StairStepper: An Adaptive Remedial iSTART Module

Cecile A. Perret, Amy M. Johnson, Kathryn S. McCarthy, Tricia A. Guerrero, Jianmin Dai,
and Danielle S. McNamara
Institute for the Science of Teaching and Learning
Arizona State University, Tempe, AZ, 85287
{cecile.perret, amjohn43, ksmccar1, taguerre, jianmin.dai, dsmcnama}@asu.edu

Perret, C. A., Johnson, A. M., McCarthy, K. S., Guerrero, T. A., & McNamara, D. S. (2017). StairStepper: An adaptive remedial iSTART module. In. E. Andre, R. Baker, X. Hu, M. M. T. Rodrigo, & B. du Boulay (Eds.), *Proceedings of the 18th International Conference on Artificial Intelligence in Education* (pp. 557-560). Wuhan, China: Springer. Published with acknowledgment of federal support.

Author's Note

This research was supported in part by the Institute of Education Sciences (R305A130124) and the Office of Naval Research (N00014140343 and N000141712300). Any opinion, conclusion, or recommendation expressed are those of the authors and do not represent views of the IES or ONR.

StairStepper: An Adaptive Remedial iSTART Module

Cecile A. Perret, Amy M. Johnson, Kathryn S. McCarthy, Tricia A. Guerrero, Jianmin Dai, and Danielle S. McNamara

Institute for the Science of Teaching and Learning
Arizona State University, Tempe, AZ, 85287

{cecile.perret, amjohn43, ksmccar1, taguerre, jianmin.dai, dsmcnama}@asu.edu

Abstract. This paper introduces StairStepper, a new addition to Interactive Strategy Training for Active Reading and Thinking (iSTART), an intelligent tutoring system (ITS) that provides adaptive self-explanation training and practice. Whereas iSTART focuses on improving comprehension at levels geared toward answering challenging questions associated with complex texts, StairStepper focuses on improving learners' performance when reading grade-level expository texts. StairStepper is designed as a scaffolded practice activity wherein text difficulty level and task are adapted according to learners' performance. This offers a unique module that provides reading comprehension tutoring through a combination of self-explanation practice and answering of multiple-choice questions representative of those found in standardized tests.

Keywords: Intelligent tutoring systems • Game-based learning • Reading comprehension • Strategy based learning • Reading assessment • System adaptivity

1 Introduction

Reading comprehension is a difficult skill to master, yet it is an essential skill for success in school, careers, and daily life [1, 2]. Unfortunately, most students are not proficiently literate, with only 37% of 12th grade students at or above proficiency [3]. Literacy rates continue to be problematic in adulthood: recent data revealed that 40% of adults in the United States scored below functional literacy levels.1

A common way of assessing literacy is through standardized tests that rely on students reading passages and answering multiple-choice questions [4]. Extended practice is necessary to achieve success on these tests. Intelligent tutoring systems (ITS) provide extended practice with individualized automated, adaptive instruction and feedback that can supplement classroom instruction [5].

One example of such a program is Interactive Strategy Training for Active Reading and Thinking (iSTART), an ITS that provides reading comprehension strategy training and practice geared toward complex texts. Within iSTART learners self-explain texts and receive automated feedback on the quality of their self-explanation (SE). This training increases inferencing skills, which supports deeper text comprehension

[6, 7]. iSTART includes instructional scaffolds such as animated lesson videos, guided practice, and game-based practice [8]. Additionally, the system utilizes natural language processing (NLP) algorithms to assess the quality of the learners' SEs and to drive feedback.

StairStepper is a recent addition to iSTART, developed to provide an additional source of instruction for less-skilled readers. In particular, Stairstepper targets adult literacy learners whose reading proficiency is comparable to students at the 3rd to 8th grade levels. By contrast, iSTART involves reading texts that are appropriate for high school and college students. Hence, one objective was to include reading practice for these less-skilled readers. A second objective was to include practice with short passages and corresponding multiple-choice questions that mimic the types of reading tasks in standardized tests. While multiple-choice questions tend not to reveal deep text comprehension, they are tasks that students are expected to perform well, and are often used in high-stakes testing to estimate reading ability. StairStepper is not intended to be used in isolation, but rather to supplement the existing SE and comprehension strategy instruction already found in iSTART, such that students receive training and practice to develop both test-taking skills and deeper text comprehension.

2 StairStepper

StairStepper is a game-based practice module in which learners ascend a staircase of increasingly difficult texts. Students climb up the stairs when they answer questions about the text successfully, and go down the stairs when they do not. Students are also asked to self-explain texts when their performance drops. The goal is to reach the top step by successfully answering comprehension questions for the most difficult texts.

StairStepper adapts to learners' performance through scaffolded SE practice and by adjusting text difficulty depending on comprehension question scores. SE prompts are scaffolded (no prompt, prompt without feedback, prompt with feedback) based on students' multiple-choice comprehension question performance. This approach is intended to help learners increase their awareness of when self-explanation may benefit comprehension, and in turn, train learners to self-explain without being prompted.

The module comprises 162 expository texts, varying in difficulty (appropriate for grades 1-12), length (ranging from 7-80 sentences), and topic matter (history, sports, science, and life-relevant). This library was extracted from public websites[1,2] to afford extended practice using a diverse set of texts found in typical reading assessments. Four raters ranked the texts into levels, iteratively sorting subsets of the texts according to their relative difficulty. Raters repeated this procedure until they agreed on the texts' rankings. This process yielded 13 levels with at least five texts per level. Texts in the first level were discarded because they were too short to support comprehension questions, yielding a final set of 12 levels. These difficulty rankings correlated with Flesch-Kincaid Grade Level (r = .79) and with Dale-Chall readability (r = .77).

1 http://www.ereadingworksheets.com/e-reading-worksheets/online-reading-tests/

2 http://mrnussbaum.com/readingpassageindex/

Each text is followed by multiple-choice comprehension questions (4 – 20 per text) retrieved from the same sites as the texts. Questions that were ambiguous or not related to text material were removed. The questions were manually categorized as text-based (N=677; i.e., answers found within one sentence), bridging (N=160; inferences made across two or more sentences), or elaboration (N=144; inferences made between text information and prior knowledge). The proportion of each type of question was 35-45% text-based, 45-55% bridging, 8-10% elaboration for texts (n = 78) at level 8 and up. The lower levels were dominated by text-based questions (70-100%) and fewer bridging and elaboration questions (around 15% each). Therfore, only the more difficult texts (levels 8 and up), were piloted via Mechanical Turk. Participants (n = 259) read and answered the multiple-choice questions for a randomized subset of 20 texts. These data were used to further revise the question set. Questions with 0% accuracy (N = 6 of 818) were removed. Experimenters then correlated text performance with participants' average scores. Three of the 78 texts were removed that did not correlate with individuals' average score due to ceiling effects. Note that each difficulty level reflected text difficulty, not the difficulty of any given question. However, there were more deep comprehension questions (bridging and elaboration) for higher level texts.

There are four parameters that can be set by the experimenter: initial text difficulty level (default value = 5); threshold for question accuracy (default value = 0.75), use of SE score in determining scaffolding/difficulty level (default is off), and threshold for SE score (default value = 2.0). Each learner begins StairStepper by reading a text at the pre-set initial difficulty level (e.g., level 5), and then answers multiple-choice comprehension questions for that text. In order to ascend a stair, learners must receive a comprehension score at or above the question accuracy threshold; in our described scenario, the learner would next read a text at difficulty level six. If the comprehension score is below this threshold, the learner would be prompted to self-explain the subsequent text and then answer corresponding questions. If on this text, the learner still does not meet the comprehension threshold, the text difficulty level would decrease and the learner would self-explain the next text with feedback. Using an NLP algorithm, the system scores SEs from 0 to 3. Learners receive 3 points for SEs that successfully bridge information across multiple sentences in the text and with prior knowledge. The score is intended to offer feedback to students about their SEs so that they might improve as they moved forward through the module. This provides additional support for deeper text comprehension, thus facilitating question answering. Regardless of whether the learner is at the scaffolding level that provides feedback, they always see their score for each SE. This process allows learners to receive sustained remedial SE practice tailored to their specific needs.

3 Conclusion

The StairStepper module was developed to help improve reading comprehension skills of adult literacy learners and prepare learners for standardized assessments. Because standardized testing is commonly used to determine graduation from and

admittance to education programs, performance on the tests is crucial to success in school, careers, and daily life. Hence, StairStepper was developed to provide learners with practice using SE to support comprehension (when needed) and practice answering reading comprehension questions similar to those observed in standardized tests.

We have begun testing the benefits of using StairStepper as a form of extended and remedial practice. Next steps include development of an NLP algorithm to reliably estimate the difficulty of new texts using the StairStepper rating system, allowing researchers and educators to add texts and have them automatically assigned to the appropriate (relative) difficulty level. Such an algorithm is necessary because texts and question difficulty in various tests are each leveled according to different standards and methods, and thus one algorithm is necessary for this particular module.

Future work will examine the relations between SE skills and comprehension accuracy for adult literacy learners. This project contributes a necessary technology to address the dearth of computer-based literacy education for adult literacy learners.

Acknowledgments. This research was supported in part by the Institute of Education Sciences (R305A130124) and the Office of Naval Research (N00014140343 and N000141712300). Any opinion, conclusion, or recommendation expressed are those of the authors and do not represent views of the IES or ONR.

4 References

1. Geiser, S., Studley, R.: UC and the SAT: Predictive validity and differential impact of the SAT I and SAT II at the University of California. University of California, Office of the President (2001)
2. Powell, P. R.: Retention and writing instruction: Implications of deliberate practice for academic performance. Contemporary Educational Psychology, 30, 96-116 (2009)
3. 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)
4. NACAC: Report of the commission on the use of standardized tests in undergraduate admission. National Association for College Admission Counseling. Retrieved from https://eric.ed.gov/?q=ED502721&id=ED502721 (2008)
5. Crossley, S. A., McNamara, D. S. (eds): Adaptive educational technologies for literacy instruction. New York: Taylor & Francis, Routledge (2016)
6. McNamara, D. S.: Self-explanation and reading strategy training (SERT) improves low-knowledge students' science course performance. Discourse Processes (2015)
7. McNamara, D. S., O'Reilly, T. P., Best, R. M., Ozuru, Y.: Improving adolescent learners' reading comprehension with iSTART. Journal of Educational Computing Research, 34, 147-171 (2006)
8. Jackson, G. T., McNamara, D. S.: Motivation and performance in a game-based intelligent tutoring system. Journal of Educational Psychology, 105, 1036-1049 (2013)