

Towards assessing persistence in reading in young learners using pedagogical agents

Caitlin Tenison¹, Beata Beigman Klebanov¹, Noah Schroeder², Shan Zhang², Michael Suhan¹, Chuyang Zhang²

¹ Educational Testing Service, ² University of Florida

Correspondence: ctenison@ets.org

Abstract

This pilot study investigated the use of a pedagogical agent to administer a conversational survey to second graders following a digital reading activity, measuring comprehension, persistence, and enjoyment. Analysis of survey responses and behavioral log data provide evidence for recommendations for the design of agent-mediated assessment in early literacy.

1 Introduction

Understanding how young learners respond to reading difficulties is important for supporting early literacy development. This paper presents a preliminary effort to use a pedagogical agent to elicit self-reflections during reading and examine how these reflections align with behavioral patterns during reading captured through process data.

2 Literature Review

2.1 Persistence and Reading

Persistence is generally defined as a student's sustained effort toward academic goals despite difficulty, confusion, or failure (Skinner et al., 2022; Wang et al., 2020). It is a key predictor of learning and long-term achievement, particularly in complex problem-solving tasks (Dweck et al., 2014; Farrington et al., 2012). In K-12 settings, persistence is often supported through instructional strategies such as productive failure (Kapur, 2008), erroneous examples (Richey et al., 2019), and motivational framing (Cook et al., 2019). Persistence is measured in various ways, including behavioral indicators (e.g., time on task, number of attempts) and self-report instruments (e.g., grit scales, mindset surveys) (Shute et al., 2013). For instance, Goh (2025) measured persistence by log-transformed time on task and found it was associated with improved performance.

In the reading context, persistence is a critical factor in reading achievement. A meta-analysis

by Toste et al. (2020) reviewed 60 studies of K-12 students and found that motivational factors like self-efficacy, task value, and goal orientation positively relate to reading outcomes, especially when they foster sustained effort and engagement. This effect was even stronger in elementary schools, suggesting motivation is especially important in early literacy. Another study defined persistence in reading as sustained cognitive effort and behavioral engagement in the face of challenges, assessed through self-reports in middle and high school students (Reschly and Christenson, 2022). Maegi et al. (2018) found that effortful control and task persistence, rated by teachers and parents, predicted reading fluency and comprehension in sixth graders. These findings suggest that persistence enables students to stay focused and overcome reading challenges.

This prior research highlights the important role of persistence in reading achievement and outlines diverse ways it has been conceptualized and measured. However, most existing studies focus on upper elementary or older students, leaving a gap in understanding how persistence develops in younger learners. This may stem partly from a reliance on self-report measures to assess persistence, whose validity is limited with younger children who are still developing the metacognitive skills needed for accurate self-reflection (Craig et al., 2020). In response to these challenges, researchers have recommended age-appropriate adaptations to self-reports and combining evidence from surveys and observed behaviors (Gascoine et al., 2017; Desoete, 2008). That said, few studies in the reading context explicitly examine how persistence is being measured using process data (e.g., behavior logs, transcript) compared to traditional methods, such as self-report or teacher rating. This limited use of multimodal approaches limits our understanding of the behavioral and cognitive dimensions of persistence, especially in early literacy development.

The current pilot study begins to address this gap by pairing age-appropriate self-reports with behavioral data in young learner's reading activities.

2.2 Pedagogical Agents

One approach to administer a survey or gain insights into a student's experience is through the use of virtual characters. Frequently called pedagogical agents when used in learning settings (Schroeder et al., 2025; Siegle et al., 2023), virtual characters have been widely used in K-12 settings to help students learn (Zhang et al., 2024a,b). However, pedagogical agents can play a wide variety of roles in a learning environment (Clarebout et al., 2002). Research has shown that it is very common to use pedagogical agents as an information source or for coaching and scaffolding, but is rare to use them to administer self report surveys (Schroeder and Gotch, 2015; Zhang et al., 2024a). While many critical questions must ultimately be addressed to validate agent-administered assessments, a foundational concern is whether students will engage meaningfully with the pedagogical agents in a testing setting. In this proof-of-concept study, we designed and piloted a pedagogical agent to administer a conversational survey to young learners. Given the complexity of surveying students at this age, our goal was to explore the feasibility and potential of integrating agent-mediated assessment within the learning process in an engaging manner.

2.3 Research Questions

We address the following research questions:

- To what extent can a pedagogical agent elicit a broad range of responses (i.e., responses are distributed across response options)?
- How do the behavioral data relate to the data collected via the pedagogical agent survey?

3 Methods

3.1 Participants

Participants were drawn from a second-grade classroom in a charter school in the Northeastern U.S.; the study was reviewed and approved by the Institutional Review Board and all participating students had documented parental consent. The students engaged in a reading activity with the App in six sessions for about 15 minutes each. In each session, the class read one story. The stories were grade-appropriate fictional narratives licensed from

Cricket Media, which publishes literary magazines for young readers. After the 3rd reading session during which the students read the story titled "Happy, The Hearing Ear Dog,"¹ the students interacted with Adam, the virtual agent that delivered the survey. The activities took place as part of normal school programming during the first two weeks of April 2025 and were led by the teacher, with technical assistance from a member of the research team during the first reading session and during the survey activity.

Of the 25 students in the class, 20 had parental consent for their reading data to be used for research. Of these, 18 completed the survey. The primary analyses reported below focus on the 18 students who participated in both components. Information about student gender was provided by the school. Our final sample consisted of 7 female and 11 male students. Additional demographic information came from an optional demographic questionnaire completed by the students' parent or guardian. The average age was 7.5 years ($SD=0.51$). Eleven students' parents reported English as their child's first language, while five indicated English was not their child's first language (2 did not respond). Ten students were identified by parents as Hispanic or Latino and 6 as Black or African American; 2 parents did not respond to this item.

3.2 Instruments

3.2.1 Reading Application

Relay Reader² is a reading and listening app developed to support readers as they transition into fluent reading (Madnani et al., 2019). Readers take turns reading stories out loud with a pre-recorded model human narrator (audiobook). The target length of reading and listening turns can be configured by the reader; for this study both were set up at 70 words as default. None of the students changed the default settings. The transition between narrator and student turns occurs on paragraph breaks. The allocation of a passage into narrator or student turn happens dynamically, where paragraphs are added to the turn as long as adding the paragraph made the passage closer to the target length (from above or from below) than not adding it. After every other student turn, before the next narrator turn starts, the student is asked two multiple choice comprehension questions. Questions were created

¹<https://www.audible.com/pd/Happy-the-Hearing-Ear-Dog-Audiobook/B0DJ9SDW68>

²<https://relayreader.org>

for approximately every 100 words of running text by researchers and research assistants experienced with developing such items, and reviewed by the senior members of the team as well as by the research institution's fairness review committee. The questions focus on the salient aspects of the plot, settings, and characters, and are surface-level, not requiring inferential reasoning. Thus, the two questions a reader would be presented with would refer to something that was mentioned within 200 words preceding the current bookmark.

As readers interact with the app, the app collects timestamped log data of the various activities (including the focal activities of reading, listening, answering questions, as well as other activities in the app such as looking at the reading history or changing the fonts or other settings). The audio recording of every student turn is processed using an in-house speech analysis system validated for this use case using data predominantly from students in grades 3-5 (Beigman Klebanov and Loukina, 2021; Loukina et al., 2019, 2017). The system produces estimates of reading accuracy and fluency for all scorable recordings. The app has been previously used with students in grades 2-8 in the US in school and summer camp contexts. Depending on the grade and the study goals, the app library – how many and which books each student has access to at any given time – is flexibly managed by the researchers. For this study, one story was put in the library for all participants for each of the six reading sessions. For sessions 1-5, all students got the same story; on day 6, we assigned a new story to students who read all five and assigned one of the stories they missed for students who missed some of the first five reading sessions. All 18 students were present during the 3rd reading session in which they engaged with the pedagogical agent and all read the same story about a dog. Students read on Kindle Fire HD 8 (12th generation) tablets provided by the researchers.

3.3 Reading Data

During the reading activity, the app collected data related to students' reading performance. We focused on four measures for analysis:

1. **RCQ:** Percent correct in the multiple choice reading comprehension questions embedded in the app (see previous section), both for the Dog story specifically and across all the six stories);

2. **SKT:** proportion of skipped turns, defined as the share of reading turns completed faster than the 90th percentile of oral reading fluency norm for the Spring of second grade³ (148 words per minute) adjusted for text-based variation in fluency estimates using the method in Beigman Klebanov et al. (2019), resulting in the cut rate of 178 words per minute;
3. **ACC:** reading accuracy, calculated as the proportion of words in the passage assigned to the reading in the current turn that were recognized as pronounced correctly by the in-house speech analysis engine (see previous section);
4. **WCPM:** words read correctly per minute, a fluency measure calculated based on the automated recognition and scaled to account for systematic error in the automated measurement (Beigman Klebanov and Loukina, 2021). Following prior research, we restricted analysis of accuracy and WCPM to recordings with accuracy ≥ 0.7 (Liceralde et al., 2022), as lower scores often reflect issues with audio recording rather than students' bona-fide performance.

3.3.1 Pedagogical Agent Design

The pedagogical agent's role was positioned as that of a proof-of-concept conversational assessment. Specifically, the pedagogical agent delivered a series of questions in a conversational style that the learner replied to by answering a multiple-choice question. Questions were presented as on-screen text and simultaneously narrated by the agent. The agent was designed to appear as a teenager, with narration provided by a teenage male speaking English. The system was built using Unity and deployed on the same Kindle Fire tablets that the students used for the reading app. Student responses to the agent were stored locally on the device.

3.3.2 Survey

To assess students' reading attitudes and experience with the reading app, the pedagogical agent administered a brief conversational survey after the 3rd reading session. The survey was designed for second-grade students and emphasized simple language, limited response options, and concrete behavioral prompts. With a total of 14 items, the survey included a mix of comprehension checks,

³<https://www.readingrockets.org/topics/fluency/articles/fluency-norms-chart-2017-update>

attitudinal measures, and behavioral self-reports (see Appendix A for more detail).

We assessed basic reading comprehension and memory of the story using four multiple-choice items based on the story they just read (Happy, The Hearing Ear Dog). These items differed from the in-app RCQ items and asked students what characters the story was about and about a key characteristic of one of the characters. Our reading self-reflection items measured students' perception of reading ease, interest, and learning. These items used simple, developmentally appropriate 3 or 4-point Likert scales (such as 'not really', 'sometimes', 'definitely').

We measured reading persistence, the focal construct of this study, through two items that asked students how they typically respond when confronting challenges when reading, specifically unfamiliar written words or unknown word meanings. Response options reflected specific behaviors associated with persistence or quitting. Students could respond that they used effortful strategies such as 'sounding it out' or 'figuring it out' to indicate persistence or avoidance behaviors such as *skipping* or *stopping* which may suggest lower persistence. We deliberately framed these as concrete, first-person behavior reports (e.g., 'What do you do if you don't know what a word means?') rather than abstract or hypothetical self-assessments typical of measures designed for older students (e.g., "I leave things unfinished"; Chernyshenko et al. 2018; Sparks and Lehman 2025).

To gauge the students' reaction to the reading and agent activities, we asked 3 subjective questions. This included two questions about their preferences related to the stories and a final item which asked the students if they wanted to interact with the agent again.

4 Results

This study reflects an early-stage pilot investigation, and the small sample size ($n = 18$) limits our use of formal statistical tests or validation procedures.

4.1 Survey

In Figure 1, we present a heat map showing student response patterns across the multiple reading and persistence constructs assessed in our survey. Responses are grouped by construct and scaled such that higher values (and darker colors) reflect greater expression of the underlying construct. Ba-

sic reading comprehension and story recall were high, with 72% of students answering all four questions correctly. This is consistent with all of the students reporting that reading aloud was 'OK' or 'easy', with no one saying that it was 'hard' (Item 9); however, 27.7% responded that there were 'lots' of words they did not know how to say (Item 11) or did not know the meaning of (Item 13) on our two follow-up questions about reading ease. While no students reported that they hated reading, 27.8% reported they didn't enjoy reading and 16.7% reported they did not feel they learned much from reading.

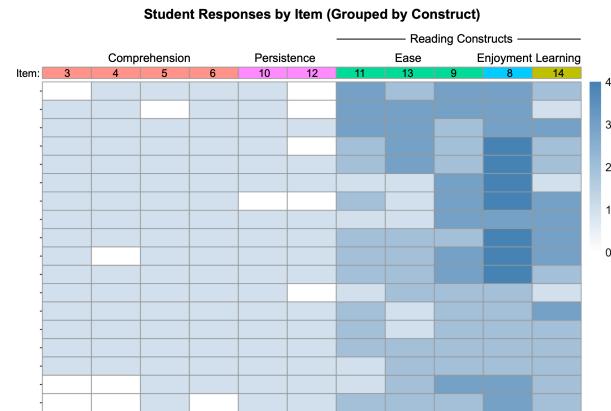


Figure 1: Heatmap of student responses to survey. Rows represent students; columns represent individual items grouped by construct. Darker colors indicate stronger expression of targeted construct. Comprehension and persistence items are coded as correct/incorrect or persistent/not persistent; reading self-reflection constructs use a Likert scale.

In response to the subjective questions about the activities, all students responded that they liked the story and when asked to recommend a story for the agent to read next, 4 students chose to not recommend any of the stories. At the end of the survey, the agent asked students if they wanted to interact again in the future. Only one student chose 'no', with 6 responding 'maybe' and 11 agreeing to a future interaction ('sure').

Student responses on our items measuring persistence showed limited variability. In response to 'What do you do if you don't know how to say a word?' 17 out of 18 students selected 'sounding it out', a behavior we would associate with persistence. Only one student selected 'I skip it and continue' while no students chose 'I ask someone' or 'I stop reading'. Similarly, the second item, 'What do you do if you don't know what a word means?', 13 students selected 'I figure it out' with the re-

maining 5 responses divided between asking for help or skipping the word. While these responses suggest high levels of self-reported persistence, the strong skew towards a single response option on each item limits our ability to differentiate students' behavior using these measures alone.

4.2 App Use

We first examined the pairwise correlations between students' RCQ scores and behavioral measures extracted from the logs of the app (Figure 2).⁴ RCQ accuracy was positively correlated with reading accuracy (ACC; $r(13) = .56$, $p=0.019$) and negatively correlated with the proportion of skipped reading turns (SKT; $r(16)=-.66$, $p=.003$), but was not significantly associated with fluency (WCPM; $r(13)=.2$, $p=.47$).

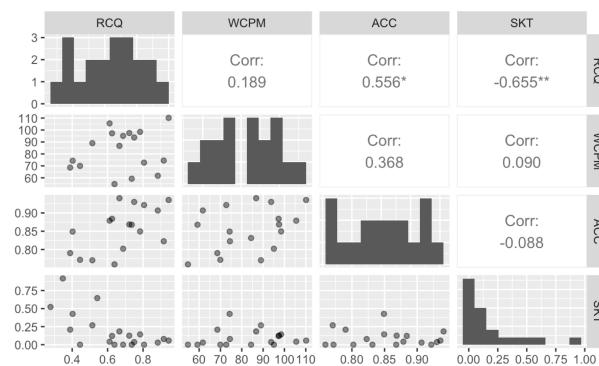


Figure 2: Pairwise correlations between RCQ scores and Relay Reader behavioral measures. Pearson r values shown above the diagonal; $p \leq .05$ (*), $p \leq .005$ (**).

4.3 Relating Survey Response to In-App Behaviors

We examined whether students' self-reported survey responses reflected the behavioral data collected about those same students during their reading activity. First, we tested whether comprehension scores on the survey related to performance on the RCQ questions presented during the reading activity for the same story. Using a Wilcoxon rank-sum test to compare students with perfect vs. non-perfect scores on the survey,⁵ we found no significant difference in app RCQ scores ($W = 23$, $p = .36$). Next, we assessed whether students' ratings of reading ease related to measured WCPM or

⁴Only 15 of our 18 students had audio recordings of adequate quality to estimate an average ACC and WCPM.

⁵We chose to collapse scores into two categories because the items assessed basic recall and understanding. Given the simplicity of the questions, incorrect responses likely reflected substantial comprehension difficulties or inattention.

ACC. We found no association with fluency ($R(16) = .15$, $p = .6$) and a marginally significant negative correlation with accuracy ($R(16) = -.45$, $p = .089$). Finally, we examined whether self-reported reading persistence was related to the proportion of skipped reading turns. A Wilcoxon rank-sum test revealed no significant difference in skipping behaviors between students who responded with 'sound it out' or 'figure it out' to both items versus those who did not ($W = 16$, $p = .11$).

5 Discussion

5.1 Relations Between Variables

For the reading data, we found that oral reading accuracy (ACC) and comprehension (RCQ) were positively correlated, while proportion of skipped reading turns and comprehension (RCQs) were negatively correlated. These relationships make sense: Reading what is on the page is necessary to answer questions based on the story, as the questions were designed so that one cannot consistently answer them correctly based on general knowledge, without actually reading the story. The negative correlation between the proportion of skipped turns and comprehension is consistent with findings with the app with older students (Beigman Klebanov et al., 2019).

While higher comprehension typically does correlate with higher fluency in the same readers in assessment data (Wise et al., 2010), it is in principle possible to read slowly but with good comprehension. This is likely in our case since students are not reading for a test and are not being urged to read fast. The scatterplot of RCQ and WCPM scores in Figure 2 shows some low fluency readers who nevertheless showed strong comprehension – there are three students with comprehension $\geq .80$ who read with fluency around 60-75 WCPM (72 WCPM is the 25th percentile in fluency, according to norms). Further analysis of these individuals revealed that while one of these readers had near perfect reading accuracy, the other two struggled, with an average reading accuracy between 72-76% and all three very rarely skipped their turn ($\leq 8\%$ turns skipped). All three reported using persistent reading strategies in the survey. Though a small sample, these students illustrate our hypothesis about persistent reading behavior; despite low fluency, they consistently complete their turns (low skip proportion), suggesting they may be encountering and overcoming difficulties reading.

In terms of the relationship between the reading data and the survey data, there are some discrepancies. In particular, 17 out of 18 students said they would sound out a word they didn't know and most admitted that there were some or lots of words they didn't know. However, the process data suggests that there were 3 students who did not spend enough time on most of their turns to read to any substantive degree ($SKT \geq 0.5$), let alone sound out difficult words. However, the interpretation may not be straightforward. First, students may have responded to the survey question as if it asked about a general habit rather than about reading in the app specifically; they may be sounding words out in other contexts but not in this reading context, which they may have perceived as more informal. Second, the teacher told us she explicitly instructs them to sound out unfamiliar words as they read, so their response to the survey may have reflected what they thought they should be doing rather than what they actually did.

5.2 Limitations and Future Research

The study was carried out in a single classroom with only 18 students who had parental consent and participated in all activities. The small sample size limited our ability to detect significant relationships between variables. Furthermore, adapting the surveys to this young population limited the number of questions and response options that we could include within a single survey administration.

Students generally reacted positively to the pedagogical agent-based activity, based on observations of the research team member who assisted the teacher, conversations with the teacher, and on the students' responses where they expressed readiness or tentative readiness to chat with the pedagogical agent again. We are thus encouraged to continue exploring the utility of the agent through co-design with teachers for expanding the agent's role beyond the self-report-based assessment function. In a preliminary focus group we conducted with three elementary school teachers to start addressing this issue, the clearest message was that the agent should try and encourage the students in their reading endeavor. In addition, we intend to implement a more flexible conversation with students, using automated speech recognition and an LLM that would help generate a larger variety of agent responses, with the caveat that strong guardrails would need to be implemented to ensure that the agent's conversation is appropriate. We envision

using an LLM to create a diverse set of responses that would be vetted and placed into a database for the agent to choose from rather than allowing students to interact with the LLM directly.

Weak correlations between self-reports and behavioral indicators for constructs such as persistence, metacognition, and self-regulation are well documented in the literature (e.g., [Craig et al., 2020](#)). These discrepancies can point to limitations in the survey design ([Desoete, 2008](#)), but they may also yield meaningful insights into students' developmental trajectory in their own self-awareness ([Andrade, 2019](#)). Reflecting on our own findings, a fruitful direction for future research would be to explore when and how we prompt young learners to self-report, such as asking why they skipped some of their reading turns when those behaviors occur. That would require better personalization of the survey so that the question is only posed to students who did skip their turns as a matter of course. A conversational agent that is connected to the process data will be able to deliver such personalization. Embedding surveying within the activity would also support capturing students' motivations and metacognitive awareness in context, providing insight into how their behaviors and beliefs relate to their reading experience as it unfolds. This type of survey administration would help support future efforts to better understand how student persistence varies over time and across learning contexts. Students' overall positive response and receptivity to the agent-delivered survey suggest this is a promising approach for integrating surveys into the activity without disrupting engagement. However, these findings stem from a small proof-of-concept study and should not be interpreted as validation of the approach itself.

6 Conclusion

We investigated the extent to which a pedagogical agent can function appropriately in an assessment or surveying role with young learners in a classroom environment. The results of our study showed that pedagogical agents hold promise for engaging students in conversational surveys following a tablet-based reading task. However, our study was limited to one classroom, and thus more research is needed to understand in what learning scenarios and for what learners pedagogical agents are appropriately positioned in a testing or surveying role.

Acknowledgments

This work is supported in part by the National Science Foundation (NSF) and the Institute of Education Sciences (IES) under Grant DRL-2229612. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author and do not necessarily reflect the views of the NSF or IES. We are grateful to Jesse Sparks and Catherine Castellano for their feedback on drafts of this paper.

References

- Heidi L Andrade. 2019. A critical review of research on student self-assessment. In *Frontiers in education*, volume 4, page 87. Frontiers Media SA.
- Beata Beigman Klebanov and Anastassia Loukina. 2021. **Exploiting structured error to improve automated scoring of oral reading fluency.** In *Proceedings of the International Conference on Artificial Intelligence in Education (AIED)*, pages 76–81.
- Beata Beigman Klebanov, Anastassia Loukina, Nitin Madnani, John Sabatini, and Jennifer Lentini. 2019. Would you? could you? on a tablet? analytics of children's ebook reading. In *Proceedings of the 9th International Conference on Learning Analytics & Knowledge*, pages 106–110, Tempe, AZ, USA. ACM.
- Oleksandr S Chernyshenko, Miloš Kankaraš, and Fritz Drasgow. 2018. Social and emotional skills for student success and well-being: Conceptual framework for the oecd study on social and emotional skills. *OECD Education Working Papers*, 173.
- Geraldine Clarebout, Jan Elen, Lewis Johnson, and Erin Shaw. 2002. Animated pedagogical agents: An opportunity to be grasped? *Journal of Educational multimedia and hypermedia*, 11(3):267–286.
- David A Cook, Becca L Gas, David R Farley, Matthew Lineberry, Nimesh D Naik, Francisco J Cardenas Lara, and Anthony R Artino Jr. 2019. Influencing mindsets and motivation in procedural skills learning: two randomized studies. *Journal of Surgical Education*, 76(3):652–663.
- Kym Craig, Daniel Hale, Catherine Grainger, and Mary E Stewart. 2020. Evaluating metacognitive self-reports: systematic reviews of the value of self-report in metacognitive research. *Metacognition and Learning*, 15:155–213.
- Annemie Desoete. 2008. Multi-method assessment of metacognitive skills in elementary school children: How you test is what you get. *Metacognition and Learning*, 3:189–206.
- Carol S Dweck, Gregory M Walton, and Geoffrey L Cohen. 2014. Academic tenacity: Mindsets and skills that promote long-term learning. *Bill & Melinda Gates Foundation*.
- Camille A Farrington, Melissa Roderick, Elaine Al-lensworth, Jenny Nagaoka, Tasha Seneca Keyes, David W Johnson, and Nicole O Beechum. 2012. *Teaching Adolescents to Become Learners: The Role of Noncognitive Factors in Shaping School Performance—A Critical Literature Review*. University of Chicago Consortium on Chicago School Research.
- Louise Gascoine, Steve Higgins, and Kate Wall. 2017. The assessment of metacognition in children aged 4–16 years: a systematic review. *Review of Education*, 5(1):3–57.
- Tiong-Thye Goh. 2025. Learning management system log analytics: the role of persistence and consistency of engagement behaviour on academic success. *Journal of Computers in Education*, pages 1–24.
- Manu Kapur. 2008. Productive failure. *Cognition and Instruction*, 26(3):379–424.
- Van Rynald T Liceralde, Anastassia Loukina, Beata Beigman Klebanov, and John R Lockwood. 2022. Beyond text complexity: Production-related sources of text-based variability in oral reading fluency. *Journal of Educational Psychology*, 114(1):16.
- Anastassia Loukina, Beata Beigman Klebanov, Patrick L Lange, Yao Qian, Binod Gyawali, Nitin Madnani, Abhinav Misra, Klaus Zechner, Zuowei Wang, and John Sabatini. 2019. **Automated estimation of oral reading fluency during summer camp e-book reading with My Turn To Read.** In *Proceedings of INTERSPEECH*, pages 21–25.
- Anastassia Loukina, Beata Beigman Klebanov, Patrick Lange, Binod Gyawali, and Yao Qian. 2017. **Developing speech processing technologies for shared book reading with a computer.** In *Proc. WOCCI 2017: 6th International Workshop on Child Computer Interaction*, pages 46–51.
- Nitin Madnani, Beata Beigman Klebanov, Anastassia Loukina, Binod Gyawali, Patrick L Lange, John Sabatini, and Michael Flor. 2019. **My Turn to Read: An interleaved e-book reading tool for developing and struggling readers.** In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pages 141–146.
- Katrin Maegi, Eve Kikas, and Piret Soodla. 2018. Effortful control, task persistence, and reading skills. *Journal of Applied Developmental Psychology*, 54:42–52.
- Amy L Reschly and Sandra L Christenson. 2022. *Handbook of research on student engagement*. Springer.
- J Elizabeth Richey, Juan Miguel L Andres-Bray, Michael Mogessie, Richard Scruggs, Juliana MAL Andres, Jon R Star, Ryan S Baker, and Bruce M McLaren. 2019. More confusion and frustration, better learning: The impact of erroneous examples. *Computers & Education*, 139:173–190.

- Noah L Schroeder, Robert O Davis, and Eunbyul Yang. 2025. Designing and learning with pedagogical agents: An umbrella review. *Journal of Educational Computing Research*, 62(8):2127–2156.
- Noah L Schroeder and Chad M Gotch. 2015. Persisting issues in pedagogical agent research. *Journal of Educational Computing Research*, 53(2):183–204.
- Valerie J Shute, Matthew Ventura, and Yoon Jeon Kim. 2013. Assessment and learning of qualitative physics in Newton’s playground. *The Journal of Educational Research*, 106(6):423–430.
- Robert F Siegle, Noah L Schroeder, H Chad Lane, and Scotty D Craig. 2023. Twenty-five years of learning with pedagogical agents: History, barriers, and opportunities. *TechTrends*, 67(5):851–864.
- Ellen A Skinner, Thomas A Kindermann, Justin W Volland, and Nicolette P Rickert. 2022. Complex social ecologies and the development of academic motivation. *Educational Psychology Review*, 34(4):2129–2165.
- J. R. Sparks and B. Lehman. 2025. Measuring persistence and academic resilience: Analytic review and operational definitions. In *Paper presented at the annual meeting of the American Educational Research Association*. Denver, CO.
- Jessica R Toste, Lisa Didion, Peng Peng, Marissa J Filderman, and Amanda M McClelland. 2020. A meta-analytic review of the relations between motivation and reading achievement for K–12 students. *Review of Educational Research*, 90(3):420–456.
- Ming-Te Wang, Jessica L Degol, Jamie Amemiya, Alyssa Parr, and Jiesi Guo. 2020. Classroom climate and children’s academic and psychological wellbeing: A systematic review and meta-analysis. *Developmental Review*, 57:100912.
- Justin C Wise, Rose A Sevcik, Robin D Morris, Maureen W Lovett, Maryanne Wolf, Melanie Kuhn, Beth Meisinger, and Paula Schwanenflugel. 2010. The relationship between different measures of oral reading fluency and reading comprehension in second-grade students who evidence different oral reading fluency difficulties. *Language, Speech, and Hearing Services in Schools*, 41(3):340–348.
- Shan Zhang, Chris Davis Jaldi, Noah L Schroeder, and Jessica R Gladstone. 2024a. Pedagogical agents in K–12 education: a scoping review. *Journal of Research on Technology in Education*, pages 1–28.
- Shan Zhang, Chris Davis Jaldi, Noah L Schroeder, Alexis A López, Jessica R Gladstone, and Steffi Heidig. 2024b. Pedagogical agent design for K–12 education: A systematic review. *Computers & Education*, page 105165.

A Appendix: Survey Interface and Script

To provide further detail on the pedagogical agent’s role in survey administration, we include a screenshot (Figure A1) of the agent interface as it appeared during the survey session with students. The image depicts the agent presenting a question with response options rendered as on-screen buttons, designed to scaffold independent interaction for early elementary learners.



Figure A1: Screenshot of the pedagogical agent delivering a survey question.

In addition to the interface example, Table A1 provides the full conversational script used by the agent during survey delivery. This includes all administered items as well as transitional dialogue interludes designed to maintain a conversational tone and scaffold student engagement. The script also includes conditional statements (e.g., varied follow-up prompts) based on students’ prior responses. This structure reflects our effort to position the survey as an interactive, child-appropriate experience rather than a traditional assessment.

See Table A1 (next page) for the complete set of agent interactions, response options, and associated constructs.

Table A1: Conversational Agent Survey Script

Q#	Agent Prompt	Student Response Options	Construct
<i>Hi, I am Adam. Let's chat about the reading you've done on the Kindle. I'm curious how that went.</i>			
1	Ready to begin?	Let's go!	-
2	Did you like the story you read today? Please tap on your answer for me.	Yes / No	Subjective Experience
<i>Let's see. Which story was that? Hmm... .</i>			
3	About a frog and a dog?	Yes / No	Comprehension
4	About a cat and a dog?	Yes / No	Comprehension
5	About a grandma and a dog?	Yes / No	Comprehension
6	What sort of grandma was she?	Very tired / Deaf / Upset / Singing	Comprehension
<i><TEACHER> told me you read stories about Willie the Donkey, Chippy's Birthday, and Grandma and the Dog.</i>			
7	Which of these should I read next?	Willie / Chippy / Grandma / None	Subjective Experience
<i>Ok, thank you! I was thinking about reading about Chippy's Birthday, because my friend's birthday is next week, you know? [If the student selected Willie or Grandma above: "But now I think I'll read the <X> story first."]</i>			
8	Do you like reading?	Very much / It's OK / Not really / I hate reading	Reading Enjoyment
<i><TEACHER> said you read out loud today.</i>			
9	How did reading out loud feel?	Easy / OK / Hard	Reading Ease
<i>I see. [If response to Q9 = Easy: "Easy-peasy!"; if Hard: "Tough going."]</i>			
10	What do you do if you don't know how to say a word?	Sound it out / Ask someone / Skip / Stop reading	Persistence
11	Were there lots of words in the story you didn't know how to say?	Yes, lots! / Only a few / None	Reading Ease
<i>Sounds like reading out loud was [insert: "pretty hard" / "not too hard" / "a breeze" based on Q11].</i>			
12	What do you do if you don't know what a word means?	Figure it out / Ask someone / Skip / Stop reading	Persistence
13	Were there lots of words in the story you didn't understand?	Yes, lots! / Only a few / None	Reading Ease
<i>[If Q13 = lots or few: "You'll figure it out!"; if Q13 = none: "You got it!"]</i>			
14	Do you feel like you learn a lot when you read?	Definitely / Sometimes / Not really	Learning
15	Thanks for chatting with me! It was fun, for me. Should we chat again sometime?	Sure / Maybe / No, thank you	Subjective Experience
<i>Bye-bye!</i>			