Predicting Performance Behaviors during Question Generation in a Game-Like Intelligent Tutoring System

Carol Forsyth, Arthur C. Graesser, Borhan Samei, Breya Walker, and Philip Pavlik, Jr.
The University of Memphis, 202 Psychology, Memphis, TN 38152
Email: cmfrsyth @memphis.edu, graesser @memphis.edu, bsamei@memphis.edu, bswlker2@memphis.edu, ppavlik@memphis.edu

Abstract: The present research investigates learning constructs predicting performance behaviors during question generation in a serious game known as Operation ARA. In a between-subjects design, undergraduate students (*N=66*) completed the three teaching modules of the game, teaching the basic factual information, application of knowledge, and finally question generation about scientific research cases. Results suggest that constructs such as time-on-task, discrimination, and generation along with type of instruction (factual vs. applied) impact student behaviors during question generation.

Introduction

Research in the learning sciences suggests that generating good questions about difficult conceptualizations contributes to deep-level learning. However, students rarely generate deep and relevant questions (for a review see Graesser, Ozuru, & Sullins, 2009). The goal of the present research is to discover predictors of performance behaviors during question generation within a serious game known as Operation ARA (Millis et al. 2011, Halpern et al., 2012). Operation ARA teaches 11 topics of research methodology using natural language tutorial conversations in a game-like atmosphere. The system includes three separate modules, teaching the basic factual information (Cadet Training module), application (Proving Ground module), and question generation about research cases (Active Duty module). Previous research on Operation ARA has shown differences between as well as modules (i.e. Cadet Training vs. Proving Ground) within the game. Specifically, differences in learning gains across these three modules are correlated with performance on three time- honored cognitive constructs known as time-on- task, generation, and discrimination (Forsyth et al., 2012). Additionally, variations in types of learning (deep vs. shallow learning) has been shown between the two modules suggesting that the Cadet Training module teaching factual information correlates with shallow learning and the application module (Proving Ground module) teaches deep-level learning (Forsyth et al., 2013) Previous research in the learning and cognitive sciences has shown relationships between these three constructs and learning in various learning environments requiring memorization of facts as well as application of knowledge (Pashler et al., 2007; Cepeda et al., 2006; Graesser, Conley & Olney, 2012). In the current study, these three cognitive constructs as well as the effect the other modules teaching factual and applied information on research methodology (i.e. Cadet Training and Proving Ground) are used to predict performance during question generation in the Active Duty Module of Operation ARA.

Figure 1. The interactive training modules of OperationARA

Methods

In the current study, 66 undergraduate students interacted with Operation ARA in a between-subjects, counter-balanced pretest-intervention-posttest design. While all of the students completed the assessments, the interaction between learning modules varied depending on the condition. The four conditions included the following combinations of modules 1) Cadet Training and Active Duty, 2) Cadet Training, Proving Ground, and Active Duty, 3) Proving Ground and Active Duty, and 4) Active Duty only. The logged data from these

interactions were used to assess the effect of the modules and cognitive constructs on behaviors during the question generation module (The Active Duty Module).

Analyses and Results

In analyzing these logged data, we first correlated the performance on the three constructs within each module as well as the presence or absence of the other two modules (i.e., Cadet Training and Proving Ground) with the performance metrics in the Active Duty module. Next, we conducted a forward stepwise regression using the significant correlates as predictors on the three cognitive constructs within the Active Duty module. The results revealed that higher word generation ($t = 3.86$, $p < .001$) and the absence of the Proving Ground module (t (66) = -5.50, $p < .001$) significantly predicted higher generation during the Active Duty module (F (2, 64) = 15.21, $p < .001$, $R^2 = .32$). Alternately, it was the absence of the Proving Ground module (t (66) = -2.714, $p < .05$), less words generated (t (66) = -2.52, $p < .05$) and less time-on-task (t (66) = -2.53, $p < .01$) in the Cadet Training Module that significantly predicted higher time-on-task in the Active Duty Module ($F(3, 63) = 11.68$, $p < .002$, $R^2 = .33$). No statistically significant predictors were discovered for discrimination in the Active Duty Module.

Discussion and Future Work

The present work discovered statistically significant predictors for the performance behaviors for generation and time-on-task during question generation. These findings are significant to the learning sciences community because it is extremely important and unfortunately uncommon for students to generate good questions. These results may help researchers and educators encourage question generation behaviors in classrooms as well as artificial environments. For example, educators may devise tasks requiring students to contribute more thoughts and information while applying knowledge rather than while simply memorizing factual information to encourage students to generate more questions. With increased question generation, students may be able to obtain a deeper-level understanding of important concepts.

References

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T. & Rohrer, D. (2006). Distributed practice in verbal recall tasks: a review and quantitative synthesis. *Psychology Bulletin,* 132*,* 354-380.

Forsyth,C.M. Pavlik, P., Graesser,A.C. Cai,Z ,Germany,M, Millis, K., Butler,H., Halpern,D. and Dolan,R.(2012). Learning gains for core concepts in a serious game on scientific reasoning. In K.Yacef,O. Zaïane, H. Hershkovitz, M. Yudelson, and J. Stamper (Eds.) (2012).*Proceedings of the 5th International Conference on Educational Data Mining* (pp.172-175). Chania, Greece: International Educational Data Mining Society.

Forsyth, C.M., Graesser, A.C., Walker, B., Millis, K., Pavlik, P, & Halpern, D. (2013). Didactic galactic: Types of knowledge learned in a serious game. In H. C. Lane, K.Yacef, J.Mostow, & P. Pavlik (Eds.), *Proceedings of Artificial Intelligence in Education: 16th International Conference (AIED 2013)* (pp. 832-835). Berlin Heidelberg: Springer Verlag.

Graesser, A. C., Conley, M. W., & Olney, A. M. (2012). Intelligent tutoring systems. In S. Graham, & K. Harris (Eds.), *APA Educational Psychology Handbook: Vol. 3. Applications to Learning and Teaching* (pp. 451-473). Washington, DC: American Psychological Association.

Graesser, A. C., Ozuru, Y., & Sullins, J. (2009). What is a good question? In M. G. McKeown & L. Kucan (Eds.), *Threads of coherence in research on the development of reading ability* (112-141). New York: Guilford.

Halpern, D. F., Millis, K., Graesser, A. C., Butler, H., Forsyth, C., & Cai, Z. (2012). Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning. *Thinking Skills and Creativity*, *7*, 93-100.

Millis, K., Forsyth, C., Butler, H., Wallace, P., Graesser, A. C., & Halpern, D. (2011). Operation ARIES! A serious game for teaching scientific inquiry. In M. Ma, A. Oikonomou, & J. Lakhmi (Eds.), *Serious games and edutainment applications* (pp.169-196). London: Springer-Verlag.

Pashler, H., Bain, P. M., Bottge, B. A., Graesser, A., Koedinger, K., and McDaniel, M. (2007). Organizing Instruction and Study to Improve Student Learning: IES Practice guide. (NCER 2007-2004). Washington, DC: National Center for Education Research.

Acknowledgements

This research was supported by the Institute for Education Sciences (R305B070349, R305C120001) and National Science Foundation (REC 0106965, ITR 0325428, HCC 0834847). The opinions expressed are those of the authors and do not represent views of the IES and NSF.