

Article

# iSTART StairStepper—Using Comprehension Strategy Training to Game the Test

Tracy Arner <sup>1</sup>, Kathryn S. McCarthy <sup>2</sup> and Danielle S. McNamara <sup>1,\*</sup>

<sup>1</sup> Department of Psychology, Arizona State University, Tempe, AZ 85281, USA; tarner@asu.edu

<sup>2</sup> Department of Learning Sciences, Georgia State University, Atlanta, GA 30302, USA; kmccarthy12@gsu.edu

\* Correspondence: dsmcnama@asu.edu

**Abstract:** Literacy skills are critical for future success, yet over 60% of high school seniors lack proficient reading skills according to standardized tests. The focus on high stakes, standardized test performance may lead educators to “teach-to-the-test” rather than supporting transferable comprehension strategies that students need. StairStepper can fill this gap by blending necessary test prep and reading comprehension strategy practice in a fun, game-based environment. StairStepper is an adaptive literacy skill training game within Interactive Strategy Training for Active Reading and Thinking (iSTART) intelligent tutoring system. StairStepper is unique in that it models text passages and multiple-choice questions of high-stakes assessments, iteratively supporting skill acquisition through self-explanation prompts and scaffolded, adaptive feedback based on performance and self-explanations. This paper describes an experimental study employing a delayed-treatment control design to evaluate users’ perceptions of the StairStepper game and its influence on reading comprehension scores. Results indicate that participants enjoyed the visual aspects of the game environment, wanted to perform well, and considered the game feedback helpful. Reading comprehension scores of students in the treatment condition did not increase. However, the comprehension scores of the control group decreased. Collectively, these results indicate that the StairStepper game may fill the intended gap in instruction by providing enjoyable practice of essential reading comprehension skills and test preparation, potentially increasing students’ practice persistence while decreasing teacher workload.

**Keywords:** reading comprehension; strategy training; game-based learning; intelligent tutoring system; feedback



**Citation:** Arner, T.; McCarthy, K.S.; McNamara, D.S. iSTART StairStepper—Using Comprehension Strategy Training to Game the Test. *Computers* **2021**, *10*, 48. <https://doi.org/10.3390/computers10040048>

Academic Editors: Carlos Vaz de Carvalho and Antonio Coelho

Received: 16 March 2021

Accepted: 6 April 2021

Published: 9 April 2021

**Publisher’s Note:** MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Copyright:** © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

Literacy refers to “the ability to understand, evaluate, use, and engage with written texts to participate in society, to achieve one’s goals, and to develop one’s knowledge and potential” [1] (p. 61). Literacy skills are not only critical for educational and career success, but the ability to read and comprehend various text types across multiple subjects is necessary to function in everyday life. However, national reading assessment data suggests that many students struggle with reading comprehension. The most recent National Assessment of Educational Progress [2] on reading skills found that 63% of twelfth-grade students were below proficient in reading. Similarly, 66% of eighth-graders and 65% of fourth-graders were also below proficiency. These numbers suggest that additional instructional support is needed to improve students’ reading achievement as they progress through grade levels.

The emphasis placed on standardized testing has increased in the last few decades to ensure that all students (i.e., non-White, lower-income) are receiving equal, high-quality instruction and to monitor adequate yearly progress (AYP) [3,4]. These goals are admirable but have fallen short of the intended targets as many students have failed to learn essential skills and strategies necessary to become part of the global workforce while preparing for

standardized assessments [4,5]. Amidst increasing pressure and limited instructional hours, teachers may resort to “teaching to the test” in an effort to demonstrate learning gains on these standardized assessments [3,5,6]. Unfortunately, this practice leads to inaccurate inferences about the knowledge and skills that students have acquired and unreliable inflation in scores on state-level standardized assessments that are not achieved on the NAEP [5,7], nor do these types of tests reflect the types of literacy tasks that the student will encounter outside of the testing room [8,9]. The result is that students fail to develop the comprehension strategies that will better serve them outside of a standardized test [10].

There are multiple barriers to building students’ reading comprehension strategies. Developing comprehension strategies requires ample opportunity for cycles of deliberate practice and targeted feedback. One issue is that providing feedback on ill-structured tasks like reading comprehension is both time- and resource-intensive for instructors and students alike. Thus, students have few opportunities to practice with one-on-one support. A second issue is that students may become disengaged from the strategy-building activities before they have mastered the skills [11].

With these issues in mind, we developed a game-based module, StairStepper, to implement an intelligent tutoring system, iSTART. StairStepper was designed to support and reward the use of reading comprehension strategies in the context of a mock standardized reading comprehension test. Our aim was to leverage the power of automated evaluation and game-based principles to offer a scalable, efficient, and fun way for students to develop their reading comprehension skills that can transfer to high-stakes testing environments.

To contextualize StairStepper, we first provide a brief background on the theoretical and educational motivation for reading comprehension strategies as well as an overview of the broader intelligent tutoring system, iSTART, in which StairStepper was built.

### *1.1. Reading Comprehension Strategies*

Theories of discourse comprehension suggest that as learners read, they construct mental models or mental representations of the text information [12]. The mental representation that readers construct as part of comprehension is comprised of multiple levels or layers, which include the (1) surface code, (2) textbase and, (3) situation, model. The surface code includes specific words and syntax but is unlikely to be retained except in cases of rote memorization. This immediate textual information gives rise to the textbase or gist meaning of the text. In the situation model, the reader goes beyond the present text to make inferences and to elaborate from prior knowledge [12–14]. Readers must develop a coherent mental model, beyond the surface level, for text comprehension and knowledge transfer [12,15].

There is ample evidence to suggest that prompting and training students’ comprehension strategies improves reading comprehension [16–20]. One such strategy shown to benefit comprehension is self-explanation [21–24]. Generating an explanation to oneself aids in the integration of new information with prior knowledge; these connections support the construction of a more elaborated and durable mental representation of the content [22]. Despite the benefit of using reading comprehension strategies, students may not adopt and use these strategies on their own [10,17,23]. However, there is ample evidence of the benefit of strategy training and practice, specifically self-explanation training [17,23–26], particularly with low-knowledge or struggling students [18,27].

Self-explanation reading training (SERT) improves students’ reading comprehension through instruction on five active reading strategies (comprehension monitoring, paraphrasing, predicting, bridging, and elaborating) that lead to generating high-quality self-explanations, which in turn, improves text comprehension [17,18]. The comprehension monitoring strategy encourages the reader to continuously evaluate whether or not they understand what they just read [22]. Comprehension monitoring is an inherent feature of generating self-explanations because if the student is not able to successfully explain what they just read, it is an indication that there is a breakdown in understanding. Thus, comprehension monitoring provides an indicator that the reader needs to employ strategies

to repair the gap in knowledge [28]. Skilled readers are more likely to engage in comprehension monitoring, notice inconsistencies in the text, gaps in understanding, and use strategies to repair gaps when they do not understand [28,29]. Paraphrasing is a frequently used strategy in which the reader restates the content of the text in their own words [17]. Although this strategy focuses mainly on developing a textbase, putting the text into one's own words is an important step toward more meaningful processing. A prediction is when the reader speculates about what they think might happen next in the text [17]. While predictions are relatively infrequent during reading, they support comprehension by encouraging the reader to consider more global aspects of the text [30,31]. The last two strategies are similar in that they are both generations of inferences, bridging and elaborative. Bridging inferences are those that connect a statement to a prior sentence or passage in the text. In contrast, an elaborative inference occurs when the reader connects the current text to prior knowledge [32,33]. Generating inferences is an essential component of reading comprehension.

Helping students to use active reading strategies is effective for elementary [34,35], middle [23,26], and high [17,18,36] school students as well as young adults (i.e., college students) [17,37]. However, simply prompting students to use these strategies or providing direct instruction about the strategies is only part of the process of improving students' reading skills. That is, in addition to instruction, students also need ample time to engage in deliberate practice where they are able to use the strategies while receiving feedback on how to improve [38,39]. Although strategy instruction and practice have pronounced benefits for literacy skills [40–42], it is sometimes difficult to keep students engaged and motivated so that they keep practicing. One potential method to encourage the training and practice of these beneficial strategies is through the use of automated intelligent tutoring systems (ITSs) that provide a mechanism for more engaging, game-based practice [39,43,44]. Given the ample evidence of the benefit of strategy training and teachers' limited time to teach strategies, intelligent tutoring systems may be useful in filling that gap as they can provide adaptive feedback inside an engaging, game-based activity that may increase students' motivation to engage in deliberate practice of reading comprehension strategies.

### 1.2. Intelligent Tutoring Systems

Computers have been used to support learning for the last few decades [45,46], initially in the form of computer-assisted instruction and, more recently, as intelligent tutoring systems (ITSs) [47]. Meta-analyses suggest that computer-assisted instruction (CAI) and ITSs have positive impacts on learning [45,47,48]. ITSs differ slightly from other CAI systems in that they attempt to emulate the one-on-one tutoring experience through adaptive instruction and more granular feedback [46,47,49]. For example, students may receive stepwise feedback (i.e., correct/incorrect, solution hint) during problem-solving (i.e., error detection), and they may also be able to engage in natural dialog with the system emulating a human tutor [37,47]. Comparisons of learning outcomes between human, CAI, and ITS systems suggest that, while the more sophisticated ITSs may be more beneficial to learning than some CAI systems, they are still not quite as effective as human, one-on-one tutoring, which is considered the "gold standard" of instruction [45,47,50]. One feature that may make an ITS system more similar to one-on-one tutoring while also providing actionable feedback and increasing student motivation is the addition of a pedagogical agent [51,52].

Pedagogical agents are characters in technology-based instructional applications designed to facilitate learning [52,53]. The interactions that the agent has with the learner may serve to provide instruction, feedback, or motivation [37,54–57]. The addition of a pedagogical agent may facilitate or increase interaction between the learner and the intelligent tutoring system [58–60]. Pedagogical agents may be a "talking head" that provides information via text or audio comments, or they may be full-body characters who have animated gestures that can be used for additional learning supports, such as signaling [58,59,61,62]. Anthropomorphizing an intelligent tutoring system with a pedagogical agent that has a

human-like figure, voice, or both (i.e., *persona effect*) [63] may further increase students' motivation to engage with the intelligent tutoring systems [64–67]. This *persona effect* can lead students to view the engagement with the pedagogical agent as a social interaction similar to what would occur with a human tutor [68]. Thus, students have more positive perceptions of the learning environment and are more accepting of instructions or feedback from the pedagogical agent, which may aid in learning or motivation to persist [59,63,65].

Feedback, broadly defined, is information provided about one's performance. It may also include the difference between one's performance and the learning objective or goal [69,70]. The influence of feedback on student learning outcomes has myriad evidence evaluating its efficacy across task type, subject area and grade levels [71–75]. The general consensus is that feedback has a positive effect on student learning outcomes through the benefit may be moderated by learners' prior knowledge, context, timing, and type of feedback [71,76–79]. Feedback provided by a pedagogical agent in an ITS might be goal-driven (i.e., response correctness), instructional (i.e., hint or strategy suggestion), or affective (i.e., positive reinforcement to continue), which may motivate the learner to continue with the task or practice in the ITS [80,81]. For example, learners with low prior-knowledge experience a greater benefit from explanatory feedback (e.g., "That answer is incorrect because...") than basic corrective feedback (e.g., right or wrong) [77,82].

Students' motivation to engage in a task or persist through struggle is positively related to their achievement [83–86]. The more motivated a student is to engage with a learning task, the more likely they are to complete the task, thereby achieving the learning goal [87]. Motivation to persist in the practice necessary to improve reading comprehension skills may be bolstered by the affordances of ITSs [48], particularly those with anthropomorphized feedback mechanisms (e.g., pedagogical agents) [68] and game-based learning and assessment [39,40,88].

### 1.3. iSTART

Interactive strategy training for active reading and thinking (iSTART) is an intelligent tutoring system (ITS) based on SERT. iSTART provides self-explanation followed by game-based practice. The iSTART system first provides overview lessons on each of the self-explanation strategies (i.e., paraphrasing, bridging and elaborative inferences, prediction and comprehension monitoring) using video instruction and modeling [89]. During the generative practice, students are given passages to read and then asked to self-explain target sentences. Students' responses are evaluated using natural language processing algorithms that detect evidence of the different comprehension strategies. This algorithm is used to provide a summative score (0–3) as well as formative feedback indicating ways to improve their self-explanation. For example, when responses are too short or too long, Mr. Evans, a pedagogical agent, provides various types of feedback that can help the student to write higher quality self-explanations [26,89,90].

Although the original system demonstrated positive impacts on learner's self-explanation and reading comprehension, it was difficult to keep students motivated in repeated rounds of guided practice. Thus, iSTART-motivationally enhanced (ME) [39,91] introduced additional motivational features via game-based practice. iSTART includes both generative and identification games. In *generative* games, students practice writing self-explanations. For example, students can play "Self-Explanation Showdown", in which they play against a CPU in a head-to-head competition. In *identification* games, students view example self-explanations and need to correctly identify the strategy. Reaching new high scores or levels earns trophies as well as additional "iBucks", the system currency units, which can be used to open and play more games or customize their player avatar. These game-based features support learning in that they may encourage students to engage in prolonged practice, which is critical for developing reading comprehension skills [39,92].

#### 1.4. StairStepper

Building upon iSTART's tradition of game-based literacy practice, StairStepper was designed to provide engaging, a game-based practice that closely approximates question types that students experience in standardized assessments. More specifically, StairStepper gamifies the use of scaffolding to challenge the student to read increasingly difficult texts. Thus, the goal of StairStepper was two-fold; to (1) provide students generative practice of self-explanation strategies that will benefit their reading comprehension skills while simultaneously (2) preparing them for the standardized assessment texts and questions that they will see throughout their educational careers.

##### 1.4.1. Reading Comprehension Strategies for Standardized Testing

Traditional standardized reading assessments are designed to isolate and evaluate reading comprehension skills. For example, The National Assessment of Educational Progress [2] reading assessment is used ubiquitously in K-12 education. The assessment for grade four consists of two texts that students read and then respond to approximately 20 questions that are either selected response (i.e., multiple-choice) or constructed response (i.e., open-ended text entry). The questions are written to assess three types of *cognitive targets* or the kinds of thinking that underlie reading comprehension: locate and recall, integrate and interpret, and critique and evaluate.

The "locate and recall" cognitive target requires students to recall content from the text to answer the question. While students do have the option to refer back to the text, data show that students frequently do not do so [2]. Each of the reading comprehension strategies that students learn about and practice in iSTART can support performance on these types of tests. In StairStepper, like the assessment, students must decide if they *definitely know* the answer to the question, *definitely do not know*, or *might know* the answer. Practicing the comprehension monitoring strategy in the StairStepper game can help students be better prepared to make a clear decision about what they do and do not know when responding to questions. The second target, "integrate and interpret," requires students to make complex inferences within and across texts to derive meaning, explain a character's motivation or action, or uncover the theme of the text. The bridging and elaboration strategies that are practiced in StairStepper are the same strategies that are used to make these complex inferences when responding to the "integrate and interpret" target questions. The third question type, "critique and evaluate", requires students to think critically about text and evaluate aspects of it using a variety of perspectives based on their knowledge of the world. The paraphrasing and elaboration strategies encourage students to think about the text in ways as if they were going to explain it to another while also using their knowledge of the world to make sense of the content. Despite the benefit of reading comprehension strategies, they are often put to the side as teachers focus on preparation for standardized tests when, in fact, these strategies can and should be leveraged in standardized testing environments.

##### 1.4.2. Text Set and Questions

The first step in designing the StairStepper game was to develop and evaluate a corpus of texts and corresponding questions that would emulate these standardized assessments. The texts and their accompanying questions were retrieved from publicly available educational resources. The text topics span multiple domains, including knowledge gained in school (i.e., science and social science) and knowledge gained in daily life (i.e., sports and pop culture). Texts range from seven to 80 sentences in length. Rather than relying on shallow measures of readability, the texts were leveled through comparative judgments made by independent raters (for description, see [93]). The initial set of 172 texts was separated into 12 levels of increasing difficulty. These rater judgments of difficulty were correlated with both Flesch–Kincaid grade level ( $r = 0.79$ ) and Dale–Chall readability ( $r = 0.77$ ) [94]. After inspection and piloting, the final text set was reduced to 162 leveled texts chosen to

mimic those students may see in reading comprehension assessments taken in classrooms every year.

The full corpus of multiple-choice questions from these texts was piloted to check for floor and ceiling effects. Some items were removed or slightly edited for clarity. The remaining items were then categorized by question type based on the source of the knowledge required to answer the questions correctly. Questions categorized as *textbase* ( $N = 677$ ) can be answered from information found in a single sentence in the text. *Bridging inference* questions ( $N = 160$ ) require the reader to combine information from two or more sentences in the text. Finally, *elaboration* questions ( $N = 144$ ) require the reader to use the information found in the text and prior knowledge to answer correctly. Lower level texts (below level 8) have a higher percentage of textbase questions (>70%), whereas the higher-level texts (level 8 and above) had fewer textbase questions (35–45%) and more bridging inference (45–55%) and elaboration (8–10%) questions.

The texts and question types used in StairStepper are representative of various standardized assessments that students are likely to encounter in their educational careers. Therefore, these questions may be useful in helping students prepare for standardized assessments. Furthermore, StairStepper also provides students a more engaging way to practice reading comprehension strategies that are supported by evidence from numerous studies [17,18,90]. The underlying benefits of practicing these strategies inside the StairStepper game are that students may be more motivated to persist in practice, and they may be more likely to transfer the strategy used to the standardized assessments taken in the future.

#### 1.4.3. Game Play

The goal of StairStepper is to ascend to the top stair by answering comprehension questions about increasingly difficult texts. Students begin the game with instructions on their task and a reminder of the strategies (comprehension monitoring, paraphrasing, prediction, bridging, elaboration) that they can use when writing self-explanations.

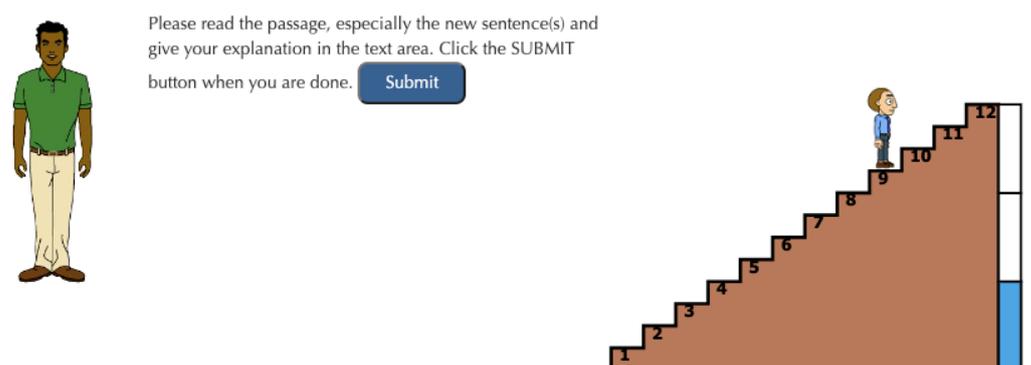
The game begins with the student's avatar on step five of twelve, where they are presented with a text of low-moderate difficulty (iSTART's default setting begins at level 5, but this is an adjustable feature). In the first text, they are not prompted to self-explain. At the end of the passage, they are asked to answer a series of multiple-choice questions about the text. Students who meet the correct response threshold (75%) on the multiple-choice questions are promoted to the next step and begin a new, slightly more difficult passage. In contrast, students who answer less than 75% of the question incorrectly receive another text at the same level. In this second text, the student is prompted to self-explain at various target sentences. In the first phase of scaffolding, students receive a score of the quality of their self-explanation on a color-coded, four-point scale ranging from Poor to Great (See Figure 1).

As depicted above, the StairStepper game includes iSTART's Mr. Evans, who serves as a guide through the game-based practice. He provides three types of information to students during gameplay; task instructions, feedback, and progress messages. The task instruction messages let students know what they need to do or what will happen next. For example, when the student begins StairStepper, Mr. Evans tells them that they will read the text and answer the questions (see Figure 2). Statements like this are provided any time there is a change in procedure, such as when a student moves between scaffolded levels.

Second, Mr. Evans provides feedback to the students. One type of beneficial feedback is motivational (i.e., praise) [47,49,52], such as telling the student, "Great, you got that one right." when they answer a multiple-choice question correctly (Figure 3). Mr. Evans will also provide metacognitive prompts that require students to think about and identify what self-explanation strategy they used (bridging, elaboration, paraphrasing) [17].



**Figure 1.** Players receive feedback on their self-explanation quality from the system.



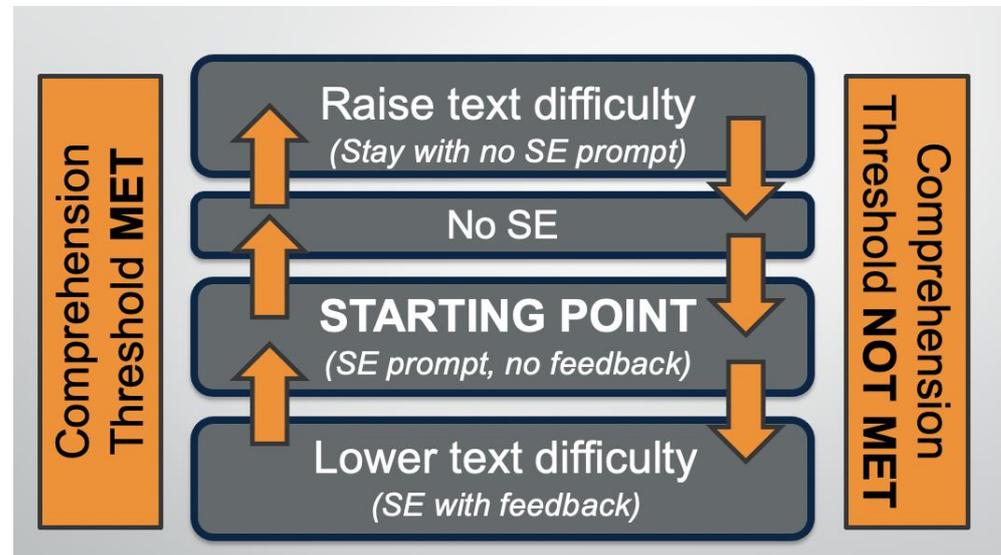
**Figure 2.** Mr. Evans gives the player instruction on what will happen next in the game.



**Figure 3.** Mr. Evans provides feedback on comprehension practice. Player avatar moving up to the next level.

After submitting their self-explanations for the entire text, students are given between 5 and 20 multiple-choice questions, depending on the text. If the student again receives a score below the threshold (75%), the next text includes prompting for self-explanation and feedback on the quality of the self-explanations with an opportunity to revise. Thus, students can receive three support levels (no SE, SE, SE + feedback; see Figure 4). If the student continues to struggle, the text difficulty is decreased, and as their comprehension improves, the subsequent texts become more challenging. Students' progress through the

game follows this same cycle of assessing comprehension at each level of text difficulty, and when the minimum is not met, students are provided scaffolded strategy training and feedback to aid in text comprehension.



**Figure 4.** Scaffolded support process in the StairStepper game.

### 1.5. Present Study

The iSTART research team continues to refine and evolve the types of game-based activities available in the ITS. The purpose of the present study was to investigate the effects of the new game-based adaptive literacy module, StairStepper. More specifically, we examined the potential benefit of the scaffolded support design of the StairStepper game on students' perceptions and motivations, as well as the effects of short-term practice with StairStepper on reading comprehension skills. We sought to answer three research questions with the present study.

1. How do students respond to the StairStepper game-based practice?
2. How will participants progress through the StairStepper game based on text adaptivity and scaffolded feedback?
3. How do iSTART training and StairStepper practice influence participant performance on a comprehension test and standardized assessment?

College students ( $n = 51$ ) completed the iSTART lesson videos and a round of Coached Practice. They then engaged in 90 min of StairStepper practice. Students were asked to complete a questionnaire about their experiences to measure their enjoyment and interest in the game-based practice and their self-reported sense of learning. To explore the efficacy of StairStepper, half of the participants ( $n = 25$ ) were assigned to a 3-day treatment condition that received a pretest, iSTART/StairStepper, and then a post-test. The other half ( $n = 26$ ) were assigned to a delayed treatment control in which they completed a pretest, a post-test, and then the iSTART/StairStepper training. We hypothesized that students in the StairStepper treatment condition would show pretest to post-test improvement on the proximal outcome of standardized Gates–MacGinitie reading test (GMRT) score and the more distal comprehension scores.

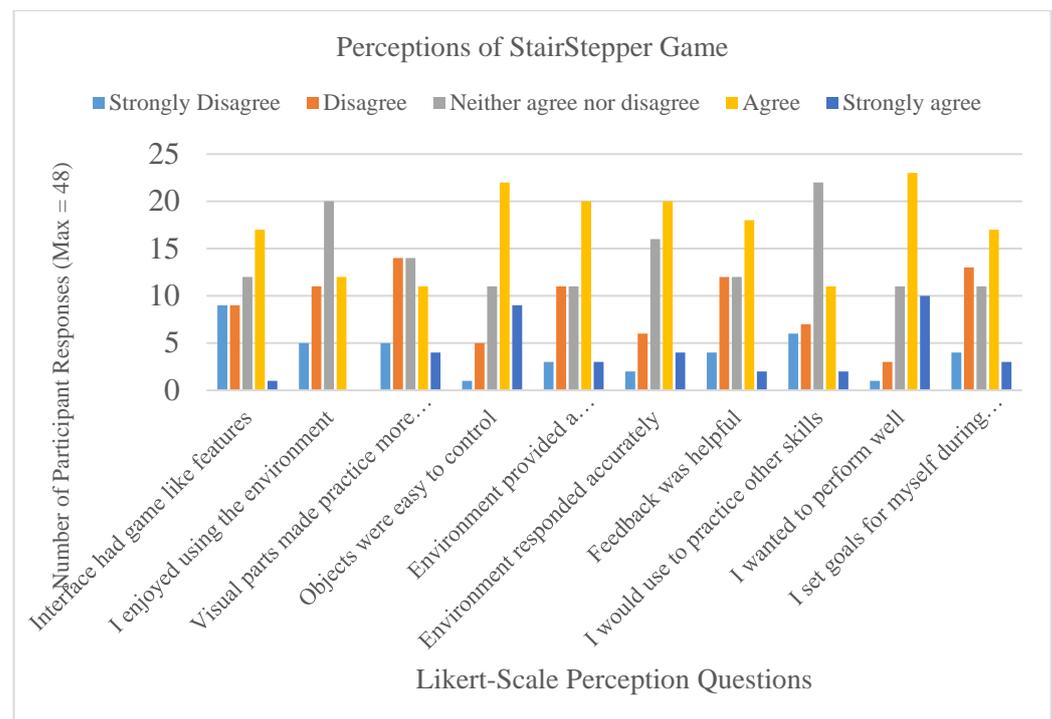
## 2. Results

### 2.1. Perceptions of the StairStepper Game

Our first question regarded students' enjoyment of StairStepper. Our purpose for building StairStepper was to include a fun and motivating test prep module in a way that aligned with the purpose of iSTART (reading comprehension strategy training). Thus,

it was important to investigate the extent to which participants enjoyed the game-based features of StairStepper. To this end, we asked students to answer survey questions regarding their experiences and perceptions of StairStepper. These analyses include 48 students, including those in the delayed treatment control, who played StairStepper after their post-test assessment. Three students did not complete the perceptions portion of the study.

As shown in Figure 5, participants rated their experience with the game interface (e.g., objects in the game) and game features (e.g., visual appearance) as well as personal attributes as they related to learning in game environments (e.g., goal setting). Overall, participants had positive attitudes about the StairStepper game as a method to practice reading comprehension strategies (see Figure 5).



**Figure 5.** Participant responses on 5-point Likert scale questions on perceptions of the StairStepper game. Three participants did not complete the perceptions survey.

We conducted Wilcoxon signed-rank tests to evaluate whether or not participant responses were significantly different from neutral. Results revealed that three items were indeed significantly positive: “Objects were easy to control” ( $p < 0.000$ ); “Environment responded accurately” ( $p = 0.02$ ); and “I wanted to perform well” ( $p < 0.000$ ). The other perceptions items were not significantly different from neutral, suggesting that the students did not have negative opinions of StairStepper.

We further analyzed participants’ perceptions of the StairStepper game as a function of reading skill using a median split on the GMRT pretest scores to determine the influence of reading skill on participants’ perceptions of the StairStepper game (see Table 1). Results indicated a significant difference on participants’ agreement with “Enjoyed the practice environment” ( $t(1, 45) = 3.45, p = 0.001$ ), “Interface had game-like features” ( $t(1, 45) = 2.16, p = 0.04$ ), and “I would use for other skills” ( $t(1, 45) = 1.98, p = 0.005$ ). These results suggest that participants who had lower reading comprehension skills found the StairStepper game more enjoyable than those who were more proficient in reading. These results may stem from proficient participants not believing that they were benefiting from the StairStepper practice module.

**Table 1.** Participant perceptions of the StairStepper game in iSTART.

Question	Reading Skill Level	N	Mean (SD)	t	p
Enjoyed the practice environment	Low	21	3.29 (0.72)	3.45	0.001 **
	High	26	2.42 (0.95)		
Feedback was helpful	Low	21	3.33 (0.91)	1.84	0.073
	High	26	2.77 (1.14)		
Interface had game-like features	Low	21	3.24 (0.94)	2.16	0.0360 *
	High	26	2.54 (1.27)		
Provided a purpose for actions	Low	21	3.33 (1.02)	0.938	0.353
	High	26	3.04 (1.11)		
I set goals during practice	Low	21	3.19 (1.12)	0.809	0.423
	High	26	2.92 (1.13)		
Visual appearance made the practice more enjoyable	Low	21	3.19 (1.03)	1.62	0.111
	High	26	2.65 (1.20)		
Objects were easy to control	Low	21	3.67 (0.97)	−0.22	0.826
	High	26	3.73 (1.00)		
I wanted to perform well	Low	21	3.71 (0.85)	−0.479	0.634
	High	26	3.85 (1.00)		
I would use for other skills	Low	21	3.24 (0.83)	1.98	0.054 *
	High	26	2.65 (1.13)		
Environment responded accurately	Low	21	3.43 (0.98)	0.287	0.775
	High	26	3.35 (0.98)		

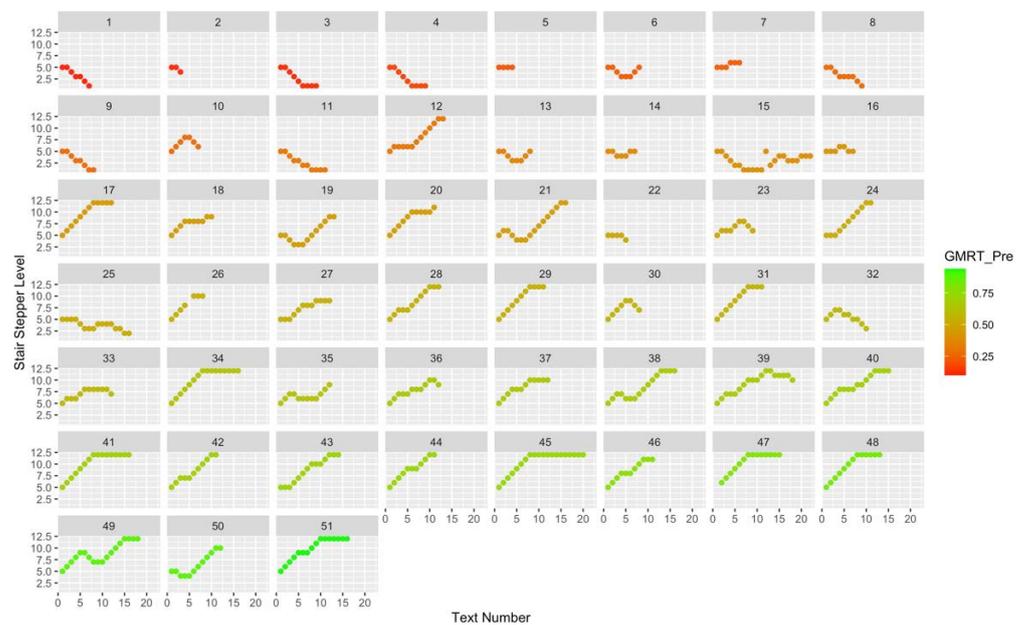
Note: \*  $p < 0.05$ , \*\*  $p < 0.01$ .

## 2.2. System Data

Our second question regarded students' traverse through the system in terms of whether they descended or ascended the "stairs" or text difficulty. To this end, we conducted a visual inspection of the log data. Figure 6 shows each participants' trajectory through StairStepper with text number along the  $x$ -axis and text difficulty along the  $y$ -axis. The participants are ordered based on their pretest GMRT score and color-coded accordingly. This visual inspection demonstrated two important findings. First, the game architecture was responsive to participants' reading skills. Participants with lower GMRT scores were given less difficult texts; more skilled readers ascended to the most difficult texts more quickly. Second, these graphs also demonstrate that many of the participants showed some decreases and increases in text difficulty, suggesting that the different amounts of scaffolding (self-explanation, feedback, text leveling) were effectively providing just-in-time support.

## 2.3. Reading Comprehension

Our third question regarded the impact of StairStepper on reading comprehension skills. Reading skills are generally impervious to relatively brief treatments, as in this study. For example, observed increases in self-explanation and comprehension skills generally have required at least 4 to 8 h of instruction and practice [17,18,39]. Yet, given that students in the StairStepper treatment condition received explicit instruction and practice on self-explanation and comprehension strategies, one of our objectives was to examine the extent to which this brief game-based practice impacted their ability to comprehend challenging science texts as well as their performance on the GMRT and comprehension of a science text. The GMRT texts are similar to the practice texts in StairStepper, whereas the science text included textbase and open-ended inference questions. Descriptive data and correlations between the measures are presented in Table 2.



**Figure 6.** Participants’ log data shows their progression through the StairStepper texts (number along the x-axis) as a function of text difficulty at each level (y-axis). These data revealed two system issues: (1) the “jump” in participant 15’s data reveals a system crash, and (2) several of the more skilled readers should have “won” the game after completing two levels 12 texts (e.g., 45). However, system settings prevented the game from ending. These issues were reported to the programmer and addressed.

**Table 2.** Descriptive statistics for comprehension measures.

Test	Mean (SD)	1	2	3	4	5	6	7	8
1. GMRT Pretest	0.50 (0.21)	1.00							
2. GMRT Posttest	0.50 (0.24)	0.89 **	1.00						
3. Pretest Mean	0.35 (0.21)	0.56 **	0.62 **	1.00					
4. Posttest Mean	0.36 (0.21)	0.65 **	0.67 **	0.63 **	1.00				
5. Pretest Textbase	0.28 (0.23)	0.44 **	0.48 **	0.85 **	0.52 **	1.00			
6. Posttest Textbase	0.41 (0.24)	0.57 **	0.60 **	0.57 **	0.87 **	0.49 **	1.00		
7. Pretest Inference	1.66 (1.01)	0.53 **	0.59 **	0.88 **	0.57 **	0.50 **	0.50 **	1.00	
8. Posttest Inference	0.32 (0.24)	0.57 **	0.57 **	0.53 **	0.87 **	0.42 **	0.51 **	0.49 **	1.00

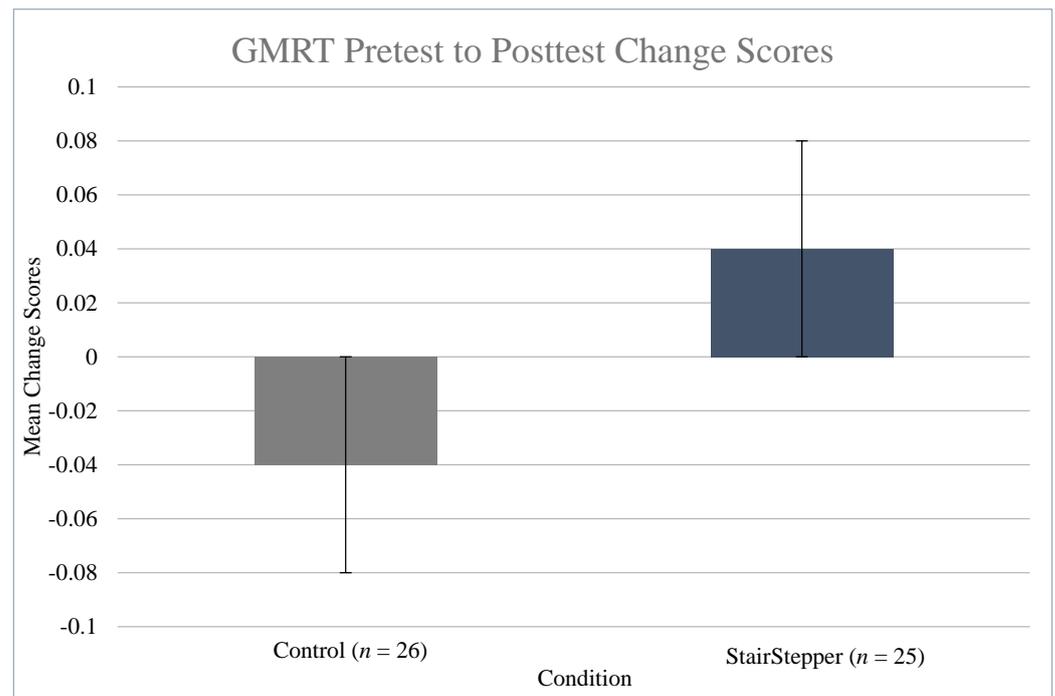
Note: pretest text is *Red Blood Cells*, and post-test text is *Cell Repair* to assess comprehension of challenging science texts. \*\*  $p < 0.01$ .

We conducted preliminary analyses to examine whether there was a significant difference in reading comprehension skills between groups. Results of an independent samples t-test conducted on participants’ GMRT pretest scores indicated that there was a significant difference ( $t(1, 53) = -2.59, p = 0.012$ ) in pretest means between the delayed treatment control ( $M = 0.44, SD = 0.22$ ) and the StairStepper training condition ( $M = 0.58, SD = 0.19$ ). Similarly, results of independent samples t-test conducted on students’ science comprehension (i.e., *Red Blood Cells*) pretest mean scores indicated a significant difference ( $t(1, 53) = -2.13, p = 0.038$ ) in mean scores between the delayed-treatment control ( $M = 0.30, SD = 0.19$ ) and the StairStepper training condition ( $M = 0.42, SD = 0.22$ ). As such, pretest scores were included as covariates in the analyses to control for prior reading skills.

### 2.3.1. GMRT

We examined the extent to which 90 min of StairStepper practice impacted performance on a standardized reading comprehension measure. To account for group differences, we conducted a t-test on pretest to post-test change scores to evaluate the impact of

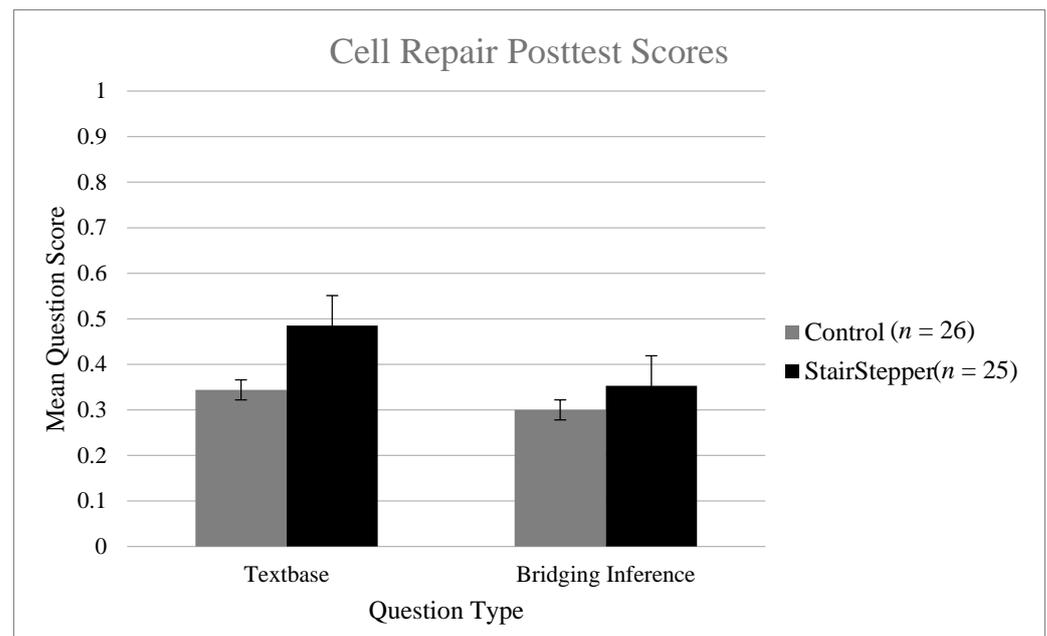
the StairStepper practice game on students' reading comprehension. Results indicated that there was a significant difference in pretest to posttest change ( $t(1,49) = -2.72, p = 0.009$ ; Figure 7) between the StairStepper training group ( $M = 0.04, SD = 0.12$ ) and the delayed treatment control ( $M = -0.04, SD = 0.09$ ). While we did not have hopes of observing substantial gains from such a short training session and single, 90 min practice session on a standardized test, such as GMRT, these results suggest that the StairStepper practice game has strong promise in helping students to improve their reading skills and performance on similar tests.



**Figure 7.** Gates-MacGinitie reading test (GMRT) pretest to post-test change as a function of condition. Error bars indicate standard error.

### 2.3.2. Science Comprehension

A 2 (question type: textbase, bridging inference) by 2 (condition: control, StairStepper training) analysis of covariance (ANCOVA) was conducted to examine the effect of StairStepper training on the two different question types. Item type was included as the within-subjects factor, condition as a between-subject factor, and performance on the pretest comprehension test was included as a covariate. There was no significant effect of question type,  $F(1, 48) = 0.580, p = 0.45$ , nor was there any effect of StairStepper training  $F(1, 48) = 1.28, p = 0.26$ . There was also no interaction effect,  $F(1, 48) = 1.21, p = 0.27$  (Figure 8).



**Figure 8.** Comprehension question scores as a function of question type and condition.

### 3. Discussion

The present study investigated a new game designed to provide students with an engaging environment to practice reading comprehension strategies while simultaneously preparing for standardized reading comprehension assessments. StairStepper, housed in the iSTART intelligent tutoring system, uses adaptive text and scaffolded feedback support to guide students through self-explaining and answering questions about increasingly challenging texts. The goal of this study was to evaluate participants' perceptions of the new game and to examine the possible benefits of StairStepper practice as measured by lab-designed comprehension measures and standardized (GMRT) performance.

Results indicated that participants had positive attitudes about the StairStepper game as a way to practice the reading comprehension strategies. Specifically, participants considered the objects in the environment to be easy to control and that the game provided an accurate reflection of their performance. In addition, of note is that a significant number of participants reported "wanting to do well" while engaging with the system. Indeed, these results align with prior work on the use of the game-based practice to support students' motivation and engagement [39]. Interestingly, the participants, who had lower reading comprehension skills, had more positive attitudes about the game environment and the game features. Furthermore, they indicated that they would use this game to practice different types of skills. One explanation for these results is that the benefit of gameplay may have been more salient to those participants who had lower reading comprehension skills. These results align with prior research indicating that the students who have lower reading comprehension scores garner a greater benefit from self-explanation reading training [18,36] and strategy training in iSTART [95]. While these studies investigated self-explanation reading training and reading comprehension training in iSTART for longer durations, the participants' perceptions of StairStepper in the present study are promising.

We evaluated the influence of iSTART training and StairStepper practice on participants' scores on a standardized reading assessment (i.e., Gates–MacGinitie reading test, GMRT). Preliminary results indicated that there was a significant difference between groups at the pretest. To account for differences between groups, we analyzed change scores from pretest to post-test on the GMRT and found that the participants in the StairStepper game condition maintained their reading comprehension score, while those in the delayed treatment control experienced a significant decrease in comprehension score from pretest to post-test. These results suggest that the StairStepper practice game benefited participant

maintenance and use of the reading comprehension strategies. This aligns with prior research suggesting that reading comprehension strategies need to be practiced in order for students to consistently adopt and use them [40,41].

This study also investigated the influence of the short iSTART training session followed by 90 min of practice on the StairStepper game on participants' reading comprehension scores. The results indicate no significant effects of training on open-ended comprehension scores. These results may reflect the need for larger sample sizes to detect effects as a function of pre-training differences. That is, pretest scores on comprehension and GMRT were strongly predictive of post-test scores. Given that less-skilled readers found the game more valuable, it may be that these students would benefit more from StairStepper and from extended practice. Indeed, these results suggest that students may need more training and practice than occurred in this study (i.e., one session of training and one session of practice). Evidence from prior studies indicates that consistent adoption of strategy use requires extended, deliberate practice [95]. Therefore, additional work is needed to investigate the number of practice sessions that may result in participants' efficient use of different types of reading strategies and the extent to which this supports performance on textbase and bridging question performance.

Taken together, these results suggest that the students who received self-explanation strategy training and StairStepper game-based practice did benefit in that their reading comprehension scores remained stable. Conversely, participants in the delayed-treatment control group experienced a significant decrease over the course of the three-day study. While we did not expect to see an increase in reading comprehension skills after a short training and practice session, these results indicate that there is a benefit to students' motivation to perform well on the test. Further work is needed to evaluate the practice dosage (i.e., number of sessions) and duration (i.e., length of sessions) that may lead to long-term improvement in reading comprehension skills. Additionally, larger studies will also allow us to more rigorously investigate how StairStepper training varies across different individual differences, such as reading skills.

The positive attitudes that participants reported about the StairStepper game and the maintenance of reading comprehension scores on a standardized assessment are promising. The goal of this work was to develop a game-based module in the iSTART intelligent tutoring system that would be engaging for students to practice using self-explanation strategies while also preparing them for the standardized assessments that they will experience throughout their educational career. Additional work is needed to investigate the dosage (i.e., how many practice sessions) and the durability (i.e., how long will strategy adoption last) that is most beneficial for this type of game-based practice. In sum, the StairStepper game-based practice module may serve an important role in students' acquisition of and long-term adoption of self-explanation strategies that contribute to reading comprehension and literacy skills.

#### *Limitations and Future Directions*

While the results of this study are promising, we acknowledge some limitations that should be considered in future work. First, the StairStepper game was designed as an engaging way for high school students to practice self-explanation strategy use while preparing for standardized assessments that are common throughout K12 education. However, our sample comprised undergraduate students who were earning course credit as part of the participant pool. As demonstrated by some ceiling effects in our data, several of our participants were skilled readers. These students are less likely to substantially benefit from this practice game in this context. However, there were also a number of undergraduates in our sample who did not immediately reach the highest level in the game. Thus, we will continue to explore the student characteristics and contexts under which StairStepper practice could be most beneficial. To this end, future work will broaden the scope of the participant pool to include a diverse sample of secondary students to evaluate the efficacy of the intervention with the target population.

Second, the current study relied on a small sample completing only 90 min of practice. We are developing additional studies in which larger, more diverse samples of students complete extended training and practice. Such studies will allow us to better detect and articulate the effects of self-explanation training and deliberate strategy practice using the StairStepper game in iSTART.

## 4. Materials and Methods

### 4.1. Participants

The participants in this study were 55 undergraduate students from a large university in the southwest. A demographic questionnaire indicated the sample was predominantly male (female = 38.2%, male = 61.8%,  $M_{age} = 19.83$  years) and the sample was 1.8% African American, 36.4% Asian, 40% Caucasian, 16.4% Hispanic and 7% identified as other. English was not the first language for 38.2% of the participants. The final analyses included 51 participants as 4 were unable to complete the study in the allotted time.

### 4.2. Learning Measures

Participants' reading comprehension was measured using the Gates–MacGinitie reading comprehension test (GMRT, grades 10–12) [96] at pre- and post-test. Forms S and T were counterbalanced across participants such that those who were given form S at pretest were given form T at post-test or the reverse, in the training and delayed treatment control.

All participants also completed pretest and post-test comprehension assessments. The pretest text was titled *Red Blood Cells*, and the post-test text was titled *Cell Repair*. The pretests and post-tests include textbase questions (i.e., those that can be answered directly from the text) and bridging inference questions (i.e., those that require students to make a bridging inference between two sentences in the text).

### 4.3. Perceptions Measures

Participants completed a survey following their interaction with iSTART and the StairStepper practice game. Participants rated their experience with iSTART and StairStepper, including their enjoyment of the game and its features. Participants were also asked to rate their performance using the system. Items were rated on a 5-point Likert scale from (1) strongly disagree to (5) strongly agree.

### 4.4. Procedure

Participants self-selected into study A (delayed-treatment control) or study B (training condition) through the SONA research participant sign-up system. Scheduling of study (A) and study (B) was counterbalanced across weeks to prevent selection bias. Participants in Study A and Study B completed the same tasks for this experiment. Participants in the training condition experienced the intervention between the pretests and post-tests (see Table 3). However, participants in the delayed-treatment control experienced the intervention, iSTART self-explanation training and playing the StairStepper game during Session 3. This design allowed us to compare conditions while not depriving the control group of instruction and practice using StairStepper.

**Table 3.** Delayed treatment and training group session activities.

Study	Session 1	Session 2	Session 3
A (delayed-treatment control)	Demographics questionnaire, GMRT pretest, self-explanation and comprehension test for <i>Red Blood Cells</i>	GMRT post-test, self-explanation and comprehension test for <i>Cell Repair</i> iSTART self-explanation instruction (60 min)	90 min playing StairStepper
B (training)	Demographics questionnaire, GMRT pretest, self-explanation and comprehension test for <i>Red Blood Cells</i> iSTART self-explanation instruction (60 min)	90 min playing StairStepper	GMRT post-test, self-explanation and comprehension test for <i>Cell Repair</i>

## 5. Conclusions

Results from standardized reading assessments suggest that many students struggle to develop proficiency in literacy skills that are critical to educational and career success. Unfortunately, these test results, in conjunction with limited time and resources, often lead instructors to focus on preparing students for the high-stakes assessments rather than helping them to develop more generalizable reading comprehension skills [3,5]. The StairStepper game in iSTART was designed to address these potentially competing objectives by offering an automated, game-based practice environment that supports students' learning of reading comprehension strategies while also preparing them for high-stakes assessments. This study sought to answer three research questions; (1) How do students respond to the StairStepper game-based practice?; (2) How will participants progress through the StairStepper game, based on text adaptivity and scaffolded feedback?; (3) How to do iSTART training and StairStepper practice influence participants' performance on a comprehension test and standardized assessment?.

This study suggests that students enjoyed the game interface, attempted to perform well, and found the gameplay to be motivating. Specifically, they believed that it had game-like features and indicated that they would use this type of game to practice other skills. Regarding research question two, system data analysis results suggested that students progressed through text difficulty levels successfully and benefited from the scaffolding and feedback process. Finally, the results suggested that the scores of students who played the StairStepper game remained stable, whereas the students who did not receive training and play the StairStepper game demonstrated a decrease in their reading comprehension scores. Collectively, these results suggest that iSTART training and StairStepper game-based practice were beneficial for students reading comprehension strategy use. This initial student suggests promise for implementing StairStepper into the classroom and into test prep and as a positive step toward helping students to excel in high-stakes testing and beyond.

**Author Contributions:** Conceptualization, K.S.M. and D.S.M.; methodology, K.S.M. and D.S.M.; formal analysis, T.A. and K.S.M.; investigation, K.S.M.; resources, D.S.M.; data curation, K.S.M.; writing—original draft preparation, T.A.; writing—review and editing, T.A., K.S.M., and D.S.M.; visualization, T.A.; supervision, D.S.M.; funding acquisition, D.S.M. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was supported by The Office of Naval Research through Grant number N00014-20-1-2623. The opinions expressed are those of the authors and do not represent views of the Office of Naval Research.

**Institutional Review Board Statement:** The study was conducted according to the guidelines of the American Psychological Association and approved by the Institutional Review Board of Arizona State University (00011488, 10 February 2020).

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Data may be accessed by emailing the first author at tarner@asu.edu.

**Conflicts of Interest:** The authors declare no conflict of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript, or in the decision to publish the results.

## References

1. Organization for Economic Cooperation and Development (OECD). *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*; OECD Publishing: Paris, France, 2013. [CrossRef]
2. Nations Report Card. NAEP Reading 2019 HIGHLIGHTS. NAEP Report Card: Reading. Available online: <https://www.nationsreportcard.gov/highlights/reading/2019/g12/> (accessed on 20 January 2021).
3. Roach, R. Teaching to the test. *Divers. Issues High. Educ.* **2014**, *31*, 32.
4. Volante, L. Teaching to the test: What every educator and policymaker should know. *Can. J. Educ. Adm. Policy* **2004**, *35*, 1–6.
5. Popham, W.J. Teaching to the Test? *Educ. Leadersh.* **2001**, *58*, 16–21.
6. Pressley, M. Effective Beginning Reading Instruction. *J. Lit. Res.* **2002**, *34*, 165–188. [CrossRef]

7. Jennings, J.L.; Bearak, J.M. "Teaching to the test" in the NCLB era: How test predictability affects our understanding of student performance. *Educ. Res.* **2014**, *43*, 381–389. [[CrossRef](#)]
8. Magliano, J.P.; McCrudden, M.T.; Rouet, J.F.; Sabatini, J. The modern reader: Should changes to how we read affect research and theory? In *Handbook of Discourse Processes*, 2nd ed.; Schober, M.F., Rapp, D.N., Britt, M.A., Eds.; Routledge: New York, NY, USA, 2018; pp. 343–361.
9. Sabatini, J.; O'Reilly, T.; Weeks, J.; Wang, Z. Engineering a twenty-first century reading comprehension assessment system utilizing scenario-based assessment techniques. *Int. J. Test.* **2020**, *20*, 1–23. [[CrossRef](#)]
10. O'Reilly, T.; McNamara, D.S. The Impact of Science Knowledge, Reading Skill, and Reading Strategy Knowledge on More Traditional "High-Stakes" Measures of High School Students' Science Achievement. *Am. Educ. Res. J.* **2007**, *44*, 161–196. [[CrossRef](#)]
11. Healy, A.F.; Schneider, V.I.; Bourne, L.E., Jr. Empirically valid principles of training. In *Training Cognition*; Healy, A.F., Bourne, L.E., Jr., Eds.; Optimizing Efficiency, Durability, and Generalizability Psychology Press: Hove, East Sussex, UK, 2012; pp. 13–39.
12. Kintsch, W. The role of knowledge in discourse comprehension: A construction-integration model. *Psychol. Rev.* **1988**, *95*, 163–182. [[CrossRef](#)] [[PubMed](#)]
13. Butcher, K.R.; Davies, S. Inference generation during online study and multimedia learning. *Inferences Dur. Read.* **2015**, 321–347. [[CrossRef](#)]
14. McNamara, D.S.; Magliano, J. Toward a comprehensive model of comprehension. *Psychol. Learn. Motiv.* **2009**, *51*, 297–384.
15. McNamara, D.S.; Kintsch, W. Learning from texts: Effects of prior knowledge and text coherence. *Discourse Process.* **1996**, *22*, 247–288. [[CrossRef](#)]
16. Guthrie, J.T.; Wigfield, A.; Barbosa, P.; Perencevich, K.C.; Taboada, A.; Davis, M.H.; Tonks, S. Increasing reading comprehension and engagement through concept-oriented reading instruction. *J. Educ. Psychol.* **2004**, *96*, 403. [[CrossRef](#)]
17. McNamara, D.S. SERT: Self-Explanation Reading Training. *Discourse Process.* **2004**, *38*, 1–30. [[CrossRef](#)]
18. McNamara, D.S. Self-Explanation and Reading Strategy Training (SERT) Improves Low-Knowledge Students' Science Course Performance. *Discourse Process.* **2017**, *54*, 479–492. [[CrossRef](#)]
19. Meyer, B.J.; Ray, M.N. Structure strategy interventions: Increasing reading comprehension of expository text. *Int. Electron. J. Elem. Educ.* **2011**, *4*, 127–152.
20. Palinscar, A.S.; Brown, A.L. Reciprocal teaching of comprehension-fostering and comprehension-monitoring activities. *Cogn. Instr.* **1984**, *1*, 117–175. [[CrossRef](#)]
21. Bisra, K.; Liu, Q.; Nesbit, J.C.; Salimi, F.; Winne, P.H. Inducing Self-Explanation: A Meta-Analysis. *Educ. Psychol. Rev.* **2018**, *30*, 703–725. [[CrossRef](#)]
22. Chi, M.T.H.; Bassok, M.; Lewis, M.W.; Reimann, P.; Glaser, R. Self-explanations: How students study and use examples in learning to solve problems. *Cogn. Sci.* **1989**, *13*, 145–182. [[CrossRef](#)]
23. Chi, M.T.; De Leeuw, N.; Chiu, M.-H.; Lavancher, C. Eliciting Self-Explanations Improves Understanding. *Cogn. Sci.* **1994**, *18*, 439–477. [[CrossRef](#)]
24. Hausmann, R.G.; VanLehn, K. Explaining self-explaining: A contrast between content and generation. *Front. Art Int.* **2007**, *158*, 417.
25. Hausmann, R.G.; VanLehn, K. The effect of self-explaining on robust learning. *Int. J. Artif. Intell. Educ.* **2010**, *20*, 303–332.
26. McNamara, D.S.; O'Reilly, T.P.; Best, R.M.; Ozuru, Y. Improving Adolescent Students' Reading Comprehension with Istart. *J. Educ. Comput. Res.* **2006**, *34*, 147–171. [[CrossRef](#)]
27. O'Reilly, T.; Taylor, R.S.; McNamara, D.S. Classroom based reading strategy training: Self-explanation vs. reading control. In *Proceedings of the 28th Annual Conference of the Cognitive Science Society*; Sun, R., Miyake, N., Eds.; Erlbaum: Mahwah, NJ, USA, 2006; pp. 1887–1892.
28. Chi, M.T.; Bassok, M. Learning from Examples via Self-Explanations. *Knowing Learn. Instr. Essays Honor Robert Glas.* **1988**. [[CrossRef](#)]
29. Wagoner, S.A. Comprehension Monitoring: What It Is and What We Know about It. *Read. Res. Q.* **1983**, *18*, 328. [[CrossRef](#)]
30. Magliano, J.P.; Baggett, W.B.; Johnson, B.K.; Graesser, A.C. The time course of generating causal antecedent and causal consequence inferences. *Discourse Process.* **1993**, *16*, 35–53. [[CrossRef](#)]
31. Magliano, J.P.; Dijkstra, K.; Zwaan, R.A. Generating predictive inferences while viewing a movie. *Discourse Process.* **1996**, *22*, 199–224. [[CrossRef](#)]
32. McNamara, D.S. Reading both high-coherence and low-coherence texts: Effects of text sequence and prior knowledge. *Can. J. Exp. Psychol. / Rev. Can. De Psychol. Expérimentale* **2001**, *55*, 51–62. [[CrossRef](#)]
33. McNamara, D.S.; Kintsch, E.; Songer, N.B.; Kintsch, W. Are Good Texts Always Better? Interactions of Text Coherence, Background Knowledge, and Levels of Understanding in Learning from Text. *Cogn. Instr.* **1996**, *14*, 1–43. [[CrossRef](#)]
34. Best, R.M.; Floyd, R.G.; McNamara, D.S. Differential competencies contributing to children's comprehension of narrative and expository texts. *Read. Psychol.* **2008**, *29*, 137–164. [[CrossRef](#)]
35. Best, R.M.; Rowe, M.; Ozuru, Y.; McNamara, D.S. Deep-Level Comprehension of Science Texts. *Top. Lang. Disord.* **2005**, *25*, 65–83. [[CrossRef](#)]

36. O'Reilly, T.; Best, R.; McNamara, D.S. Self-explanation reading training: Effects for low-knowledge readers. In *Proceedings of the 26th Annual Cognitive Science Society*; Forbus, K., Gentner, D., Regier, T., Eds.; Erlbaum: Mahwah, NJ, USA, 2004; pp. 1053–1058.
37. Fang, Y.; Lippert, A.; Cai, Z.; Hu, X.; Graesser, A.C. A Conversation-Based Intelligent Tutoring System Benefits Adult Readers with Low Literacy Skills. In *Proceedings of the Lecture Notes in Computer Science*; Metzler, J.B., Ed.; Springer: Cham, Switzerland, 2019; pp. 604–614.
38. Healy, A.F.; Clawson, D.M.; McNamara, D.S.; Marmie, W.R.; Schneider, V.I.; Rickard, T.C.; Bourne, L.E., Jr. The long-term retention of knowledge and skills. *Psychol. Learn. Motiv.* **1993**, *30*, 135–164.
39. Jackson, G.T.; McNamara, D.S. Motivation and performance in a game-based intelligent tutoring system. *J. Educ. Psychol.* **2013**, *105*, 1036–1049. [[CrossRef](#)]
40. Jackson, G.T.; McNamara, D.S. *The Motivation and Mastery Cycle Framework: Predicting Long-Term Benefits of Educational Games In Game-Based Learning: Theory, Strategies and Performance Outcomes*; Baek, Y., Ed.; Nova Science Publishers: New York, NY, USA, 2017; pp. 759–769.
41. Kellogg, R.T.; Whiteford, A.P. Training Advanced Writing Skills: The Case for Deliberate Practice. *Educ. Psychol.* **2009**, *44*, 250–266. [[CrossRef](#)]
42. Wijekumar, K.; Graham, S.; Harris, K.R.; Lei, P.-W.; Barkel, A.; Aitken, A.; Ray, A.; Houston, J. The roles of writing knowledge, motivation, strategic behaviors, and skills in predicting elementary students' persuasive writing from source material. *Read. Writ.* **2018**, *32*, 1431–1457. [[CrossRef](#)]
43. Huizenga, J.; Admiraal, W.; Akkerman, S.; Dam, G.T. Mobile game-based learning in secondary education: Engagement, motivation and learning in a mobile city game. *J. Comput. Assist. Learn.* **2009**, *25*, 332–344. [[CrossRef](#)]
44. Rieber, L.P. Seriously considering play: Designing interactive learning environments based on the blending of micro worlds, simulations, and games. *Educ. Technol. Res. Dev.* **1996**, *44*, 43–58. [[CrossRef](#)]
45. Kulik, J.A.; Fletcher, J.D. Effectiveness of intelligent tutoring systems: A meta-analytic review. *Rev. Educ. Res.* **2016**, *86*, 42–78. [[CrossRef](#)]
46. VanLehn, K. The behavior of tutoring systems. *Int. J. Artif. Intell. Educ.* **2006**, *16*, 227–265.
47. VanLehn, K. The Relative Effectiveness of Human Tutoring, Intelligent Tutoring Systems, and Other Tutoring Systems. *Educ. Psychol.* **2011**, *46*, 197–221. [[CrossRef](#)]
48. Ma, W.; Adesope, O.O.; Nesbit, J.C.; Liu, Q. Intelligent tutoring systems and learning outcomes: A meta-analysis. *J. Educ. Psychol.* **2014**, *106*, 901–918. [[CrossRef](#)]
49. Mitrovic, A.; Ohlsson, S.; Barrow, D.K. The effect of positive feedback in a constraint-based intelligent tutoring system. *Comput. Educ.* **2013**, *60*, 264–272. [[CrossRef](#)]
50. Aleven, V.; Roll, I.; McLaren, B.M.; Koedinger, K.R. Help helps, but only so much: Research on help seeking with intelligent tutoring systems. *Int. J. Artif. Intell. Educ.* **2016**, *26*, 205–223. [[CrossRef](#)]
51. Moreno, R.; Mayer, R.; Lester, J. Life-like pedagogical agents in constructivist multimedia environments: Cognitive consequences of their interaction. In *EdMedia+ Innovate Learning*; Association for the Advancement of Computing in Education (AACE): Morgantown, WV, USA, 2000; pp. 776–781.
52. Veletsianos, G.; Russell, G.S. Pedagogical Agents. In *Handbook of Research on Educational Communications and Technology*; Spector, J., Merrill, M., Elen, J., Bishop, M., Eds.; Springer: New York, NY, USA, 2014; pp. 759–769.
53. Craig, S.D.; Gholson, B.; Driscoll, D.M. Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features and redundancy. *J. Educ. Psychol.* **2002**, *94*, 428–434. [[CrossRef](#)]
54. Chen, Z.-H.; Chen, S.Y. When educational agents meet surrogate competition: Impacts of competitive educational agents on students' motivation and performance. *Comput. Educ.* **2014**, *75*, 274–281. [[CrossRef](#)]
55. Craig, S.D.; Driscoll, D.M.; Gholson, B. Constructing knowledge from dialog in an intelligent tutoring system: Interactive learning, vicarious learning, and pedagogical agents. *J. Educ. Multimed. Hypermedia* **2004**, *13*, 163–183.
56. Duffy, M.C.; Azevedo, R. Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. *Comput. Hum. Behav.* **2015**, *52*, 338–348. [[CrossRef](#)]
57. Moreno, R.; Mayer, R.E.; Spires, H.A.; Lester, J.C. The Case for Social Agency in Computer-Based Teaching: Do Students Learn More Deeply When They Interact with Animated Pedagogical Agents? *Cogn. Instr.* **2001**, *19*, 177–213. [[CrossRef](#)]
58. Johnson, A.M.; Ozogul, G.; Reisslein, M. Supporting multimedia learning with visual signalling and animated pedagogical agent: Moderating effects of prior knowledge. *J. Comput. Assist. Learn.* **2015**, *31*, 97–115. [[CrossRef](#)]
59. Moreno, R.; Mayer, R.E. Role of Guidance, Reflection, and Interactivity in an Agent-Based Multimedia Game. *J. Educ. Psychol.* **2005**, *97*, 117–128. [[CrossRef](#)]
60. Schroeder, N.L.; Adesope, O.O.; Gilbert, R.B. How effective are pedagogical agents for learning? A meta-analytic review. *J. Educ. Comp. Res.* **2013**, *49*, 1–39. [[CrossRef](#)]
61. Li, W.; Wang, F.; Mayer, R.E.; Liu, H. Getting the point: Which kinds of gestures by pedagogical agents improve multi-media learning? *J. Educ. Psychol.* **2019**, *111*, 1382–1395. [[CrossRef](#)]
62. Mayer, R.E.; Dapra, C.S. An embodiment effect in computer-based learning with animated pedagogical agents. *J. Exp. Psychol. Appl.* **2012**, *18*, 239–252. [[CrossRef](#)] [[PubMed](#)]

63. Lester, J.C.; Converse, S.A.; Kahler, S.E.; Barlow, S.T.; Stone, B.A.; Bhogal, R.S. The persona effect: Affective impact of animated pedagogical agents. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*; Edwards, A., Pemberton, S., Eds.; Association for Computing Machinery: New York, NY, USA, 1997; pp. 359–366.
64. Atkinson, R.K.; Mayer, R.E.; Merrill, M.M. Fostering social agency in multimedia learning: Examining the impact of an animated agent's voice. *Contemp. Educ. Psychol.* **2005**, *30*, 117–139. [[CrossRef](#)]
65. Baylor, A.L. The design of motivational agents and avatars. *Educ. Technol. Res. Dev.* **2011**, *59*, 291–300. [[CrossRef](#)]
66. Gulz, A.; Haake, M. Design of animated pedagogical agents—A look at their look. *Int. J. Hum. Comput. Stud.* **2006**, *64*, 322–339. [[CrossRef](#)]
67. Kim, M.; Ryu, J. Meta-Analysis of the Effectiveness of Pedagogical Agent. In *Proceedings of ED-MEDIA 2003–World Conference on Educational Multimedia, Hypermedia & Telecommunications*; Lassner, D., McNaught, C., Eds.; Association for the Advancement of Computing in Education (AACE): Honolulu, HI, USA, 2003; pp. 479–486.
68. Schroeder, N.L.; Romine, W.L.; Craig, S.D. Measuring pedagogical agent persona and the influence of agent persona on learning. *Comput. Educ.* **2017**, *109*, 176–186. [[CrossRef](#)]
69. Hattie, J.; Timperley, H. The power of feedback. *Rev. Educ. Res.* **2007**, *77*, 81–112. [[CrossRef](#)]
70. Sadler, D.R. Formative assessment and the design of instructional systems. *Instr. Sci.* **1989**, *18*, 119–144. [[CrossRef](#)]
71. Hattie, J.A.C. *Visible Learning*; Routledge, Taylor and Francis Group: London, UK; New York, NY, USA, 2009.
72. Havnes, A.; Smith, K.; Dysthe, O.; Ludvigsen, K. Formative assessment and feedback: Making learning visible. *Stud. Educ. Eval.* **2012**, *38*, 21–27. [[CrossRef](#)]
73. Kluger, A.N.; DeNisi, A. The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychol. Bull.* **1996**, *119*, 254. [[CrossRef](#)]
74. Kulik, J.A.; Kulik, C.-L.C. Timing of Feedback and Verbal Learning. *Rev. Educ. Res.* **1988**, *58*, 79–97. [[CrossRef](#)]
75. Mory, E.H. Feedback research revisited. In *Handbook of Research on Educational Communications and Technology*; Jonassen, D.H., Ed.; Lawrence Erlbaum Associates Publishers: Mahwah, NJ, USA, 2004; pp. 745–783.
76. Butler, A.C.; Karpicke, J.D.; Roediger, H.L., III. The effect of type and timing of feedback on learning from multi-ple-choice tests. *J. Exp. Psychol. Appl.* **2007**, *13*, 273. [[CrossRef](#)] [[PubMed](#)]
77. Fyfe, E.R.; Rittle-Johnson, B. Feedback both helps and hinders learning: The causal role of prior knowledge. *J. Educ. Psychol.* **2016**, *108*, 82–97. [[CrossRef](#)]
78. Fyfe, E.R.; Rittle-Johnson, B. The benefits of computer-generated feedback for mathematics problem solving. *J. Exp. Child. Psychol.* **2016**, *147*, 140–151. [[CrossRef](#)]
79. Shute, V.J. Focus on Formative Feedback. *Rev. Educ. Res.* **2008**, *78*, 153–189. [[CrossRef](#)]
80. Cabestrero, R.; Quirós, P.; Santos, O.C.; Salmeron-Majadas, S.; Uria-Rivas, R.; Boticario, J.G.; Arnau, D.; Arevalillo-Herráez, M.; Ferri, F.J. Some insights into the impact of affective information when delivering feedback to students. *Behav. Inf. Technol.* **2018**, *37*, 1252–1263. [[CrossRef](#)]
81. Robison, J.; McQuiggan, S.; Lester, J. Evaluating the consequences of affective feedback in intelligent tutoring systems. In *Proceedings of the 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*; Institute of Electrical and Electronics Engineers (IEEE): Piscataway, NJ, USA, 2009; pp. 1–6.
82. Moreno, R. Decreasing cognitive load for novice students: Effects of explanatory versus corrective feedback in discovery based multimedia. *Instr. Sci.* **2004**, *32*, 99–113. [[CrossRef](#)]
83. Blumenfeld, P.C.; Kempler, T.M.; Krajcik, J.S. Motivation and Cognitive Engagement in Learning Environments. In *The Cambridge Handbook of the Learning Sciences*; Cambridge University Press (CUP): Cambridge, UK, 2005; pp. 475–488.
84. A Fredricks, J.; Blumenfeld, P.C.; Paris, A.H. School Engagement: Potential of the Concept, State of the Evidence. *Rev. Educ. Res.* **2004**, *74*, 59–109. [[CrossRef](#)]
85. Pintrich, P.R.; De Groot, E.V. Motivational and self-regulated learning components of classroom academic performance. *J. Educ. Psychol.* **1990**, *82*, 33. [[CrossRef](#)]
86. Zimmerman, B.J.; Bandura, A.; Martinez-Pons, M. Self-Motivation for Academic Attainment: The Role of Self-Efficacy Beliefs and Personal Goal-Setting. *Am. Educ. Res. J.* **1992**, *29*, 663–676. [[CrossRef](#)]
87. Guthrie, J.T.; Wigfield, A.; You, W. Instructional contexts for engagement and achievement in reading. In *Handbook of Research on Student Engagement*; Reschly, A.L., Sandra, L.C., Wylie, C., Eds.; Springer: Boston, MA, USA, 2012; pp. 601–634.
88. Phillips, V.; Popović, Z. More than Child's Play: Games have Potential Learning and Assessment Tools. *Phi Delta Kappan* **2012**, *94*, 26–30. [[CrossRef](#)]
89. McNamara, D.S.; Levinstein, I.B.; Boonthum, C. iSTART: Interactive strategy training for active reading and thinking. *Behav. Res. Methods Instrum. Comput.* **2004**, *36*, 222–233. [[CrossRef](#)]
90. McNamara, D.S.; O'Reilly, T.; Rowe, M.; Boonthum, C.; Levinstein, I.B. iSTART: A web-based tutor that teaches self-explanation and metacognitive reading strategies. In *Reading Comprehension Strategies: Theories, Interventions, and Technologies*; McNamara, D.S., Ed.; Erlbaum: Mahwah, NJ, USA, 2007; pp. 397–421.
91. Jackson, G.T.; McNamara, D.S. Motivational impacts of a game-based intelligent tutoring system. In *Proceedings of the 24th International Florida Artificial Intelligence Research Society (FLAIRS) Conference*; Murray, R.C., McCarthy, P.M., Eds.; AAAI Press: Menlo Park, CA, USA, 2011; pp. 519–524.
92. Anderson, J.R.; Conrad, F.G.; Corbett, A.T. Skill acquisition and the LISP tutor. *Cogn. Sci.* **1989**, *1*, 467–505. [[CrossRef](#)]

- 
93. Balyan, R.; McCarthy, K.S.; McNamara, D.S. Applying Natural Language Processing and Hierarchical Machine Learning Approaches to Text Difficulty Classification. *Int. J. Artif. Intell. Educ.* **2020**, *30*, 337–370. [[CrossRef](#)]
  94. Perret, C.A.; Johnson, A.M.; McCarthy, K.S.; Guerrero, T.A.; Dai, J.; McNamara, D.S. StairStepper: An Adaptive Remedial iSTART Module. In *Proceedings of the Lecture Notes in Computer Science*; Springer Science and Business Media LLC: Berlin, Germany, 2017; Volume 10331, pp. 557–560.
  95. Jackson, G.T.; Boonthum, C.; McNamara, D.S. *The Efficacy of iSTART Extended Practice: Low Ability Students Catch Up*. *International Conference on Intelligent Tutoring Systems*; Springer: Berlin/Heidelberg, Germany, 2010; pp. 349–351.
  96. MacGinitie, W.H.; MacGinitie, R.K. *Gates-MacGinitie Reading Tests*, 4th ed.; Houghton Mifflin: Iowa City, IA, USA, 2006.