children Conference (London, United Kingdom) (IDC '20)* (pp. 361–371). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3392063.3394418>
- Xu, Y., He, K., Vigil, V., Ojeda-Ramirez, S., Liu, X., Levine, J., Cervera, K., & Uci, M. W. (2023). “Rosita reads with my family”: Developing a bilingual conversational agent to support parent-child shared reading. In *Proceedings of the 22nd annual ACM interaction design and children conference* (pp. 160–172). <https://doi.org/10.31219/osf.io/wj89p>
- Yan, L., Greiff, S., Teuber, Z., & Gašević, D. (2024). Promises and challenges of generative artificial intelligence for human learning. *Nature Human Behaviour*, 8(10), 1839–1850. <https://doi.org/10.1038/s41562-024-02004-5>
- Yip, J., Clegg, T., Bonsignore, E., Gelderblom, H., Rhodes, E., & Druin, A. (2013). Brownies or bags-of-stuff? domain expertise in cooperative inquiry with children. In *Proceedings of the 12th International Conference on Interaction Design and Children (New York, New York, USA) (IDC '13)* (pp. 201–210). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/2485760.2485763>
- Yuan, A., Coenen, A., Reif, E., & Ippolito, D. (2022). Wordcraft: Story writing with large language models. In *27th International Conference on Intelligent User Interfaces (IUI '22)* (pp. 841–852). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3490099.3511105>
- Zhang, C., Guo, S., Davis, A., & Koh, E. (2026). Narrix: Remixing narrative strategies from examples for story writing. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)* (p. 24). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3772318.3790813>
- Zhang, C., Ju, K., Han, Z., Yen, Y.-C. G., & Rzeszutowski, J. M. (2025). Synthia: Visually interpreting and synthesizing feedback for writing revision. In *Proceedings of the 38th Annual ACM Symposium on User Interface Software and Technology (UIST '25)* (p. 16). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3746059.3747703>
- Zhang, Z., Xu, Y., Wang, Y., Yao, B., Ritchie, D., Wu, T., Yu, M., Wang, D., & Li, T. J.-J. (2022). StoryBuddy: A human-AI collaborative chatbot for parent-child interactive storytelling with flexible parental involvement. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems (CHI '22)* (pp. 1–21). New York, NY. Association for Computing Machinery. <https://doi.org/10.1145/3491102.3517479>

About the authors

Ariel Han is a Postdoctoral Fellow III in Instructional Technology and Learning Sciences at Utah State University. Her research focuses on child-AI interaction, human-centered AI, and understanding the impact of artificial intelligence on learning and teaching.

Shenshen Han is a PhD candidate in Informatics at the University of California, Irvine. His research and technical expertise span software development across web, mobile, embedded, and extended reality (XR) platforms, with interests in the design and development of interactive technologies.

Zhenyao Cai is a PhD candidate in the School of Education at the University of California, Irvine. Her research examines cognitive processes involved in learning through communication, with particular interest in learning through interactions with AI and AI-mediated communication between people.

Joey Huang is an Assistant Professor of Learning, Design, and Technology at North Carolina State University. Her research focuses on computational thinking, educational technology and design, and STEM learning, with an emphasis on understanding and designing technologies and learning experiences that support learners.

Franceli Cibrian is an Assistant Professor of Electrical Engineering and Computer Science in the Fowler School of Engineering at Chapman University. Her research focuses on the design, development, and evaluation of ubiquitous interactive technologies that support children's development, particularly children with special needs.

Kylie Peppler is a Professor of Informatics and Education at the University of California, Irvine. Her research focuses on designing new technologies to support learning and participation among traditionally minoritized populations, particularly at the intersection of the arts, STEM learning, and learning ecosystems.

Appendix A

Table A.1. Comparative analysis of models of writing education and theoretical frameworks of writing support.

Models of writing education	Theoretical frameworks
Emphasis on writing for varied purposes and genres (Coker et al., 2016)	Situated writing practices; writing as a goal-directed and self-regulated cognitive activity (Graham et al., 2012)
Teaching writing strategies (Graham et al. (2012; 2019)	Flower & Hayes' model: Planning, translating, revising cycle (Flower & Hayes, 1981)
Emphasis on writing structure and development (Bradford et al., 2016)	Cognitive writing models emphasizing recursive structure (Flower & Hayes, 1981)
Support for writing fluency and independence (Crowhurst, 1990; Graham et al., 2017; Graham & Perin, 2007b; Harris & Graham, 2009)	Social-cognitive view: writing as meaning-making and audience negotiation (Nordlof, 2014))

Table A.2. Opinion writing prompt engineering strategy samples.

Component	Prompt strategy
Introduction	<i>You are a supportive writing tutor. The student is working on an opinion essay about {teacher-initiated-topic}. Start by asking the student to clearly state their opinion and briefly explain why they hold that view.</i>
Body	<i>The student has written: {student-argument}. Ask the student to elaborate by explaining why they believe this and to provide at least one supporting reason. Prompt deeper thinking with questions like: "What makes this issue important to you?" or "Can you think of an example that supports this idea?"</i>
Reason	<i>The student has listed: {student-reasons}. Encourage them to expand by asking: "What makes this a strong reason?" and "How might someone with a different opinion respond?" Invite them to explore the counterargument for a balanced perspective.</i>
Evidence	<i>The student has included: {student-evidence}. Ask clarifying questions to strengthen their argument, such as: "Where did you find this information?" or "Is there a real-life example or fact that supports this point?" Encourage consideration of alternative viewpoints with: "What might someone who disagrees say?"</i>
Conclusion	<i>Help the student wrap up their essay by restating their opinion and summarizing the reasons. Prompt them with: "Let's restate your claim and go over the reasons you gave." Follow up with: "Can you clearly say your opinion again and briefly explain why you believe it?"</i>
Vocabulary	<i>You are an English teacher reviewing student writing: {student-check-sentence}. Identify any words that appear too frequently for a student in grade {grade #}, and suggest five age-appropriate synonyms to encourage more varied word choice.</i>