Thinking Skills and Creativity 7 (2012) 93–100

Contents lists available at SciVerse ScienceDirect

Thinking Skills and Creativity

journal homepage: http://www.elsevier.com/locate/tsc

Operation ARA: A computerized learning game that teaches critical thinking and scientific reasoning

Diane F. Halpern[a,*], Keith Millis[b], Arthur C. Graesser[c], Heather Butler[d], Carol Forsyth[c], Zhiqiang Cai[c]

[a] Claremont McKenna College, Claremont, CA, USA
[b] Northern Illinois University, DeKalb, IL, USA
[c] University of Memphis, Memphis, TN, USA
[d] Claremont Graduate University, Claremont, CA, USA

ARTICLE INFO

Article history:
Received 14 February 2012
Received in revised form 22 March 2012
Accepted 27 March 2012
Available online 5 April 2012

Keywords:
Computerized learning
Critical thinking
Scientific reasoning
Science of learning
Learning games
Serious games

ABSTRACT

Operation ARA (Acquiring Research Acumen) is a computerized learning game that teaches critical thinking and scientific reasoning. It is a valuable learning tool that utilizes principles from the science of learning and serious computer games. Students learn the skills of scientific reasoning by engaging in interactive dialogs with avatars. They are tutored by avatars with tutoring sessions that vary depending on how well students have responded to questions about the material they are learning. Students also play a jeopardy-like game against a feisty avatar to identify flaws in research and then generate their own questions to determine the quality of different types of research. The research examples are taken from psychology, biology, and chemistry to help students transfer the thinking skills across domains of knowledge. Early results show encouraging learning gains.

1. Introduction

Fitness shoes with rounded soles and an orthopedic appearance promise "more shapely legs and a better butt" in addition to weight loss, reduced cellulite, and improved circulation (University of California Wellness Letter, 2012). Sound too good to be true? Not to consumers who spent $1.1 billion dollars on these shoes in 2010. (They have been on the market since 2000.) The United States Trade Commission investigated these unbelievable claims and because Reebok, one of the manufacturers, could not back up its claim that wearing these shoes would "result in more strength and tone in the buttocks muscles than traditional walking shoes," Reebok agreed to settle the case with the Trade Commission and refund $25 million to consumers (ElBoghdady, 2012).

Bogus claims about toning shoes may seem like a frivolous concern, but this example is one of countless similar scams that represents massive public disregard for the value of assessing claims, seeking evidence, and understanding scientific thinking. Millions of dollars are spent annually on products that are "packaged" as scientific, even though the "research" or evidence is highly suspect or absent. For example, millions of dollars are spent on homeopathic remedies that lack scientific evidence for their effectiveness. Scientific inquiry is important in the workplace, especially as jobs become increasingly technical, require specialized skills, and require domain-specific problem-solving, reasoning, and decision making (National

* Corresponding author at: Department of Psychology, Claremont McKenna College, 850 Columbia Avenue, Claremont, CA 91711, USA. Tel.: +1 909 607 9647; fax: +1 909 621 8419.
E-mail address: Diane.Halpern@cmc.edu (D.F. Halpern).

1871-1871/$ – see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.tsc.2012.03.006

Research Council, 2011). Often, the same techniques that are used to sell miraculous products are also used to create or enhance prejudice and hostilities toward other groups and to justify wars and policies that cause harm and sometimes, even death (Lilienfeld, Ammirati, & Landfield, 2009). Ideally, everyone should know how to critically evaluate the products of science and evidence-based claims, even when the topic seems distant from our usual notions of what constitutes science.

1.1. Need for scientific reasoning/critical thinking

We want to educate our students and the general public to be wise consumers of research. We also want them to apply the principles of scientific thinking in their daily interactions and in contexts that bear little resemblance to scientific studies. This is a common goal for critical thinking instruction. To achieve this goal, we conceptualize science as a process, as a way of thinking about and obtaining information that informs the conclusions that we make. For example, if our students were embroiled in an argument over who was more influential to hip-hop, James Brown or Stevie Wonder, we would want them to respond that the argument can be resolved by using what scientists call an operational definition. If someone complains about a colleague who is always late, we hope that they will ask about the sample size (how often was the colleague late?) that supports this conclusion. These are examples in which the principles of scientific thinking are important in daily life.

Of course, we hope that we are also educating future scientists, helping them to "think like a scientist." Scientific thinking is needed for everyone—future scientists and future consumers of science—because we are all exposed to causal claims made by scientists and nonscientists on a daily basis. These claims are made on the Internet, television, news reports, advertisements, and the wide range of social media. Does wine really lower the risk of heart attacks? Are the screening devices at airports really safe? Can I teach my toddler to read? Will copper bracelets cure arthritis? The list of claims is endless. Operation ARA (Acquiring Research Acumen) was designed to teach students how to determine what and whom to believe.

1.2. The best of E-learning

We took up the challenge of creating the best tools for learning by applying what we know about the science of learning and adapting those principles for a generation of "digital natives," a term that describes today's young adults who have never known a world without computers and the internet. They are more likely to get their information from the vast array of nonprint media available to them than from textbooks. According to a report assembled by Frontline (Public Broadcast System, 2010), digital natives aged 13–17 average 1742 text messages a month, 91% use their profiles on social media websites to stay in touch with friends, and they spend an average of 4.5 h a day viewing screen media (internet, television, etc.), excluding games. Not surprisingly, less time is spent learning from books, and more of today's students are finding it difficult to learn from printed text. We can lament the loss of textbook skills and design educational interventions that help students achieve better learning from books, *and* we can also provide high quality learning experiences based on how students actually learn, including computerized learning games. There are benefits to learning from computerized games, which include improvements in visual attention and faster response times (Dye, Green, & Bavelier, 2009).

We developed a computerized learning game, Operation ARA that incorporates the best findings from the science of learning. Mayer (2008) reminded educators that no matter how advanced the technology is for educational content, if the fundamental principles of learning are not incorporated into the design of the materials, then quality learning is unlikely to occur. Here is a sampling of some of the learning principles we used in developing this program (Halpern & Hakel, 2003).

1.2.1. Active engagement

Meaningful learning of complex material is NOT a spectator sport. Deep learning requires that the learner be actively engaged with the material. Operation ARA requires that learners demonstrate their learning consistently throughout the program. There are three modules in Operation ARA: *Basic Training* in which students use an interactive e-book to learn basic principles of scientific thinking, such as the need for control groups and the difference between correlational and causal research designs; *Proving Ground* where students play a jeopardy-like game against an avatar who assumes the role of a competing student; and *Active Duty* in which students ask questions in order to determine whether a reported study is reliable or flawed. Early in the program, students respond to questions, but by the end of the program they are asking questions. Unlike a standard book, students cannot merely move their eyes across text or wipe a highlighter across a page. They cannot move through the program without demonstrating their learning at each step.

1.2.2. Response generation

Recent research has shown that repeated testing, particularly when the test requires the generation of information, enhances learning. This practice produces gains in understanding and memory when the tests are aligned with important content (Karpicke & Roediger, 2007; Roediger, Agarwal, McDaniel, & McDermott, 2011). According to standard "memory trace" theories of how people remember, the act of remembering strengthens some memory traces and weakens, or perhaps fails to strengthen, others, a fact that should influence how we test students. There are complex mathematical models and functions that describe the course of remembering (or forgetting) over time (Anderson et al., 2004; Oberauer & Lewandowsky, 2011). When students practice retrieval, the strength of the memory trace is increased and the shape of the forgetting curve is altered to reflect the fact that the just-retrieved information is now more likely to be retrieved (i.e., less likely to be

forgotten) in the future. In more common language, this principle is sometimes called the "testing effect," which refers to the finding that by responding to a question, the learner is practicing retrieval, thus making later retrieval of the same material more likely to be successful. As students progress through Operation ARA their understanding of the concepts and principles of science are repeatedly tested. In *Basic Training*, students answer a number of multiple choice questions at the end of each chapter and then participate in trialog discussions with avatars where the understanding of the material from the chapter is clarified and reinforced. Students evaluate 11 cases in the *Proving Ground*, and for each case they must apply their understanding of scientific concepts to determine whether the research is reliable or flawed. In *Active Duty* students generate their own questions about abbreviated research descriptions in order to determine whether the research is flawed. In this module, students actively recall the scientific concepts in order to generate questions, with little scaffolding at this point in the game.

1.2.3. Spacing effect

Learning and retention occur over time, so it is not surprising that much of the research literature on these topics is concerned with time intervals—the time spent on the initial learning, the time between acquisition (initial learning) and recall, and the length of the time intervals between recall trials. One of the primary "laws" of learning is that spaced practice is generally superior to massed practice. Spaced schedules of studying and testing produce better long-term retention than single or few study sessions or tests (Karpicke & Bauernschmidt, 2011). Students work through Operation ARA over time and repeatedly practice and apply concepts in different contexts.

1.2.4. Individualized tutoring and reciprocal teaching

As students move through the interactive chapters during *Basic Training*, they receive computer-generated tutoring that varies depending on how well the student responds. The mixed-initiative dialogue used for tutoring is similar to Auto Tutor, which was designed by Art Graesser and numerous colleagues. Students using Auto Tutor have shown considerable learning gains comparable to one-on-one human tutoring (Graesser et al., 2004; VanLehn et al., 2007). The type of tutoring that students receive following each chapter is determined by the number of questions they answer correctly about the chapter they just read. Effective tutors need to both correctly gauge and adapt to the student's current level of understanding. A successful adaptive tutor chooses problems that specifically address the level of the student's prior knowledge and take previous test scores into consideration (Graesser, Conley, & Olney, 2012).

In compliance with these criteria, students are adaptively placed in one of three tutoring conditions based on their current level of knowledge, which is gauged by their scores on the previous multiple choice tests. If the human students demonstrate a low-level understanding of the concept, they receive the *vicarious learning* trialog in which they observe a virtual teacher tutoring a virtual student, with limited active participation. Vicarious learning conditions have shown significant learning gains specifically for low prior-knowledge students (Driscoll et al., 2003). In order to maintain engagement during vicarious learning, the students are asked to respond to questions about the tutoring situation. For example, the virtual teacher might ask the human student whether the virtual student understands the concept or whether the virtual student's answer was correct. If the learners demonstrate a moderate understanding of the concept, they receive the *standard tutoring trialog*, and they are tutored by the virtual teacher. For example, the virtual teacher might ask the human student to provide a definition of the concept and scaffold the student using hints, prompts, feedback, and misconception correction. If the learners demonstrate good understanding of the concept, they interact with the *teachable agent trialog*, and the human student tutors a virtual student. For example, virtual students might tell the human student that they do not really understand the concept and offer an incorrect explanation. The human student would then have to explain to the virtual students what the concept is and why they were incorrect.

1.2.5. Feedback as knowledge of results

Standard learning texts always emphasize the need for feedback or knowledge of results as an essential component for enhancing learning, but more recent reviews have shown that feedback is a complex variable that sometimes hinders learning and sometimes enhances it (Kluger & DeNisi, 1996, 1998; van der Kleij, Eggen, Timmers, & Veldkamp, 2012). When data from many studies were combined in a meta-analysis, the overall effect size for feedback was a healthy $d = .41$, but this sizable effect masks the fact that over 1/3 of the studies showed that feedback had negative effects on learning and performance. Feedback is important in that it provides information to the learner about his or her own performance, but the learner still has to derive meaning from it and it may be that the way learners interpret feedback is what determines when it will be beneficial. What does feedback really mean to a learner? Is successful learning indicative of the intelligence or skill of the learner, the difficulty of the learning task, the way learning performance is assessed, or the amount of effort expended? Learning settings need to be designed so that the type of feedback is matched to the intended reason for the learning task. Even a formerly simple concept like feedback needs to be understood in terms of its intended use so that learners do not interpret feedback as punishment, which leads to resistance to the information provided by the feedback.

Most research reviews conclude that feedback should be given at appropriate intervals, probably with increasing length of the intervals as learning progresses (Winstein & Schmidt, 1990). Thus, early in the learning process, more frequent feedback is needed than later in the learning process. In this way, learners can become judges of their own performance and rely less upon external knowledge of results than they would with more constant feedback. We used this principle in the design of

Operation ARA. As learners progress through the program, they receive feedback that is increasingly less frequent and less detailed.

1.2.6. Variability enhances transfer

Variability during learning has been shown to be important for improving long-term recall and enhancing knowledge transfer. The theme of variability during learning underscores the causal and sometimes contradictory relationship between what happens during learning and long-term retention. By using examples from different science domains—psychology, biology, and chemistry, students need to concentrate on the underlying scientific principles, which actually makes the learning more difficult. Learners need to expend more effort when learning conditions are variable, but the hard work of effortful learning pays large dividends when long-term retention is assessed.

Like practice at retrieval, varied learning conditions pay high dividends for the effort exerted. In the jargon of cognitive psychology, when learning occurs under varied conditions, key ideas have "multiple retrieval cues" and thus are more "available" in memory. For example, educational research suggests that significant learning gains can occur when different types of problems and solutions are mixed in the same lesson, even though the initial learning can take significantly longer (Perry, Samuelson, Malloy, & Schiffer, 2010).

1.3. Using a game format to enhance motivation and engagement

In order to learn well, students need to be motivated and engaged in the process. According to a report from the Bureau for Labor Statistics (2011), time spent playing video games varies greatly with age. On an average weekend day, 15–19-year-old Americans read an average of 6 min and play video games 1 h and 6 min. At the time of this writing, the game *Angry Birds* has been downloaded tens of millions of times. There are numerous international gaming competitions with entries from most countries in the world. Thus, if we could harness only a small percentage of the energy and enthusiasm spent on computerized games so that game playing resulted in learning, we would have made great strides in creating quality learning environments. Here are some of the properties of the best selling games that we applied to Operation ARA.

1.3.1. An intriguing story line

Just as Operation ARA utilizes the best of what we know about how people learn, it also borrows from the growing literature on serious games. In Operation ARA learning is embedded in an intriguing story line. The game begins when players join the fictitious Federal Bureau of Science where they learn about an evil extraterrestrial plot. Apparently, the extraterrestrials are publishing poorly designed research in a variety of media outlets with the goal of confusing humans as to what constitutes good science. However, the aliens are disguised as human beings. Therefore, the player is given the task of learning the scientific method so that he can help the FBS identify the extraterrestrials. Along the way, an avatar protagonist learns that his brother is being held captive by Fuaths from the planet Thoth and apparently has broken under the stress. To complicate matters, our protagonist finds that he is falling in love with a Fuath, thus blending traditional action and romance scripts. But, all is not what it seems. We later learn that the extraterrestrials are stealing our natural resources. The plot has surprising twists and ultimately it is up to the learner to save the world by identifying the aliens among us. There are double-agents, humor, romance, political intrigue, and more all taking place within a "green" theme. Additional descriptions about Operation ARA (formerly known as Operation ARIES!) are available in Butler, Forsythe, Halpern, Graesser and Millis (2010) and Millis et al. (2011).

1.3.2. Points and other incentives

As students progress through the program, they receive incentives for responding correctly, which include earning points in a competition against a feisty and somewhat irritating student avatar. Later in the game, the learner has to decide if particular research is reliable or flawed, and if they are correct, human researchers are set free and alien researchers are sent to jail. Errors can be disastrous because they would allow dangerous aliens to be free and innocent humans would go to jail. Ultimately, good decisions will save the world, which we believe to be a motivating outcome.

There are yet other psychological principles that are built into the game to optimize motivation and engagement. These include giving the student control through choices, presenting problems that are at the optimal zone of challenge (not too easy and not too difficult), high amounts of interactivity between the game and the student, feedback on their contributions, and engaging media. All of the design decisions in developing ARA were motivated by scientific principles to enhance cognition or motivation.

Of course, we hope that readers are looking for evidence that students actually learn the skills of scientific thinking by playing Operation ARA. Although we have been working on this learning game for several years, data collection is still in its early stage, with initial results looking positive. This paper presents some early findings. Results from several other preliminary studies of the effectiveness of Operation ARA are available from the first author upon request. As the number of students with different sorts of backgrounds who use Operation ARA increases, we expect to have many more studies completed over the next several years.

Table 1
Seventeen concepts taught in Operation ARA that were used in the pretest and posttest.

1. Theories and hypotheses
2. Science and pseudoscience
3. Operational definitions
4. Independent and dependent variables
5. Reliability, accuracy, and precision
6. Validity
7. Replication of results
8. Experimental control
9. Control groups
10. Random assignment to condition
11. Attrition and mortality
12. Samples are representative of populations
13. Sample size
14. Experimenter bias
15. Conflict of interest
16. Making causal claims
17. Generalizability

2. Study 1: knowledge gains from basic training

2.1. Participants

Students from three qualitatively different colleges and universities (17 community college students, 66 state university students, 53 private liberal arts college students) participated in the study. A majority of the sample was female (67.6% women, 32.4% men) and lower division students (47.1% freshman, 27.2% sophomore, 9.6% junior, 11.8% senior, 4.4% other). The sample was ethnically diverse, 34.6% Caucasian, 27.9% Asian, 17.6% Latino, 8.1% Bi-racial, 5.1% African-American, and 6.7% of the sample reported other ethnicities or declined to state their ethnicities. Approximately 87% of the sample had taken or were currently taking a course in Introduction to Psychology.

2.1.1. Materials

All students took a pretest and posttest that assessed their knowledge of research methods and scientific reasoning. Two versions of the tests were created and administration of the tests was counter-balanced. The test consisted of two case studies, 21 short answer questions, and 21 multiple choice questions. Proportional learning gains were computed for each pretest and posttest. The formula for computing the proportional learning gains was (posttest – pretest/1 – pretest).

Recall that the first module is taught with an interactive e-text in which students receive different types of automated tutoring based on their performance. We compared pretest and posttest scores on 17 different concepts that are taught in Operation ARA and compared the scores to a control group that was enrolled in college, but did not engage in Operation ARA training.

The seventeen concepts are listed below in Table 1.

2.1.2. Procedure

After taking the pretest, students either played Operation ARA ($n = 58$) or participated in a control condition that did not play the game ($n = 78$), and then all students took the posttest at approximately the same time during the semester of testing. Each concept was assessed in the pretest and posttest with a constructed response and a multiple choice question. There were two forms of the test, with half of all students taking Form A as a pretest and Form B as a posttest and the reverse for the other half of the students. Each item on the test was scored from 0 to 2, with 0 indicating an incorrect response, 1 and 1.5 indicating a partially correct response, and 2 indicating a correct response on both questions that correspond to each concept.

2.2. Results

Students who played Operation ARA had higher proportional learning gains ($M = .193$, SE = .031) than students who did not play the game ($M = -.101$, SE = .032), $F(1, 130) = 43.279$, $p = .001$, $d = 1.40$. There was no difference in proportional learning gains between the colleges, nor was there an interaction between type of college and whether or not they played the game; all the p-values were not statistically significant.

Mean scores on the pretest and posttest are shown in Fig. 1.

In addition to comparing proportional learning games for students who played Operation ARA with a control group, we also assessed the effectiveness of the different types of tutoring conditions.

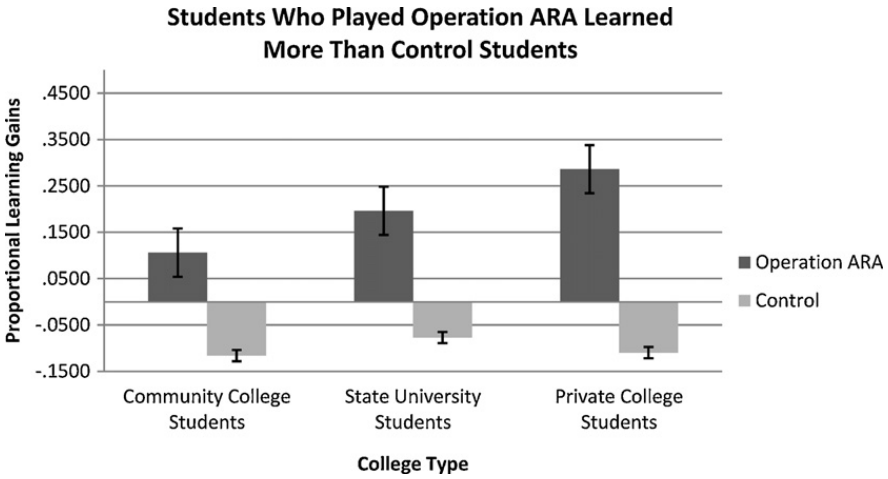

Fig. 1. Mean learning gains on 17 scientific reasoning concepts. Each concept was scored from 0 to 2, with 0 = incorrect answer, 1 and 1.5 = partially correct answer, and 2 = correct answer. Error bars represent standard errors.

3. Study 2: knowledge gains for different tutoring conditions

3.1. Participants

Two hundred fifteen students from a large Midwestern university were assigned at random to one of four different types of tutoring. Because we were interested in the effects of prior knowledge on learning from Operation ARA we used a tertiary split based on a pre-test to form "high" and "low" prior knowledge groups. The tutoring groups were: (a) a no-tutoring control group ($n = 32$), (b) a vicarious tutoring group in which the learner mostly watched an avatar student being tutored by an avatar teacher ($n = 28$), (c) an adaptive tutoring group in which the human learner was tutored by an avatar teacher ($n = 27$), and (d) a teaching tutoring group in which the human learner tutored the avatar student ($n = 29$). The sample was 57% female.

3.2. Procedure

Students responded to the multiple choice question to five chapters of Operation ARA in their own time at a laboratory at the university. When they responded to each set of questions about the main concepts, one of the four types of tutoring was initiated. Participants had the same type of tutoring for all of the concepts.

3.3. Results

Finally, to test for long-term learning gains, we compared the type of tutoring students received (using a no-tutoring control group) and tested them immediately on a short answer test after the tutoring sessions and also after a week delay. Percent correct was calculated for each participant. The pattern of means is shown in Fig. 2. They were submitted to a mixed ANOVA with trialog condition and prior knowledge as between-participants factors and topic and delay as

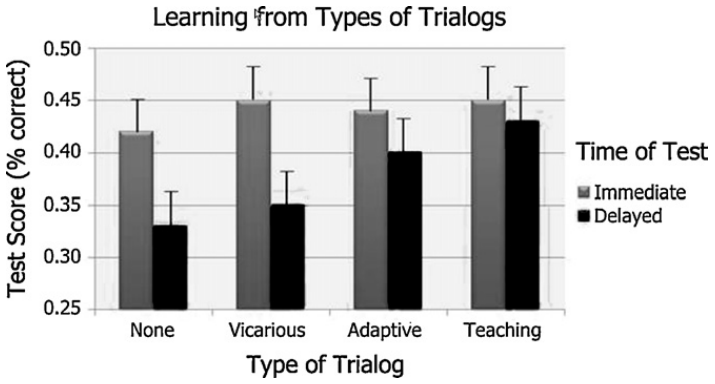

Fig. 2. Percentage of questions correct on a test of scientific reasoning principles for students who received one of four types of tutoring: a no-tutoring control group, a vicarious tutoring group in which the learner mostly watched an avatar student being tutored by an avatar teacher, an adaptive tutoring group in which the human learner was tutored by an avatar teacher, and a teaching tutoring group in which the human learner tutored the avatar student. Error bars are standard errors.

within-participants factors. Generally we found that the type and presence of tutoring had little impact when tested immediately. This makes sense because the material would be still "fresh" in the participants' memory. We had expected the amount of decay observed over the week to be significantly smaller for the tutoring conditions because of the active engagement. However, the delay by tutoring condition interaction was only marginally significant, $F(3, 108) = 1.99$, $p = .06$ (one-tailed). When we compared the tutoring conditions with the least amount of active engagement (vicarious) to the most (teaching), the delay by condition interaction was significant according to a two-tailed test, $F(1, 53) = 5.41$, $p < .05$. These findings indicate greatest durable learning for tutoring conditions that required active engagement, as expected from the learning principle of engaged learning. Although there was a large main effect of prior knowledge, $F(1, 108) = 49.20$, $p < .01$, and topic $F(4, 432) = 63.03$, $p < .01$, they did not interact with any of the other factors. Presumably, participants with differing levels of prior knowledge benefit equally from the trialogs embedded within the game.

4. General discussion

Although these results are encouraging, we recognize that they are not the sort of well-controlled studies that are needed to make strong claims that Operation ARA leads to superior learning of scientific reasoning skills. Multiple additional studies are currently underway and many more are planned for the next several years. Although our research with Operation ARA is still in its early stages, we have reasons to believe that the positive early results will be replicated in more comprehensive studies. For example, prior research with computerized tutoring programs similar to the one used here has shown impressive learning gains. Graesser and his colleagues have found an average effect size of $d = .8$ across 18 experiments (Graesser et al., 2004; VanLehn et al., 2007) and other intelligent tutoring systems report effect sizes of $d = 1.0$ (Anderson, Corbett, Koedinger, & Pelletier, 1995; Dodds & Fletcher, 2004; Graesser et al., 2012; VanLehn et al., 2007).

Thus far, data from Operation ARA show that students in open-admissions community colleges, state universities, and highly selective liberal arts colleges all make learning gains, thus suggesting that it is appropriate for multiple levels of student achievement. Adaptive tutoring sessions were designed so that students with varying levels of background knowledge would benefit from its use. Operation ARA can be accessed on the Internet with any computer with Internet connectivity. It does not assume any computer knowledge beyond what is needed to log onto a web site. Although not yet operable, it will have the capacity to read all on-screen materials for users with low vision and the ability to enhance the size of all fonts. Thus, we believe that it will be appropriate for a broad range of users.

We have many research-related questions about the use of Operation ARA. For example, other gaming materials have suffered from a lack of emotional connectivity with the avatars (D'Mello, Dale, & Graesser, in press). We have not yet assessed this aspect of Operation ARA, although the avatars were created to facilitate an emotional reaction, with each avatar having a distinct personality (e.g., the feisty component, the love interest between two of the avatars, and the feelings that main protagonist has for his brother who is a captured double-agent). Emotional aspects of Operation ARA are among a long list of future studies that are planned.

5. Conclusion

It is time that learning materials catch up with the kinds of information that the students in our classes, the digital natives, are already using. Operation ARA is a serious computerized learning game that teaches the principles of scientific reasoning. It is among the first to explicitly target important thinking skills using a game format. We believe that better thinking can be an educational outcome, and that a game format may be the best mode for enhancing long-term retention and transfer of these critical skills, although we realize that many unanswered questions still remain as to its effectiveness as a tool for learning scientific reasoning/critical thinking.

Acknowledgments

This research was supported by the Institute of Education Sciences of the U.S. Department of Education (R305B070349) to Northern Illinois University. The opinions expressed are those of the authors and do not represent views of the Institute or the U.S. Department of Education.

Operation ARA is being marketed in 2012 by Pearson Higher Education, which pays royalties to three of the authors and/or their university. The reported research was conducted on the IES research grant prior to the marketing of the game.

References

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. (2004). An integrated theory of the mind. *Psychological Review*, *111*(4), 1036–1060. http://dx.doi.org/10.1037/0033-295X.111.4.1036

Anderson, J. R., Corbett, A. T., Koedinger, K. R., & Pelletier, R. (1995). Cognitive tutors: Lessons learned. *Journal of the Learning Sciences*, *4*, 167–207. http://dx.doi.org/10.1207/s15327809jls0402_2

Butler, H. A., Forsyth, C., Halpern, D. F., Graesser, A. C., & Millis, K. (2010). Secret agents, alien spies, and a quest to save the world: Operation ARIES! Engages students in scientific reasoning and critical thinking. In R. L. Miller, R. F. Rycek, E. Amsel, B. Kowalski, B. Beins, K. Keith, & B. Peden (Eds.), *Promoting student engagement. Vol. 1: Programs, techniques and opportunities*. Syracuse, NY: Society for the Teaching of Psychology. Available from the STP web site:. http://teachpsych.org/ebooks/pse2011/vol1/index.php

Bureau of Labor Statistics. (2011, June 22). *American Time Use Survey Summary*. USDL-11-0919. http://www.bls.gov/news.release/atus.nr0.htm

D'Mello, S. K., Dale, R., & Graesser, A. C. (in press). Disequilibrium in the mind, disharmony in the body. *Cognition and Emotion*, 26, 362–374. http://dx.doi.org/10.1080/02699931.2011.575767.
Dodds, P. V. W., & Fletcher, J. D. (2004). Opportunities for new smart learning environments enabled by next generation web capabilities. *Journal of Education Multimedia and Hypermedia*, *13*(4), 391–404.
Driscoll, D., Craig, S. D., Gholson, B., Ventura, M., Hu, X., & Graesser, A. (2003). Vicarious learning: Effects of overhearing dialog and monolog-like discourse in a virtual tutoring session. *Journal of Educational Computing Research*, *29*, 431–450.
Dye, M. G., Green, C., & Bavelier, D. (2009). Increasing speed of processing with action video games. *Current Directions in Psychological Science*, *18*, 321–326. http://dx.doi.org/10.1111/j. 1467-8721.2009.01660.x
ElBoghdady, E. (2012). Reebok to refund $25M to customers who bought EasyTone, RunTone shoes. *The Washington Post with Bloomberg Business*. http://www.washingtonpost.com/realestate/reebok-to-refund-25m-to-customers-who-bought-easytone-runtone-shoes/2011/09/28/gIQATmUo4K_story.html
Graesser, A. C., Conley, M., & Olney, A. (2012). Intelligent tutoring systems. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA educational psychology handbook. Vol. 3. Applications to learning and teaching* (pp. 451–473). Washington, DC: American Psychological Association.
Graesser, A. C., Lu, S., Jackson, G. T., Mitchell, H., Ventura, M., Olney, A., et al. (2004). AutoTutor: A tutor with dialogue in natural language. *Behavioral Research Methods, Instruments, and Computers*, *36*, 180–193.
Halpern, D. F., & Hakel, M. D. (2003). Applying the science of learning to the university and beyond: Teaching for long-term retention and transfer. *Change*, (July–August), 2–13.
Karpicke, J. D., & Bauernschmidt, A. (2011). Spaced retrieval: Absolute spacing enhances learning regardless of relative spacing. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *37*(5), 1250–1257. http://dx.doi.org/10.1037/a0023436
Karpicke, J. D., & Roediger, H. L. (2007). Repeated retrieval during learning is the key to long-term retention. *Journal of Memory and Language*, *57*, 151–162. http://dx.doi.org/10.1016/j.jml.2006.09.004
Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin*, *119*(2), 254–284. http://dx.doi.org/10.1037/0033-2909.119.2.254
Kluger, A. N., & DeNisi, A. (1998). Feedback interventions: Toward the understanding of a double-edged sword. *Current Directions in Psychological Science*, *7*(3), 67–72. http://dx.doi.org/10.1111/1467-8721.ep10772989
Lilienfeld, S. O., Ammirati, R., & Landfeld, R. (2009). Giving debiasing away: Can psychological research on correcting cognitive errors promote human welfare? *Perspectives on Psychological Science*, *4*, 390–398. http://dx.doi.org/10.1111/j. 1745-6924.2009.01144.x
Mayer, R. E. (2008). Applying the science of learning: Evidence-based principles for the design of multimedia instruction. *American Psychologist*, *8*, 760–769. http://dx.doi.org/10.1037/0003-066X.63.8.760
Millis, K., Forsyth, C., Butler, H., Wallace, P., Graesser, A., & Halpern, D. F. (2011). Operation ARIES!: A serious game for teaching scientific inquiry. In M. Ma, A. Oikonomou, & L. Jain (Eds.), *Serious games and edutainment applications* (pp. 169–195). UK: Springer-Verlag.
National Research Council. (2011). *Assessing 21st century skills: Summary of a workshop*. Washington, DC: The National Academies.
Oberauer, K., & Lewandowsky, S. (2011). Modeling working memory: A computational implementation of the time-based resource-sharing theory. *Psychonomic Bulletin & Review*, *18*(1), 10–45. http://dx.doi.org/10.3758/s13423-010-0020-6
Perry, L. K., Samuelson, L. K., Malloy, L. M., & Schiffer, R. N. (2010). Learn locally, think globally: Exemplar variability supports higher-order generalization and word learning. *Psychological Science*, *21*(12), 1894–1902. http://dx.doi.org/10.1177/0956797610389189
Public Broadcast System. (2010, February 2). *Frontline*. Digital_nation: Life on the virtual frontier. http://www.pbs.org/wgbh/pages/frontline/digitalnation/extras/digital_native.html
Roediger, H., Agarwal, P. K., McDaniel, M. A., & McDermott, K. B. (2011). Test-enhanced learning in the classroom: Long-term improvements from quizzing. *Journal of Experimental Psychology: Applied*, *17*, 382–395. http://dx.doi.org/10.1037/a0026252
University of California Wellness Letter. (2012, January). Fitness shoes: Toning down the claims. 28, 6.
van der Kleij, F. M., Eggen, T. M., Timmers, C. F., & Veldkamp, B. P. (2012). Effects of feedback in a computer-based assessment for learning. *Computers & Education*, *58*(1), 263–272. http://dx.doi.org/10.1016/j.compedu.2011.07.020
VanLehn, K., Graesser, A. C., Jackson, G. T., Jordan, P., Olney, A., & Rose, C. P. (2007). When are tutorial dialogues more effective than reading? *Cognitive Science*, *31*, 3–62.
Winstein, C. J., & Schmidt, R. A. (1990). Reduced frequency of knowledge of results enhances motor skill learning. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *16*(4), 677–691. http://dx.doi.org/10.1037/0278-7393.16.4.677