TEACHING GROUNDED READING SKILLS VIA AN INTERACTIVE ROBOT TUTOR

A PREPRINT

Kayleigh Bishop
Department of Computer Science
University of Colorado Boulder
Boulder, CO 80302
kaleb.bishop@colorado.edu

Bradley Hayes
Department of Computer Science
University of Colorado Boulder
Boulder, CO 80302
bradley.hayes@colorado.edu

Alessandro Roncone
Department of Computer Science
University of Colorado Boulder
Boulder, CO 80302
alessandro.roncone@colorado.edu

February 18, 2021

1 Introduction

Reading comprehension is a critical skillset for student success in education. The U.S. Common Core guidelines expect students to be able to identify key details, main topics, and connections between events in multi-paragraph informational texts as early as the second grade. These skills are vital to tackling educational texts and gaining full reading proficiency as students get older [1]. Those who struggle at this transition point from concrete decoding to conceptual understanding are often very likely to continue to struggle with reading skills as their education progresses [2].

Despite a wide variety of new initiatives and educational practices aimed at improving reading skills in elementary education [3], progress has largely stagnated in improving ELA education in the United States. Scores on the National Assessment of Educational Progress long-term exams show only small improvements in performance for students at ages 9 and 13 from 1971 to 2013 – but no statistically significant change at all for students at age 17 over the same time period [4]. In addition, there continues to be a significant racial achievement gap in reading skills on standard assessments, most prominently between white and black students, which begins at least as early as age 9 and tends to persist or even widen by age 17 [2].

One promising avenue in educational research is the realm of cognitive learning theory. In recent decades researchers of cognitive science and language have proposed and elaborated upon the thesis that reading comprehension is embodied – that is, all cognitive processes involved in reading comprehension, however abstract they may appear, are grounded in sensorimotor neural activity. This description has led to the development of a number of embodiment-centered practices in reading education, several of which have seen promising results on the scales of entire schools or school districts (see [5, 6]). While these practices vary in their particular approaches, they share a focus on grounding key comprehension processes, such as main idea extraction and conclusion drawing, in multimodal experiences, mental "images," and behaviors. The preliminary success of these practices when compared to standard approaches indicates the vast potential of leveraging embodied learning in the classroom.

Given the value of multimodal, sensorimotor engagement in building strong reading comprehension skills, robot tutoring systems are a prime candidate for implementing personalized, embodiment-centered techniques in the classroom. Robots have significant potential to work with students in personalized 1-on-1 settings, in ways that are often not viable for teachers – due to either their high numbers of students, or to students' reluctance to make mistakes in front of a teacher in such a close setting [7, 8]. While robot tutoring systems have seen impressive results in the domains of math and STEM tutoring [9], second language learning [10], and social skills [11], as of yet few to no studies have touched on the potential of applying robots to the development of abstract reading comprehension skills.

In this paper, we propose a classroom-based social robot tutoring system that targets reading comprehension skills, using methods inspired by successful embodied reading pedagogical practices. This system will incorporate skill tracing of reading comprehension skills in order to identify problem areas and help students ground their reading in concrete experiences and ideas. We begin by providing background on embodiment-based pedagogy in the domain of reading comprehension, as well as the state of the art in robot tutoring and skill tracing methods. We then outline our proposed implementation of such a system, and conclude by described our proposed evaluation setting and experimental setup.

2 Background and Related Work

2.1 Pedagogical approaches to reading comprehension

The embodied account of language understanding and reading comprehension has been incorporated into pedagogical practices in a number of ways. In particular, one of the most well-extended specific embodied theories with respect to literacy is Dual Coding Theory (DCT), which maintains that connections between verbal and nonverbal mental simulations account for the comprehension of both concrete and abstract concepts. DCT has been implemented in multiple forms of classroom practices aiming to strengthen these connections through multimodal integration, with substantial success across grade levels. Key examples most relevant to our proposed work include Bell's dialogue-based *Visualizing and Verbalizing* (VV) [12] and Block et al.'s motor-focused *Comprehension Process Motions* (CPM) [13]. Notably, a key aspect of the above techniques, as well as related ones like the successful *Extending Concept through Language Activities* (ECOLA) [14], is social interaction and dialogue. Verbal and non-verbal social interaction is an embedded aspect of the embodiment-centered learning process, as both the pedagogical channel and the concrete experience in which reading skills are grounded. Contrary to what might be expected, having students perform analogous reading exercises alone, rather than in pairs, results in inferior gains in independent reading skills [14, 15].

2.2 Robotic tutoring

Current applications. Robotic tutoring has seen unprecedented success in application domains ranging from STEM areas [9] to second-language learning [10, 16] to social therapies [11]. Despite these preliminary successes, applying current robotic tutoring techniques to reading comprehension remains a largely unaddressed challenge. The most closely related prior work to our objective is likely that conducted within the domain of second-language learning; however, the majority of such applications have focused on concrete vocabulary or grammar rule acquisition, rather than higher-order comprehension skills [10, 17].

Technical approaches to skill tracing. A key aspect of robot tutoring is the robot's ability to track the progress of the student: having a clear model of the mastered and not-mastered skills of the particular student, from which it can make decisions about its next tutoring action. Partially Observable Markov Decision Processes (POMDPs) and variants have been frequently used as a core pedagogical module with considerable success [18]. However, Rafferty et al. found POMDPs are typically on par with Maximum Information Gain (MIG) algorithms in terms of the resulting action-policy, until the skill space reaches a critical size. One of the most historically popular approaches to skill tracing is Bayesian Knowledge Tracing (BKT), a variant of Hidden Markov Models consisting of observable "signal" variables and latent "skill" variables. Many variants of BKT methods exist, from those that use affect as part of their skill belief-updates to those that integrate many modeled skills into a single dynamic Bayesian network. Of particular note is Schodde et al.'s Adaptive BKT-based tutoring model [17], which integrates models of the robot's own actions to solve for the best action at any given belief state. For our proposed study targeting only a few key skills with many potential actions, BKT appears to be a fitting approach.

Social robotics for embodied learning. The success of embodied learning practices has flung open the door to potential robotics applications. Embodied AI agents have many demonstrated advantages over screen-based agents in tutoring contexts [19, 16, 8]. Part of the success of robots in learning domains can be traced to their unique role as a physically present, social agent. Researchers have proposed that robots' impressive value in working with children with Autism Spectrum Disorder (ASD) stems from robots' ability to be inherently rewarding to interact with, while encouraging students to engage socially through their design and AI. Further, for both children and adults, robots have the ability to shape group social interactions and influence prosocial behavior through their decisions and displays of vulnerability [20]. This success as social agents positions robots as particularly well-suited to collaborative learning environments, being able to encourage positive collaboration and engagement with the material.

Embodied learning practices are a prime domain in which to exploit the strengths of embodied agents to allow for more engaging and effective tutoring. Robots also can be designed as inherently collaborative partners, allowing for learning that is maximally multimodally engaging. Like many novel instructional techniques, a major potential roadblock to any large-scale implementation of embodiment-centered practices is the amount of individual attention they might demand from already overextended teachers. Robots in the classroom provide an exciting way to surpass this obstacle and instate these effective learning practices on a larger scale.

3 Methods

3.1 Tutoring goals

Using the theory and practices of *Visualizing and Verbalizing* (VV) as our central inspiration and the U.S. Common Core standards to inform our goals, we aim to target some of the following skills at the second- and/or third-grade levels:

- **Concrete detail identification**: Answering *where, who, what* questions about the text.
- **Action/event detail identification**: Answering *why, how* questions about events or actions in the text.
- **Sequence explanation**: Understanding the order of events or steps described in the text.
- **Using context clues**: Inferring the meaning of an unfamiliar word or phrase in a relevant text.
- **Main topic/argument understanding**: Understanding the primary argument or "main idea" of a text.

While this is by no means a comprehensive list of skills that could be targeted by such a system, these skills all depend to varying degrees on strengthening what Bell [12] describes as "Gestalt understanding" of concepts conveyed in text. This makes them particularly well suited to using VV-inspired techniques to develop proficiency in mental visualization and grounding.

3.2 Robot tutoring for reading skills

Skill assessment. The system will begin the task starting from its ground truth: a grade-level-appropriate text passage annotated for content and semantic role. From this annotated data, the system will produce an image response (see figure) for each associated comprehension prompt question. The student, given this text and a prompt question, will create their own response using a touch interface on a connected tablet computer. The tutoring robot will then be able to compare a vectorized representation of this response with its own, both to assess its correctness and to identify how the student might have come to a particular incorrect response. The benefit of this setup is that it allows us to assess abstract reading skills in as concrete a way as possible, while allowing for the system to model the incorrect "rules" a student might be using that is resulting in a wrong answer. Understanding these hidden "rules" will allow the model to make more informed decisions about the next steps that should be taken, given a particular incorrect response. The goal behind prompting an image response is to reinforce a correct grounding of abstract comprehension techniques in visual and tactile experiences.

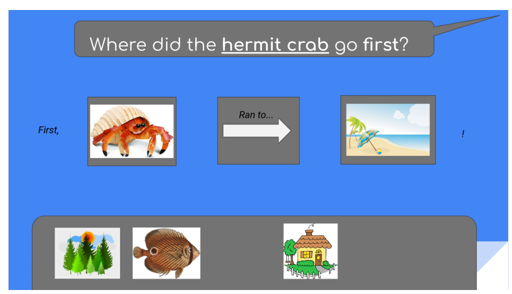

Figure 1: A mockup image of the tablet interface for the interactive robot. Students will be tasked with choosing and arranging the provided images to answer the prompt posed by the robot partner.

Knowledge tracing. Our approach to skill modelling will be based on Bayesian Knowledge Tracing (BKT), a common approach among robot tutoring implementations. Although standard BKT implementations typically use separate networks for each skill, we propose using a a single comprehensive dynamic Bayesian network to represent beliefs over all skill levels. Our reasoning for this is that reading comprehension skills as we hope to characterize them are highly inter-dependent. This comprehensive implementation would also allow for reasoning over which order skills may be learned best. Each skill will have an associated latent variable S, representing the skill belief in one of 4-5 bins describing relative mastery (the exact amount of bins will likely depend on the skill and the associated practice texts available). Common incorrect "rules" (e.g. assuming the first sentence of a paragraph is always the main idea, or that the first character mentioned is the main character) that students may use will form additional hidden nodes that have their own associated emission probabilities. These emission probabilities can be assessed by the system using its response production rules.

Selecting next example. The system should aim to present examples that will maximally clarify the robot's belief of the student's skill state – as well as maximally help the student understand the core strategies of reading comprehension through grounded examples. For this reason, the system will select the next action using a maximum information gain (MIG) approach over the associated skill belief states.

3.3 Embodied implementation

In order to make the engagement with the robot social in nature, we plan to replicate (or approximate) with the robot the social feedback behaviors most frequently identified in educator feedback [17]: eye contact, joint attention, and nodding. The robot will be programmed to execute these behaviors consistently when giving instructions, prompts, or feedback. An ideal embodied agent for this task would be able to perform all of these behaviors in addition to some other social or motor output, such as smiling or gesturing. Based on these criteria, research robots such as the Jibo, Nao, or DarwinOP [21, 22, 23] would be good candidates for the hardware implementation of our system.

4 Evaluation

4.1 Interaction Structure

The evaluation of this system will take place following this basic protocol (after [17] and other similar HRI user studies):

1. **Opening**: The interaction should begin with an introduction for the student to the robot and its basic capabilities.
2. **Game Setup**: The basic structure of the reading skills activity, along with the tablet interface, should be explained. The student will be able to ask clarifying questions.
3. **Test Run**: A test run of the prompt and response interaction will be conducted to ensure the instructions and interface can be understood.
4. **Practice**: The main reading skills practice activity can now begin, and be run for roughly 3-4 rounds. Each response from the student is followed by spoken verbal feedback from the robot tutor, followed by a chance to retry if appropriate.
5. **Closing**: The interaction will end as positively as possible; the robot can thank the student for participating and give praise for practice they put into developing their skills.

4.2 Data Collection

The data to be collected from this evaluation will include the following:

- **Performance data**: The system will store performance over the interaction tasks for each student.
- **Pre- and post-test data**: Prior to the experimental interaction, students will take a brief pre-test evaluating their mastery of the skills practiced in the tutoring interaction. The same (or a very similar) test will be administered 48-72 hours after the experimental interaction. This is to test both the effectiveness and the retention of the practiced skills.
- **Robot perception survey results**: Immediately following the the interaction, students will be asked to complete the Robotic Social Attributes Scale [24] with the goal of evaluating students' perceptions of the robots' warmth, competence, and associated discomfort.
- **Backchanneling, speech, and social gestures**: Any nods, smiles, gestures, or speech directed towards the robot will be recorded based on collected video data.

5 Next Steps and Conclusions

We hope to make significant progress on this system before the end of the current academic year, with a rough goal of a working implementation of our interface, automated assessment, and skill tracing systems within the next three to four months. We believe that this system and its evaluation will forward the field of robot tutoring by bringing it into a new domain, with new challenges to face in creating engaging and effective pedagogical interactions. As such, through our implementation we will be tackling the difficulty of taking the relatively abstract skills of reading comprehension and grounding them in a social and computational framework – a challenge that, perhaps ironically, mirrors the very challenge students take on in learning these skills. Through this work we aim to open the door to future investigation of interactive paradigms and technical approaches that can be used to develop the vital skills of reading comprehension.

References

[1] Common core standards by state, subject, and grade. *U.S. Department of Education*, Nov 2018.

[2] Charles T Clotfelter, Helen F Ladd, and Jacob L Vigdor. The academic achievement gap in grades 3 to 8. *The Review of Economics and Statistics*, 91(2):398–419, 2009.

[3] Sebastian P. Suggate. A meta-analysis of the long-term effects of phonemic awareness, phonics, fluency, and reading comprehension interventions. *Journal of Learning Disabilities*, 49(1):77–96, 2016. PMID: 24704662.

[4] Thomas D. Snyder and Sally A. Dillow. Digest of education statistics, 2013. *National Center for Education Statistics*, May 2015.

[5] Mark Sadoski and Victor L Willson. Effects of a theoretically based large-scale reading intervention in a multicultural urban school district. *American Educational Research Journal*, 43(1):137–154, 2006.

[6] Mark Sadoski. Reading Comprehension is Embodied: Theoretical and Practical Considerations. *Educational Psychology Review*, 30(2):331–349, 2018.

[7] Iris Howley, Takayuki Kanda, Kotaro Hayashi, and Carolyn Rosé. Effects of social presence and social role on help-seeking and learning. In *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '14, page 415–422, New York, NY, USA, 2014. Association for Computing Machinery.

[8] Daniel Leyzberg, Samuel Spaulding, Mariya Toneva, and Brian Scassellati. The Physical Presence of a Robot Tutor Increases Cognitive Learning Gains. *Proceedings of the Annual Meeting of the Cognitive Science Society*, 34(34), 2012.

[9] Aditi Ramachandran, Chien Ming Huang, Edward Gartland, and Brian Scassellati. Thinking Aloud with a Tutoring Robot to Enhance Learning. *ACM/IEEE International Conference on Human-Robot Interaction*, pages 59–68, 2018.

[10] Daniel Leyzberg, Aditi Ramachandran, and Brian Scassellati. The Effect of Personalization in Longer-Term Robot Tutoring. *ACM Transactions on Human-Robot Interaction*, 7(3), 2018.

[11] Brian Scassellati and Marynel Vázquez. The potential of socially assistive robots during infectious disease outbreaks. *Science Robotics*, 5(44):1–3, 2020.

[12] N Bell. Visualizing and verbalizing for learning comprehension and thinking, 2007.

[13] Cathy Collins Block, Sheri R Parris, and Cinnamon S Whiteley. Cpms: A kinesthetic comprehension strategy. *The reading teacher*, 61(6):460–470, 2008.

[14] Haerazi and Lalu Ari Irawan. The effectiveness of ECOLA technique to improve reading comprehension in relation to motivation and self-efficacy. *International Journal of Emerging Technologies in Learning*, 15(1):61–76, 2020.

[15] Anita Varga. Metacognitive perspectives on the development of reading comprehension: a classroom study of literary text-talks. *Literacy*, 51(1):19–25, 2017.

[16] Eun Ja Hyun, So Yeon Kim, Siekyung Jang, and Sungju Park. Comparative study of effects of language instruction program using intelligence robot and multimedia on linguistic ability of young children. *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN*, pages 187–192, 2008.

[17] Thorsten Schodde, Kirsten Bergmann, and Stefan Kopp. Adaptive Robot Language Tutoring Based on Bayesian Knowledge Tracing and Predictive Decision-Making. *ACM/IEEE International Conference on Human-Robot Interaction*, Part F1271:128–136, 2017.

[18] Aditi Ramachandran, Chien-Ming Huang, and Brian Scassellati. Toward Effective Robot–Child Tutoring. *ACM Transactions on Interactive Intelligent Systems*, 9(1):1–23, 2019.

[19] Jeonghye Han, Miheon Jo, Sungju Park, and Sungho Kim. The educational use of Home Robots for children. *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication*, 2005:378–383, 2005.

[20] Sarah Strohkorb Sebo, Margaret Traeger, Malte Jung, and Brian Scassellati. The ripple effects of vulnerability: The effects of a robot's vulnerable behavior on trust in human-robot teams. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, pages 178–186, 2018.

[21] Project overview: Jibo social robotic research platform.

[22] Nao the humanoid and programmable robot: Softbank robotics.

[23] Darwin-op humanoid research robot - deluxe edition.

[24] Colleen M Carpinella, Alisa B Wyman, Michael A Perez, and Steven J Stroessner. The robotic social attributes scale (rosas) development and validation. In *Proceedings of the 2017 ACM/IEEE International Conference on human-robot interaction*, pages 254–262, 2017.