



Colin: A Multimodal Human-AI Co-Creation Storytelling System to Support children's Multi-Level Narrative Skills

Lyumanshan Ye

Shanghai Jiao Tong university
Shanghai, China
2623706775@qq.com

Jiandong Jiang

Shanghai Jiao Tong University
Shanghai Jiao Tong University
Shanghai, China
jjdjiang@sjtu.edu.cn

Yuhan Liu

Department of Science and
Engineering
Southern University of Science and
Technology
Shenzhen, Guangdong, China
12111811@mail.sustech.edu.cn

Yihan Ran

School of design
Shanghai Jiao Tong University
Shanghai, China
ran.yh@sjtu.edu.cn

Danni Chang*

School of Design
Shanghai Jiao Tong university
Shanghai, China
dchang1@sjtu.edu.cn

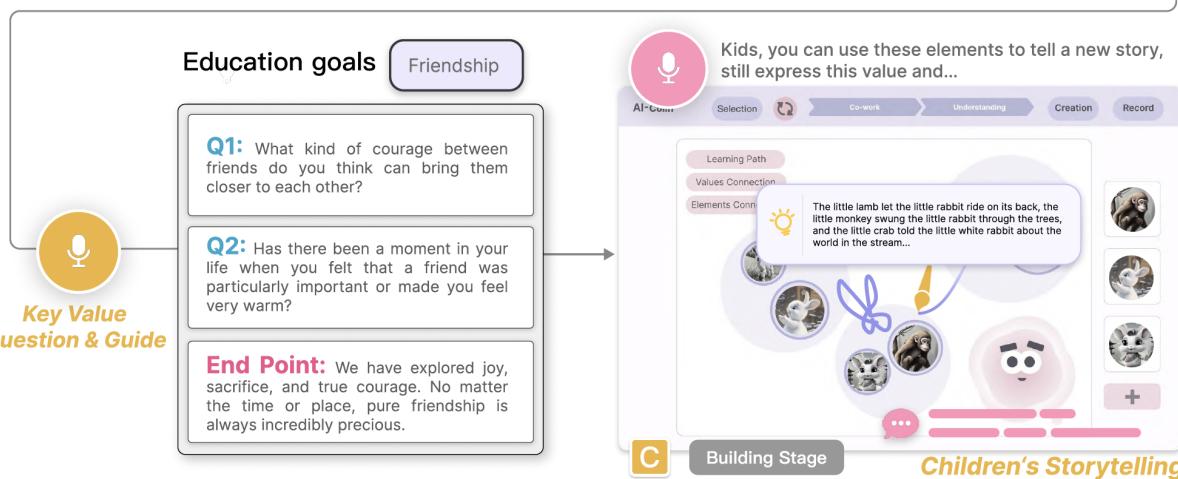
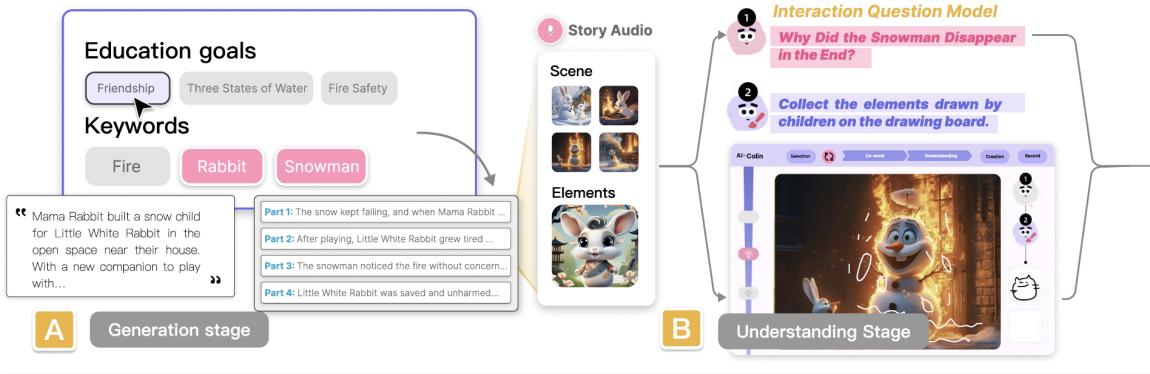


Figure 1: Interaction Flow of Colin

Abstract

Children develop narrative skills by understanding and actively building connections between elements, image-text matching, and consequences. However, it is challenging for children to clearly grasp these multi-level links only through explanations of text or facilitator's speech. To address this, we developed Colin, an interactive storytelling tool that supports children's multi-level narrative skills through both voice and visual modalities. In the generation stage, Colin supports facilitator to define and review generated text and image content freely. In the understanding stage, a question-feedback model helps children understand multi-level connections while co-creating stories with Colin. In the building phase, Colin actively encourages children to create connections between elements through drawing and speaking. A user study with 20 participants evaluated Colin by measuring children's engagement, understanding of cause-and-effect relationships, and the quality of their new story creations. Our results demonstrated that Colin significantly enhances the development of children's narrative skills across multiple levels.

CCS Concepts

- Human-centered computing → Empirical studies in HCI; Interactive systems and tools.

Keywords

Children, AI, Large Language Model, Story-Reading, Interaction, Human-AI Interaction, Personalization, Guided Conversation, Multi-Modal Interaction

ACM Reference Format:

Lyumanshan Ye, Jiandong Jiang, Yuhan Liu, Yihan Ran, and Danni Chang. 2025. Colin: A Multimodal Human-AI Co-Creation Storytelling System to Support children's Multi-Level Narrative Skills. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '25), April 26–May 01, 2025, Yokohama, Japan*. ACM, New York, NY, USA, 11 pages. <https://doi.org/10.1145/3706599.3719837>

1 Introduction

Interactive storytelling is a flexible and effective educational method that integrates multi-level narrative skill with children's participation. In interactive dialogue, facilitator pose questions to children about the story or pictures in the book and provide scaffolded feedback based on children's responses. Feedback strategies include implicit correction, repetition, expansion, and modeling of answers in various forms. Within dialogic reading, facilitator encourage children to participate in the creation of story plots by designing questioning methods and articulating their understanding of the

*Corresponding author: dchang1@sjtu.edu.cn.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

CHI EA '25, Yokohama, Japan

© 2025 Copyright held by the owner/author(s).

ACM ISBN 979-8-4007-1395-8/25/04

<https://doi.org/10.1145/3706599.3719837>

key ideas contained within these storylines. Facilitator often encourage children to visualize their ideas through drawing in addition to using language, as this approach may support their creative and cognitive development.

However, providing real-time guidance to facilitator based on children's feedback during interactive storytelling poses numerous challenges. These challenges include expanding new story lines based on children's ideas and integrating them with the story background, visualizing children's ideas through drawing, and further explaining the key ideas of the story based on children's understanding. To address these challenges, this paper makes the following contributions:

- We designed and developed Colin, a multi-modal child-story creation system, which includes a question-scaffolding feedback framework, to enhance children's narrative multi-level skill.
- We involved facilitator's preview in story generation stage.
- We conducted a user study, summarizing the learning effectiveness and learning experience of Colin.

2 Related Works

2.1 AI-Based Storytelling System

Digital storytelling systems now commonly integrate AI[49] to encompass both the act of telling a story through narration and the depiction of the story's events in visual form. OSOS[30] leverages AI technology by profiling a child's language environment with pervasive language modeling, extracting priority words for learning, and generating bespoke storybooks that incorporate these words[30]. The digital tool Communics[41] is utilized to facilitate children in collaborative comic-based digital storytelling. These systems also empower children to actively participate in and influence story development through decision-making[11], thereby providing more opportunities for storytelling. Interactive storytelling has been demonstrated as an effective educational method [10, 15, 22, 31]. It has been proven that it can improve children's language and narrative skills [5, 34, 50, 51, 54, 56]. However, these studies focus on how to improve the performance of interactive storytelling system, but do not pay attention to the learning process of children's interaction with the story system. With the emergence of large language models, recently, some studies have incorporated large language models into children's interactive storytelling activities. Nisha uses automated planning in natural language text generation to create stories that are coherent and believable [12, 13, 44, 45]. The qualitative research of Yuling Sun proves that parents have a need for interactive story systems [46].

2.2 Children-AI Collaborative Story Generation

Previous studies incorporate LLMs into educational technology [29]. Generation methods include virtual reality avatars, voice interface creation stories [7], through drawing creation, tangible interface [14, 33, 55] and graphic interface [35, 43]. Such tools highlight the significance of multi-modal creative support, allowing children to express themselves through various mediums. However, few studies have addressed the core understanding of children's

Table 1: Design Goals and Challenges

Challenges	Design Goals
C1: Narrative skills development needs elderly facilitator guidance	DG1: Involve elderly facilitator's prior experience in generation stage.
C2: Expanding and integrating children's ideas timely based on existing story framework	DG2: Assist in co-creating children's stories using a question-scaffold approach.
C3: Clarifying multi-level narrative with visual-voice hint	DG3: Support children's understanding and development in multifaceted narrative skills through a multimodal approach.

grasp of stories through participatory creation. Benton presents a storytelling system for narrative co-creation through a game [9], Zarei came up with an enactment-scaffolded narrative authoring tools [36]. But these studies do not integrate the educational goals and understanding processes of stories. Some studies have noted that interactive systems improve children's math [40, 53], coding skill [16, 17, 28, 39] and narrative skills [32]. These adaptations demonstrate the application of LLMs in creating educational content [18, 21], improving student engagement and interaction [3, 6, 47], and providing personalized learning experiences [24, 42]. LLMs have been used to generate children's narratives [4, 20, 23]. In different research trajectories, some scholars have used LLMs to create intelligent learning partners that are able to collaborate with humans [26], provide feedback [27], and encourage students [19, 47]. A common application involves employing LLMs as dialogue partners in written or oral form, such as in the context of task-oriented dialogues that provide opportunities for language practice [19, 52], or as story continuation partners to collaborate with children in creating stories [20, 37].

In contrast to previous studies, Colin incorporates novel design features aimed at enhancing children's visual comprehension and narrative association skills. While the systems previously discussed are predicated on aiding children in the articulation of narrative richness, Colin, through a model that interlinks pictorial elements, encourages children to contemplate and expand upon discrete elements, thereby cultivating the completeness and connectivity within narrative skills.

3 System Design and Development

3.1 Design Goals

To generate the design principles of the participatory story interactive system for children, we recruited five education experts and three parents to conduct semi-structured interviews. Each participant had more than three years of experience in educating or raising children. These interviews focused on three points: educational purposes, benefits of interactive storytelling for children, and problems that facilitator currently encounter in interactive storytelling. We summarized the experts' viewpoints and derived the following three design goals in table 1.

3.2 System Development Integrating Multi-modal LLMs

Colin uses multi-modal LLMs to help children expand on the contributions they wish to add, and effectively weaves the key ideas

and knowledge points into the development of the storyline. It has five interactive steps in interactive storytelling process: 1) Question Guidance: The system generates open-ended questions to elicit narrative contributions from children; 2) Storyline Expansion: Based on children's feedback, the system further develops the storyline to support continued narrative development; 3) Story Element Completion: After children complete some story elements through drawing, the system assists them in completing the story content; 4) Key idea Questions: The system generates questions related to the key ideas of the story, based on children's participation in the creation process, to assess their understanding of the central idea; 5) Explanation of key ideas: The system will affirm children's answers and further explain the key ideas expressed in the story to help them deeply understand the moral and knowledge points of the story. Figure 2 shows the overview of the Colin process.

A round of dialogue introduces the story and provides background information. The Colin asks the children questions and allows them to decide the story's progression at this point.

To accommodate children's short attention spans, limiting questions to two may help maintain focus. At the story's end, the AI poses additional questions and clarifies reasoning. Conversations can occur in multiple rounds; if a child's response contradicts key ideas, the AI provides guidance. Children interact with Colin by pressing a voice button, which lights up during recording. Pressing it again stops the recording, and Colin continues the story based on the child's input while upholding key ideas. After contributing, children can access a drawing board to illustrate elements of their creation. Upon story completion, Colin asks two questions to assess understanding of knowledge and values. Children can engage in multiple voice interactions, and even incorrect responses prompt Colin to patiently explain key ideas. In the final assessment, children use the drawing tool to create or modify elements and manipulate story components, aiming to craft a new story that conveys the original values.

Colin is a web-based app implemented in the React based on JavaScript. For the backend server, we implemented on Flask with Python. We adopted GPT-4 to generate text. Besides, we use ChatGLM [8] to generate images. Colin consists of five key interaction processes 3.

Human-feedback Preview In story generation, parents first input the texts that children use. Then Colin uses outline generation prompt(Appendix A.1.1) to combine the story outline with the character and keywords selected by parents to generate a new story outline. After obtaining the story outline, we use the few shot method to hint the LLM to split the story correctly(Appendix

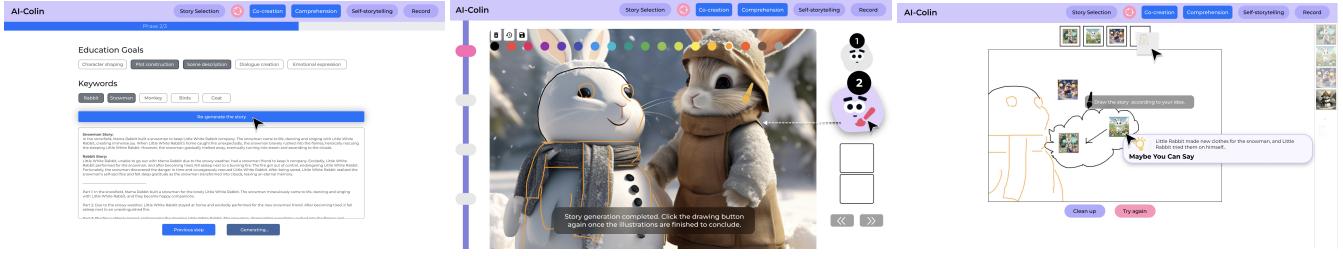


Figure 2: Interaction Process

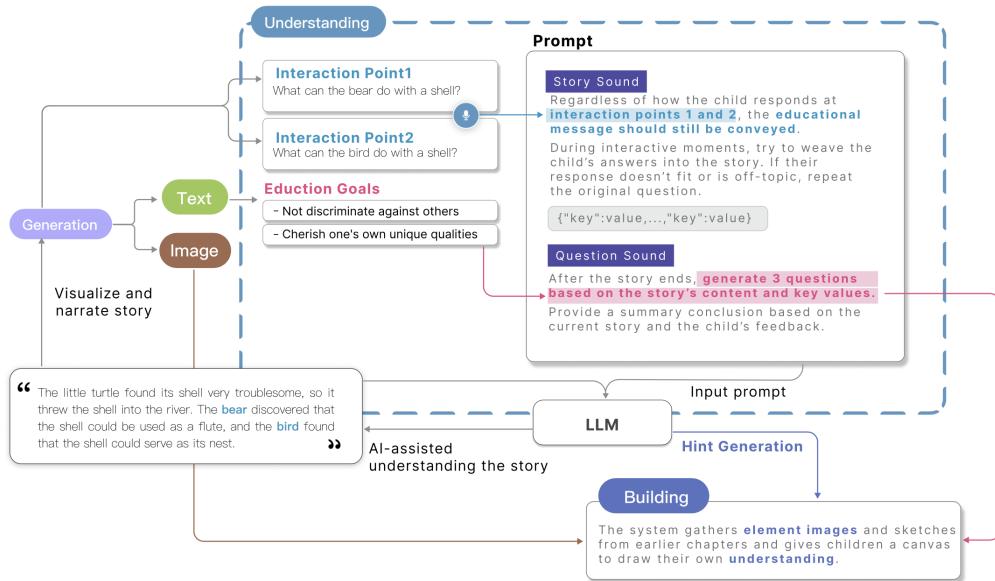


Figure 3: System Prompt Flow Design

A.1.2), and then the LLM divided the story into multiple parts. Once the outline and story fragments are generated, parents can choose to manually re-generate them to ensure the quality of the story text. Then Colin will extract story background and character description through extract prompt(Appendix A.1.3, A.1.4). We use the ChatGLM[8] API to generate background pictures and character pictures from extracted description information. Parents can click on the unsatisfactory existing image to regenerate images. Through LLM and human feedback, we get story text fragments and images used in story interaction.

Story Co-Creation In the story co-creation session, the story fragments are expanded into a complete plot of each chapter through the story generation prompt (Appendix A.1.5), and the LLM is prompted to generate interactive questions in the middle of the chapter to realize the children's ability to co-create the story plot. We save the conversation context in whole process to ensure LLM achieve consistency in storytelling in later story chapters. Colin uses interface provided by iFlytek[25] to implement story text to voice and children voice to text transformation. After the child

clicks on the speak button, the device turns on the built-in microphone for recording. When the child clicks on the speak button again, the audio file is generated and sent to the back-end voice module for parsing into response text and for continuing the story. The child's response will be used to build feedback to generate an extended continuation of the story for the next chapter (Appendix A.1.6). Colin also provides the ability to draw sketches by offering the drawing board function based on p5.js[2], providing a variety of color brushes for children to draw.

Story Connection Build Once co-creation is complete, Colin allows children to express their understanding of the story by clicking a corresponding button in the navigation bar. The system gathers character images and sketches from previous chapters, providing a blank drawing board for children to illustrate their interpretations. They can place image elements onto the board and use a brush tool to enhance their drawings.

3.3 Feasibility of LLM Integration

Evaluation Metrics As shown in Table 4, we designed evaluation metrics including two modules: text generation and question generation. The text generation module is divided into plot extension text and truth extension text, and the question generation module is divided into plot question and truth question. In this method, evaluators were shown stories where one is generated by GPT-4 and another one is generated by a human. They were asked to choose from each pair based on the evaluation dimension. We asked three evaluators to choose from each pair of generated text based on the different dimensions given in the index system. All assessments were conducted by three university student research assistants with more than one year of experience in the education of children.

Results As the figure 4 demonstrates, GPT-4 matches human-generated content in intelligibility and surpasses it in logical rationality, attraction and overall preference.

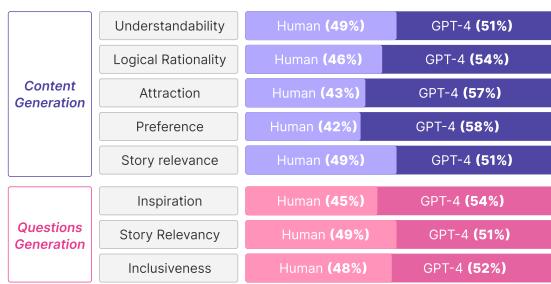


Figure 4: Evaluation Results of GPT-Human

The evaluation result shows that human-generated questions scored lower on story relevance, possibly due to participants associating questions with various life situations rather than focusing on the story. GPT-4 generated questions were more inspiring and inclusive, better promoting story development and gathering meaningful feedback from children. Human participants sometimes created questions more related to daily issues or educational goals rather than the story's core content, often resulting in yes/no questions that limited children's response opportunities. In contrast, GPT-4's questions were more closely aligned with the story's plot and encouraged more comprehensive answers from children. Therefore, GPT-4 can assist in expanding and training children's thinking skills in the field of children's storytelling education, making it suitable for deployment in interactive storytelling systems.

4 Evaluation of Colin

4.1 Participants and Procedure

We recruited 20 child participants (11 female, 9 male, aged 5 to 13 years old, M=7.97, SD=2.81) from child training institutions and kindergartens through offline recruitment. This experiment was approved by the Institutional Review Board (IRB) of the affiliated university. Each participant received a 20CNY compensation for the 60-minute session.

The pre-test phase aimed to assess children's understanding of the story's knowledge before interacting with Colin. Before the experiment, a demographic questionnaire was administered to the children to collect basic information. During the pretest phase, the children were asked questions about the fundamental knowledge in the story.

At the beginning of the experiment, the experimenter introduced each participant on how to use the system and provided 5 minutes for self-practice and inquiries. Participants utilized Colin through the experimenter's browser on their personal computers, with their computer interfaces being recorded. Parents first read the text, selecting the main theme they wished to convey through the story. After reading the text and visually assisted story content, children were required to narrate a story with a similar theme but different plot, based on elements randomly generated by the system. They were also expected to create informational and logical associations between the elements provided by the system. During the children's narration, the experimenter recorded the children's recognition and use of the system's prompted elements (characters or scenes) to assess their ability to map and link visual information. Additionally, the time interval from when the children viewed all the elements to the start of their narration, as well as the proportion of pauses and self-correction time during the narration process, were recorded. These metrics were used to evaluate their capacity to process and match the acquired visual information. Finally, the children's expressed content was analyzed for the use of conjunctions and coherence of sentences to assess their handling of logical associations between matched elements. After completing the story narration, participants were asked to fill out a questionnaire assessing their level of engagement during the interaction with the system. Subsequently, they were invited to take part in a 20-minute semi-structured interview regarding their user experience and usability of Colin.

4.2 Evaluation Metrics

To evaluate how Colin influenced understanding and narrative skill improvement process, we referenced the Language Arts definition in "U.S. Common Core State Standards" [1], and narrative metrics in previous research [38] to evaluate narrative skills through storytelling. We evaluate three level narrative skill in four dimensions, knowledge, internal response, key ideas, and narrator evaluation through a pre-test and post-test questions scored.

The three different levels in the evaluation are:

- **Level-1:** Children identify characters in the story, assigning the basic knowledge of characters in the story.
- **Level-2:** Children can understand the internal response of characters and the key ideas of the whole story in their own expression.
- **Level-3:** Children express their own story, transferring elements from the original story, and narrate to justify why an action or event took place.

To prevent children from merely repeating their pre-test answers in the post-test, the wording of the post-test questions was slightly modified. A comprehensive list of sample questions can be found in the appendix 3.

To measure children's enjoyment in interacting with artificial intelligence or humans, we adapted a questionnaire consisting of 4 items based on the prior work of Waytz [48]. The experimenters verbally presented questions to the participants and guided them to indicate their level of agreement using a graphical Likert scale. The 5-point scale was visualized as five circles, each depicting a different facial expression: "Strong Agree", "Agree", "Neither Disagree Nor Agree", "Disagree", and "Strong Disagree". The circles varied in size, with the smallest representing the least happiness (crying face) and the largest representing the greatest happiness (laughing face). Researchers pointed to each circle when describing the corresponding option. The Cronbach's alpha for internal consistency was 0.77, indicating acceptable reliability.

4.3 Result analysis

4.3.1 Learning Outcome. To calculate the scores on this scale, we recruited three college students with experience in children's education to evaluate the children's performance according to the defined indicators as figure 6. The knowledge and key idea dimensions are scored on a binary scale: 0 for wrong and 1 for right. The internal response and narrator evaluation dimensions Scoring is defined as follows: 2 indicates the ability to apply the key idea to a new story, 1 indicates the ability to retell the original story in their own words, and 0 indicates an inability to complete the task.

Since the data did not follow a normal distribution, pre-test and post-test comparisons were conducted using the paired-sample Wilcoxon signed-rank test. For the "Key Ideas" dimension, all questions and the total score had p-values < 0.05, showing major differences. This suggests that the process of co-creating stories with Colin enhances children's understanding of the key ideas conveyed in the stories. The Internal Response" dimension also showed significant differences in all questions and total scores before and after the test, indicating that Colin not only promotes children's understanding of the key ideas but also helps them abstract and generalize these truths into broader concepts. In the "Narrator Evaluation" dimension, eight participants successfully completed the task of creating a new story that conveyed the same core idea as the original story shortly after hearing it. This result indicates that many children were able to quickly Narrator Evaluation the concept to a new scenario, effectively applying their background knowledge. Nine participants were able to retell the original story to convey the newly learned truth, while only three of the youngest participants(aged 5-6) were unable to construct their own stories.

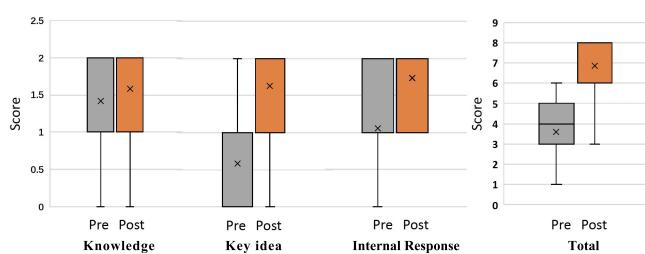


Figure 5: Pre-Post Test Among four Dimensions

As shown in Figure 5, the median of all indicators increased, with notable improvements in the pre-test and post-test scores for key ideas and Internal Response. The total score, calculated from the four learning effect indicators, also showed an increase in the median, demonstrating that Colin is beneficial for multi-level narrative skill development.

Table 2: Wilcoxon signed rank test for paired samples of different dimensions

Pre-test and post-test Pair Comparison	P value	Cohen's d value
Total score of Knowledge	0.180	0.392
Total score of Key idea	0.000***	2.048
Total score of Internal Response	0.000***	1.949
Narrator Evaluation	0.001***	1.798

As shown in Table 2, significant differences were observed in three of the four dimensions: "key ideas", "Internal Response", and "Narrator Evaluation" between pre-test and post-test ($p < 0.05$). However, the overall score in the "knowledge" dimension had p-values greater than 0.05, indicating no significant difference. This is likely due to the familiarity of the story elements—common animals, food, and plants—among the participants, which limited the impact of the Colin system on this dimension.

4.3.2 Learning Experience. Behavioral observation of Engagement: The observations include whether participants showed happiness (16/20), interacted with Colin in a positive tone (16/20), repeatedly used certain Colin features (12/20), and whether they were interested in continuing the discussion about Colin after the experiment (8/20).

Evaluation of Enjoyment: The figure7 shows the children's responses to four questions (Q1 to Q4) about the pleasantness of the system. Each bar represents the percentage of responses ranging from "strongly disagree" to "strongly agree." For Q1, a majority (84%) strongly agreed. Feedback from Q2 was more balanced, with 42% agreeing. Q3 shows that a majority of 21% neither agree nor disagree (gray) and 42% agree. 57% of Q4 respondents agreed, with the rest of the responses spread across other categories. The result shows that children have an overall positive perception of learning story content through collaborative storytelling. There is a general high level of engagement and satisfaction: children's willingness and satisfaction in using the AI story assistant were assessed using a five-dimensions scale ranging from 1 to 5. Among the 20 participants, 16 gave a rating of 5 for all four dimensions on the scale. Children expressed high satisfaction with the combination of voice and visual modalities. Two participants, aged 5-6, accurately articulated the value of the AI story assistant without guidance from experimenters: "It can save my parents' energy because it is a machine and can tell me new stories unlimited times" and "My dad is very busy with work, he only tells me once, but AI can tell me many times".

5 Discussion and Limitations

There are great differences in the development of narrative ability and logic ability of children in different ages. Future research could considerate design different story interaction modes with different

difficulty levels according to children of different ages and abilities. (For example, for children with particularly strong narrative skills, co-creation questions can increase to 3-5). Improving children's ability to understand the key ideas of stories through interactive stories takes time to develop habits. Future story systems should record historical data about each child's interaction, matching varying levels of difficulty to the data the child retains in the system.

The participants in our study engaged in only a single session of AI-assisted collaborative storytelling, which may not fully capture the changes that occur with prolonged use of the storytelling system. Future research should explore the long-term effects of extended interaction with the system, offering children more opportunities to engage with the AI could show changes in their interaction patterns, learning habits, and cognitive processes.

6 Conclusions

In this paper, we introduce Colin, a child-story creation system that emphasizes key story ideas through the use of a question-scaffolding feedback framework. The storytelling system adopt an interactive model of "Question-Feedback-Story Generation", presenting open-ended questions to elicit narrative contributions from children. Based on children's ideas and the story framework, the system expand details to support further story development. The system evaluation reveals that Colin enhance multi-level narrative skill for children in story understanding.

References

- [1] 2024. English Language Arts Standards – Common Core State Standards Initiative. <https://www.corestandards.org/ELA-Literacy/>
- [2] 2024. p5.js. <https://p5js.org/>
- [3] Rania Abdelghani, Yen-Hsiang Wang, Xingdi Yuan, Tong Wang, Pauline Lucas, Hélène Sauzéon, and Pierre-Yves Oudeyer. 2023. Gpt-3-driven pedagogical agents to train children's curious question-asking skills. *International Journal of Artificial Intelligence in Education* (2023), 1–36.
- [4] Amal Alabdulkarim, Siyan Li, and Xiangyu Peng. 2021. Automatic story generation: Challenges and attempts. *arXiv preprint arXiv:2102.12634* (2021).
- [5] Maya Andelina Anggryadi. 2014. The effectiveness of storytelling in improving students' speaking skill. *Jakarta: UHAMKA* (2014).
- [6] Minhui Bao. 2019. Can home use of speech-enabled artificial intelligence mitigate foreign language anxiety—investigation of a concept. *Arab World English Journal (AWEJ) Special Issue on CALL* 5 (2019).
- [7] Heather Barber and Daniel Kudenko. 2010. Generation of dilemma-based interactive narratives with a changeable story goal. In *2nd International Conference on Intelligent Technologies for interactive enterTAINment*.
- [8] Ltd. Beijing Zhipu Huazhang Technology Co. 2024. ChatGLM. <https://chatglm.cn>
- [9] Laura Benton, Asimina Vasalou, Daniel Gooch, and Rilla Khaled. 2014. Understanding and fostering children's storytelling during game narrative design. In *Proceedings of the 2014 conference on Interaction design and children*. 301–304.
- [10] Calvin Bonds. 2016. *Best storytelling practices in education*. Pepperdine University.
- [11] Marc Cavazza, Fred Charles, and Steven J Mead. 2002. Planning characters' behaviour in interactive storytelling. *The Journal of Visualization and Computer Animation* 13, 2 (2002), 121–131.
- [12] Jiaju Chen, Yuxuan Lu, Shao Zhang, Bingsheng Yao, Yuanzhe Dong, Ying Xu, Yunyao Li, Qianwen Wang, Dakuo Wang, and Yuling Sun. 2023. FairytaleCQA: Integrating a Commonsense Knowledge Graph into Children's Storybook Narratives. *arXiv preprint arXiv:2311.09756* (2023).
- [13] Jiaju Chen, Yuxuan Lu, Shao Zhang, Bingsheng Yao, Yuanzhe Dong, Ying Xu, Yunyao Li, Qianwen Wang, Dakuo Wang, and Yuling Sun. 2023. FairytaleCQA: Integrating a Commonsense Knowledge Graph into Children's Storybook Narratives. *arXiv:2311.09756 [cs.CL]* <https://arxiv.org/abs/2311.09756>
- [14] Jean-Péic Chou, Alexa Fay Siu, Nedim Lipka, Ryan Rossi, Franck Dernoncourt, and Maneesh Agrawala. 2023. TaleStream: Supporting Story Ideation with Trope Knowledge. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23)*. Article 52, 12 pages.
- [15] Fiona Collins. 1999. The use of traditional storytelling in education to the learning of literacy skills. *Early Child Development and Care* 152, 1 (1999), 77–108.
- [16] Griffin Dietz, Jimmy K Le, Nadin Tamer, Jenny Han, Hyowon Gweon, Elizabeth L Murnane, and James A. Landay. 2021. StoryCoder: Teaching Computational Thinking Concepts Through Storytelling in a Voice-Guided App for Children. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Article 54, 15 pages.
- [17] Griffin Dietz, Nadin Tamer, Carina Ly, Jimmy K Le, and James A. Landay. 2023. Visual StoryCoder: A Multimodal Programming Environment for Children's Creation of Stories. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Article 96, 16 pages.
- [18] Ramon Dijkstra, Zülküf Genç, Subhradeep Kayal, Jaap Kamps, et al. 2022. Reading Comprehension Quiz Generation using Generative Pre-trained Transformers.. In *iTextbooks@ AIED*. 4–17.
- [19] Reham El Shazly. 2021. Effects of artificial intelligence on English speaking anxiety and speaking performance: A case study. *Expert Systems* 38, 3 (2021), e12667.
- [20] Min Fan, Xinyue Cui, Jing Hao, Renxuan Ye, Wanqing Ma, Xin Tong, and Meng Li. 2024. StoryPrompt: Exploring the Design Space of an AI-Empowered Creative Storytelling System for Elementary Children. In *Extended Abstracts of the 2024 CHI Conference on Human Factors in Computing Systems (CHI EA '24)*. Article 303, 8 pages.
- [21] Ebrahim Gabajiwala, Priyav Mehta, Ritik Singh, and Reeta Koshy. 2022. Quiz maker: Automatic quiz generation from text using NLP. In *Futuristic Trends in Networks and Computing Technologies: Select Proceedings of Fourth International Conference on FINCT 2021*. Springer, 523–533.
- [22] Kathleen Marie Gallagher. 2011. In search of a theoretical basis for storytelling in education research: Story as method. *International Journal of Research & Method in Education* 34, 1 (2011), 49–61.
- [23] Jian Guan, Fei Huang, Zhihao Zhao, Xiaoyan Zhu, and Minlie Huang. 2020. A knowledge-enhanced pretraining model for commonsense story generation. *Transactions of the Association for Computational Linguistics* 8 (2020), 93–108.
- [24] Michael A. Hedderich, Natalie N. Bazarova, Wenting Zou, Ryun Shim, Xinda Ma, and Qian Yang. 2024. A Piece of Theatre: Investigating How Teachers Design LLM Chatbots to Assist Adolescent Cyberbullying Education. In *Proceedings of the CHI Conference on Human Factors in Computing Systems (CHI '24)*. Article 668, 17 pages.
- [25] IflyTek. 2024. Iflytek. https://www.xfyun.cn/doc/tts/online_tts/API.html
- [26] Hyangeun Ji, Insook Han, and Yujung Ko. 2023. A systematic review of conversational AI in language education: Focusing on the collaboration with human teachers. *Journal of Research on Technology in Education* 55, 1 (2023), 48–63.
- [27] Qinjin Jia, Jialin Cui, Yunkai Xiao, Chengyuan Liu, Parvez Rashid, and Edward F Gehring. 2021. All-in-one: Multi-task learning bert models for evaluating peer assessments. *arXiv preprint arXiv:2110.03895* (2021).
- [28] jiawen liu. 2023. GPT-4. <https://arxiv.org/abs/2402.04975>
- [29] Enkelejda Kasneci, Kathrin Seßler, Stefan Küchemann, Maria Bannert, Daryna Dementieva, Frank Fischer, Urs Gasser, Georg Groh, Stephan Günemann, Eyke Hüllermeier, et al. 2023. ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and individual differences* 103 (2023), 102274.
- [30] Jungeun Lee, Suwon Yoon, Kyoosik Lee, Eunae Jeong, Jae-Eun Cho, Wonjeong Park, Dongsun Yim, and Inseok Hwang. 2024. Open Sesame? Open Salami! Personalizing Vocabulary Assessment-Intervention for Children via Pervasive Profiling and Bespoke Storybook Generation. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–32.
- [31] Lana S Leonard. 1990. Storytelling as experiential education. *Journal of Experiential Education* 13, 2 (1990), 12–17.
- [32] Hui Liang, Jian Chang, Shujie Deng, Can Chen, Ruofeng Tong, and Jianjun Zhang. 2015. Exploitation of novel multiplayer gesture-based interaction and virtual puppetry for digital storytelling to develop children's narrative skills. In *Proceedings of the 14th ACM SIGGRAPH International Conference on Virtual Reality Continuum and its Applications in Industry*. 63–72.
- [33] Meng Liang, Yanhong Li, Thomas Weber, and Heinrich Hussmann. 2021. Tangible interaction for children's creative learning: A review. In *Proceedings of the 13th Conference on Creativity and Cognition*. 1–14.
- [34] Marzuki Marzuki, Johannes Ananto Prayogo, and Arwijati Wahyudi. 2016. Improving the EFL learners' speaking ability through interactive storytelling. *Dinamika Ilmu* 16, 1 (2016), 15–34.
- [35] John Murray, Michael Mateas, and Noah Wardrip-Fruin. 2017. Proposal for analyzing player emotions in an interactive narrative using story intention graphs. In *Proceedings of the 12th International Conference on the Foundations of Digital Games*. 1–4.
- [36] Nilofar Francis Quek, Sharon Lynn Chu, and Sarah Anne Brown. 2021. Towards designing enactment-scaffolded narrative authoring tools for elementary-school children. In *Proceedings of the 20th Annual ACM Interaction Design and Children Conference*. 387–395.
- [37] Zhenhui Peng, Xingbo Wang, Qiushi Han, Junkai Zhu, Xiaojuan Ma, and Huamin Qu. 2023. Storyfier: Exploring Vocabulary Learning Support with Text Generation Models. In *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology*. 1–16.

- [38] Douglas B Petersen, Sandra Laing Gillam, and Ronald B Gillam. 2008. Emerging procedures in narrative assessment: The index of narrative complexity. *Topics in language disorders* 28, 2 (2008), 115–130.
- [39] Filipa Rocha, Filipa Correia, Isabel Neto, Ana Cristina Pires, João Guerreiro, Tiago Guerreiro, and Hugo Nicolau. 2023. Coding Together: On Co-located and Remote Collaboration between Children with Mixed-Visual Abilities. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. 1–14.
- [40] Sherry Ruan, Jiayu He, Rui Ying, Jonathan Burke, Dunia Hakim, Anna Wang, Yufeng Yin, Lily Zhou, Qianyao Xu, Abdallah AbuHashem, et al. 2020. Supporting children’s math learning with feedback-augmented narrative technology. In *Proceedings of the interaction design and children conference*. 567–580.
- [41] Carolina Beniamina Rutta, Gianluca Schiavo, Massimo Zancanaro, and Elisa Rubegni. 2020. Collaborative comic-based digital storytelling with primary school children. In *Proceedings of the interaction design and children conference*. 426–437.
- [42] Michael Sailer, Elisabeth Bauer, Riikka Hofmann, Jan Kiesewetter, Julia Glas, Iryna Gurevych, and Frank Fischer. 2023. Adaptive feedback from artificial neural networks facilitates pre-service teachers’ diagnostic reasoning in simulation-based learning. *Learning and Instruction* 83 (2023), 101620.
- [43] Bingchan Shao, Yichi Zhang, Qi Wang, Xinchu Xu, Yang Zhou, Guihuan Feng, and Fei Lv. 2022. StoryIcon: A Prototype for Visualizing Children’s Storytelling. In *Proceedings of the Ninth International Symposium of Chinese CHI (Online, Hong Kong) (Chinese CHI ’21)*. Association for Computing Machinery, New York, NY, USA, 152–157. doi:10.1145/3490355.3490377
- [44] Nisha Simon. 2024. Does Robin Hood Use a Lightsaber?: Automated Planning for Storytelling. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 38. 23421–23422.
- [45] Nisha Simon and Christian Muise. 2022. TattleTale: storytelling with planning and large language models. In *ICAPS Workshop on Scheduling and Planning Applications*.
- [46] Yuling Sun, Jiali Liu, Bingsheng Yao, Jiaju Chen, Dakuo Wang, Xiaojuan Ma, Yuxuan Lu, Ying Xu, and Liang He. 2024. Exploring Parent’s Needs for Children-Centered AI to Support Preschoolers’ Storytelling and Reading Activities. *arXiv preprint arXiv:2401.13804* (2024).
- [47] Tzu-Yu Tai and Howard Hao-Jan Chen. 2023. The impact of Google Assistant on adolescent EFL learners’ willingness to communicate. *Interactive Learning Environments* 31, 3 (2023), 1485–1502.
- [48] Adam Waytz, Joy Hefner, and Nicholas Epley. 2014. The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of experimental social psychology* 52 (2014), 113–117.
- [49] Wendy R Williams. 2019. Attending to the visual aspects of visual storytelling: using art and design concepts to interpret and compose narratives with images. *Journal of Visual Literacy* 38, 1-2 (2019), 66–82.
- [50] Ying Xu, Dakuo Wang, Mo Yu, Daniel Ritchie, Bingsheng Yao, Tongshuang Wu, Zheng Zhang, Toby Jia-Jun Li, Nora Bradford, Branda Sun, Tran Bao Hoang, Yisi Sang, Yufang Hou, Xiaojuan Ma, Diyi Yang, Nanyun Peng, Zhou Yu, and Mark Warschauer. 2022. Fantastic Questions and Where to Find Them: FairytaleQA – An Authentic Dataset for Narrative Comprehension. *arXiv:2203.13947 [cs.CL]*
- [51] Bingsheng Yao, Dakuo Wang, Tongshuang Wu, Zheng Zhang, Toby Li, Mo Yu, and Ying Xu. 2022. It is AI’s Turn to Ask Humans a Question: Question-Answer Pair Generation for Children’s Story Books. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Smaranda Muresan, Preslav Nakov, and Aline Villavicencio (Eds.). Association for Computational Linguistics, Dublin, Ireland, 731–744. doi:10.18653/v1/2022.acl-long.54
- [52] Ann Yuan, Andy Coenen, Emily Reif, and Daphne Ippolito. 2022. Wordcraft: Story Writing With Large Language Models. In *Proceedings of the 27th International Conference on Intelligent User Interfaces (IUI ’22)*. 841–852.
- [53] Chao Zhang, Xuechen Liu, Katherine Ziska, Soobin Jeon, Chi-Lin Yu, and Ying Xu. 2024. Mathemyths: Leveraging Large Language Models to Teach Mathematical Language through Child-AI Co-Creative Storytelling. *arXiv preprint arXiv:2402.01927* (2024).
- [54] Zheng Zhang, Ying Xu, Yanhao Wang, Bingsheng Yao, Daniel Ritchie, Tongshuang Wu, Mo Yu, Dakuo Wang, and Toby Jia-Jun Li. 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)*. Association for Computing Machinery, New York, NY, USA, Article 218, 21 pages. doi:10.1145/3491102.3517479
- [55] Yijun Zhao, Yiming Cheng, Shiyi Ding, Yan Fang, Wei Cao, Ke Liu, and Jiacheng Cao. 2024. Magic Camera: An AI Drawing Game Supporting Instantaneous Story Creation for Children. In *Proceedings of the 23rd Annual ACM Interaction Design and Children Conference (IDC ’24)*. 738–743.
- [56] Mukminatus Zuhriyah. 2017. Storytelling to improve students’ speaking skill. *English Education: Jurnal Tadris Bahasa Inggris* 10, 1 (2017), 119–134.

A APPENDIX

A.1 Prompt Design

A.1.1 Outline Generate Prompt. You will receive a story. Use the provided keywords to simplify the story Keywords: {keyword} Story: {story_text} Target: {target} Where story is the story input, target is the educational goal of the story, and keyword is the character element that needs to appear in the story. You need to generate a 3-4 sentences story outline, called a summary, in which only one character appears in each paragraph and describes its actions.

A.1.2 Story Split Prompt. You will receive a story. Please divide the story into four parts: the beginning, key plot point 1, key plot point 2, and the ending. For each part, describe the actions of the characters in the scene.

The story is: {story_text} The four parts are identified with part1, part2, part3, and part4 as keys. Each part only needs 1-2 sentences, each part only needs to describe the plot, do not add additional description. Here is a complete breakdown of the process: The little tortoise found his shell troublesome, so he threw it into the river. The little bear found that the shell could play the flute. The little bird found that the shell could be used as its nest. They all thanked the little turtle, the little turtle felt very happy ", this story educates children "do not discriminate against others", "cherish their own characteristics" Return Json format: "part1": "Once upon a time, there was a little turtle who always felt that his shell was heavy and hard, which was very troublesome. One day, he decided to take off his shell and throw it into a nearby river", "part2": "Soon after, Little Bear found the shell while playing by the river. Little Bear ponders the function of this turtle shell", "part3": "Little Bear picks up the shell, finds it round and has many holes, and suddenly has a good idea. It took the turtle shell as a special instrument and played a melodious sound, just like playing the flute. My friends in the forest were attracted to the beautiful music.", "part4": "The bird uses the turtle shell as a natural skateboard, gliding between the branches, attracting the envious eyes of other birds, everyone wants to try this special skateboard. The little turtle felt very warm and proud when he heard that his shell had brought so much happiness to his friends." Note: Only one character can appear in each part and do some actions. Don’t have multiple character descriptions and interactions in one part

A.1.3 Background Extract Prompt. You will receive a story’s part. Identify all characters and specify what they are. For example, If a character is Wukong, who is a monkey, refer to him as Wukong the monkey. Assign each character characteristics that describe what they look like, considering the story’s context. For humans, include descriptions of their clothing. Write a single sentence that depicts an image for this page in the following format:

[subject], [doing action], [adjective], [background subject], [scenery]
For example: Little monkey with brown fur, curious eyes, entering the cornfield, tall corn stalks, golden field, looking at the big corn, plentiful corn stalks, sunlit field. The story is: {story_text}

A.1.4 Character Extract Prompt. You will receive a story. Find all of the characters within this story. Don’t just say the name, include what they actually are, without including adjectives. If the character is wukong, who is a monkey, output it as wukong the monkey.

Output them in the following format: Character 1 | Character 2 | Character 3 | . The story is : {story_text}

A.1.5 Story Expand And Interaction Question Generate Prompt. I need to tell stories to children aged 3-8. Please expand and tell the story. {story_split_text} In the story, you design interaction point 1 and interaction point 2, so that interaction point 1 appears after part2, and interaction point 2 appears after part3. Each expansion is less than 250 words and conveys an educational message regardless of the feedback from interaction points 1 and 2.

In the interaction point, try to embed any of the child's answers in the development of the story. If the answer does not fit into the story at all, or has nothing to do with the question, repeat the question generated at the interaction point. If the child has strong resistance to the story telling activity itself, use a patient tone of persuasion and repeat the questions generated by the interaction points.

Now when we get to interaction point 1 in Parts 1 and 2, you need to stop and wait for my feedback. Your reply should be returned in Json format {"key":value,... , "key":value} :

1. The Json tag name corresponding to the story text is story 2. The text of the question raised in interaction point 1 corresponds to the Json label interact 3. Avoid suggestive text in parentheses 4. Animals, people do not appear names, directly use nouns

A.1.6 Interaction Feedback Prompt. For interaction point, the answer is message, please generate feedback, and pay attention to the following points when generating feedback:

1. timely to the child affirmation and praise 2. Help your child fill in the details of this answer

After feedback is generated, talk about next part until interaction point stops, waiting for the child's feedback. Your reply should be returned as a pure json string in the format {"key":value,... , "key":value} :

1. The story text and feedback corresponding to the story is named Story 2. The text of the question raised in point is labeled interact 3. Avoid suggestive text in parentheses 4. Animals, people do not appear names, directly use nouns

A.1.7 Story End And Recall Question Generate Prompt. After generating feedback, proceed directly to the ending of part4 After the ending, please generate three questions according to the material of the story and the truth conveyed. After each question is asked, you need to stop and wait for my feedback, and give feedback according to the answer content, which is controlled within 200 words.

In the follow-up, the first question given by children will be question 1 (Q1), followed by question 2 (Q2), and question 3 (Q3). After the answer of Q3, the guidance will describe the summary and feedback of all links.

If the child's answer does not correctly understand the truth conveyed by the story, you need to patiently guide the child to understand the truth and ask again. Make sure your child understands, then ask the next question. Your reply should be returned in Json format

A.1.8 Recall Feedback Prompt. For the current problem, the child responds as message, generates feedback and waits for the child's response to the next question (Q1, Q2, Q3), and the return should be returned in json format. The format is {"key":value,... , "key":value}

1. The feedback generated after answering the questions is labeled guidance 2. The leading text for the next question is labeled interact 3. Avoid suggestive text in parentheses 4. Animals, people do not appear names, directly use nouns Your return contains only json strings and nothing else

A.1.9 Post-Study.

A.2 Evaluation Metrics and Result

Table 3: List of pre and post test questions

Evaluation dimension	Problem list(Correctly marked as 1 point, incorrectly marked as 0 points)
Knowledge	1. Is the turtle's shell light or heavy? 2. From the following words, please choose the characteristics of the turtle shell: light/soft/heavy.
Key idea	3. Why did the little turtle dislike his shell at first and then become happy again? 4. If the baby turtle is sad because its shell is heavy, how do you comfort it?
Internal Response	5. What does it mean to cherish? 6. What is the definition of characteristics?
Evaluation dimension	Problem list(This dimension is scored 0 points in the pre-test, and 0, 1, and 2 points in the post-test respectively)
Narrator Evaluation	7. Use the elements in the picture (turtle, firefly, cat, fish, bird's nest) to tell a story and express "cherish your own characteristics". 8. Using the elements in the picture (turtle, firefly, cat, fish, bird's nest), tell a story about "finding the good in your friends".

Table 4: Question and Context Generation Evaluation Metrics

Evaluation dimension	Detailed explanation of metrics
Story Relevancy	The questions generated relate to the story plot
Inspiration	The generated question stimulates children's thinking and provokes them to articulate their thoughts
Inclusiveness	The questions generated are open-ended, allowing children to answer freely
Evaluation dimension	Detailed explanation of metrics
Preference	Whose story/principle do you like better?
Understandability [53]	Who explains the key ideas more clearly and concretely for children?
Logical Rationality	Whose story is more logical, with plot develops more rationally?
Attraction	Whose story sounds more creative, contains more twists and turns and more vivid details?
Story Relevancy	Which explanations are more in line with the story's core educational goals?

<i>Evaluation dimension</i>	<i>Standard</i>	<i>Requirement</i>	<i>Example</i>
Knowledge	Whether the child understands the specific knowledge mentioned in the story?	Explaining the meaning of factual information in the story.	<i>What is the characteristic of a turtle's shell?</i>
Key idea	Whether the child understands the key idea expressed by the story?	Analyzing the characters' action and consider how they would advance the plot.	<i>The turtle thinks its shell is heavy, how would you comfort it?</i>
Generalization	Summarize and simplify general vocabulary after understanding the main ideas from specific story plots.	Understand and abstract words mentioned in stories that generalize.	<i>What is the definition of characteristics?</i>
Transfer	Whether children apply the story's value to a new situation and explain it in their own words?	Use random new story elements to create a new story while keeping the core idea.	—————

Figure 6: Evaluation Metrics Definition

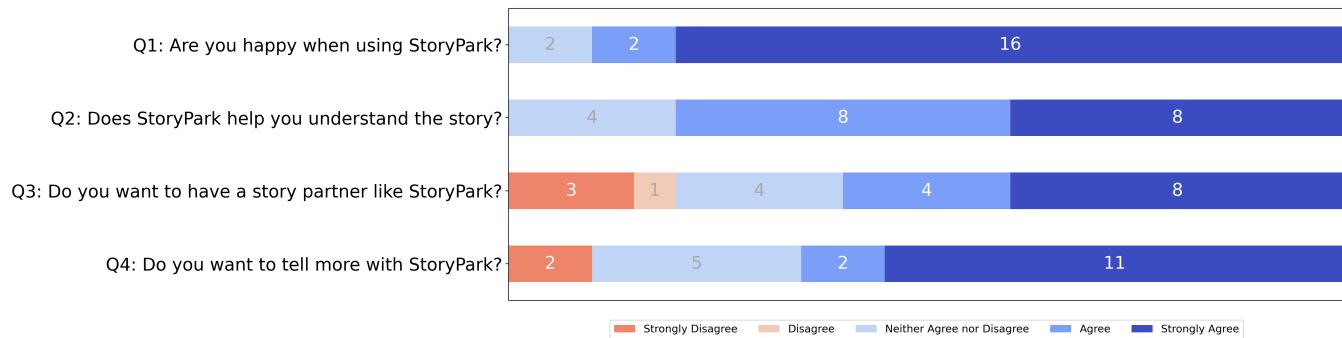


Figure 7: Results of children's Enjoyment in the System