iSTART Therefore I Understand: But Metacognitive Supports Did Not Enhance Comprehension Gains

Kathryn S. McCarthy[1], Matthew E. Jacovina[1], Erica L. Snow[2], Tricia A. Guerrero[1], & Danielle S. McNamara[1]

[1]Institute for the Science of Teaching and Learning,
Arizona State University, Tempe, AZ, USA
{ksmccar1, Matthew.Jacovina, Tricia.Guerrero, dsmcnamara}@asu.edu

[2]Center for Technology in Learning, SRI International, Menlo Park, CA, USA
Erica.Snow@sri.com

iSTART Therefore I Understand: But Metacognitive Supports Did Not Enhance Comprehension Gains

Kathryn S. McCarthy[1], Matthew E. Jacovina[1], Erica L. Snow[2], Tricia A. Guerrero[1], & Danielle S. McNamara[1]

[1]Institute for the Science of Teaching and Learning,
Arizona State University, Tempe, AZ, USA
{ksmccar1, Matthew.Jacovina, Tricia.Guerrero, dsmcnamara}@asu.edu

[2]Center for Technology in Learning, SRI International, Menlo Park, CA, USA
Erica.Snow@sri.com

Abstract. iSTART is an intelligent tutoring system designed to provide self-explanation instruction and practice to improve students' comprehension of complex, challenging text. This study examined the effects of extended game-based practice within the system as well as the effects of two metacognitive supports implemented within this practice. High school students (n = 234) were either assigned to an iSTART treatment condition or a control condition. Within the iSTART condition, students were assigned to a 2x2 design in which students provided *self-assessments* of their performance or were transferred to Coached Practice if their performance did not reach a certain *performance threshold*. Those receiving iSTART training produced higher self-explanation and inference-based comprehension scores. However, there were no direct effects of either metacognitive support on these learning outcomes.

Keywords. Intelligent Tutoring Systems • Reading Comprehension • Strategy-based Learning • Metacognition • Game-based practice

1 Introduction

Though reading is an everyday aspect of the classroom, students often struggle to successfully comprehend the kinds of informational texts they encounter in school [1]. Self-explaining, or explaining the meaning of a text to one's self, and self-explanation training have been shown to help students better comprehend these types of text [2]. The current work explores the effects of a self-explanation training system, iSTART, on two learning outcomes: self-explanation score and reading comprehension test performance. More specifically, it investigates the benefits of extended game-based practice and the effect of implementing two metacognitive supports within this practice.

1.1 iSTART

Interactive Strategy Training for Active Reading and Thinking

(iSTART) is an intelligent tutoring system (ITS) designed to improve students' reading comprehension abilities through self-explanation training [3]. iSTART teaches five self-explanation strategies (comprehension monitoring, paraphrasing, prediction, elaboration, and bridging) through video lessons, demonstration, and game-based practice [4]. During practice, iSTART uses a natural language processing algorithm to provide students with feedback on the quality of their self-explanations [5].

iSTART training encourages the generation of both bridging inferences that connect information from different parts of the text and elaborative inferences that connect information from the text to prior knowledge. The construction of these inferences supports the development of a more elaborate and coherent mental representation that is necessary for successful reading comprehension [6]. As such, iSTART has been shown to increase the quality of inferences during self-explanation [7] and increase comprehension for both high school and college students [8,9].

1.2 Game-based Practice

A common shortcoming of ITSs is that students may lose interest or motivation during the extended training and practice necessary to yield benefits. One way in which this issue has been addressed is through the addition of game-like components to increase engagement and motivation [10]. With this in mind, iSTART has been adapted to include game-based practice activities [11]. Practice includes both identification and generative games. In the *identification practice games*, students are presented with a self-explanation and must determine which strategy was used. For example, in the game *Balloon Bust*, the different strategies are presented on balloons that float around the screen. To earn points, the student must not only identify the correct strategy, but also "throw" the dart at the balloon that represents this strategy (Fig. 1).
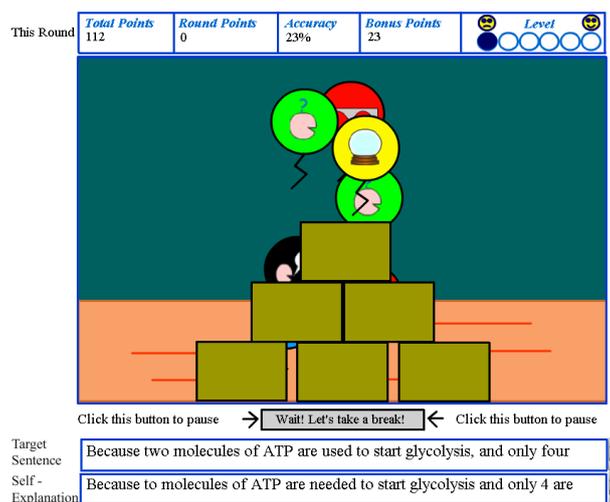

Fig. 1. Balloon Bust

More pertinent to the current study are the *generative practice games.* In these two games, *Map Conquest* and *Showdown,* students construct their own self-explanations. In *Showdown*, for example, the student competes against a computer opponent to construct the best self-explanation (Fig. 2).
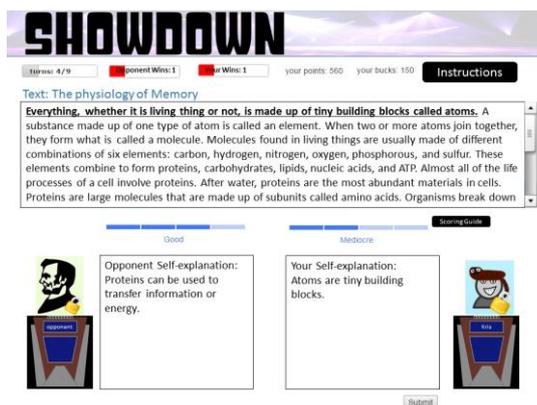
Fig. 2 Showdown

The game-based version of iSTART enhances enjoyment and motivation relative to a non-game-based version [12], and performance within game-based generative practice is correlated with posttest and transfer comprehension scores [13]. The current study furthers the investigation by examining learning outcomes (self-explanation score and comprehension question performance) for students receiving extended game-based practice as compared to students in a non-training control condition.

1.3   Metacognitive Features

Students generally have poor metacognition [14]. Metacognition refers to a person's ability to reflect upon their own knowledge as well as their understanding of a task or task goals. Metacognition can support successful comprehension as it regulates a learner's strategies and efforts [14]. Prompting metacognition has been shown to improve comprehension. Consequently, researchers have encouraged the inclusion of metacognitive supports to increase the efficacy of intelligent tutoring systems [15-17]. These metacognitive supports occur at the global level by prompting self-reflection after the task is completed, but they are also assumed to be beneficial at the local level such as when students are prompted to reflect on their performance during the task [18]. Instructional lessons within iSTART provide instruction on comprehension monitoring, encouraging students to recognize when they do or do not understand the text or parts of the texts; however, there is currently no explicit prompting within the system to encourage students to monitor their performance during self-explanation practice. With this in mind, we developed two interventions designed to support metacognition at both the local and global levels during extended game-based practice.

The first is a *performance threshold* that encourages self-reflection at the global level. Within each generative game, students write between 4 and 10 self-explanations. On each self-explanation, the participant receives a score (0-3). At the end of each game, the student's average self-explanation score is compared to an experimenter-set threshold (2). If this threshold is not met, the student is notified that the score is too low and is then transitioned to Coached Practice, in which a pedagogical agent provides explicit feedback to improve self-explanations. This threshold performance feature has been shown to increase average self-explanation score on the subsequent generative game [19].

The second feature, *self-assessment,* supports local metacognition as it asks the student to reflect on performance on each trial during the task. The self-assessment asks students to rate the quality of each of their self-explanations before receiving feedback from the system. Prior work with this feature indicates that students tend to overestimate their performance on self-explanations, though students with high prior knowledge tend to be more accurate [20].

This study is the first study to examine the effect of both metacognitive supports on posttest performance following extended practice in iSTART. Prior investigations with the performance threshold and self-assessment features have been limited to single sessions in which training was too brief to observe measurable posttest gains. Therefore, the focus of this study is on post-training learning outcomes.

1.3   Current Study

The current study investigated potential comprehension benefits from extended practice in iSTART self-explanation training. It also follows up on previous work exploring the implementation of two metacognitive support features: a performance threshold and a self-assessment rating during practice.

High school students were assigned to either a control condition with no training (n = 116) or an iSTART training condition (n = 118). Within the iSTART condition, we employed a 2 (performance threshold: off, on) x 2 (self-assessment: off, on) between-subjects design.

We compared the quality of participants' self-explanations at pretest and posttest and as well as their comprehension test question performance at pretest, posttest, and on a transfer task. We had two sets of predictions. The first set regarded the use of iSTART compared to the no training condition. We predicted that the extended practice in iSTART's game-based environment would yield improved self-explanation scores from pretest to posttest. We also predicted this practice would yield increased comprehension test performance on both the posttest and transfer test and that this benefit would be most evident for inference-based comprehension questions that assess deeper comprehension.

The second set of predictions pertained to the effects of the performance threshold and self-assessment features embedded within the iSTART training condition. Theories of metacognition generally state that as students gain more information concerning their performance during learning, they are better situated to adapt or change their future

learning behaviors and strategies [21]. Accordingly, it might be hypothesized that students exposed to both metacognitive supports would be best situated to adapt or change their behaviors and strategies, and subsequently show superior performance on the posttests. A second competing hypothesis might suggest that the two metacognitive prompts would be redundant. As such, when combined they would not provide unique insights for the student relative to having only one [22]. A third (null) hypothesis comes from skill acquisition theories [23,24] which place a greater emphasis on the development of the skills necessary to complete the task, rather than on explicit metacognitive interventions. Based on this hypothesis, there were be an overall effect of iSTART in comparison to the control condition, but no effects of metacognitive support conditions.

2 Method

2.1 Participants

Participants were 234 current high school students and recent high school graduates (147 female, 87 male) from the southwestern United States who were financially compensated for their participation in the study. They were, on average 15.90 years old (range 13-20). The sample was 48.7% Caucasian, 23.1% Hispanic, 10.7% African-American, 8.5% Asian, and 9.0% identified as other ethnicities.

2.2 Design and Materials

The study employed a 2(threshold: off, on) x 2(self-assessment: off, on) between-subjects design within those participants who received iSTART as well as a no training control, resulting in five treatment conditions: 1) threshold only (n = 28), 2) self-assessment only (n = 29), 3) threshold and self-assessment (n = 30), 4) neither threshold nor self-assessment (iSTART control, n = 31), and 5) no iSTART training (no training control, n = 116).

Performance Threshold. The performance threshold was designed to support global metacognition. After each self-explanation, participants receive a score of *poor, fair, good,* or *great*, which reflects a numeric score from 0 to 3. Lower scores (zero or one) indicate that the learner has produced an self-explanation that is too short to be of substance or is a restating or paraphrasing of the target sentence. Scores of two or higher reflect that the reader has demonstrated integration of prior knowledge into their response [11]. Given that inferencing and integrating is critical for successful comprehension, the performance threshold was set at 2. This threshold score is consistent with the previous implementation of this feature in iSTART. If the participant's average self-explanation score fell below this threshold at the end of a generative game, a pop-up message would appear (Fig. 3) and they were directed back to Coached Practice for remediation.

Fig 3. Performance threshold pop-up notification

Self-Assessment. The self-assessment feature was designed to encourage local meta-cognition. After each self-explanation, the participant was prompted to predict the quality of the self-explanation as *poor*, *fair*, *good*, or *great* (again reflected numerically as 0-3) and to rate their confidence in this prediction on the same scale. After making this selection, participants were given the actual self-explanation score.

Pretest and Posttest. The pretest and posttest consisted of two science texts: Red Blood Cells and Heart Disease. The presentation order of the texts as pretest or posttest was counterbalanced across participants. These texts were approximately 300 words and were matched for linguistic difficulty. In each text, participants were prompted to self-explain nine target sentences. After reading, they took a constructed response comprehension test that included four text-based and four bridging inference comprehension questions. Designed to assess more shallow comprehension, text-based questions have answers that can be found in a single sentence in the text. In contrast, bridging inference questions require the reader to connect information across two or more sentences in the text to derive the answer, indicative of deeper comprehension.

Transfer Test. The transfer test was designed to assess the extent to which students could apply the strategies they had learned to a new context. The transfer text, Plant Growth, was longer and more difficult than the Red Blood Cells and Heart Disease texts from the pretest and posttest. Importantly, participants were not prompted to self-explain while they read the transfer text. After reading, participants took another comprehension test that consisted of 10 text-based and 8 bridging inference comprehension questions about this text.

2.3   Procedure

Participants in the iSTART training conditions came into the lab for five sessions. In the first pretest session, participants completed the pretest, including the self-explanations and comprehension questions. Participants were also asked to give basic demographic information and to answer a battery of questions that included prior science knowledge. During the three days of training (three two-hour sessions), participants watched the iSTART video lessons that introduced them to the purpose of self-explanation and the five strategies. After the lessons, participants were transitioned to Coached Practice, a non-game-based activity in which students practice writing self-explanations and receive detailed, formative feedback. After one round of Coached Practice, the participants were allowed to move freely throughout the system to interact with videos, Coached Practice, and the generative and identification games for the remainder of the training sessions. It was during these three training sessions that participants in the performance threshold conditions were transitioned back to Coached Practice if they did not meet the performance threshold and participants in the self-assessment conditions were prompted to rate their self-explanation quality. In the final session, participants completed the posttests and transfer test.

Those in the no training control condition came into the lab for the pretest session and then returned to the lab after a few days ($M = 3.64$, $SD = .95$) to take the posttest.

2.3   Scoring

Using the same scoring algorithm employed within iSTART, each self-explanation in the pretest and posttest (nine in each) was automatically scored from 0-3. We then calculated an average self-explanation score for the entire text.

Three raters scored a subset of 20% of the constructed response comprehension questions for each text achieving high intra-class correlations for all three texts, (Red Blood Cells = .90, Heart Disease = .93, Plant Growth = .94). These raters then scored the remainder of the questions.

3   Results

Analyses of the posttest data indicated no effects of the threshold or self-assessment features on self-explanation scores, posttest comprehension scores, or transfer test comprehension scores (all $Fs < 1.00$, ns). Consequently, the following analyses compare all those who were provided iSTART training to those in the no training control.

3.1   Self-Explanations

We first compared pretest average self-explanation scores for those in the iSTART training condition to those in the control condition. Though the difference between the two conditions was not significant, $t(232) = 1.67$, ns, we conducted a two-level analysis of covariance (ANCOVA) which controlled for average self-explanation score at pretest. As shown in Table 1, this analysis indicated that those who received iSTART

training received higher average self-explanation scores than those in the no training control condition, $F(1, 231) = 29.78$, $p < .001$, $\eta^2_p = .11$.

Table 1. Means and standard deviations of self-explanation scores at pretest and posttest

	Self-Explanation Scores	
	Pretest	Posttest
iSTART (N=116)	2.30(.55)	2.43(.47)
Control (N=118)	2.18(.56)	2.02(.63)

3.2 Comprehension Tests

Though students were randomly assigned to condition, there was a significant difference in overall pretest comprehension score between the two conditions, $t(232) = 2.17$, $p < .05$. Consequently, we conducted a 2(treatment: iSTART, control) x 2(question type: text-based, bridging inference) ANCOVA that controlled for overall pretest comprehension score. This analysis indicated no main effect of treatment on posttest comprehension score, $F < 1.00$, *ns.* There was no main effect of question type, $F < 1.00$, nor was there a significant interaction, $F < 1.00$ (Table 2). Essentially, when the participants were instructed to self-explain, there was no effect of training on the immediate posttest.

To investigate the effect of iSTART treatment on the transfer comprehension test, we conducted a similar 2(treatment: iSTART, control) x 2(question type: text-based, bridging inference) ANCOVA that controlled for overall pretest comprehension score. This analysis revealed no main effect of treatment condition, $F < 1.00$, *ns,* but a significant main effect of question type, $F(1, 231) = 11.85$, $p < .01$, $\eta^2_p = .05$, such that students had higher average comprehension scores for the text-based questions than for the inference questions. This was qualified by a significant treatment by question type interaction, $F(1, 231) = 6.65$, $p < .01$, $\eta^2_p = .03$. As shown in Table 2, there was no effect of iSTART training on text-based comprehension question performance, but those who received iSTART training had a significantly higher average score on the inference comprehension questions than those who received no iSTART training.

Table 2. Means and standard deviations of comprehension test scores from pretest, posttest, and transfer test as a function of question type

| | Comprehension Scores | | | | | |
| | Pretest | | Posttest | | Transfer Test | |
	Text-based	Inference	Text-based	Inference	Text-based	Inference
iSTART (N=116)	.55(.27)	.58(.26)	.58(.38)	.58(.27)	.35(.20)	.27(.20)
Control (N=118)	.49(.31)	.50(.24)	.51(.31)	.55(.28)	.33(.19)	.21(.18)

Comparing iSTART training to a no-training control, iSTART increased the quality of participants' self-explanations at posttest, but had no effect on comprehension test performance. However, in a transfer task in which participants were not explicitly prompted to self-explain, those who had iSTART training yielded deeper comprehension as indicated by higher scores on bridging inference questions.

Consistent with the null hypothesis predicted by a skill acquisition account, there was no effect of either metacognitive support, in isolation or in tandem, on self-explanation score, posttest comprehension, or transfer test comprehension scores.

4   Conclusions

This study explored the benefits of extended game-based practice in iSTART on two posttest learning measures: self-explanation and comprehension. Additionally, it examined the effects of two metacognitive supports implemented within this extended practice.

Consistent with previous research, iSTART training improved high school students' self-explanation quality. Comprehension test scores indicated no effect of iSTART training on a comparable posttest text that prompted for self-explanations. This is a bit surprising, given previous demonstrations of comprehension gains using iSTART [25, 26]. Nonetheless, and perhaps more importantly, there were significant benefits of iSTART training on a more difficult transfer text in which participants were not prompted to self-explain. More specifically, iSTART training enhanced deep comprehension, as indicated by higher scores on bridging inference questions.

The implementation of the metacognitive supports in a 2 x 2 design allowed for the testing of three potential outcomes: an additive effect of having both supports prompting metacognition, an interactive effect in which having both supports would be no more beneficial than having only one support, and a null effect in which the metacognitive supports would show no benefits above and beyond the regular iSTART training. The findings of this study support this final hypothesis, as there were no effects of either the performance threshold or self-assessment. This suggests that the gains in self-explanation quality and comprehension performance are related to consistent practice, rather that the metacognitive interventions that we implemented in this study.

Given that the implementation of these metacognitive supports did not harm performance and have previously showed in-system benefits [19], it is worth continuing to implement their use and to conduct further investigations into how they affect training. One possibility is that the metacognitive supports have an indirect effect such that the performance threshold and self-assessment affect the way the readers interact with the system. Thus, we are investigating how these supports affected in-system performance and how these differences may in turn relate to posttest learning outcomes. For example, these features may increase motivation or enjoyment, which can encourage students to persevere and engage in the long-term practice needed to master complex reading comprehension skills [11, 12, 27]. We are currently analyzing the log data collected during training to explore how these supports affected interactions with the system, such as which games were played or time spent in off-task behaviors, and how differences in these interactions relate to self-explanation and comprehension gains. Importantly however, the manipulation of the metacognitive supports did not affect post-training performance, regardless of students' abilities or reported motivation. Hence, this study provides important information regarding the potential impact of these types of scaffolds, particularly in the context of intelligent tutoring systems that provide adaptive tutoring grounded in skill acquisition theories [28].

5 Acknowledgments

6 References

1. NAEP: The Nation's Report Card: Mathematics and Reading at Grade 12. Retrieved from https://www.nationsreportcard.gov/reading_math_g12_2015/#reading/acl#header (2015)
2. McNamara, D.S.: SERT: Self-explanation reading training. Discourse Processes 38, 1-30. (2004)
3. McNamara, D.S., Levinstein, I.B., & Boonthum, C.: iSTART: Interactive strategy trainer for active reading and thinking. Behavioral Research Methods, Instruments, & Computers 36, 222-233 (2004)
4. Snow, E.L., Jacovina, M.E., Jackson, G.T., & McNamara, D.S.: iSTART-2: A reading comprehension and strategy instruction tutor. In McNamara DS, Crossley SA (eds) Adaptive educational technologies for literacy instruction. Taylor & Francis, Routledge: NY, pp 104-121 (2016)
5. McNamara, D.S., Boonthum, C., Levinstein, I.B., & Millis, K.: Evaluating self-explanations in iSTART: Comparing word-based and LSA algorithms. In Landauer, T. McNamara, D.S., Dennis, S., Kintsch, W. (eds) Handbook of Latent Semantic Analysis. Mahwah, NJ: Erlbaum, pp 227-241 (2007)

6. Kintsch, W., Britton, B.K., Fletcher, C.R., Kintsch, E., Mannes, S.M., Nathan, M.J.: A comprehension-based approach to learning and understanding. Psychology of Learning and Motivation: Advances in Research and Theory 30: 165-214 (1993)

7. Kurby, C.A., Magliano, J.P., Dandotkar, S., Woehrle, J., Gilliam, S., McNamara, D.S.: Changing how students process and comprehend texts with computer-based self-explanation training. Journal of Educational Computing Research 47: 429-459 (2012)

8. Magliano, J.P., Todaro, S., Millis, K., Wiemer-Hastings, K., Kim, H.J., McNamara, D.S.: Changes in reading strategies as a function of reading training: A comparison of live and computerized training. Journal of Educational Computing Research 32: 185-208 (2005)

9. O'Reilly T., Sinclair, G.P., McNamara, D.S.: iSTART: A web-based reading strategy intervention that improves students' science comprehension. In Kinshuk Sampson, D.G. Isaías, P. (eds), Proceedings of the IADIS International Conference Cognition and Exploratory Learning in Digital Age: CELDA 2004. Lisbon, Portugal: IADIS Press, pp 173-180 (2004)

10. Dickey, M.D.: Engaging by design: How engagement strategies in popular computer and video games can inform instructional design. Educational Technology Research and Development 53: 67-83 (2005)

11. Jackson, G.T., McNamara, D.S.: Motivational impacts of a game-based intelligent tutoring system. In Murray, R.C., McCarthy, P.M. (eds), Proceedings of the 24th International Florida Artificial Intelligence Research Society (FLAIRS) Conference. Menlo Park, CA: AAAI Press, pp. 519-524 (2011)

12. Jackson, G.T., McNamara, D.S.: Motivation and performance in a game-based intelligent tutoring system. Journal of Educational Psychology 105: 1036-1049 (2013)

13. Jackson, G.T., Snow, E.L., Varner, L.K., McNamara, D.S.: Game performance as a measure of comprehension and skill transfer. In Boonthum-Denecke C., Youngblood, G.M. (eds), Proceedings of the 26th Annual Florida Artificial Intelligence Research Society (FLAIRS) Conference. Menlo Park, CA: The AAAI Press, pp 497-502 (2013)

14. Theide, K.W., Dunlosky, J.: Toward a general model of self-regulated study: an analysis of selection of items and self-paced study. Journal of Experimental Psychology: Learning, Memory, and Cognition 4: 1024-1037 (1999)

15. Azevedo, R.: Computer environments as metacognitive tools for enhancing learning. Educational Psychologist 40: 193-197 (2005)

16. Azevedo, R., Hadwin, A.F.: Scaffolding self-regulated learning and metacognition–Implications for the design of computer-based scaffolds. Instructional Science 33: 367-379 (2005)

17. Roll, I., Aleven, V., McLaren, B.M., Koedinger, K.R.: Improving students' help-seeking skills using metacognitive feedback in an Intelligent Tutoring System. Learning and Instruction 21: 267-280 (2011)

18. Schraw, G.: The effect of metacognitive knowledge on local and global monitoring. Contemporary Educational Psychology 19:143-154 (1994)

19. Snow, E.L., McNamara, D.S., Jacovina, M.E., Allen, L.K., Johnson, A.J., Perret, C.A., Dai, J., Jackson, G.T., Likens, A.D., Russell, D.G., Weston, J.L.: Promoting metacognitive awareness within a game-based intelligent tutoring system. In Mitrovic, A., Verdejo, F., Conati, C., Heffernan, N. (eds), Proceedings of the 17th International Conference on Artificial Intelligence in Education (AIED 2015). Madrid, Spain: Springer, pp 786-789 (2015)

20. Snow, E.L., Jacovina, M.E., McNamara, D.S.: Promoting metacognition within a game-based environment. In Mitrovic A, Verdejo F, Conati C, Heffernan N (eds), Proceedings of the 17th International Conference on Artificial Intelligence in Education (AIED 2015). Madrid, Spain: Springer, pp 864-867 (2015)

21. Flavell, J.H.: Metacognition and cognitive monitoring: A new areas of cognitive-developmental inquiry. American Psychologist 34: 906-911 (1979)

22. Roll, I., Aleven, V., McLaren, B.M., Koedinger, K.R.: Designing for metacognition – applying cognitive tutor principles to the tutoring of help seeking. Metacognition and Learning 2: 125-140 (2007)

23. Healy, A.F., Clawson, D.M., McNamara, D.S., Marmie, W.R., Schneider, V.I., Rickard, T.C., Curtcher, R.J., King, C., Ericsson, K.A., Bourne, L.E.: The long term retention of knowledge and skills. Psychology of Learning and Motivation – Advances in Research and Theory 30: 135-164 (1993)

24. Ericsson, K.A., Krampe, R.T., Tesch-Römer, C.: The role of deliberate practice in the acquisition of expert performance. *Psychological Review 100*: 363-406 (1993)

25. 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)

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

27. Jackson, G.T., Boonthum, C., McNamara, D.S.: iSTART-ME: Situating extended learning within a game-based environment. In Lane HC, Ogan A, Shute V (eds), *Proceedings of the Workshop on Intelligent Educational Games at the 14th Annual Conference on Artificial Intelligence in Education.* Brighton, UK: AIED, pp 59-68 (2009)

28. McNamara, D. S., Jacovina, M. E., Snow, E. L., & Allen, L. K.: From generating in the lab to tutoring systems in classrooms. American Journal of Psychology, 128, 159-172 (2015)