ISCTE IUL
REPOSITÓRIO
INSTITUTO UNIVERSITÁRIO DE LISBOA

Repositório ISCTE-IUL

Deposited in *Repositório ISCTE-IUL*:
2025-06-06

Deposited version:
Accepted Version

Peer-review status of attached file:
Peer-reviewed

Citation for published item:
Carmo, I., Costa, P. & Santana, P. (2024). Boosting children's reading motivation with LLM-generated story crossovers. In Anabela Marto, José Creissac Campos (Ed.), Proceedings of the International Conference on Graphics and Interaction (ICGI). Vila Real: IEEE.

Further information on publisher's website:
https://gpcg.pt/icgi2024/

Publisher's copyright statement:
This is the peer reviewed version of the following article: Carmo, I., Costa, P. & Santana, P. (2024). Boosting children's reading motivation with LLM-generated story crossovers. In Anabela Marto, José Creissac Campos (Ed.), Proceedings of the International Conference on Graphics and Interaction (ICGI). Vila Real: IEEE.. This article may be used for non-commercial purposes in accordance with the Publisher's Terms and Conditions for self-archiving.

Use policy

Creative Commons CC BY 4.0
The full-text may be used and/or reproduced, and given to third parties in any format or medium, without prior permission or charge, for personal research or study, educational, or not-for-profit purposes provided that:

- a full bibliographic reference is made to the original source
- a link is made to the metadata record in the Repository
- the full-text is not changed in any way

The full-text must not be sold in any format or medium without the formal permission of the copyright holders.

Serviços de Informação e Documentação, Instituto Universitário de Lisboa (ISCTE-IUL)
Av. das Forças Armadas, Edifício II, 1649-026 Lisboa Portugal
Phone: +(351) 217 903 024 | e-mail: administrador.repositorio@iscte-iul.pt
https://repositorio.iscte-iul.pt

Boosting Children's Reading Motivation with LLM-Generated Story Crossovers

Inês Carmo
Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal
indco@iscte-iul.pt

Patrícia Costa
Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal; ISTAR—Information Sciences and Technologiesand Architecture Research Center, Lisboa, Portugal
patricia.dinis.costa@iscte-iul.pt

Pedro Santana
Instituto Universitário de Lisboa (ISCTE-IUL), Lisboa, Portugal; ISTAR—Information Sciences and Technologiesand Architecture Research Center, Lisboa, Portugal
pedro.santana@iscte-iul.pt

***Abstract*—It is widely recognised that reading is essential for children's education. Nevertheless, the hours children should spend reading are increasingly being replaced by hours in front of tablets, phones, or computer screens. To counter this trend, we developed a tool that allows educators to create stories that combine children's interests with literary classics. The goal is to motivate the reading indirectly through children's interests. The tool exploits the generative capabilities of ChatGPT to create compelling crossovers between literature classics and children's interests. The tool was developed under the design participatory paradigm. A user study conducted in Lisbon, Portugal with 20 children shows a positive experience with the tool and an increase in motivation to read the classics after using it. This paper presents the design, development, and evaluation of the tool.**

Keywords: Reading Motivation, Large Language Models (LLMs), Educational Mobile Tool, Children's Interests, Tool Design and Development, System Usability Scale (SUS).

I. Introduction

Reading is a fundamental skill that promotes intellectual and emotional growth, and is essential for children's education [1]. It is also a form of play that expands the imagination [2]. For children to continue to develop their reading skills into their adult years, it is important that they enjoy reading and are motivated to continue reading during their younger years [3]. However, recent years have seen the appearance of Generation Z that is inseparable from computers, tablets and the internet [4], and whose number of hours spent in front of a screen is increasing, while reading hours is decreasing [5]. One of the problems children face when trying to develop their literacy skills is identity, both as a reader and finding connections to the literature they read [6].

Reading motivation increases when there is enjoyment in reading [7]. Based on Self Determination Theory's notion of intrinsic motivation, which consists on the inherent enjoyment of doing a certain action without pressures from external sources [8], children might feel more motivated if their own interests are involved. This addition of familiarity and relatability to otherwise boring stories in their perspective, may increase their interest in books that previously did not interest them. Hence, given the motivation children exhibit towards digital content, a reading tool on an electronic device with engaging stories is likely to help to combine these two areas. Creating stories using Artificial Intelligence (AI) around the children's interests might have a big impact on their reading motivation. Moreover, Artificial Intelligence in Education has been a relevant and much researched topic in recent years, and it is predicted to only grow over time [9]. Bearing this in mind, we have designed, developed, and evaluated a tool, in which children can read stories that combine their interests with characters from literary classics and which are created by their educators using AI in a clear and straight forward way. The tool's development was contributed with the children's education professional's suggestions and the stories, which were created through ChatGPT-4o with a curated list of prompts, were also critiqued by them. An experiment conducted with children showed promising results in their enjoyment of the tool and their reading motivation. This paper is organized as follows: Section II will describe the already existing related work on the field of children's reading motivation, e-books and Artificial Intelligence in this context; Section III presents the tool development process; Section IV details the experiment design conducted with children; Section V describes the analysis and the results of the experiment; finally, Section VI draws some conclusions and presents some future research directions.

II. Related Work

Children's reading engagement and enjoyment are connected to their environment. Specifically, the number of books in their home significantly impacts their reading achievement. Additionally, their school environment and their peers' influence, who in turn have been influenced by their own parents, are important for fostering reading habits [10].

Children's motivation to read can be influenced by their perception on its importance and value. Some may not recognise it beyond independent reading skill acquisition [7]. It

This work was partially supported by Fundação para a Ciência e Tecnologia (FCT) / Ministério da Ciência, Tecnologia e Ensino Superior (MCTES) under ISTAR projects UIDB/04466/2020 and UIDP/04466/2020.

is suggested that e-books with a better reading experience might motivate reluctant readers more [11]. Numerous studies have explored children's reading experiences with e-books. Kobayashi et al. [12] studied different paragraph layouts on a reading device to facilitate reading comprehension. Xu et al.'s research stated several challenges children face on the simple swipe motion to turn the page of an e-book [13].

Rubegni et al.'s study focused on finding the elements that make a story on a tablet engaging for children and was remarked that the tablet itself contributed to the children's enjoyment of the stories [14]. Although the importance of the support of an adult in children's reading is undeniable, Korat et al. [15] showed that e-books with expansions through visual and audio channels in the story content can assist children's story comprehension independently. When it comes to the analysis of elementary school student's reading behaviors, Umarji et al. [16] showed how useful user-logs can be to better understand them.

ChatGPT, an AI chatbot based on Large Language Models (LLMs), has already been shown to have great potential in supporting teachers and their students. Its capability to create texts comprehensible to fourth graders was supported, but it still exists the need to have teachers to give contextualisation to the story [17]. Additionally, it was argued that due to its unpredictability, there should be more studies conducted on ChatGPT's capability to cater to a younger audience [18]. Miniscalco et al. [19] verify that posing story related questions may be a good strategy to better children's narrative comprehension who have Attention Deficit Hyperactivity Disorder (ADHD) or Autism Spectrum Disorder (ASD). Furthermore, it was found that for less-proficient readers (fourth graders in the study) answering questions during reading negatively affects comprehension outcomes compared to after reading [20]. ChatGPT's potential and adaptability in generating questions to evaluate the reading comprehension skills of children has been confirmed, and its questions were well tailored for the target demographic's cognitive abilities [21]. Davey's study [22], which focused on autistic children and their special interests, found that these children enjoyed having another person taking interest in their interest and that additionally, these interests can be useful to provide opportunities to learn.

Throughout our literature review, a gap was found regarding the impact of stories with the children's interests on their reading motivation for books that do not appeal to them. Additionally, there is a lack of existing literature concerning the use of ChatGPT for generating stories based on children's specific preferences. In this paper, we present a tool designed to evaluate its effect by analyzing elementary school children's motivation to read with LLM-generated story crossovers. This is done by combining children's interests and literary classics, adequate to these children's age group. LLM-generated questions related to the stories are also created, in order to analyze children's comprehension of them. Questionnaires were applied to students on their background and reading habits, and usability tests of the application were also performed.

III. Tool Design and Development

A. Conceptualization

The created tool "Crossover Stories" has two target users. One includes third and fourth grade children, which roughly correspond from ages 8 to 10 years old, as it is the age where they start developing their autonomous reading ability [23]. The other target user entails children's education professionals that would create new stories with quality for them. Thus, this tool has two phases: the authoring of the stories and their reading.

In the authoring phase, the stories' creator can add and edit the reader's interests (Fig. 2) and literary classics. Automatically through the LLM-based tool, they can also create the crossover stories (Fig. 3) and its questions (Fig. 4), and later edit them in order to verify its quality. On creating the story, they can choose the story's size (less than 5 paragraphs or more than 5 paragraphs) and the age range. The text in all images was translated from Portuguese into English for this paper.

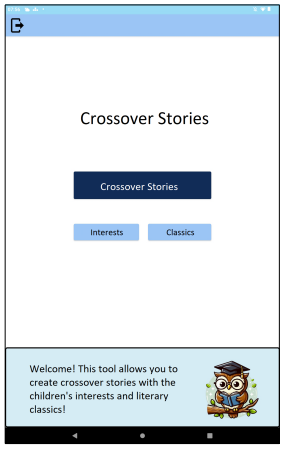


Fig. 1: Creator's home screen

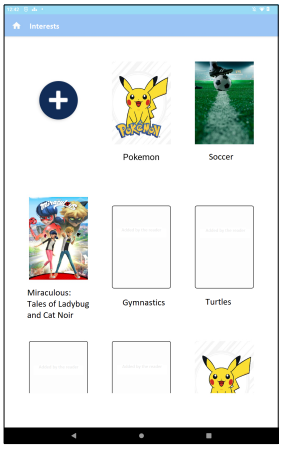


Fig. 2: Creator's screen with the children's interests

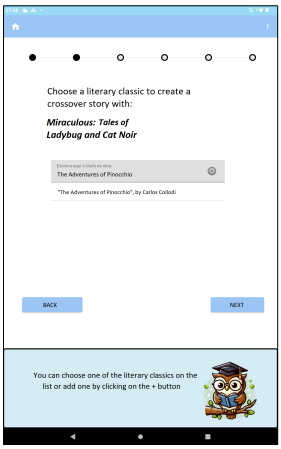


Fig. 3: Creator's screen when choosing a classic to create a crossover story with

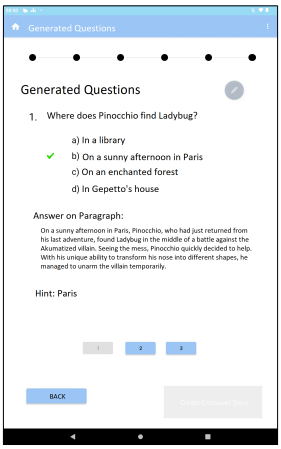


Fig. 4: Creator's screen after the crossover questions are generated

In the reading phase, the reader can choose one of the available interests on the tool (Fig. 5), which will then lead to

a screen that shows crossover stories with that specific interest. After choosing a story to read, another screen would appear with the hints for the story's questions, if the hints variable had been chosen before logging in.

As in common reading tools, there are several ways to navigate the story: by scrolling on the seekbar, by clicking on the buttons on either side of the seekbar, and by touching the sides of the screen (Fig. 6).

When finishing the story, the readers would have to answer three multiple choice questions about it (Fig. 7). With each correct answer, they win a star, in order to provide positive reinforcement [24]. If they answered incorrectly, the tool would jump to the relevant paragraph, allowing them to read it again and answer correctly. To keep motivation high, there were no penalties, and they would always have three stars for completing all the questions. Then, it was asked if they would want to know more about the literary classic that the crossover story they read combined. If so, they would be able to read the summary for that classic (Fig. 8).

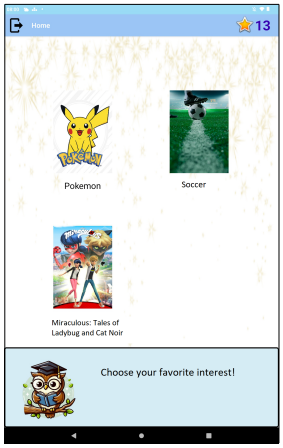


Fig. 5: Reader's home screen



Fig. 6: Paragraph of a crossover story

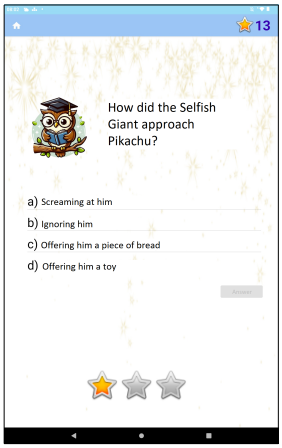


Fig. 7: Crossover story's questions

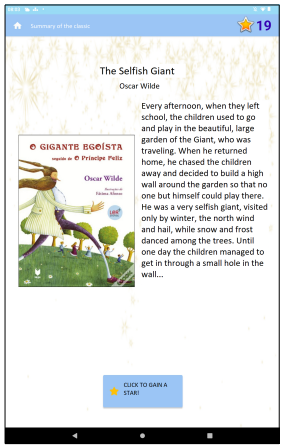


Fig. 8: Reader's screen of the literary classic's summary

When selecting books for reading for leisure, children usually read its abstract and quickly flip through it [25]. Thus, when choosing this option, it indicates that their curiosity was peaked. A button would appear only after three seconds of being on the screen, allowing them to collect another star. This time limit was selected in order to ensure the child stayed on the screen and actually read the summary, decreasing the option of them gaining a star without reading it. When they leave that screen or if they chose not to read the classic's summary, it is asked if they would want to add another interest for later use by the educators when authoring new crossover stories.

Seven children's educators, including children librarians, psychologists and elementary teachers recruited from Lisbon's elementary schools and libraries, played a major part in the development of the tool's concept and design through interviews and usability tests, allowing for a participatory design to be developed. Their role in the tool development consisted on getting their opinion about the study's concept, verifying the stories and questions' quality for the specific age range (vocabulary, complexity, logical progression) and testing the tool's design from the story's creators point of view.

All interviewed professionals reported a very positive opinion about the concept. They stated that incorporating familiar characters and using a device the children were comfortable with, could make them more open-minded about reading. However, three participants expressed doubts about motivating children to read literary classics, noting that it might only work if a child was already curious about reading; otherwise, they felt skeptical. Their feedback helped not only to design the tool, but also to develop the children's reading habits survey and to improve the study's analysis of children's reading motivation.

B. Design and development

Several mock-ups of the prototype's design were made before testing it with professionals through usability tests. On these tests, it was asked for them to perform five tasks on the tool, which included: 1. "Introduce the new interest Pokemon", 2. "Delete an interest", 3. "Edit the name of the author of a literary novel", 4. "Create a crossover story", and depending on the iteration, 5.1 "Delete an already created story", 5.2 "Edit the text of an already created story" or 5.3 "Edit the questions of an already created story". When they were evaluated against the usability tests, the following measures were observed: completion of the task, completion time of the task, number of touches on the screen and level of effort. These measures were chosen in order to effectively assess whether the flow of actions on the design were the most optimal for the participants.

All prototypes focused on task-independent evaluation of the interface by checking its compliance with Nielsen's Usability Heuristics [26]. The interface and interaction were also tested on a task-centred way by running Cognitive Walkthroughs. Each iteration of the prototype was evaluated by different education professionals.

The first prototype tested was a non-functional low-fidelity prototype designed in Figma [27] and was tested with two education professionals. Due to the prototype's lack of cus-

tomizability on Figma, there was some confusion with its interaction and usability. However, all tasks were done successfully, although the level of effort was especially high on the first task. The subsequent prototypes were high-fidelity. The high-fidelity version was implemented to run on a 10" tablet and was programmed using Java on Android Studio, due to its ease of integration with databases and easy to design for handheld electronic devices. The database stored data about the children's reading on the tool. The second iteration of the prototype was tested with one professional. Although the participant completed all tasks successfully, some elements affected the overall workflow of the tool and were therefore resolved for the next iteration.

The third iteration of the prototype was done with one elementary school teacher after the children had also tested the tool, who expressed no difficulties in navigating the tool from the reader's point of view. The iteration included the data added on the "Interests" screen by the children. There was an issue encountered due to the visual clutter this new data created, which overwhelmed the user. Thus, the design was changed to resolve this issue. The fourth and final iteration was tested with three children's education professionals, but the problem with the visual clutter persisted. In addition to a design modification to address the issue, a menu option to delete an interest was added, along with a navigation adjustment.

C. Creating the stories with ChatGPT

The educators create the crossover stories through the tool, by informing it of the children's literary and non-literary interests. Then, under the hood, the tool automatically creates a set of prompts to request ChatGPT-4o [28] to generate the story and its title, questions and hints. ChatGPT's response is automatically parsed by the tool and presented to the user. Then, the user uses the editing features of the tool to curate and edit the automatically generated story.

The set of prompts was constructed by studying ChatGPT's response to several prompts to create the best outcome possible. The list of prompts that gave the best results was obtained using the prompt-chaining technique [29], which involves breaking the main tasks into smaller sub tasks. The LLM will then use the response of the sub task as part of the input to the subsequent prompt.

One prompt on this list (which was translated to English from Portuguese for this paper) states: "Create a story with less than 5 paragraphs and less than 600 characters of a specific moment between *chosen literary classic* and *chosen interest*". The text in italic is inserted by the program. The list consists of eight prompts, where prompts one through three create the story adequate for the age range; prompt four creates a title for the story, prompts five through seven create the questions and their answers and finally, prompt eight gives hints for the created questions.

Stories with visual queues are more enjoyed by children and help to capture the children's attention, as stated by their educators. Therefore, for each paragraph of the story, a related image is included (Fig. 7). These images are created with either ChatGPT or through Designer [30], an AI-powered tool that creates images out of a chosen text prompt. For the crossover stories, the main focus of each paragraph is extracted and put into either tool manually, and the generated image that fit the paragraph the best is chosen. The process of generating the images is done outside of the tool, and these can be added to each paragraph after the story is created.

To evaluate the overall quality of the generated stories and its questions, they were shown to and critiqued by the children's educators. All confirmed that they were clear and had adequate vocabulary for the age range. They appreciated the stories' clear moral and logical flow, but noted that some questions were a bit long, and all answer options should have equal detail. Overall, they all had a positive opinion of both the stories and the questions created. The prompt list was iteratively adapted with the educator's suggestions.

IV. Experiment Design

A. Surveys

In order to better understand children's reading habits before and after their first contact with the tool, a survey was developed and applied to the children. It was based on the most recent cycle of the Progress in International Reading Literacy Study's (PIRLS) [23] survey, which was conducted in 2021, and several questions were added on by the children's educators' suggestions.

It was also given to the children an adaptation of the System Usability Scale (SUS) survey to better understand their opinion about the tool itself. The questions were adapted as Putnam et al. [31] suggested to make them more understandable for children in the chosen age range. Putnam [31] had two adaptations, one for children from ages 7 to 8 and another from 9 to 11. Because our chosen age range is from 8 to 10 years old, the younger option was chosen. The survey structure was adapted to be in compliance with the previous survey, which followed the PIRLS structure, meaning that its answer scale was inverted (most positive answer on the left and most negative on the right).

B. Procedure

The experiment was conducted in Lisbon, Portugal and the participants were recruited from two elementary schools, one private and one public, in Lisbon. The children that wanted to participate on the experiment had their legal guardians give their explicit written consent. The participants were randomly chosen out of the ones that had accepted.

The experiment was done in groups of either 2 or 5 children in the same room, depending on their time availability and the school's room size that was given to run the experiment. Participants were first briefed on the experiment's purpose. Then, they were asked about their reading habits and literary preferences. After completing the reading habits survey, they explored the tool by reading at least one story and answering its questions There were 3 stories available: one combining the literary classic "Adventures of Pinocchio" with soccer,

one combining "The Selfish Giant" with Pokemon and one combining "The Selfish Giant" with the series Miraculous: Tales of Ladybug and Cat Noir. The interests and literary classics were chosen based on the educators suggestions. Additionally, these classics were on the National Reading Plan [32]. The study also wanted to analyse if the children paid more attention and were more motivated to read the story if they knew they had to answer a few questions afterwards. As such, half of the participants were randomly assigned to the "hints variable" group, where a screen would appear before they start reading their chosen story warning them they had to answer 3 questions and showing the hints for them. Finishing their exploration, they would fill one final page of the reading habits survey and finally the SUS survey. In the end, it was given a debriefing document in the style of a participation diploma. Since only one tablet was available for the children to use the tool, the other participants took part in a design exercise, by choosing a mascot that represented 'reading' and 'sharing knowledge'. They evaluated four characters and created stories for them. The favorite was an owl, appreciated for its vibrant colors, symbolism, and detailed elements. All ethic's guidelines from the institution's ethical commission were followed.

V. Results

The experiment was conducted with 20 elementary school children from the third and fourth grade. Out of the participants, 11 were boys and 9 were girls, and 14 were 10 years old and 6 were 9 years old (AVG=9.7, STD=0.458, where AVG stands for Average and STD for Standard Deviation hereafter).

The results obtained from both surveys allowed us to perform a summative evaluation of the tool and its goal of motivating children to read. These results were transposed into excel anonymously. To check the normality assumption it was conducted the Shapiro-Wilk statistic test [33], which revealed that none of the variables were normally distributed. Thus, non-parametric tests were conducted.

A. Reading habits

Most of the children always speak portuguese at home (n=16) and had more than one full shelf of books there. Half did not know if their parents had a degree. Some (n=8) participants read less than 30 minutes per day outside of school, and the majority read for fun and things they chose themselves at least once or twice a week (n=16 and n=18, respectively). They had a wide range of literary tastes, with at least half reading stories, comics or books that explain things at least once or twice a week. Only 2 participants agreed they read only when they were forced to and only 1 thought that reading was boring. Their overall opinion about reading was very positive, with 19 participants agreeing that they like to read. Most (n=13) were uncomfortable reading on a tablet and preferred to read books physically, and went to the library to borrow books at least once or twice a month. When it comes to their opinions of the classics available on the tool before reading the crossover stories, there was a wide variety. Some (n=9) did not want to read "Adventures of Pinocchio" (n=5 knew it), while 9 participants did not know about "The Selfish Giant", with 5 not wanting to read it.

B. SUS questionnaire

The SUS questionnaire results, which used a five-point Likert scale ranging from 1 to 5, are summarized in Table I. All statements refer to the tool. This analysis refer to their level of agreement with the statements. The results indicate a positive experience with the tool.

Code	Statement (Abbreviated)	Bias	AVG	STD
SQ01	I would continue to use it if I had time	+	3.90	1.070
SQ02	The app was difficult to understand	-	1.55	0.887
SQ03	I thought the app was easy to use	+	4.68	0.478
SQ04	I would need help to continue	-	1.85	1.140
SQ05	I knew what to do next	+	4.40	0.995
SQ06	I sometimes felt confused	-	1.95	1.100
SQ07	I had to do complicated things	-	1.65	0.993
SQ08	It will be easy to use by my friends	+	4.37	0.684
SQ09	I'm proud on how I used the app	+	4.40	0.995
SQ10	There were a lot of things to learn	-	2.55	1.360

TABLE I: Results of the adapted SUS questionnaire, where the "Bias" column indicates the direction of a favourable scoring for each statement, where a positive bias (+) indicates that a higher score represents a more favourable response, whereas a negative bias (-) indicates that a higher score represents a less favourable response. In other words, success is higher when higher scores are obtained in positive statements and lower scores are obtained in negative statements. AVG stands for Average and STD for Standard Deviation

The participants slightly agreed they wanted to continue to use the app if they had more time (AVG=3.90, STD=1.070). They thought that the app was not difficult to understand (AVG=1.55, STD=0.887) and agreed that it was easy to use (AVG=4.68, STD=0.478). They disagreed that they would need help to continue to use the app (AVG=1.85, STD=1.140) and agreed that they knew what to do next while using it (AVG=4.40, STD=0.995), also stating they did not feel confused (AVG=1.95, STD=1.100) or had to do complicated things (AVG=1.65, STD=0.993) while using it. They agreed that the app would be easy to use by their friends (AVG=4.37, STD=0.684) and that they are proud on how they used it (AVG=4.40, STD=0.995). Besides the very positive results, they tended to be neutral about thinking there were a lot of things to learn to use the tool (AVG=2.55, STD=1.360). The overall usability score of the tool stands at 80.5 out of 100. Although the questions used were an adaption of the SUS questionnaire, we consider the score to be positive within Sauro's [34] benchmark of 68, that states a perceived usability above average.

The results comparing the SUS score by grouping them based on the chosen reading related variables are shown on Table II. The first line can be read as "11 participants did not have hints and had an average SUS score of 80.75, with a standard deviation of 14.634".

The Kruskal-Wallis Test [35] was conducted to the SUS score's average according to the Hints categories. No significant differences (p = 0.879) were found among the two categories, which included with or without hints.

Children who read more than 30 minutes a day had a higher usability score average (n=11, AVG=84.32, STD=12.707)

Grouped by	Statement	*n*	AVG	STD
Hints	Did not have hints	11	80.75	14.634
	Had hints	9	80.25	16.791
Time Reading Per Day	Less than 30 minutes	8	75.93	18.580
	More than 30 minutes	11	84.32	12.707
Read for Fun	Not everyday or almost	8	75.93	18.411
	Everyday or almost	10	83.25	13.041
Uncomfortable Reading on Devices	Disagree	7	76.78	15.042
	Agree	13	82.50	15.600
Adventures of Pinocchio	Doesn't want to read it	9	75.00	13.017
	Doesn't know but wants to read it	4	94.37	5.694
How Soon will Read the Classics	Sooner	9	84.17	15.000
	Later	9	75.56	16.500
Liked Reading Crossover Stories	Disagree	3	69.17	8.250
	Agree	17	82.50	15.790
More Interested in the Classics	Disagree	5	72.00	12.390
	Agree	13	83.04	16.100

TABLE II: Results of the usability score grouped by the chosen reading related variables, where *n* stands for Sample Size, AVG for Average and STD for Standard Deviation

compared to children who read less (*n*=8, AVG=75.93, STD=18.580). Furthermore, children who read for fun everyday or almost every day (*n*=10, AVG=83.25, STD=13.041) also had a higher usability score average compared to children who read for fun less (*n*=8, AVG=75.93, STD=18.411). Children who did not know the classic "Adventures of Pinocchio" but still were curious to read it had a higher enjoyment of the tool (*n*=4, AVG=94.37, STD=5.694), compared to children who were not curious to read it (*n*=9, AVG=75.00, STD=13.017). However, it is clear that its enjoyment affects children's motivation on how soon they will read the literary classics. The more they enjoyed the tool, the sooner they want to read the classics, meaning they chose either to read the classic "Still this week" or "Next week" (*n*=9, AVG=84.17, STD= 15.000). A similar statement can be made regarding their enjoyment of the crossover stories - the more they enjoyed the tool, the more they enjoyed reading the stories (*n*=17, AVG=82.50, STD=15.790) and the more interested they got in reading the classics (*n*=13, AVG=83.04, STD=16.100). An interesting fact is that it seems children who were uncomfortable reading on an electronic device (*n*=13, AVG=82.50, STD=15.600) actually liked the tool more than children who were not (*n*=7, AVG=76.78, STD=15.042).

Overall, the children who liked the tool the most according to their SUS score were children who already had an established reading habit and who got more curious about the classics.

1) Correlations: There were two missing values on the SUS questionnaire and eighteen on the reading habit's survey. To make a correlation matrix with the obtained results, these missing values were replaced by the median of each variable, since all were not normally distributed and were skewed either towards the minimum or the maximum value. The correlation matrix was calculated with Pearson's Correlation Coefficient. Out of the 42 significant correlations between the SUS survey questions and the children's reading habits, 10 correlations were chosen to highlight, due to their connection with reading motivation, on Table III.

SUS variable	Reading habits variable (Abbreviated)	Correlation	Significance
SQ01	Selected reading	0.619	**
SQ01	Likes reading	0.591	**
SQ01	Liked reading crossover stories	0.652	**
SQ01	Wants more crossover stories	0.542	*
SQ02	Only reads when forced	0.471	*
SQ07	"The Selfish Giant"	-0.498	*
SQ09	Books that explain things	0.706	***
SQ09	Liked reading crossover stories	0.739	***
SQ09	Wants more crossover stories	0.583	**
SQ09	More interested in the classics	0.583	**

TABLE III: Significant correlations between tool usability and reading behavior. Note: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The more they agreed to want to play the tool for more time, the more they read books they choose themselves and they liked to read the crossover stories and in general. The more they thought the app difficult to understand, the more they're forced to read. The more they thought they had to do complicated things on the app, the less they wanted to read the classic "The Selfish Giant". To finalise, how proud they felt while using the tool played a major role in their enjoyment of crossover classics and their curiosity of the classics.

C. Children's reading motivation

To assess if the child's motivation rose after exploring the tool, the questions on Section 4 of the Reading habits' survey, which were answered after their contact with the tool, were analysed and are shown on Table IV.

Code	Statement (Abbreviated)	Bias	AVG	STD
RQ410	When do I want to read the classics	-	2.61	0.979
RQ421	I liked to read the stories	+	3.25	0.716
RQ422	I want to read more crossover stories	+	3.00	0.725
RQ423	I am more interested in the classics	+	3.00	0.745
RQ424	The stories would be better if they had other of my favorite interests	-	3.15	0.875
RQ425	I will recommend this app to a friend	+	3.15	0.875

TABLE IV: Results of Section 4 of the reading habits survey, where the "Bias" column indicates the direction of a favourable scoring for each statement, where a positive bias (+) indicates that a higher score represents a more favourable response, whereas a negative bias (-) indicates that a higher score represents a less favourable response. In other words, success is higher when higher scores are obtained in positive statements and lower scores are obtained in negative statements. AVG stands for Average and STD for Standard Deviation. Question RQ410 had 4 options, while questions 421 through 425 used 4-point Likert scale, ranging from 1 to 4

Question RQ410 did not follow the 4-point Likert Scale like the other five questions, but had the following 4 options: "Still this week", "Next week", "Next month" and "Never". The options are ordered from the most positive to the most negative for this study, making the statement have a negative bias. The participants tended to be neutral on how soon they would want to read the classics (AVG=2.61, STD=0.979). However,

	Motivation *Indicator*	Usability Score
n	Negative	5
	Positive	15
AVG	Negative	76.00
	Positive	82.00
STD	Negative	11.800
	Positive	17.300

TABLE V: Results of the usability score grouped by the Motivation *Indicator* (maximum value = 3), where Negative stands for $Indicator < 1.5$ and Positive for $Indicator \geq 1.5$, and *n*, AVG and STD stand for Sample Size, Average and Standard Deviation, respectively

they did like to read the stories on the tool (AVG=3.25, STD=0.716), and would like to read more stories like them (AVG=3.00, STD=0.725), but would prefer if they had more interests of theirs (AVG=3.15, STD=0.875). They agreed that they got more interested in reading the classics (AVG=3.00, STD=0.745) and they would recommend the tool to a friend (AVG=3.15, STD=0.875).

From these six questions pertaining reading motivation, it was then calculated the average for each participant, which will correspond to their Motivation *Indicator* hereafter. Each participant's Motivation *Indicator* (maximum value = 3) was divided into "Negative" ($Indicator < 1.5$) or "Positive" ($Indicator \geq 1.5$) and the results are shown on Table V. A positive result translates into the participant liking the stories they read on the tool and wanting to read the classics more.

The reading motivation and engagement by using the tool was positive. Although 5 participants had a negative result ($Indicator < 1.5$), their enjoyment of the tool was still high (AVG=76.00, STD=11.800). As expected, the children who had a positive result ($Indicator \geq 1.5$) and subsequently liked reading the stories and had their curiosity peaked on the classics, had a higher enjoyment of the tool (AVG=82.00, STD=17.300). It was thus found a significant positive correlation ($p < 0.05$) between the Motivation *Indicator* and the Usability Score.

Table VI shows the most interesting results when grouping the Motivation *Indicator* (reminding the maximum value is 3) with the chosen variables. There is a significant difference ($p \leq 0.01$) between boys and girls when it comes to their enjoyment of the stories and motivation to read the classics, with boys being the ones with the higher average score (AVG=2.11, STD=0.394). Children who read for fun and who read books that explain things at least once a week (AVG=1.88, STD=0.581 and AVG=2.01, STD=0.441, respectively) had higher motivation to read the classics and enjoyed the crossover stories more. Children who agreed to like to talk about what they were reading with other people (AVG=1.91,STD=0.480) and who would like to have more time to read (AVG=1.91, STD=0.556) also had a higher average score. The children who went to the library to borrow books at least once a month had a higher Motivation *Indicator* average (AVG=2.00, STD=0.494). Like with children's enjoyment of the tool, the existence or lack thereof hints makes no significant ($p > 0.05$) difference in their reading enjoyment or motivation. Finally, the children who read the classic's summary had a higher Motivation *Indicator* average (AVG=2.01, STD=0.381) than those who did not read it (AVG=1.54, STD=0.534).

Grouped by	Statement	*n*	AVG	STD
Gender	Boy	11	2.11	0.394
	Girl	9	1.46	0.549
Read for Fun	Less than once a week	2	1.58	0.25
	At least once a week	16	1.88	0.581
Read Books that Explain Things	Less than once a week	4	1.25	0.534
	At least once a week	14	2.01	0.441
Likes Talking about Books	Disagrees	5	1.55	0.715
	Agrees	15	1.91	0.480
More Time to Read	Disagrees	4	1.75	0.300
	Agrees	15	1.91	0.556
Borrows Books from Library	Sometimes per year or never	5	1.27	0.455
	At least once a month	14	2.00	0.494
Hints	Did not have hints	11	1.81	0.605
	Had hints	9	1.83	0.522
Read Classic's Summary	Did not read	10	1.54	0.534
	Read	8	2.01	0.381

TABLE VI: Results of the Motivation *Indicator* grouped by the chosen reading related variables, where *n* stands for Sample Size, AVG for Average and STD for Standard Deviation. "Read Classic's Summary" sample size is only 18 due to the existence of 2 outliers.

Again, the Kruskal-Wallis test [35] was conducted to the Motivation *Indicator* and all other variables on Table VI. The only one with a significant difference was the variable gender ($p \leq 0.01$) with higher values of motivation for boys.

D. Stories Read

After the experiment, some older children mentioned that while they enjoyed the stories, they found them too simple and wished for longer narratives. Most children (*n*=16) agreed that the stories could be improved by incorporating their favorite interests, such as "Winx", comics, and "One Piece." The children read on average 1.28 stories, excluding 2 outliers that read 5 stories each. It is important to remind that there were only 3 stories available and a time constraint. Out of 34 stories read, the summary of the classic was read 16 times and out of the 102 questions answered, the children only got 3 wrong.

VI. Conclusion

This paper presented the design, development, and evaluation of a didactic tool developed by the authors to boost children's reading motivation. The tool allows children to read stories that combine their own interests with literary classics and for their educators to create more stories in that format through LLMs.

To evaluate the developed tool, we conducted a user study with 20 children. The results indicate a positive user experience with it and a positive opinion on the generated stories. Moreover, the results indicate a relevant potential of combining children's interests with classics to increase their reading motivation. Additionally, the results also show that children

who already had an established habit of reading (i.e., read for fun at least once a week, borrows books from the library at least once a month) are more easily receptive of stories they are not familiar with. Furthermore, the results also show that a higher enjoyment of the tool translated into a higher motivation to read. Its System Usability Scale score, which stood at 80.5 out of 100 overall, show a positive experience with the tool itself, which was designed and developed across multiple iterations with children's education professionals.

As future work we intend to run additional experiments with larger sample sizes and to track the long-term impact of the tool in the children's reading habits. We also intend to keep improving the tool in order to improve the user experience in both target groups (i.e., children and educators). Additionally, we intend to improve the tool in order to enable the generation of more complex stories, with automatically added generated images (i.e., without having to leave the tool).

REFERENCES

[1] K. Palani, "Promoting Reading Habits and Creating Literate Society", Researchers World, vol. 3, pp. 90, 2012.

[2] C. Clark and K. Rumbold, "Reading for Pleasure: A Research Overview," National Literacy Trust, Nov. 2006. Accessed: Apr. 18, 2024. [Online]. Available: https://eric.ed.gov/?id=ED496343

[3] Z. Hercigonja, "The influence of modern media on raising children," in Front impress. 2018, Varaždin.

[4] Z. Bigdeli, G. Haidari, A. HajiYakhchali, and R. Basirianjahromi, "Gamification in library websites based on motivational theories," vol. 13, Jun. 2016.

[5] M. Müller, "Reading Habits of Young People in the Context of Digital Progress: An Example of Research of Republic Croatia," 2021 10th International Conference on Educational and Information Technology (ICEIT), Chengdu, China, 2021, pp. 219-225, doi: 10.1109/ICEIT51700.2021.9375553.

[6] E. Rhodes and G. Walsh, "Recommendations for Developing Technologies that Encourage Reading Practices Among Children in Families with Low-literate Adults," in Proceedings of the The 15th International Conference on Interaction Design and Children, Manchester United Kingdom: ACM, Jun. 2016, pp. 125–136. doi: 10.1145/2930674.2930709.

[7] M. K. Merga and S. Mat Roni, "Children's perceptions of the importance and value of reading," Australian Journal of Education, vol. 62, no. 2, pp. 135–153, Aug. 2018, doi: 10.1177/0004944118779615.

[8] R. M. Ryan and E. L. Deci, "Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions," Contemporary Educational Psychology, vol. 25, no. 1, pp. 54–67, Jan. 2000, doi: 10.1006/ceps.1999.1020.

[9] X. Chen, H. Xie, D. Zou, and G.-J. Hwang, "tool and theory gaps during the rise of Artificial Intelligence in Education," Computers and Education: Artificial Intelligence, vol. 1, p. 100002, 2020, doi: https://doi.org/10.1016/j.caeai.2020.100002.

[10] M. M. Chiu and C. McBride-Chang, "Gender, context, and reading: A comparison of students in 43 countries", en, Sci. Stud. Read., vol. 10, no. 4, pp. 331-362, Oct. 2006.

[11] S. Maynard, "The Impact of e-Books on Young Children's Reading Habits," Pub Res Q, vol. 26, no. 4, pp. 236–248, Dec. 2010, doi: 10.1007/s12109-010-9180-5.

[12] J. Kobayashi and T. Kawashima, "Paragraph-based Faded Text Facilitates Reading Comprehension," in Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow Scotland Uk: ACM, May 2019, pp. 1–12. doi: 10.1145/3290605.3300392.

[13] Y. Xu, J. C. Yau, and S. M. Reich, "The Added Challenge of Digital Reading: Exploring Young Children's Page Turning Behaviors," in Proceedings of the 18th ACM International Conference on Interaction Design and Children, Boise ID USA: ACM, Jun. 2019, pp. 432–438. doi: 10.1145/3311927.3323121.

[14] E. Rubegni and M. Landoni, "Evaluating Engagement in Reading: Comparing Children and Adult Assessors," in Proceedings of the The 15th International Conference on Interaction Design and Children, Manchester United Kingdom: ACM, Jun. 2016, pp. 113–124. doi: 10.1145/2930674.2930696.

[15] O. Korat, M. Tourgeman, and O. Segal-Drori, "E-book reading in kindergarten and story comprehension support," Read Writ, vol. 35, no. 1, pp. 155–175, Jan. 2022, doi: 10.1007/s11145-021-10175-0.

[16] O. Umarji, S. Day, Y. Xu, E. Zargar, R. Yu, and C. Connor, "Opening the black box: user-log analyses of children's e-Book reading and associations with word knowledge," Reading and Writing, vol. 34, no. 3, pp. 627–657, Mar. 2021, doi: 10.1007/s11145-020-10081-x.

[17] E. Murgia, M. S. Pera, M. Landoni, and T. Huibers, "Children on ChatGPT Readability in an Educational Context: Myth or Opportunity?," in Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization, Limassol Cyprus: ACM, Jun. 2023, pp. 311–316. doi: 10.1145/3563359.3596996.

[18] E. Murgia, Z. Abbasiantaeb, M. Aliannejadi, T. Huibers, M. Landoni, and M. S. Pera, "ChatGPT in the Classroom: A Preliminary Exploration on the Feasibility of Adapting ChatGPT to Support Children's Information Discovery," in Adjunct Proceedings of the 31st ACM Conference on User Modeling, Adaptation and Personalization, Limassol Cyprus: ACM, Jun. 2023, pp. 22–27. doi: 10.1145/3563359.3597399.

[19] C. Miniscalco, B. Hagberg, B. Kadesjö, M. Westerlund, and C. Gillberg, "Narrative skills, cognitive profiles and neuropsychiatric disorders in 7–8-year-old children with late developing language," International Journal of Language and Communication Disorders, vol. 42, no. 6, pp. 665–681, 2007, doi: 10.1080/13682820601084428.

[20] van den Broek, P., Tzeng, Y., Risden, K., Trabasso, T., & Basche, P. (2001). Inferential questioning: Effects on comprehension of narrative texts as a function of grade and timing. Journal of Educational Psychology, 93(3), 521–529. https://doi.org/10.1037/0022-0663.93.3.521

[21] X. Chen, H. Xie, D. Zou and F. L. Wang, "ChatGPT for generating stories and mind-maps in storytelling," 2023 10th International Conference on Behavioural and Social Computing (BESC), Larnaca, Cyprus, 2023, pp. 1-8, doi: 10.1109/BESC59560.2023.10386441.

[22] L. Davey, "Using the special interests of autistic children to facilitate meaningful engagement and learning," 2020.

[23] "IAVE Relatórios." Accessed: Jul. 07, 2024. [Online]. Available: https://iave.pt/relatorios/

[24] H. Desurvire and C. Wiberg, "Game Usability Heuristics (PLAY) for Evaluating and Designing Better Games: The Next Iteration," in Online Communities and Social Computing, A. A. Ozok and P. Zaphiris, Eds., Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, pp. 557–566.

[25] L. Colombo and M. Landoni, "Towards an engaging e-reading experience," in Proceedings of the 4th ACM workshop on Online books, complementary social media and crowdsourcing, Glasgow Scotland, UK: ACM, Oct. 2011, pp. 61–66. doi: 10.1145/2064058.2064074.

[26] J. Nielsen and R. Molich, "Heuristic evaluation of user interfaces," in Proceedings of the SIGCHI conference on Human factors in computing systems Empowering people - CHI '90, Seattle, Washington, United States: ACM Press, 1990, pp. 249–256. doi: 10.1145/97243.97281.

[27] "Figma," Figma. Accessed: Mar. 25, 2024. [Online]. Available: https://www.figma.com/file/rJuROxALyUt9uiXTSfc9w2/Material-Design-Icons-Community

[28] "Hello GPT-4o." Accessed: Jul. 07, 2024. [Online]. Available: https://openai.com/index/hello-gpt-4o/

[29] "Prompt Engineering Guide." Accessed: Jul. 02, 2024. [Online]. Available: https://www.promptingguide.ai/techniques/prompt_chaining

[30] "Bing," Bing. Accessed: Jul. 07, 2024. [Online]. Available: https://www.bing.com/images/create/

[31] C. Putnam, M. Puthenmadom, M. A. Cuerdo, W. Wang, and N. Paul, "Adaptation of the System Usability Scale for User Testing with Children," in Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu HI USA: ACM, Apr. 2020, pp. 1–7. doi: 10.1145/3334480.3382840.

[32] P. N. de Leitura 2027, "Catálogo PNL." Accessed: Jul. 07, 2024. [Online]. Available: https://www.pnl2027.gov.pt//np4/catalogo_blx.html

[33] S. S. Shapiro and M. B. Wilk, "An analysis of variance test for normality (complete samples)," Biometrika, vol. 52, no. 3–4, pp. 591–611, Dec. 1965, doi: 10.1093/biomet/52.3-4.591.

[34] J. Sauro, "Measuring Usability with the System Usability Scale (SUS) – MeasuringU.", 2011, Accessed: Jul. 07, 2024. [Online]. Available: https://measuringu.com/sus/

[35] W. H. Kruskal and W. A. Wallis, "Use of Ranks in One-Criterion Variance Analysis," Journal of the American Statistical Association, vol. 47, no. 260, pp. 583–621, 1952, doi: 10.1080/01621459.1952.10483441.