Timing Game-Based Practice in a Reading Comprehension Strategy Tutor

Matthew E. Jacovina[1], G. Tanner Jackson[2], Erica L. Snow[3], and Danielle S. McNamara[1]

[1]Institute for the Science of Teaching & Learning, Arizona State University, Tempe, AZ, 85287
{Matthew.Jacovina, Danielle.McNamara}@asu.edu

[2]Cognitive Science, Educational Testing Service, Princeton, NJ, 08541
gtjackson@ets.org

[3]SRI International, Menlo Park, CA, 94025
erica.snow@sri.com

Jacovina, M. E., Jackson, G. T., Snow, E. L., & McNamara, D. S. (2016). Timing game-based practice in a reading comprehension strategy tutor. In A. Micarelli, J. Stamper, & K. Panourgia (Eds.), *Proceedings of the 13th International Conference on Intelligent Tutoring Systems (ITS 2016)* (pp. 80-89). Zagreb, Croatia: Spinger. Published with acknowledgment of federal support.

Author's Note

This research was supported in part by the Institute for Educational Sciences (IES R305A130124). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES. We thank the many colleagues and students who have contributed to this work, and extend a special thanks to Tricia Guerrero for her help in coding data for this project.

Timing Game-Based Practice in a Reading Comprehension Strategy Tutor

Matthew E. Jacovina[1], G. Tanner Jackson[2], Erica L. Snow[3], and Danielle S. McNamara[1]

[1]Institute for the Science of Teaching & Learning, Arizona State University, Tempe, AZ, 85287
{Matthew.Jacovina, Danielle.McNamara}@asu.edu
[2]Cognitive Science, Educational Testing Service, Princeton, NJ, 08541
gtjackson@ets.org
[3]SRI International, Menlo Park, CA, 94025
erica.snow@sri.com

Abstract. Game-based practice within Intelligent Tutoring Systems (ITSs) can be optimized by examining how properties of practice activities influence learning outcomes and motivation. In the current study, we manipulated *when* game-based practice was available to students. All students (*n*=149) first completed lesson videos in iSTART-2, an ITS focusing on reading comprehension strategies. They then practiced with iSTART-2 for two 2-hour sessions. Students' first session was either in a game or nongame practice environment. In the second session, they either switched to the alternate environment or remained in the same environment. Students' comprehension was tested at pretest and posttest, and motivational measures were collected. Overall, students' comprehension increased from pretest to posttest. Effect sizes of the pretest to posttest gain suggested that switching from the game to nongame environment was least effective, while switching from a nongame to game environment or remaining in the game environment was more effective. However, these differences between the practice conditions were not statistically significant, either on comprehension or motivation measures, suggesting that for iSTART-2, the timing of game-based practice availability does not substantially impact students' experience in the system.

Keywords: Game-based learning, Intelligent Tutoring Systems, Comprehension, Motivation

1 Introduction

Intelligent Tutoring Systems (ITSs) have produced positive outcomes for students across a number of domains [1]. The individualized instruction offered by ITSs is most successful when students engage in extended practice. Unfortunately, students often become disengaged and bored while using ITSs [2]. Enhancing students' motivation to persist in their use of these systems without sacrificing educational benefits has thus been an ongoing challenge for developers. Implementing educational games and game-like features is one method for increasing students' interest in practicing within tutoring systems [3]. Games aim to leverage students' enjoyment to both increase persistence in practice and encourage deep and meaningful interactions with the content of the game [4]. Research on the addition of game features to nongame environments in order to improve user experience has become an increasingly hot field of study. Despite attracting attention from fields such as marketing and health, however, there are many gaps left to be filled in understanding the impact of game-features [5].

The study of ITSs has not reached a consensus on the efficacy of educational games and game-like features. Clearly, games are not a panacea for all educational goals and contexts, and their use must be tested broadly and with multiple implementations. For example, the influence of games has been studied in contexts ranging from classrooms [6] to military training [7], all with some degree of success. Unsurprisingly, though, not all game features are equally compelling or appropriate for different goals [3, 8]. Moreover, game features may serve to distract some students from the pedagogical goals of a system [9-11].

An important aspect of testing the effectiveness of educational games is determining which specific properties of the gaming experience are important for educational outcomes and motivation. This can allow developers to make informed decisions about how to implement game features. For example, one study examined the effect of making an educational game single-player or multiplayer, and found no differences on knowledge acquisition or perceptions of the activity [12]. Design analyses of popular games can also be conducted to extract key properties of positive gaming experiences. In an analysis of the puzzle game *Candy Crush Saga*, for example, Varonis and Varonis [13] identified several important aspects of the game, such as the requirement for iterative innovation, providing immediate feedback, giving bonuses for exemplary performance, and allowing players to engage in alternative activities in between engaging with the main game.

In addition to game *features*, the *timing* of game-based practice availability may be an important factor. Given the mixed results in the literature on the effectiveness of educational games [14], one possibility is that game-based practice best serves students at particular time points. For example, a game designed to teach the programming concept of *loops* was found to be more effective when played before a more traditional assignment on the topic than after the traditional assignment [15]. This game was tightly integrated with the learning material, potentially making it immediately effective. For systems that add game features to educational activities that may distract from learning, having all features available immediately may be undesirable.

1.1 iSTART-2

The current study was conducted using the Interactive Strategy Training for Active Reading and Thinking-2 (iSTART-2) system. iSTART-2 is a game-based tutoring system that provides reading comprehension instruction by teaching self-explanation strategy lessons and strategy practice games [16, 17]. iSTART-2 provides 8^{th} grade through college students with strategies designed to help them construct deep and meaningful text representations. This is an important academic skill and one that is difficult for many readers [18]. Although the strategy lessons and practice activities are the driving forces in helping students improve, other system features (e.g., game-based practice) may help to motivate students and indirectly improve comprehension.

However, the game features do not directly teach self-explanation skills. Thus, a key goal for iSTART-2 is to include game features that increase motivation but do not distract from practicing self-explanation strategies.

Previous work has compared a game-based version of iSTART to a nongame based system, and found that students equally benefitted from the two versions of the system [19]. Another study showed that across time, a game-based version of iSTART yielded higher enjoyment and motivation than a nongame version [16]. This research suggests iSTART-2's game-based practice may be appropriately tuned to enhance motivation without decreasing learning. However, these findings do not confirm that learning and motivation have been optimized. Varying the availability of game and nongame activities may further enhance outcomes. Specifically, early exposure to nongame practice followed by access to game-based practice may afford students an uninterrupted introduction to practice activities, and then introduce motivational features that motivate their continued effort.

1.2 Current Study

In this study we aimed to determine *when* to make game-based practice available to students within the iSTART-2 practice environment. All students in this study began by watching lesson videos and answering checkpoint questions for each. During two subsequent study sessions, students practiced within iSTART-2 for two hours. Students were randomly assigned to begin their first practice session in either a game or nongame environment. During students' second practice session, they either continued in the same environment or switched to the alternate environment. This created a total of four conditions across the 2 (Initial Practice: Game or Nongame) x 2 (Practice Consistency: Switch or Stay) experimental design.

The game and nongame practice environments differed primarily on the presence of game features within the practice activities. In both environments, students had access to one generative activity and one identification activity (see Figure 1).

In generative activities, students read science texts and write self-explanations in response to predefined target sentences. After submitting their self-explanation, students receive an automated score for the quality of their response [20]. *Map Conquest*, in the game environment, allowed students to use the points they earned through their self-explanations to attempt to "conquer" a game board against computer opponents. *Coached Practice*, in the nongame environment, assigned scores to students' self-explanations, but these scores did not relate to a game activity. However, Coached Practice did offer additional feedback and suggestions to improve the quality of students' self-explanation in the form of verbal responses from a pedagogical agent.

In identification activities, students read self-explanations that are ostensibly written by other students. The students' task is to identify which iSTART-2 strategy was used to generate that self-explanation. All students receive feedback on the accuracy of their choices. *Bridge Builder*, in the game environment, also gives points to students, with point bonuses for consecutive correct answers. A simple narrative also plays out as students give correct answers, allowing an explorer to cross a bridge in search of treasure. *Strategy Identification*, in the nongame environment, only gives accuracy feedback. Within the game environment, students can use the points they earned to modify the background color of the site, purchase new hair styles and colors for an avatar, and track their achievements through a list of trophies that they win in the games. These features are not available in the nongame environment.

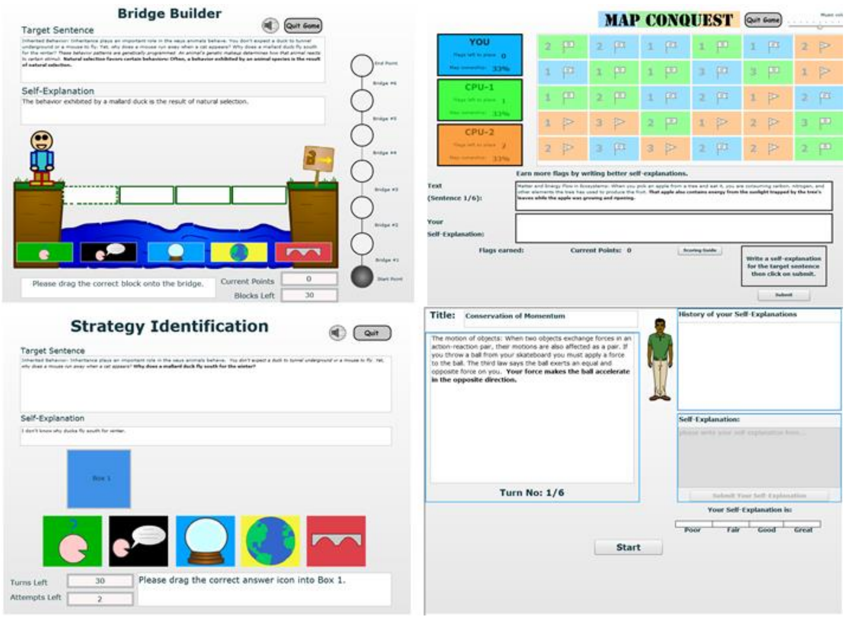

Fig. 1. The practice activities in the game (top row) and nongame (bottom row) environments

Both practice environments were thus nearly identical in terms of the educational content, but the game environment includes features intended to enhance students' experience. Learning was measured comparing pretest and posttest performance on open-ended comprehension questions for a science text. Half of the comprehension questions were textbase questions and half were bridging inference questions. Text-base questions require readers to remember information that was directly stated in one sentence of the text, whereas bridging inference questions require that readers integrate information across multiple sentences in a text.

Our first hypothesis (H1) was that students would perform better at posttest than pretest across both question types and across all practice conditions. Improvement from pretest to posttest on the comprehension measure is indicative of the benefits of iSTART on students' ability to comprehend challenging content-area texts.

Two alternative hypotheses center on how the timing of availability for these features might influence learning. Hypothesis 2a (H2a) was that practice in the game environment would on average be more beneficial than practice in the nongame environment. This hypothesis is plausible given that past research has shown benefits for game-based practice [6, 7]. Hypothesis 2b (H2b) was that the timing of game-based practice would influence pretest to posttest gain. This hypothesis is based on findings that game-based practice may impede performance [9-11]. Specifically, this leads to the prediction that larger benefits would be observed when students switch from a nongame to game environment, because students receive unadulterated practice early in the learning process, and then obtain access to motivating game features. Smaller benefits would be observed, however, when switching from a game to nongame environment, because during the second session, the system fails to meet students' expectations for game-based practice.

Our third and fourth hypotheses center on how the timing of availability for game-based features influenced dimensions of motivation, which should be related to posttest performance. Three dimensions of motivation were measured at posttest: students' reported effort exerted while using iSTART-2, their perception of their performance quality, and their emotional state at posttest. To confirm that these dimensions were related to performance, hypothesis 3 (H3) was that each dimension would correlate with posttest performance. Hypothesis 4 (H4) was that an interaction would emerge between initial practice environment and practice consistency. Specifically, we predicted that students would report higher scores on the motivational dimensions when their second session was a game environment, and report the lowest scores when switching from a game to a nongame environment. Thus, our prediction was that beginning in a nongame environment and switching to a game-environment would, overall, be the optimal condition. This condition initially provides practice without the distraction of games, and follows with game-based practice in the second practice session when students' motivation may have decreased.

2 Method

2.1 Participants

This study included 149 high school students and recent high school graduates from the Southwest United States. These students were, on average, 16.22 years of age (range: 13-20 years), with the majority of students reporting their grade level as high school seniors or sophomores. Of the 149 students, 55% self-identified as female; 43.6% self-identified as Caucasian, 32.2% as Hispanic, 8.7% as African-American, 7.4% as Asian, and 8.1% as another ethnicity. Seven students dropped out of the study before the final session and their data were not included in these analyses; one additional student's data were removed from analyses due to technical problems with the pretest survey.

2.2 Materials

The pretest and posttest included measures of reading comprehension skill and motivation. Reading comprehension skill was assessed through comprehension questions based on two science passages. The presentation order of the texts (pretest or posttest) was counterbalanced across students. The texts and questions were modified from those used in previous research [16, 21]. The texts were selected for their similar length (311 and 283 words), Flesch-Kincaid grade level (8 and 9), and linguistic features as measured by the natural language processing tool, Coh-Metrix [22]. While reading each text, students were prompted to self-explain 9 sentences. For each text, there were 8 open-ended questions, including 4 textbase questions and 4 bridging inference questions. The text was not on screen while students answered these questions. The answers to textbase questions were found within a single sentence of the text, whereas the answers to bridging inference questions required students to integrate information between two or more sentences. Each question could receive a maximum of 1 point, with some questions allowing for partial credit. Two coders independently scored at least 14% of the responses for each question, resolved discrepancies, and iterated on this process until they achieved 95% exact agreement (all kappa values above 0.8). After achieving agreement on a question, one coder completed the scoring.

Pretest motivation was assessed using the learning intentions, self-efficacy, and emotional state dimensions of a modified version of the Online Motivation Questionnaire [OMQ; 23]. Posttest motivation was also assessed using an adapted version of the OMQ, and included the dimensions of reported effort, result assessment, and emotional state.

2.3 Procedure

This project was part of a 5-session study, which lasted approximately 8.5 hours in total. Each session was completed on a different day to avoid fatigue. In session 1, students completed demographic surveys and a writing task that is unrelated to the current study. During session 2, students completed pretest measures, including the reading comprehension questions and the pretest OMQ questions. Students then completed the iSTART-2 lesson videos. During both sessions 3 and 4, students engaged with the iSTART-2 practice interface for 2 hours. These practice sessions were controlled for time and not the activities with which students engaged. The initial practice environment (game or nongame) and practice consistency (whether the environment switched or stayed the same between session 3 and 4) varied depending on students' randomly assigned condition. During session 5, students completed a posttest, which included the reading comprehension test and the OMQ questions.

3 Results

Analyses were conducted to examine the effects of initial practice (game or nongame) and practice consistency (switch or stay) on comprehension scores and motivation measures.

3.1 Comprehension Scores

To determine the effects of iSTART-2 training and practice condition on comprehension scores, a mixed ANOVA was conducted with test (pretest, posttest) and question type (textbase, bridging) as within-participant factors, and initial practice (game, nongame), and practice consistency (switch, stay) as between-participant factors.

Comprehension scores are reported as the percentage of total possible points that a student achieved (see Table 1). A main effect of question type emerged such that students scored higher on textbase questions than on bridging questions [$F(1, 137) = 26.99$, $p < .001$, $\eta_p^2 = .165$]. This finding serves as a confirmation that the bridging questions were more difficult to answer. A main effect of test also emerged such that students scored higher at posttest than at pretest [$F(1, 137) = 5.02$, $p = .027$, $\eta_p^2 = .035$]. This finding thus supported H1.

Table 1. Pretest and posttest means (and *SD*) for textbase and bridging questions.

	Pretest (*SD*)	Posttest (*SD*)	Mean (*SD*)
Textbase questions	48.2% (29.9)	51.6% (31.0)	49.9% (26.3)
Bridging questions	39.8% (24.7)	44.3% (25.1)	42.1% (21.6)
Mean (*SD*)	44.0% (24.2)	47.9% (25.0)	

No main effects or interactions involving the two practice condition factors, initial practice or practice consistency, were significant, failing to support H2a or H2b. The lack of interactions involving test, question type and practice conditions suggests that gains for both textbase and bridging questions were similar across conditions. Table 2 displays the pretest and posttest mean scores for each condition as well as the effect size of the pretest to posttest improvement. Although an interaction did not emerge between the conditions, the pretest to posttest gain were highest when students remained in a game environment (i.e., $\eta_p^2 = .081$) or switched from a nongame environment to a game environment (i.e., $\eta_p^2 = .069$), and lowest when students switched from a game environment to a nongame environment (i.e., $\eta_p^2 = .003$). This pattern is partially consistent with H2b in that switching from a game to a nongame environment led to a lower gain while switching from a nongame to game environment led to a higher gain. However, this may be attributable to pretest differences.

Table 2. Partial eta squared values for the pretest to posttest gain for each of the four conditions.

	Initial practice: Game			Initial practice: Nongame		
	Pretest	Posttest	Effect Size	Pretest	Posttest	Effect Size
Switch practice environments	47.1%	48.4%	$\eta_p^2 = .003$	46.0%	51.2%	$\eta_p^2 = .069$
Stay in practice environment	40.0%	46.0%	$\eta_p^2 = .081$	43.4%	46.3%	$\eta_p^2 = .025$

3.2 Motivation Measures

Table 3 displays correlations between posttest comprehension scores and the pretest and posttest OMQ dimensions of interest. All OMQ motivation dimensions were significantly correlated with posttest performance, supporting H3. To test the effects of practice condition on posttest motivation, between-participant ANCOVAs were conducted with the three posttest OMQ dimensions serving as dependent variables:

reported effort, performance assessment, and emotional state. Initial practice (game, non-game) and practice consistency (switch, stay) were between-participant factors. Pretest OMQ dimensions (learning intentions, self-efficacy, and emotional state) served as covariates to control for pretest differences across conditions that emerged despite random assignment. However, no main effects or interactions emerged for initial practice or practice consistency (all *F*s < 2.6, *p*s > .10). This suggests that practice condition did not influence these dimensions of posttest motivation, failing to support H4.

Table 3. Correlations between comprehension scores and motivation measures.

Measure	1	2	3	4	5	6	7
1. Post comprehension	-						
2. Pre learning intentions	.28**	-					
3. Pre self-efficacy	.28**	.35**	-				
4. Pre emotional state	.17*	.22**	.27**	-			
5. Post reported effort	.39**	.28**	.23**	.16	-		
6. Post result assessment	.33**	.25**	.42**	.22**	.57**	-	
7. Post emotional state	.32**	.32**	.34**	.43**	.39**	.37**	-

*p<.05, ** p < .01

4 Conclusions

In this study we examined how the timing of game-based practice availability influenced performance and motivation. After completing instructional videos, students spent 2 two-hour sessions in iSTART-2 practice environments. Students were randomly assigned to begin in a game-based or nongame environment; half of the students stayed in the same environment during the second practice session and half switched to the other environment. Overall, we found that students' scores on comprehension questions improved from pretest to posttest, supporting H1. Consistent with past work, these results support the notion that iSTART-2 benefits students' reading comprehension. No effects of initial practice or practice consistency emerged, failing to support H2a or H2b. Students' overall benefits were approximately equivalent regardless of whether they began in a game or nongame practice environment, or whether they switched or stayed in the same environment. However, the effect sizes of the pretest to posttest gain were partially consistent with H2b, in that switching from a game to a nongame environment was least effective, while switching from a nongame to game environment was more effective. Remaining in a game environment also led to a large effect size. All motivation dimensions were positively correlated with posttest comprehension performance, supporting H3. This suggests that testing for effects of practice condition on these motivation measures is worthwhile. However, the dimensions of motivation were not influenced by condition, failing to support H4. Reports of effort and performance quality, and posttest emotional state did not seem to be influenced by the timing of game-based feature availability.

These results align with a past study using iSTART-2 that compared students' self-explanation quality after 45 minutes of practicing in a game-like or less game-like activity, and found no overall difference [24]. For iSTART-2, one possibility is that the impact of individual game features is small compared to the overall impact of a system that affords students *agency* over their learning through choices of practice activities [17]. An additional possibility is that the outcome measures included in this study were not sufficiently sensitive. Future analyses examining interaction patterns within iSTART-2 may uncover differences between practice conditions. Moreover, posttest motivation was measured during a separate session to capture students' overall experience. Testing students' motivation more frequently, perhaps during and at the conclusion of each session, may capture changes in motivation over time that this study could not. In classrooms, behavioral measures may serve as proxies for motivation, such as how frequently students practice outside of class assignments.

Overall, the findings in the current project provide support for the effectiveness of iSTART-2. Although the results do not provide strong evidence for when game-based practice should be made available in iSTART-2, the pretest to posttest gains across conditions suggest that students should either be provided consistent access to games or should begin with nongame practice and then transition to game-based practice. Future work will continue to explore the features of game-based practice and its timing, perhaps over longer periods of time that include the gradual release of more than two games, in order to optimize students' experience within iSTART-2.

Acknowledgments. This research was supported in part by the Institute for Educational Sciences (IES R305A130124). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the IES. We thank the many colleagues and students who have contributed to this work, and extend a special thanks to Tricia Guerrero for her help in coding data for this project.

References

1. Steenbergen-Hu, S., Cooper, H.: A meta-analysis of the effectiveness of intelligent tutoring systems on college students' academic learning. Journal of Educational Psychology 106, 331–347 (2014)
2. D'Mello, S., Olney, A., Williams, C., Hays, P.: Gaze tutor: A gaze-reactive intelligent tutoring system. International Journal of Human-Computer Studies. 70, 377–398 (2012)
3. McNamara, D.S., Jackson, G.T., Graesser, A.C.: Intelligent tutoring and games (ITaG). In: Baek, Y.K. (ed.) Gaming for classroom-based learning: Digital roleplaying as a motivator of study. IGI Global, Hershey (2010)
4. Gee, J.P.: What Video Games Have to Teach Us About Learning and Literacy. Palgrave Macmillan, New York (2003)
5. Richter, G., Raban, D.R., Rafaeli, S.: Studying gamification: The effect of rewards and incentives on motivation. In: Reiners, T., Wood, L. (eds.) Gamification in education and business, pp. 21-46. Springer International Publishing (2015)
6. Papastergiou, M.: Digital Game-Based Learning in high school Computer Science education: Impact on educational effectiveness and student motivation. Computers & Education 52, 1–12 (2009)

7. Belanich, J., Orvis, K.L., Sibley, D.E.: PC-based game features that influence instruction and learner motivation. Military Psychology 25, 206–217 (2013)
8. Amory, A., Naicker, K., Vincent, J., Adams, C.: The use of computer games as an educational tool: Identification of appropriate game types and game elements. British Journal of Educational Technology 30, 311–321 (1999)
9. Adams, D.M., Mayer, R.E., MacNamara, A., Koenig, A., Wainess, R.: Narrative games for learning: Testing the discovery and narrative hypotheses. Journal of Educational Psychology 104, 235-249 (2012)
10. Rieber, L.P., Noah, D.: Games, simulations, and visual metaphors in education: Antagonism between enjoyment and learning. Educational Media International 45, 77–92 (2008)
11. Jackson, G.T., Dempsey, K.B., McNamara, D. S.: Game-based practice in reading strategy tutoring system: Showdown in iSTART-ME. In Reinders, H. (ed.), Computer games, pp. 115-138. Bristol, UK: Multilingual Matters (2012)
12. Tsai, F.H., Tsai, C.C., Lin, K. Y.: The evaluation of different gaming modes and feedback types on game-based formative assessment in an online learning environment. Computers & Education 81, 259-269 (2015)
13. Varonis, E.M., Varonis, M.E. (2015). Deconstructing Candy Crush: What instructional design can learn from game design. The International Journal of Information and Learning Technology 32, 150-164 (2015)
14. Wouters, P., van Nimwegen, C., van Oostendorp, H., van der Spek, E.D.: A meta-analysis of the cognitive and motivational effects of serious games. Journal of Educational Psychology 105, 249–265 (2013)
15. Eagle, M., Barnes, T.: Evaluation of a game-based lab assignment. In Proceedings of the 4th International Conference on Foundations of Digital Games (FDG '09). ACM, New York, NY, 64-70 (2009)
16. Jackson, G.T., McNamara, D.S.: Motivation and performance in a game-based intelligent tutoring system. Journal of Educational Psychology 105, 1036–1049 (2013)
17. Snow, E.L., Allen, L.K., Jacovina, M.E., McNamara, D.S.: Does agency matter?: Exploring the impact of controlled behaviors within a game-based environment. Computers & Education 26, 378-392 (2014)
18. McNamara, D.S., Magliano, J.P.: Towards a comprehensive model of comprehension. In Ross, B. (ed.) The psychology of learning and motivation, vol. 51. Elsevier Science, New York (2009)
19. Jackson, G.T., Varner (Allen), L.K., Boonthum-Denecke, C., McNamara, D.S.: The Impact of individual differences on learning with an educational game and a traditional ITS. International Journal of Learning Technology 8, 315-336 (2013)
20. Jackson, G.T., Guess, R.H., McNamara, D.S.: Assessing cognitively complex strategy use in an untrained domain. Topics in Cognitive Science 2, 127–137 (2010)
21. McNamara, D.S., O'Reilly, T., Best, R., Ozuru, Y.: Improving adolescent students' reading comprehension with iSTART. Journal of Educational Computing Research 34, 147-171 (2006)
22. McNamara, D.S., Graesser, A.C., McCarthy, P., Cai, Z.: Automated evaluation of text and discourse with Coh-Metrix. Cambridge: Cambridge University Press (2014)
23. Boekaerts, M.: The on-line motivation questionnaire: A self-report instrument to assess students' context sensitivity. New Directions in Measures and Methods 12, 77–120 (2002)
24. Jacovina, M.E., Snow, E.L., Jackson, G.T., McNamara, D.S.: Game features and individual differences: Interactive effects on motivation and performance. In Mitrovic, A., Verdejo, F., Conati, C., Heffernan, N. (eds.) Proceedings of the 17th International Conference on Artificial Intelligence in Education. Madrid, Spain (2015)