

Storiza: A Platform to Support Children's Oral Reading Fluency Development with Generative AI

Walter L. Leite, Stephanie Hammerschmidt-Snidarich, A. Corinne Huggins-Manley, Hatice K. Karakis, Holly Lane, Seyedahmad Rahimi, Qian Shen, Xinyi Tai
walter.leite@coe.ufl.edu, s.snidarich@coe.ufl.edu, amanley@coe.ufl.edu, hkarakis@ufl.edu, hlane@ufl.edu, srahimi@coe.ufl.edu, Quian.shen1@ufl.edu, xinyitai@ufl.edu
University of Florida

Matthew Schmidt, University of Georgia, matthew.schmidt@uga.edu

Abstract: Oral reading fluency (ORF) is critical for reading comprehension and success in academic and professional goals. Children in elementary school frequently struggle to develop ORF, particularly those with specific learning disabilities, such as dyslexia. This study describes the initial development of an app to enable students to improve ORF by practicing reading stories created with generative AI according to the student's interests and cultural background and aligned with a reading curriculum. The study describes the collection of the teacher, parent and school administrator interviews and qualitative analyses that provided information to create the app's prototype. The app follows universal design principles and learning experience design, incorporating a range of assistive technologies to ensure the success of students with learning differences. Future directions for development are discussed.

Introduction

Ensuring that elementary children, both with and without disabilities, achieve a high level of oral reading fluency (ORF) is vital to reading proficiency and reading comprehension (Paige, 2020). ORF is the ability to read quickly, accurately, and with expression. It is especially critical for child success in STEM disciplines and positive outcomes across an individual's life span. Specifically, developing ORF can have a positive trickle-down effect on children, increasing the likelihood that they will graduate from high school, become employed, and earn a living wage—all factors that can impact overall health and well-being (Lansford et al., 2016). Developing children's ORF eases the task of reading and frees cognitive resources that a child can then use for comprehending text (Cooper et al., 2024). Rather than laboriously decoding text word-by-word, a fluent reader can concentrate on extracting meaning, which is the ultimate purpose of reading. Thus, ORF is a particularly important instructional target.

Although it is optimal for children to work directly with a teacher or tutor on ORF whenever possible, the reality is that public schools frequently have scarce resources (e.g., teaching or tutoring staff, teacher instructional minutes, budget dollars) to provide sufficient opportunities for many children to adequately develop the level of reading proficiency that would enable them to experience academic success across their lifespan. Technology applications that provide children with individual-specific feedback to support ORF development can help fill a critical resource gap.

The objective of this paper is to describe the initial development of Storiza, an app that engages the creativity of elementary school children to facilitate reading by enabling students to create and illustrate their own stories. Storiza uses the GPT-4o large language model (OpenAI., 2024) to provide children with stories that are aligned with a reading curriculum (Lane & Contesse, 2022) that is used at scale nationally and internationally (Gage, 2023). The app fills a void in ORF technology for students with specific learning disabilities, such as dyslexia. It distinguishes itself through its ability to foster student creativity and maintain engagement. Storiza follows universal design for learning principles and learning experience design (Schmidt & Tawfik, 2022), incorporating a range of assistive technologies to ensure the success of students with learning differences. Storiza is designed for in-school or at-home use. In-school use can be done using a range of modalities, including whole-class, small-group, or individual instruction.

Theoretical framework

ORF is a critical reading skill important for reading comprehension. Moreover, the ability to read with ease and accuracy can enhance reading enjoyment and engagement. Students can improve their ORF by simply reading connected text such as passages or books (Maki & Hammerschmidt-Snidarich, 2022). Instructional techniques that support ORF development include Repeated Reading, which involves repeatedly reading a passage of text

multiple times to “practice” the text, or continuous reading, which consists of reading continuously without repetition (Hammerschmidt-Snidarich et al., 2019). Meta-analyses and research syntheses support the effectiveness of these methods (Maki & Hammerschmidt-Snidarich, 2022; Therrien, 2004). Overall, reading practice strategies result in positive ORF growth with moderate to large effect sizes.

Despite the importance of developing ORF, many students do not get enough reading practice. Reasons can include inattention, disinterest, or simply avoidance because of the effort involved (Stanovich, 1980). Teachers need reading resources (e.g., passages, stories) that help to increase students’ reading interest, engagement, and, ultimately, their success. For example, books on topics that represent students’ interests and cultural backgrounds can increase interest, which can, in turn, lead to increased reading engagement. In addition, given a student’s current decoding skills, texts must be appropriately difficult. Texts that are too difficult can be frustrating, while texts that are too easy provide little challenge or opportunities for growth.

The recent development of large language models (LLM) (Wang et al., 2024) has allowed the possibility of developing advanced learning technologies where students provide prompts to create stories they want to read at the right difficulty level. The architectural framework for the development of Storiza is the one shared by Intelligent Tutoring Systems (ITS) (D’Mello & Graesser, 2024), which can provide personalized learning materials, turn-by-turn scaffolding, and just-in-time feedback (VanLehn, 2011). There are extensively studied ITS focusing on reading comprehension (e.g., iSTART; ITSS) (McNamara et al., 2022; Wijekumar et al., 2020), but few target ORF (Mostow et al., 2013), and none use generative AI to provide stories that align with students’ interests and cultural backgrounds and attempt to boost engagement with reading materials. Allowing children to create their own stories for reading practice also addresses the limitation of current ITS that rely on algorithms that do not fully account for child preferences (Mostow et al., 2013). Furthermore, personalization of stories for reading practice promotes culturally relevant instruction because the linguistic and cultural backgrounds of children can be reflected in the stories they create. Therefore, personalization of stories can increase engagement and maintain child motivation. It also encourages children’s creativity and indirectly helps with writing because choosing story topics is an aspect of writing. This study addresses the following research question: What challenges encountered by reading teachers, coaches, tutors, and specialists can be addressed by an AI-powered reading app?

Methodology

A multidisciplinary team was tasked with investigating how generative AI could support student’s development of ORF within a school context. The first step implemented by this team was to develop a set of hypotheses that they aimed to evaluate. These hypotheses are shown in Table 1 in the results section. Data to evaluate the hypotheses were obtained from 30-minute semi-structured interviews with elementary school teachers, reading specialists, tutors, and coaches about the methods and challenges of teaching reading. A few interviews with parents of dyslexic children were also conducted to understand the challenges parents face in supporting children’s reading practice at home. The team also interviewed school administrators to understand the technology adoption process in schools and districts. The team developed protocols for these interviews with broad, open-ended questions that were presented to all participants, as well as specific follow-up questions to be used as needed (1). Detailed notes were taken from these interviews, but they were not recorded.

Sample

There were 85 participants, including 30 teachers, 12 reading specialists, six reading tutors, nine special educational resource teachers, six coaches, 16 school administrators, and six parents. These participants were geographically dispersed in 11 states in the U.S., including participants from Hawaii and Alaska, as well as Canada and Hong Kong. The teachers and reading tutors were recruited by posting an announcement on the Facebook group page of the UFLI Foundations curriculum (Lane & Contesse, 2022), with over 250 thousand members at the time the posting was made. School administrators were recruited for the interview through the contacts of the team members.

Analyses

To analyze the qualitative data collected through semi-structured interviews, we utilized MAXQDA (Kuckartz & Rädiker, 2019) for thematic analysis (Braun & Clarke, 2006). Each hypothesis was systematically coded into predefined categories based on teacher responses and their alignment with the research questions. This process involved multiple iterations to ensure the reliability of codes. For instance, statements that reflected challenges in

engaging students were tagged under "Challenges for ORF\Students lack motivation and concentration," while suggestions for improving ORF practice were mapped to "Suggestions\increase engagement." These codes were further aggregated to get insights into how teachers perceive the potential of generative AI and technology integration in supporting ORF development. In addition to manual coding, we leveraged MAXQDA's frequency analysis tools to identify the prevalence of key themes. Codes such as "Practice for ORF\Use of technology tools" and "Practice for ORF\Gamified learning strategies" frequently emerged, indicating strong interest in leveraging technology and interactive methods for ORF practice. Code co-occurrence analysis allowed us to uncover relationships between challenges and strategies, revealing actionable insights for future tool development. The notes were also examined to identify quotes representative of support for each hypothesis.

We also analyzed the interview notes with BERTopic, which is a topic modeling algorithm that combines pre-trained language models with clustering analysis models (Grootendorst, 2022). To analyze the topics in our interview texts using BERTopic, we first preprocessed the interview texts and split all the texts into individual sentences. Next, we used the pre-trained language model all-MiniLM-L6-v2 as the sentence embedding model for BERTopic and separately analyzed two main categories of interview questions: questions about general topics related to oral reading fluency education and questions about the use of online platforms in oral reading fluency education. During this process, we continuously adjusted BERTopic's hyperparameters to achieve higher coherence scores.

Results

Table 1
Hypotheses and Summary Results of Interviews

Hypotheses	% support	Example quote
1. Teachers have trouble engaging students in oral reading fluency practice, particularly those with a disability (e.g., dyslexia, ADHD, autism spectrum);	40%	"For kids in special education, there is not enough support. For dyslexic kids, I use explicit instruction, sound wall and sound articulations."
2. Students will be more engaged in reading materials they helped create;	41.17%	"If the kids come up with the topic/story it could help to keep them engaged."
3. Teachers welcome the use of technology to improve oral reading fluency practice;	63.52%	"Storiza would help with engagement and providing decodable stories."
4. Teachers will actively search for resources to help with oral reading fluency practice;	56.47%	"I reach out to literacy department in district, to classroom teachers, to other reading specialists to help in finding resources- others reach out to me similarly."
5. Teachers struggle to find a variety of reading materials at the appropriate reading level for students;	41.17%	"...That would be great!! We also have reading a - z and they are supposed to be at their level, but not all the books are really at their level."
6. Teachers would use an AI reading app to create stories with the whole class using the projector;	38.82%	"Projecting the app would be very nice because would build up oral discourse as well."
7. Teachers would use an AI reading app to create stories for small-group instruction;	37.64%	"I could do a lot of group and whole class instruction work. Like introducing a new concept!"
8. Teachers would assign using an AI reading app to create stories as homework.	21.17%	"It would be a great thing for them to make a story in class and read it at home or make a story at home and come to school and read it in class."

Table 1 shows the percentage of interviews that supported each hypothesis and quotes exemplifying the support. Teachers shared that their biggest challenge is keeping students engaged, especially those with dyslexia, dysgraphia or learning difficulties, mentioned 15 times. Another 15 mentions highlighted how hard it is to find materials that work for students at different levels because resources are often not engaging, kid-friendly, or flexible. Limited school support and the requirement to stick to prescribed materials added to the difficulties. To tackle this, teachers frequently turned to gamified teaching strategies, transforming instructional activities into engaging games. Examples included card games, reward charts, and other incentive-based systems to capture students' interest. Many teachers (21 mentions) emphasized finding decodable passages and using platforms like IReady and Lexia to locate more engaging content. Seventeen teachers talked about personalized instruction,

using the UFLI curriculum, assigning homework, and modeling correct reading with error correction. Some teachers also provided suggestions for improving tools like Storiza: Ten teachers recommended incorporating visual elements, such as images, to keep students engaged, while five emphasized focusing on specific ORF skills like prosody. More than ten teachers stressed the importance of a simple, intuitive software design that could be used both at school and at home, accommodating both students and teachers. These insights shed light on practical ways to enhance ORF practices and the tools supporting them.

The results of the BERTopic analysis confirm that teachers have trouble engaging students in oral reading fluency practice, particularly those with specific learning disabilities (hypothesis 1). Using games or high-engagement activities may be one way for teachers to deal with this type of problem. Students may be more interested in materials that are interesting and interactive, but the "materials students helped create" mentioned in the hypothesis did not explicitly appear in BERTopic results (hypothesis 2). Teachers will actively search for technology, resources, and various reading materials at the appropriate reading level for students to improve oral reading fluency practice (hypotheses 3, 4, 5). Meanwhile, teachers are likely to use an AI reading app to create stories collaboratively with the whole class using a projector, for small-group instruction, and as homework assignments (hypotheses 6, 7, 8).

Beyond the hypotheses we mentioned, the results of the BERTopic analysis also show other current situations: Teachers often share and obtain oral reading teaching resources through social media and attach great importance to the assessment and recording of students' oral reading fluency. Limited time, budget, and resources may limit teachers' use of AI reading app. They have clear needs for specific technological equipment and tools and may need more user-friendly and economical solutions. It is crucial to design attractive content when using AI to assist teaching, whether designing classroom teaching or homework. In addition, some students lack learning resources at home, which affects the effectiveness of homework.

Discussion and conclusion

Based on the results of the interviews, the team (2) created a list of features of the Storiza reading app mentioned by teachers and prioritized them according to both frequency in interviews and the previous literature on the effectiveness of strategies to improve ORF. Using learning experience design (Schmidt & Tawfik, 2022), we created personas of elementary school teachers, special education resource teachers, and reading specialists who would use the reading app with their students, mapping their motivators, goals, and behaviors. We then used GPT-4o to create a prototype of Storiza, which creates and illustrates stories based on child-generated prompts. Users can choose to generate stories between 75 and 400 words. Several teachers mentioned the importance of decodable stories, so we attempted to generate decodable stories aligned with lessons 10 to 128 of the UFLI Foundations curriculum (Lane & Contesse, 2022). This ensures that teachers adopting the curriculum can use the stories generated by Storiza to support student reading practice in the same sequence used in the classroom. We also added assistive technologies to Storiza, such as word highlighting, text-to-speech (Wood et al., 2018) and text chunking (Casteel, 1990), to support the success of children with learning differences.

A Storiza user session goes through the following steps: (1) Prompt teachers, tutors, or parents to select a target lesson from the UFLI Foundations curriculum; (2) Prompt students to write story topics and keywords from their own interests and cultural backgrounds; (3) Use generative AI to create and illustrate a story that matches the student's interests and is aligned with the scope and sequence of the curriculum; and (4) Prompt students to read their story and offer support for students with specific learning disabilities, such as chunking of text and modeling.

Storiza extends traditional ITS design through its ability to provide highly personalized reading materials, although it does not incorporate a full suite of learning supports as found in more mature ITS (Xu et al., 2019). Preliminary tests indicate that the stories created incorporate the phonemes being targeted in each lesson of the UFLI Foundations curriculum. However, the performance seems to vary across lessons, and the story size requested, with shorter stories working best for earlier lessons. Development of additional features that would enable Storiza to provide support aligned with modern ITS technologies, such as formative measurement of ORF and fine-grained support while reading, is possible in the future. Thus, the Storiza app is a promising tool for providing children with effective reading practice to develop ORF (3).

Endnotes

- (1) The semi-structured interview protocol can be accessed at <https://osf.io/aq9rj/>
- (2) Co-authors at the University of Florida who worked on the Storiza project were listed in alphabetical order by last name.
- (3) More information about the Storiza project is available at <https://virtuallearninglab.org/research/storiza/>

References

- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/https://doi.org/10.1191/1478088706qp063oa>
- Casteel, C. A. (1990). The effects of chunked material on reading comprehension of high and low ability readers. *Reading Improvement*, 27(4), 269-275.
- Cooper, S., Hebert, M., Goodrich, J. M., Leiva, S., Lin, X., Peng, P., & Nelson, J. R. (2024). Effects of automaticity training on reading performance: A meta-analysis. *Journal of Behavioral Education*, 33(1), 23-52.
- D'Mello, S., & Graesser, A. C. (2024). Intelligent Tutoring Systems: How Computers Achieve Learning Gains that Rival Human Tutors. In P. A. Schutz & K. R. Muis (Eds.), *Handbook of Educational Psychology* (4th ed., pp. 603-629). Taylor & Francis.
- Gage, N. (2023). *Districtwide pilot study of UFLI Foundations*. <https://www.wested.org/wp-content/uploads/2023/05/Districtwide-Pilot-Study-of-UFLI-Foundations.pdf>
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv preprint*.
- Hammerschmidt-Snidarich, S. M., Maki, K. E., & Adams, S. R. (2019). Evaluating the effects of repeated reading and continuous reading using a standardized dosage of words read. *Psychology in the Schools*, 56(5), 635-651.
- Kuckartz, U., & Rädiker, S. (2019). *Analyzing Qualitative Data with MAXQDA*. Springer.
- Lane, H., & Contesse, V. (2022). *UFLI Foundations: An explicit and systematic phonics program*. Ventris Learning.
- Lansford, J. E., Dodge, K. A., Pettit, G. S., & Bates, J. E. (2016). A public health perspective on school dropout and adult outcomes: A prospective study of risk and protective factors from age 5 to 27 years. *Journal of Adolescent Health*, 58(6), 652-658.
- Maki, K. E., & Hammerschmidt-Snidarich, S. (2022). Reading fluency intervention dosage: A novel meta-analysis and research synthesis. *J Sch Psychol*, 92, 148-165. <https://doi.org/10.1016/j.jsp.2022.03.008>
- McNamara, D. S., Arner, T., Butterfuss, R., Fang, Y., Watanabe, M., Newton, N., McCarthy, K. S., Allen, L. K., & Roscoe, R. D. (2022). iSTART: Adaptive Comprehension Strategy Training and Stealth Literacy Assessment. *International Journal of Human-Computer Interaction*, 39(11), 2239-2252. <https://doi.org/10.1080/10447318.2022.2114143>
- Mostow, J., Nelson-Taylor, J., & Beck, J. E. (2013). Computer-Guided Oral Reading versus Independent Practice: Comparison of Sustained Silent Reading to an Automated Reading Tutor That Listens. *Journal of Educational Computing Research*, 49(2), 249-276. <https://doi.org/10.2190/EC.49.2.g>
- OpenAI. (2024). GPT-4 Technical Report. *arXiv*(2303.08774). <https://www.openai.com/research/>
- Paige, D. D. (2020). Reading Fluency: A Brief History, the Importance of Supporting Processes, and the Role of Assessment. In.
- Schmidt, M., & Tawfik, A. A. (2022). Activity Theory as a Lens for Developing and Applying Personas and Scenarios in Learning Experience Design. *Journal of Applied Instructional Design*, 11(1), 1-12.
- Stanovich, K. E. (1980). Toward an interactive-compensatory model of individual differences in the development of reading fluency. *Reading Research Quarterly*, 16(1), 32-71.
- Therrien, W. J. (2004). Fluency and comprehension gains as a result of repeated reading: A meta-analysis. *Remedial and Special Education*, 25(4), 252-261.
- VanLehn, K. (2011). The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educational Psychologist*, 46(4), 197-221. <https://doi.org/10.1080/00461520.2011.611369>
- Wang, Z., Chu, Z., Doan, T. V., Ni, S., Yang, M., & Zhang, W. (2024). History, development, and principles of large language models: an introductory survey. *AI Ethics*. <https://doi.org/https://doi.org/10.1007/s43681-024-00583-7>
- Wijekumar, K., Meyer, B. J., Lei, P., Beerwinkle, A. L., & Joshi, M. (2020). Supplementing teacher knowledge using web-based Intelligent Tutoring System for the Text Structure Strategy to improve content area reading comprehension with fourth- and fifth-grade struggling readers. *Dyslexia*, 26(2), 120-136. <https://doi.org/10.1002/dys.1634>
- Wood, S. G., Moxley, J. H., Tighe, E. L., & Wagner, R. K. (2018). Does Use of Text-to-Speech and Related Read-Aloud Tools Improve Reading Comprehension for Students With Reading Disabilities? A Meta-Analysis. *J Learn Disabil*, 51(1), 73-84. <https://doi.org/10.1177/0022219416688170>
- Xu, Z., Wijekumar, K., Ramirez, G., Hu, X., & Irey, R. (2019). The effectiveness of intelligent tutoring systems on K-12 students' reading comprehension: A meta-analysis. *British Journal of Educational Technology*, 50(6), 3119-3137. <https://doi.org/10.1111/bjet.12758>