Adaptive Reading and Writing Instruction in iSTART and W-Pal

Amy M. Johnson, Kathryn S. McCarthy,
Kristopher J. Kopp, Cecile A. Perret, Danielle S. McNamara
Institute for the Science of Teaching and Learning, Arizona State University, Tempe, AZ
amjohn43@asu.edu; ksmccar1@asu.edu; kristopher.kopp@asu.edu; cperret@asu.edu; dsmcnama@asu.edu

Abstract

Intelligent tutoring systems for ill-defined domains, such as reading and writing, are critically needed, yet uncommon. Two such systems, the Interactive Strategy Training for Active Reading and Thinking (iSTART) and Writing Pal (W-Pal) use natural language processing (NLP) to assess learners' written (i.e., typed) responses and provide immediate, accurate feedback. The current paper reports on efforts to implement adaptive instruction and task selection into both systems. In iSTART, we developed a new practice module, in which learners' past performance data governs two adaptive functionalities: 1) the use of self-explanation scaffolding and 2) the increase or decrease of difficulty of practice texts. In W-Pal, adaptivity is implemented by triggering targeted instructional support on the basis of deficits identified in learners' essays. In this paper, we describe the need for adaptive reading and writing instruction, along with the design and development of adaptivity in the two systems.

Introduction

Effective Intelligent Tutoring Systems (ITSs) that offer the opportunity for learners to engage in deliberate practice of reading and writing are essential because teachers have very limited time to offer such opportunities in the classroom. However, ITSs have traditionally been developed for well-defined, mathematically precise domains, such as algebra and programming (Anderson, Corbett, Koedinger, & Pelletier, 1995; Mitrovic, Suraweera, Martin, & Weerasinghe, 2004). In contrast, the use of ITSs for ill-defined domains, such as reading and writing, is less common. To assess learner performance in systems for these less-structured domains, natural language processing (NLP) is often employed (Graesser, Chipman, Haynes, & Olney, 2005; Jackson & McNamara, 2013; McNamara, Boonthum, Levinstein, & Millis, 2007). Interactive Strategy Training for Active Reading and Thinking (iSTART) and Writing Pal (W-Pal) are two such systems that provide instruction and extended practice of reading and writing, utilizing NLP to deliver feedback on learners' written responses. The purpose of this paper is to report development efforts to make these two systems more adaptive to learners' developing skills during system interactions.

iSTART was developed to provide reading strategy instruction and opportunities for extended practice of the strategies. Specifically, iSTART provides instruction in self-explanation, an active reading strategy that can improve comprehension of difficult texts (McNamara, 2004; 2017). iSTART includes instructional videos that train self-explanation strategies and practice modules that provide immediate feedback on self-explanations using NLP. Empirical research has shown that iSTART can improve learners' ability to construct quality self-explanations and improve reading comprehension (Jackson & McNamara, 2013; McNamara, O'Reilly, Best, & Ozuru, 2006).

W-Pal teaches effective writing strategies and offers learners the opportunity to practice writing persuasive essays. Persuasive essay writing is a skill assessed in many standardized tests (Roscoe & McNamara, 2013). Instruction in W-Pal focuses on writing skills related to the following stages of the writing process: free writing, planning, introduction building, body building, conclusion building, cohesion building, paraphrasing, and revising. Empirical studies have demonstrated the efficacy of W-Pal; the system leads to increased knowledge of the writing strategies and higher quality writing (Roscoe, Allen, Cai, Weston, Crossley, & McNamara, 2011; Roscoe, Brandon, Snow, & McNamara, 2013).

iSTART and W-Pal provide adaptive feedback to performance, but are not currently adaptive to the learners' developing skills as they progress. Essentially, they are both adaptive with the inner loop of tutoring, but not the outer loop (VanLehn, 2006). The work described in this paper is aimed at increasing adaptivity within both systems, particularly in terms of the outer loop, which selects appropriate tasks as students progress through training.

In iSTART, we are developing a new practice module, called StairStepper, in which learners' performance data determines both the availability of scaffolds and the difficulty of the texts used. In W-Pal, targeted instructional modules are presented to learners based on the deficiencies detected in their essays. The primary goal of this paper is to describe the design and development of these adaptive capabilities within the two systems.

Background

Intelligent Tutoring Systems

Providing opportunities for deliberate practice of reading and writing strategies is logistically challenging in traditional learning contexts. Teachers have limited time to instruct learners on these strategies and offer practice opportunities. Due to these time constraints, teachers often opt for multiple-choice measures of reading comprehension and cannot realistically provide feedback on multiple learner compositions (e.g., essays). Therefore, ITSs are uniquely positioned to fill a crucial instructional role for learners needing to improve reading comprehension and writing skills. Using NLP tools, ITSs can provide immediate and accurate feedback on learners' self-explanations and/or writing. Additionally, by using a model of the learners' reading and writing skills based on past performance, the systems can provide adaptive instruction. ITSs construct an ever-evolving profile of learners' performance and select tasks based on estimations of skills and anticipated instructional needs (Shute & Psotka, 1996; VanLehn, 2006; Woolf, 2010).

Commonly, ITSs identify learners' skills and select appropriate tasks. Familiar ITS elements include expert models, learner models, and instructional models. The expert model is a representation of the knowledge and skills that an expert would have in the instructional domain. The learner model is a representation of the learners' knowledge of the domain, given past learner behavior. The instructional model is used to modify the instructional content and tasks, given inferences about learner knowledge and skills. A learner model often is constructed using overlay models that represent a subset of an expert model, indicating the differences between novice and expert knowledge and skills. Several ITSs select instructional content on the basis of such overlay models (Woolf, 2010).

Using various techniques to update the learner model (e.g., constraint-based models and Bayesian belief networks), systems have often been created for well-defined domains, including mathematics (Cognitive Tutors, Anderson et al., 1995) and information technology (Mitrovic et al., 2004). Such ITSs demonstrate notable learning effects (~1.0 sigma). Comparatively, there are far fewer systems that apply to less structured tasks, like reading and writing. Among these are the Intelligent Essay Assessor (Foltz, Gilliam, & Kendal, 2000), iSTART (McNamara et al., 2006), and W-Pal (Roscoe & McNamara, 2013). iSTART and W-Pal utilize NLP, another feature which is less common in ITSs (Graesser, McNamara, & VanLehn, 2005). In this context, NLP techniques attempt to assess the quality of learner responses (e.g., self-explanations and essays), using semantic and syntactic characteristics, and by comparing content contained in the response to the content in the source text (e.g., the text being explained). Compared to well-defined domains, adapting instruction and practice in systems for less structured topics can prove more difficult. However, as the current paper describes, the iSTART and W-Pal projects endeavor to tackle this challenge.

iSTART

iSTART provides self-explanation strategy instruction to improve reading comprehension. Research demonstrates that actively engaging with a text by explaining it to oneself improves comprehension (Chi et al., 1989; 1994). iSTART offers strategy instruction and extended strategy practice in self-explanation to improve learners' ability to generate good self-explanations, thereby enhancing comprehension of difficult texts. The system instructs learners on five comprehension strategies: comprehension monitoring, paraphrasing, prediction, bridging, and elaboration. By applying these strategies, learners are able to decipher which parts of the text they understand, direct additional processes toward parts they do not understand, and recognize relationships among sentences and prior knowledge, thus improving comprehension and retention.

McNamara et al. (2006) found that learners who received iSTART training produced better self-explanations and had better performance on comprehension questions, compared to a control group of learners who received a short introduction to the concept of self-explanation. Jackson and McNamara (2013) compared the non-game version of iSTART to a version that included game-based practice of self-explanation. The results revealed that both versions improved learners' self-explanation performance, but learners in the game-based version reported higher levels of enjoyment and motivation. These results suggest that the enhanced motivation associated with game-based practice may promote persistence in the environment (Jackson & McNamara, in press).

Given these benefits, iSTART utilizes non-game and game-based generative practice, in which learners produce their own self-explanations, and game-based identification practice, in which learners attempt to identify which strategy is being used in example self-explanations. In generative practice, learners type responses and receive feedback using an NLP algorithm that evaluates the degree to which the self-explanation goes beyond the sentence being explained. In identification games, learners see exemplar self-explanations to a particular segment of text and select which of the trained strategies were used in the examples.

As we describe later in the paper, increasing adaptivity is the current vision of our development efforts for iSTART. We expect the addition of adaptive instruction into iSTART to further enhance its benefits in improving self-explanation ability and reading comprehension.

W-Pal

Effective writing instruction necessitates direct strategy instruction, opportunities for extended practice, and formative feedback (Graham & Perin, 2007; Hillocks, 1984; Kellogg & Raulerson, 2007). Based on this assumption, W-Pal is an interactive system designed to teach writing strategies and provide deliberate practice opportunities for adolescent developing writers. W-Pal focuses on the skill of writing persuasive essays similar to the types of essays used for standardized testing (Roscoe & McNamara, 2013). W-Pal provides instruction on writing strategies that target specific stages of the writing process (i.e., free writing, planning, introduction building, body building, conclusion building, cohesion building, paraphrasing, and revision). The amount of content and essay practicing opportunities allows a learner to use the system over multiple sessions.

One challenge of any ITS is to sustain learner engagement. Games can improve learners' motivation to participate by leveraging their intrinsic enjoyment of gaming (Orbach, 1979; Shank & Neaman, 2001). In W-Pal, each lesson is connected to game-based practice activities in an attempt to sustain learner engagement. These practice games allow learners to apply specific writing strategies and reinforce strategy knowledge through generative and identification tasks.

W-Pal includes several writing practice opportunities, which are offered at the end of each strategy module. Once a learner submits an essay, the system provides immediate summative and formative feedback. W-Pal provides a holistic score (on a 1-6 scale) along with targeted feedback. The targeted feedback is explicitly connected to the strategies taught in the lessons and encourages the use of the strategies during revision. In order to provide the targeted feedback, W-Pal uses NLP tools to extract linguistic data from the essays and implements a series of algorithms to assess essay quality and guide the feedback delivery.

Research has shown that W-Pal has a positive influence on writing quality. Learners who have used W-Pal demonstrate more knowledge of writing strategies (Roscoe et al., 2013) and produce higher quality essays (Roscoe et al., 2011). We hypothesize that the implementation of adaptivity in W-Pal will enhance the benefits to learners. We describe later how we have introduced this adaptivity by providing just-in-time instruction based on the weaknesses detected in learners' essays.

Adaptivity within iSTART and W-Pal

iSTART - StairStepper

One of our goals was for the instruction in iSTART to take into account learners' past system performance to identify the appropriate subsequent reading task. To this end, we are developing a new practice module, StairStepper. StairStepper is a generative self-explanation practice and multiple-choice recognition game. The StairStepper interface is depicted in Figure 1.

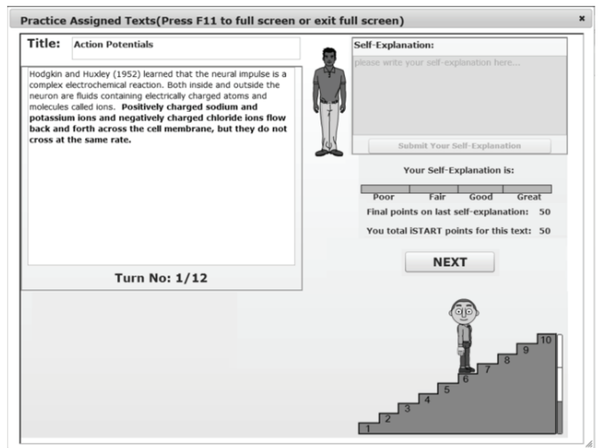

Figure 1. StairStepper interface.

In StairStepper, learners read texts and answer multiple-choice questions designed to mimic the types of questions that students encounter in standardized reading assessments such as the Gates MacGinitie or Nelson Denny reading tests. These questions ask textbased questions (answers require information directly found in the text), bridging questions (answers require the learner to make inferences across several sentences), and elaboration questions (answers require learners to make inferences across the text and their prior knowledge) to assess comprehension.

The goal of the game is to reach the highest stair by providing quality self-explanations and successfully answering comprehension questions, wherein the difficulty of the texts iteratively increases at each level. StairStepper is adaptive in two ways. First, it scaffolds the use of self-explanation (no self-explanation prompts, self-explanation prompt without feedback, self-explanation prompt with feedback) to increase both self-explanation and comprehension question scores. Second, the game adjusts the difficultly level of the texts being read based on the participant's previous scores. At the start of the game, a participant begins with a text at a pre-set difficulty. The default difficulty level is 5 (out of 13), but this starting point is a parameter that can be changed in the system. The learner is asked to read the text and answer a series of multiple-choice questions with the text available to them for reference (mimicking standard reading assessments). The default threshold is a comprehension question accuracy score (Q Score) of 75% correct, but the threshold is a parameter that can be changed in iSTART. Participants who meet this threshold are next directed to read a text at the next higher level of difficulty, and answer multiple-choice questions. Participants who do not meet the threshold are directed to read a text at the same level of difficulty, and are prompted to self-explain. If the participant continues to not meet the threshold, they are prompted to self-explain and are given feedback on their self-explanations, and the difficulty of the text decreases. When the text difficulty increases or decreases, the StairStepper character climbs or falls down the stairs, respectively. The flow of the game is depicted in Figure 2.

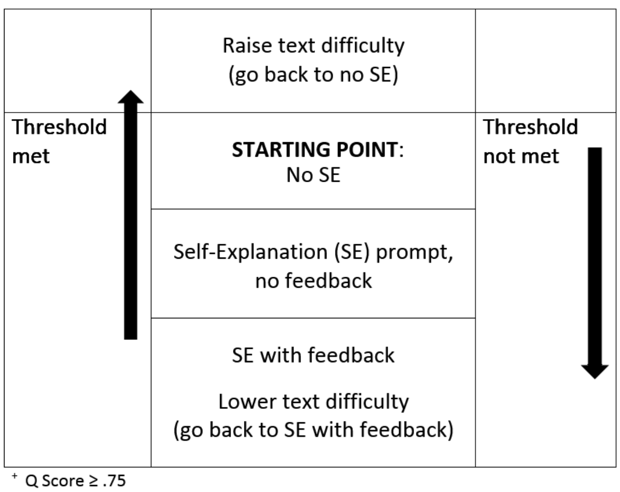

Figure 2. Flow of the StairStepper game.

To develop the game, we needed to build a set of texts with a range of difficulty levels. We selected 162 texts from two sets available on public websites. These expository texts varied in topics including science, history, and sports and were intended to be appropriate for grades 1-12. The suggested grade levels provided by the websites were not consistent across the two texts sets, nor were they consistent with Flesch-Kincaid Grade Level. StairStepper was developed to mimic common standardized tests that tend to rely on word and syntax features, rather than cohesion or content. Consequently, standard measures in Coh-metrix also did not intuitively capture the differences in difficulty across the texts.

To better assess text difficulty, human raters were used to develop a new system that ranked the texts by levels rather than on a continuum. A group of four raters were instructed to compare the texts holistically. As an initial coarse sort, the raters together arranged the entire set of essays from least to most difficult. This continuum was then broken into thirds (easy, medium, and difficult). The individual raters then categorized each text as being "less difficult" or "more difficult" relative to the other texts in that third, yielding six levels. Adjacent levels were combined and resorted into less or more difficult texts until there was agreement across raters that no text in the grouping was more or less difficult than the other texts in the group. This process continued until there was agreement that the set of texts could not be further separated. This protocol yielded thirteen difficulty levels with at least five texts in each level. These ratings had a strong correlations with Flesch-Kincaid grade level (r = .79) and with Dale-Chall readability (r = .77) suggesting that this human rating is consistent, but not redundant with existing measures of readability. An on-going project is to use this text difficulty rating system to develop an NLP algorithm that automatically assesses the level of difficulty for new texts as they are introduced into iSTART.

Adaptive W-Pal

One goal of the W-Pal system is to be able to automatically assess essays and provide immediate and direct feedback that would be beneficial for a revision. The adaptive system aims to better target strategies that would benefit the learners, based on their initial essays, and provide instructional support on those targeted strategies. In the current system, the learner goes through a module on a specific strategy (e.g., introduction building) and once they have completed all of the activities in the module, they have the opportunity to practice writing the essay. In the adaptive system, the learner begins by writing an essay and then receives feedback on the initial essay. Next, they are directed to an instructional module related to this feedback. For example, if a learner writes an essay that lacks cohesion, the feedback messages would indicate the lack of cohesion, and the learner would be directed to the 'Cohesion Building' instructional module. Thus, the system adapts to the needs of the learner based on the assessment of their written essays. After completing the appropriate instructional module, they revise their essay. After revision, they receive feedback on the second draft and then do an activity from a module relevant to the feedback on the revision. See Figure 3 for a depiction of the cycle.

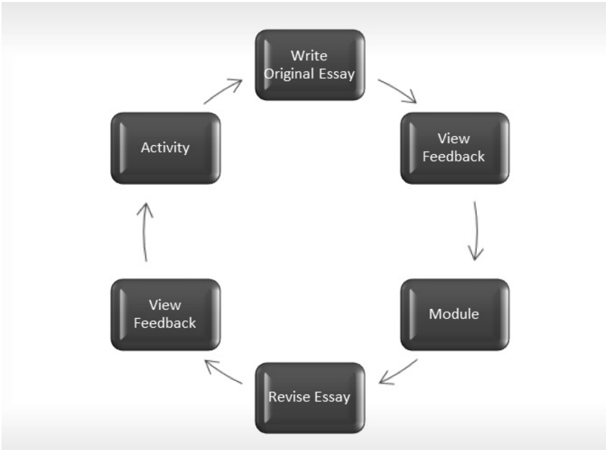

Figure 3. Representation of adaptive W-Pal cycle.

The feedback generation system works as follows. The series of algorithms are designed to assess the essays for different categories of potential flaws associated with the writing strategies from the lesson modules (e.g., length, poor introduction). When the algorithm indicates a flaw, the system presents a topic from the category.

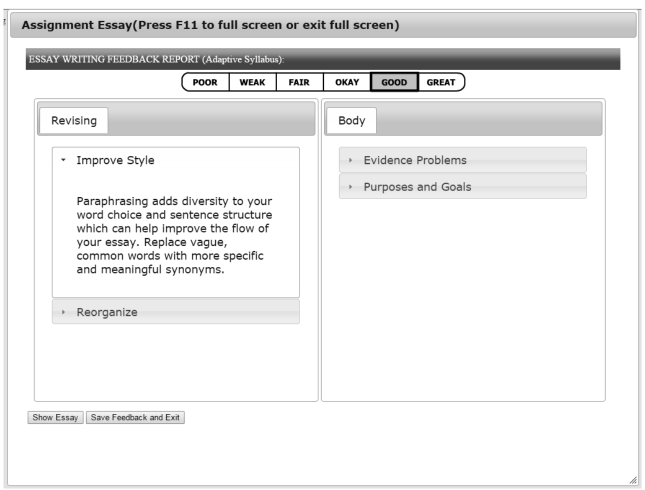

Figure 4. Screenshot of W-Pal feedback.

The current system contains multiple feedback categories and various topics within each category. Currently, due to the iterative nature of W-Pal, a learner may receive the same feedback on multiple essays. Another goal of the adaptive feedback system is to improve the feedback delivery method to adapt to learner performance and reduce repetition across multiple sessions. The feedback in the new system has been modified so that the topic message for each category is streamlined. The learner is only able to see up to two topics for each feedback category and the topic messages are not repeated. That is, after seeing a particular topic message, if they trigger the same category, new topic messages will be provided. Furthermore, if the learner triggers the same category over multiple sessions and that module has already been completed, the system moves on and the learner completes the module for an alternative category that was triggered. See Figure 4 for an example feedback screen.

Conclusion

The iSTART and W-Pal systems harness the benefits of ITSs and NLP to provide one-on-one training, practice, and feedback in ways that would be impossible in traditional reading and writing instruction given constraints of both time and resources in the classroom. Recent development efforts aim to increase adaptivity, by adjusting how each learner progresses through and interacts with the system to better address their individual needs. Our anticipation is that, by adaptively responding to student performance, the systems will lead to better comprehension and writing skills outcomes.

There are potential drawbacks to the new adaptive elements in the two systems. First, an inevitable consequence of implementing system task selection is a reduction in learner agency. Research shows that educators believe that permitting student choice in educational contexts promotes motivation and learning (Flowerday & Schraw, 2000). Although the iSTART text selection mechanism used in StairStepper diminishes learner choice within the module, learners are still fundamentally free to select among all generative and identification practice activities. Likewise, the adaptivity in W-Pal directs learners to specific instructional modules, but learners are free to choose the practice activities within the modules. Nonetheless, examining the appropriate balance of learner agency and individualized instruction is an important direction for research on iSTART and W-Pal. Another possible effect of playing the StairStepper module is a reduction in motivation and self-efficacy if the student continues to perform poorly and moves down the steps. Evaluations of the module will include pretest/posttest surveys assessing motivation and self-efficacy to determine if design modifications will be necessary.

Moving forward, our first step is to test the new adaptive features for usability and efficacy with both student learners and adult literacy learners. The results of these investigations will inform future modifications to the systems. As an example, if learner perceptions indicate that forcing modules in W-Pal negatively affects enjoyment or other motivation attributes, the system may instead offer suggestions for remediation in particular modules. We also plan to iteratively test and refine the appropriateness of the W-Pal feedback loops. Student essays from W-Pal will be scored by humans to determine whether the directed modules are fitting, given features of the produced essays. Revision essays will also be examined to assess the extent to which remediating modules positively impact the relevant features (e.g., cohesion, introduction) of the essays. Furthermore, iterative testing of iSTART may inform the appropriate settings for initial text difficulty and multiple choice accuracy threshold used for scaffolding.

We plan to refine and further implement the adaptivity algorithm from StairStepper into the other iSTART practice activities, so that each reader receives texts at the appropriate difficulty and is moved through the system features in ways that target his or her specific needs.

Given the fundamental relationship between reading and writing skills, we are also taking strides towards integrating features of both iSTART and W-Pal into one system that iteratively provides adaptive instruction based on the needs of the learner. Our ultimate goal is to adapt instruction as training progresses, such that students' literacy can be enhanced both efficiently and comprehensively.

Acknowledgments

The authors would like to recognize the support of the Institute of Education Sciences, U.S. Department of Education, through Grants R305A130124 and R305A120707, and the Office of Naval Research, through Grant N00014140343, to Arizona State University. The opinions expressed are those of the authors and do not represent views of the Institute, the U.S. Department of Education, or the Office of Naval Research.

References

Anderson, J. R.; Corbett, A. T.; Koedinger, K. R.; and Pelletier, R. 1995. Cognitive tutors: Lessons learned. *The Journal of the Learning Sciences* 4(2): 167-207.

Chi, M. T.; Bassok, M.; Lewis, M. W.; Reimann, P.; and Glaser, R. 1989. Self-explanations: How learners study and use examples in learning to solve problems. *Cognitive Science* 13(2): 145-182.

Chi, M. T.; De Leeuw, N.; Chiu, M. H.; and LaVancher, C. 1994. Eliciting self-explanations improves understanding. *Cognitive Science* 18(3): 439-477.

Flowerday, T.; and Schraw, G. 2000. Teacher beliefs about instructional choice: A phenomenological study. *Journal of educational psychology* 92(4): 634.

Foltz, P. W.; Gilliam, S.; and Kendall, S. 2000. Supporting content-based feedback in on-line writing evaluation with LSA. *Interactive Learning Environments* 8(2): 111-127.

Graham, S.; and Perin, D. 2007. A meta-analysis of writing instruction for adolescent learners. *Journal of Educational Psychology* 99: 445-476.

Graesser, A.C.; McNamara, D.S.; and VanLehn, K. 2005. Scaffolding deep comprehension strategies through Point&Query, AutoTutor, and iSTART. *Educational Psychologist* 40: 225-234.

Hillocks, G. 1984. What works in teaching composition: a meta-analysis of experimental treatment studies. *American Journal of Education* 93: 133-170.

Jackson, G. T.; and McNamara, D. S. 2013. Motivation and performance in a game-based intelligent tutoring system. *Journal of Educational Psychology* 105(4): 1036.

Jackson, G.T.; and McNamara, D.S. in press. Exploring the Potential Longer-Term Benefits of Educational Games: Motivation and Mastery Framework. In Y. Baek (Ed.) *Game-Based Learning: Theory, Strategies and Performance Outcomes*.

Kellogg, R.; and Raulerson, B. 2007. Improving the writing skills of college learners. *Psychonomic Bulletin and Review* 14: 237-242.

McNamara, D. S. 2004. SERT: Self-explanation reading training. *Discourse Processes* 38(1): 1-30.

McNamara, D. S. 2017. Self-explanation and reading strategy training (SERT) improves low-knowledge students' science course performance. *Discourse Processes*.

McNamara, D. S.; O'Reilly, T. P.; Best, R. M.; and Ozuru, Y. 2006. Improving adolescent learners' reading comprehension with iSTART. *Journal of Educational Computing Research* 34(2): 147-171.

Mitrovic, A.; Suraweera, P.; Martin, B.; and Weerasinghe, A. 2004. DB-suite: Experiences with three intelligent, web-based database tutors. *Journal of Interactive Learning Research* 15(4): 409.

Orbach, E. 1979. Simulation games and motivation for learning: A theoretical framework. *Simulation and Games* 10: 3-40.

Roscoe, R.D.; Allen, L.K.; Cai, Z.; Weston, J.L.; Crossley, S.A.; and McNamara, D.S. 2011. Internal usability testing of automated essay feedback in an intelligent writing tutor. In R. C. Murray & P. M. McCarthy (Eds.), *Proceedings of the 24th International Florida Artificial Intelligence Research Society (FLAIRS) Conference,* 543-548. Menlo Park, CA: AAAI Press.

Roscoe, R. D.; Brandon, R.; Snow, E.; and McNamara, D. S. 2013. Game-based writing strategy practice with the Writing Pal. In K. Pytash & R. Ferdig (Eds.), *Exploring Technology for Writing and Writing Instruction*, 1-20. Hershey, PA: IGI Global.

Roscoe, R.D.; Crossley, S.A.; Weston, J.L.; and McNamara, D.S. 2011. Automated assessment of paragraph quality: Introductions, body, and conclusion paragraphs. In R. C. Murray & P. M. McCarthy (Eds.), *Proceedings of the 24th International Florida Artificial Intelligence Research Society (FLAIRS) Conference*, 281-286. Menlo Park, CA: AAAI Press.

Roscoe, R. D.; and McNamara, D. S. 2013. Writing pal: Feasibility of an intelligent writing strategy tutor in the high school classroom. *Journal of Educational Psychology* 105(4): 1010.

Shank, R.; and Neaman, A. 2001. Motivation and failure in educational systems design. In K. Forbus & P. Feltovich (Eds.), *Smart Machines in Education*. Cambridge, MA: AAAI Press/ MIT Press.

Shute, V. J.; and Psotka, J. 1996. Intelligent Tutoring Systems: Past, Present and Future. D. Jonassen (ed.) *Handbook of Research on Educational Communications and Technology*.

Vanlehn, K. 2006. The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education* 16(3): 227-265.

Woolf, B. P. 2010. *Building Intelligent Interactive Tutors: Learner-Centered Strategies for Revolutionizing E-learning*. Morgan Kaufmann.