NORTHERN ILLINOIS UNIVERSITY

The Effects of a Serious Game on Learning

A Thesis Submitted to the

University Honors Program

In Partial Fulfillment of the

Requirements of the Baccalaureate Degree

With University Honors

Department of

Psychology

By

Gary Timmnis

Dekalb, Illinois

Spring 2012

University Honors Program

Capstone Approval Page

Capstone Title: (print or type):

The Effects of a Serious Game on Learning.

Student Name (print or type): Gary Timmins

Faculty Supervisor (print or type): Keith Millis

Faculty Approval Signature:

Department of (print or type): Psychology

Date of Approval (print or type): 4-21-11

HONORS THESIS ABSTRACT

THESIS SUBMISSION FORM

AUTHOR: Gary Timmins

THESIS TITLE: The Effects of a Serious Game on Learning

ADVISOR: Keith Millis

ADVISOR'S DEPT: Psychology

DICIPLINE: Psychology

YEAR: 2011

PAGE LENGTH: 36

BIBLIOGRAPHY:

ILLUSTRATED:

PUBLISHED: No

LIST PUBLCIATION: N/A

COPIES AVAILABLE:

ABSTRACT: This study tested the hypothesis that video game attributes (e.g., points, competition, pedagogical agents) affect learning and engagement in a computerized learning environment. Thirty-two undergraduate psychology students were randomly assigned to two computerized learning environments meant to teach aspects of scientific inquiry (e.g., the need for control groups). Both groups read, critiqued and identified flaws in short descriptions of research. One group used a program that simulates a video game environment while the other used a traditional computer-assisted instructional format. Both conditions were given a pre-test and post-test. An interaction between time of test and the game was found, indicating that more learning occurred in the game condition than in the no-game condition. This suggests that games do increase learning, however there is no indication based on other results that they increase motivation and engagement.

The Effects of a Serious Game on Learning

Gary Wayne Timmins

Faculty Mentor: Keith Millis

Northern Illinois University

Abstract

This study tested the hypothesis that video game attributes (e.g., points, competition, pedagogical agents) affect learning and engagement in a computerized learning environment. Thirty-two undergraduate psychology students were randomly assigned to two computerized learning environments meant to teach aspects of scientific inquiry (e.g., the need for control groups). Both groups read, critiqued and identified flaws in short descriptions of research. One group used a program that simulates a video game environment while the other used a traditional computer-assisted instructional format. Both conditions were given a pre-test and post-test. An interaction between time of test and the game was found, indicating that more learning occurred in the game condition than in the no-game condition. This suggests that games do increase learning, however there is no indication based on other results that they increase motivation and engagement.

The Effects of a Serious Game on Learning

Computers and video games are changing the way instructors teach and how students learn. Digital methods of education are everywhere in the world in many forms, and they are not going away anytime soon. Flight schools use airplane simulator software such as *Flight Simulator 2000* to teach basic pilot procedures. Foreign language software, such as *Rosetta Stone,* teaches over 30 languages through a process of immersion. Even the United States Army has taken advantage of this educational revolution with the creation of the popular video game known as *America's Army* in which the player becomes a virtual solider and learns various tactical skills. According to the Entertainment Software Association (ESA), 67% of American households play computer and video games, and 27% of those Americans are under the age of 18 (ESA, 2010). Indeed, instruction is changing to accommodate the next generation of students.

Recent research has shown that educational video games are associated with increased learning. Educational video games are defined as: "applications using the characteristics of video and computer games to create engaging and immersive learning experiences for delivering specified learning goals, outcomes, and experiences" (de Freitas, 2006, p. 10). Some characteristics include the use of a story, interaction, pedagogical agents, and competition. A series of recent studies reported by Blunt (2007) examined the difference in academic achievement in students who did and did not use video games in their course work. They compared test performance of students attending courses that used a video game to students attending courses that did not. They examined performance in Introduction to Business and Technology course, a third year Economics course, and a third year Management course. As an example, the students in the

introductory business course played a game called *Industry Giant II*, in which they were put in charge of a corporation and had to deal with everything from the land that the company sits on to the financial management of the company. Blunt (2007) found that students using the games had significantly higher course test scores than students who did play the games. He also tested whether gender, ethnicity and age moderated the effect. Interestingly the only significant difference was age in which students 40 years and under scored significantly higher with the game play than students 41 and older, but there was no difference between the age groups in the traditional classes. In another study, Duque, Fung, Mallet, Posel, and Fleiszer (2008) tested the effects of a video game designed to teach geriatric procedures to medical students. Fourth year medical students played a video game that simulated a virtual house visit to a geriatric patient in which they would have to identify factors that would be considered important to a house call and could compromise the patient's safety. Students were given a pre- and post-test and the results were significant. Compared to a control condition that received traditional instruction, students who played the game demonstrated significantly greater improvement in their knowledge on geriatric procedures during a home visit. These studies indicate the usefulness of games in a variety of fields.

Instructors are continuously trying to engage students in the learning process. In one study, Coller and Shernoff (2009) examined levels of engagement during a game-based engineering course. An engineering class was taught with a video game designed to be both engaging and to teach authentic engineering problems. In this video game, students created mathematical formulas that would cause a simulated car to drive around a track. What they found was that students in this course experienced significantly more

intellectual intensity, intrinsic motivation, positive affect, and overall engagement in the learning process than comparable students in a traditional-taught course.

While there is evidence in favor of video games in education, there are some reasons to question whether they increase learning and engagement in all cases. One reason is that video games, by their very nature, induce distractions from the learning environment. These distractions might include activities that are irrelevant to the content. For example, if players are competing against each other, they might focus on the competition rather than on the to-be-learned content or skills. Another type of distraction is cognitively based. Students' working memory loads might be overloaded if the content is carried over the same medium, which is the same content is presented through two different modalities. For example, students who hear a text read at the same time that this same text and a video are displayed will show greater cognitive decrements compared to when the orally presented voice is absent (Mayer, 2001).

One important aspect to note here is that Coller and Shernoff (2009), Duque et al. (2008), and Blunt (2007) used quasi-experimental designs. In both of these studies, they did not randomly assign students to conditions, but rather relied on pre-existing classes. The problem with quasi-experimental designs is that they offer less internal validity then true experimental designs where random assignment to conditions is conducted. Therefore, goal of the current study was to compare gaming and nongaming environments on engagement and learning. Unlike the research summarized above, this study used a pre-post true experimental design. Below is a summary of characteristics that video games tend to share and how they were implemented into the study.

Characteristics of Video Games

It is impossible to give a single definition on what makes a video game. There is, however, a common set of characteristics video games tend to share. It is by all means unnecessary for a video game to contain every characteristic. Nonetheless, the more a video game has, the more game-like it is.

Story. A story is an account of a series of events. Action adventure, first person shooters, and strategy games all have some concept of story. Conversely, many genres, such as racing and sports games, are just not conducive to a story. That is not to say they would not benefit from having a story. Stories can provide a context for learning, and perhaps increase the motivation of the player. It immerses the player into the game further. Most video games by nature are repetitive. A well-developed story can help counter repetition by adding variety.

Interaction and Feedback. Players have the ability to influence the state of the game by just taking action, and, in return, the game adapts and provides feedback (Whitton, 2010). Interactions can be as simple as stating whether or not an answer is correct during a trivia game, or they can be highly complex as seen in Massive Multiplayer Online Role Playing Games (MMORPG), such as *World of Warcraft* in which millions of players interact not only with each other but with a virtual world in real time. Feedback is one type of interaction. Feedback can be corrective (was the user correct or incorrect?), formative (the game tries to elicit a longer and correct response), and summative (the game provides a summary of the player's performance). Whatever type of feedback is chosen, immediate feedback has been shown to be much more effective in producing learning gains than delayed feedback (Shute, 2008).

Pedagogical Agents. These are animated computer characters that are embodied with an artificial intelligence (Slater, 2000). They can look like human beings, animals, or inanimate objects (e.g., Microsoft's "Clippy"), and they can communicate with the player by talking, revealing text, pointing, and showing emotion. They respond to external stimuli such as keyboard input, mouse position, and mouse clicks. Their behavior is often dictated by having them adapt to a player's particular skill level. They can motivate by offering encouragement and giving feedback. They can engage by taking on whatever personality they are designed to. Furthermore, they may be constantly updated with the latest information about the player's performance, and therefore, be able to give timely feedback.

Points. These are non-tangible arbitrary numbers that can be awarded to a player to give a sense of accomplishment. They can become both positive and negative in conjunction with performance. Points are sometimes a vital component of competition, as a sense of measurement.

Competition. This occurs when the goal of the activity is beat other player(s). Some games involve competition in that they are competing against other players (both human and computerized) in real time. Other games involve competing against themselves (high scores). Competition has the ability to both discourage and motivate players depending on how it is applied.

Operation ARIES!

The proposed study will use a component of *Operation ARIES!*. ARIES is an acronym for Acquiring Research Investigative and Evaluative Skills. It is a computerized educational video game designed to teach scientific inquiry. It consists of three modules.

Throughout the first module, players read and are tested on online texts. In the course of the second module, players evaluate potentially flawed research articles, and during the final module, players learn question-asking skills (Millis et al., in press). However, for the purpose of this study, only the second module, called Case Studies, will be used.

Many of the attributes previously discussed that create a video game are present in the Case Studies module. It contains a story: the Federal Bureau of Science is a government agency that has been tasked to indentify and capture alien spies known as the Fuaths. They have been stealing Earth's resources and publishing faulty research in order to confuse humans about the scientific method. There is interaction - players receive feedback when answering questions both incorrectly and correctly through the use of pedagogical agents. Competition and points are used in the Case Studies module when the player competes against a computerized player named Tracy in identifying flaws in research. Points are subtracted when a player incorrectly indentifies a flaw and are awarded when a player correctly identifies a flaw.

The way in which *Operation ARIES!* manages this player interaction is based on the Intelligent Tutoring System known as AutoTutor. AutoTutor was designed to help students learn by communicating through a series of dialogues with a pedagogical agent. (Graesser, Person, & Harter, 2001) The pedagogical agent might ask a question and based on the learner's response, might give hints to point the learner in the right direction, or assess that the learner adequately knows the material and will move on. This is key to *Operation ARIES!*'s tutoring and immediate feedback function.

Overview of Study and Hypotheses

Participants learned about aspects of scientific inquiry by interacting with a game (Case Studies) or nongame computerized learning environment. Specifically, students learned about attributes of flawed research. These include the lack of a control condition, invalid dependent variables, subject bias, premature generalization, etc. (see Table 1 for a list of these flaws). In both conditions, students read short descriptions of research that contain one or more of these flaws. Each research description is called a case. Table 2 provides a sample case and the flaws that are present. In both conditions, students read the cases and wrote down flaws that they saw, and were provided feedback. The differences between the two conditions are that the game condition had pedagogical agents, competition, and points.

According to a Game Increases Learning Hypothesis, educational video games increase learning. The hypothesis assumes that a story, interaction, pedagogical agents, competition, and points increase learning because they should increase motivation and engagement. Furthermore, the game version has increased interaction with the user, in which the player sometimes must answer questions for them to identify and express a flaw. Dialog has been shown to increase learning because conversations require the speaker and listener to retrieve information, put ideas into language, formulate questions, and explain the material, most of which contribute to increased learning (VanLehn, Grasser, Jackson, Jordon, Olney & Rose, 2007). Therefore, this hypothesis predicts that students in the game condition will show greater learning than in the no-game condition partly based on the assumption that they report significantly greater engagement and motivation.

Method

Overview

All participants and procedures described in this investigation adhered to the guidelines set forth by the American Psychological Association and the research was approved by the local Institutional Review Board. Participants in the experiment were asked to critique a series of cases, half playing Case Studies and half interacting with a more traditional educational format. Learning gains were measured through the use of pre and post-tests.

Participants

The participants were 32 students from Northern Illinois University enrolled in a single section of a research methods class. Full completion of this study was 15% of their overall grade. They were randomly assigned into two groups. Half was assigned to the game environment condition, while the other half was assigned to the no-game condition.

Apparatus and Materials

Eight personal computers with separate keyboards, headphones, and monitors were used. The game condition consisted of the Case Studies module in *Operation ARIES!*. The non-game condition used E-Prime (2000). Each computer was kept individually in a separate 4x8 foot sound attenuated room.

Both, the Case Studies and the E-Prime version had a total of 18 cases. A case is a short description of research written in the form of newspapers articles, journal articles, and letters. Each case had between one and four flaws. See Table 2 for an example of a case.

A pre-test and a post-test were created to measure the participants' ability to use their knowledge of research methods to detect the presence or absence of flaws. Each test consisted of three flawed studies and was administered in the first and last sessions. Both the pre- and post-test contained one of each of the following content domains: psychology, biology and chemistry (see Appendix A). The instructions to the participant were to read each study, and to write down any flaws with the study or interpretation. There were two forms of the pre- and post-test, form A and form B. Each had three cases for the participant to identify and write down flaws. One-half of the participants received form A as the pre-test, and form B as the post-test. Hence, form was counterbalanced across participants. A scorecard was created to assess the cued response portion of the pre- and post-tests (see Appendix B).

Two questionnaires were created by the experimenter to assess engagement. One was given after the first computer session and was designed to measure the participant's interest and enjoyment in the tasks. The second was given on the last session and was designed to measure the participant's interest and enjoyment in the tasks as well as previous gaming experience and enjoyment. See Appendix C for the pre- and post-questionnaire.

Design

A 2 (Game: game vs. nongame) x 2 (Time of Test: pre vs. post-test) mixed factorial design was used. Game was the between-participants factor, whereas Time of Test was the within-participants factor.

Procedure

On the first day, participants were asked to read and sign an informed consent. They were given a pre-test booklet with three Case Studies enclosed. They then were asked to read and identify any flaws within the research design and list them in the space below, one case per page. Afterwards, they were given a grid sheet with all of the possible flaws for each case and asked to check whether any that they have identified as present. See Table 1 for the flaw list.

On the second day, the participants watched two five minute videos. The videos described aspects of good research and flaws found in poor research. Participants in the game condition also watched an additional video introducing them to the case study program. The purpose of the videos was to provide a basic knowledge of research flaws that was relevant to performance in both conditions.

Afterwards, both groups engaged in a computer program designed to provide experience with critiquing experiments in order to identify flaws in research. The control group received a standard computer presentation of instructions, cases to critique, and feedback all in black font on a white background using the E-Prime software (see Figure 1). After reading and performing each set of instructions they were prompted to press F12 to continue to the next page. A screen shot of the case study program is also presented in Figure 1. After selecting and reading a case, the player types in flaws they see in the research and are immediately given feedback after each. If they answer correctly they are given points and if they answer incorrectly they lose points. The player and the pedagogical agent known as Tracy take turns identifying flaws. Participants in both conditions both had three cases to complete during the first session. After the completion of all three cases they were given the pre-questionnaire.

During the third through fifth days, the participants engaged in tasks similar to that of the second day. However, instead of three cases, they were given five to complete.

On the final day, the participants were given a post-test booklet with three Case Studies enclosed and were asked again to read and identify any flaws within the research design and list them in the space below. Afterwards they were given a grid sheet with all of the possible flaws for each case and were asked to check the ones they have identified were present. Once the post-test was completed, they were given the post-questionnaire.

Results

The independent variables were the condition with two levels: game vs. no game, time with two levels: pre- vs. post-test, and test form with two levels: form A vs. form B. The dependent variables were the pre- and post-test results and survey results. The pre- and post-test results were split up into two types: free response and cued response (ie. score sheet). To score the free responses, the experimenter compared what the participant had written to the actual flaws present and made a judgment to whether or not they had correctly identified each flaw. To score the cued responses, the experimenter compared what the participant had checked on the score sheet to the list of actual flaws present. To score the questionnaire results, the mean score for each question was calculated and separated by condition and time.

Flaw Identification Scores (FIS)

For each participant, the average hit (correct flaw noted by the participant) and false alarm rate (incorrect flaw noted by the participant) were averaged over the three cases in both the pre- and post-test. An ideal participant would have a hit rate of 1.0 and

a false alarm rate of 0. A participant who could not discriminate between the presence and absence of flaws would have a hit rate of approximately 0.5 and a false alarm rate of 0.5. A difference score was computed for each participant, for both the pre- and post-tests, as defined by FIS (hit rate – false alarm rate difference). These scores will be called "flaw identification scores" in that the higher the score, the better a person can accurately identify a flaw in a case. The same procedure was done on the pre- and post-tests that used the cued responses.

Free Response

The flaw identification scores were submitted to a 2 (Game: game vs. no game), x 2 (Time: pre- vs. post-test) x 2 (Test form: form A vs. form B) mixed analysis of variance, with time as the within-participant factor. On the pre-test, the mean FIS for the game and no-game condition were 0.09 ($SD = 0.10$) and 0.07 ($SD = 0.13$). On the post-test, the mean FIS were 0.43 ($SD = 0.15$) and 0.32 ($SD = 0.15$). As predicted, the interaction between time of test and condition was significant, $F\ (1, 27) = 3.01$, $MSE = 0.012$, $p < 0.05$ (one-tailed). The pattern indicates more learning in the game condition than the no-game condition.

There was also a time by test form interaction, $F\ (1, 27) = 12.12$, $MSE = 0.012$, $p < .05$. The means indicate that form B led to less learning when it served as the pre-test ($M = 0.03$, $SD = 0.03$) than form A ($M = 0.13$, 0.04), yet contributed to greater learning when it served as the post-test ($M = 0.43$) than form A ($M = 0.33$). However, test form did not show a main effect, nor did it interact with game (p's < 0.90).

The above scores collapse over the content domain in the pre- and post-tests. In order to examine whether the game had equal impacts on learning to indentify flaws in

the different domains, a 3 (Content: psychology, chemistry, biology) x 2 (Time: pre- vs. post-test) x 2 (Condition: game vs. no-game) mixed ANOVA was conducted, with time and content as the within-participant factors. As before, the condition by time interaction was significant ($p < .05$, one-tailed). However, there were no main effects involving Content. Therefore, the game increased learning for all three contents to about the same extent.

Cued Response

For the scored flaws the flaw identification scores were also submitted to a 2 (Game: game vs. no game) x 2 (Time: pre- vs. post-test) x 2 (Test form: form A first vs. form B first) mixed analysis of variance, with time as the within-participant factor. The average pre-test and post-test FIS were 0.25 and 0.42, respectively. On the pre-test, the mean FIS for the game and no-game condition were 0.34 ($SD = 0.20$) and 0.16 ($SD = 0.19$). On the post-test, the mean FIS were 0.44 ($SD = 0.14$) and 0.40 ($SD = 0.18$). Although this resulted in a significant interaction, $F(1, 28) = 5.30$, MSE = .014, $p < .05$, one can see that there was a large difference between the two conditions on the pre-test scores (.34 vs. .16). This is difference was statistically significant $t(30) = 2.54$, $p < .05$. Because of the differences between the conditions on the pre-test, it is difficult to interpret these results.

Survey Results

Table 3 contains the means of the survey for the game and no-game condition. Recall that levels of engagement, motivation, interest, challenge, frustration, and choice were gathered twice, once after the first session, and again after the last. These measures were submitted to a 2 (Game: game vs. no-game) x 2 (Time: pre- vs. post-test) mixed

analysis of variance, with time as the within-participant factor. There were no significant effects for engagement, enjoyment, motivation, challenge and choice. However, significant effects emerged for frustration and interest. For frustration, the game produced ($M = 3.00$, $SD = 1.18$) higher scores than the nongame version ($M = 1.96$, $SD = .78$), $F(1, 30) = 8.44$, $MSE = 2.06$, $p < .05$.. For Interest, the pattern was reversed, with the game condition ($M = 3.25$, $SD = 1.04$) resulting in higher scores than the non game condition ($M = 2.65$ $SD = .72$), $F(1, 30) = 3.47$, $MSE = 1.624$, $p < .05$ (one-tailed). T-tests were used to compare the measures that were collected only after the last session. These included measures of perceived learning, past gaming experience, gaming enjoyment, and recommendation, which showed no significant differences.

Discussion

The purpose of the study was to investigate the effectiveness of computerized games on learning and engagement. When students were asked to freely identify flaws, the students in the game-condition showed a significantly higher improvement in learning than students in the no-game condition. This supports the hypothesis that educational based video games increase learning. However, when students were asked to identify flaws with the aid of a score sheet, there was only a slight increase in learning from the game-condition, but it was not a significant difference. One possible reason is that there was an abnormally high pre-test score in the control condition, which there is no explanation for at this time. Another possibility is that recognition is easier then free-recall, which might have attenuated the differences. At least on the basis of the free response data, the findings support Blunt (2007) and Duque et al. (2008) in that educational video games improve learning.

Somewhat surprisingly, there was no difference between the conditions on motivation, enjoyment, engagement, choice and challenge. This was indeed surprising, especially given that games should be enjoyable and fun to play. Apparently, the case study module is not that enjoyable to play, at least in its present form. These findings did not support Coller and Shernoff (2009) who report educational video games increase engagement and motivation. However, there are some important differences between that and the present study. Their invention had different goals, that is, to create a game using engineering problems as opposed to playing a game. The way that feedback was given was also different between the studies. In Coller and Shernoff (2009) students received feedback by having a virtual car crashing into a track, whereas in the Case Studies module students received corrective and immediate feedback as well as the addition or subtraction of points. Lastly, the overall context was different between the studies. The Coller and Shernoff (2009) study allowed students to receive live instructional support as opposed to students in the current study that did not.

The survey results suggest that the game condition had an increase of both frustration and interest. When asked in an informal focus group held in the research methods course a few weeks afterwards, several students expressed several reasons for the increased frustrations in the game condition. One was that the voices used for the pedagogical agents sounded too "computerized". The voices resembled something similar to that of *Microsoft Sam*, perhaps causing a distraction to the main task. The text speed of the computerized opponent also frustrated students. They felt that a lot of the time was wasted on waiting for the text to appear on the screen, which simulated someone typing slowly. Another possible reason, however, which was not reported by

students, was that frustration arose due to the feedback based on their input. When a student typed in an incorrect flaw, the computer told them that they were incorrect and they were docked points. This could be seen as very frustrating, as typically people do not like being told that they are wrong. Another source of frustration may have risen if the game could not match their input to a flaw category. For example, if a person typed in "participants were impartial" instead of "subject bias" for the case in Table 2, they would receive "No Match!" and lose 30 points. The matching was based on a keyword match. The participant would feel frustrated if he or she felt that the computer did not understand the input.

The implications of the study show that when designed correctly, computerized games can be used for instructional purposes and have a positive effect on learning. It also suggests that engagement is not necessary for learning. Participants in the game condition showed to be as engaged in the task then participants in the control condition, yet the game condition showed a significant increase on learning.

A potential limitation of the study was that the students that participated were psychology students already enrolled in a research methods class. This would definitely be a confounding variable for the main effect of learning across the pre- and post-test. According to the instructor, however, the material taught during the study did not include the tested flaws. Of course, there are no assurances that students did not seek out the information presented in the learning environments from the course textbook. However, the class material could not account for the interaction because participants were all from the same class and randomly assigned to the conditions. Another limitation was that with the low number of participants in each condition, these results can be nothing short of

preliminary. This would be especially true for the null results found on the survey data. Perhaps a larger sample size would produce significant results. In addition, future studies should focus on how well the Case Studies module improves the ability of non-psychology students to identify flaws in research. It would also be interesting to see the difference between learning gains in upper level high school students as opposed to college students.

It should also be noted that participants in the game condition did not go through the training course in *Operation ARIES!*. They only received requisite knowledge through a brief video. Additional training on how to identify flaws in research, which is an intricate part of the game, could have influenced the results further. Another important note is that the game condition did not have a story line integrated into it. As stated in the Introduction, a story plays a vital role in making games more compelling and engaging.

In summary, although there was not a significant difference in engagement between the two conditions, students in the game condition did learn significantly more than students in the non-game condition. This not only shows that educational games can increase learning, but that engagement, at least in some cases is not necessary for learning. With an increasing amount of today's youth growing up in an ever so increasing computer age, it is essential that the educational field take advantage of these advances in technology. Additionally, because students are increasingly more technologically savvy, educational games give them the opportunity to learn in a more controlled, individualized, and innovative manner.

References

Blunt, R. (2007). Does Game-Based Learning Work? Results from Three Recent Studies. *Advanced Distributed Learning*, 1-11.

Coller, B. D., & Shernoff D. J. (2009). Video Game-Based Education in Mechanical Engineering: A Look at Student Engagement. *International Journal of Engineering Education*, *25*, 308-317.

Duque, G., Fung, S., Mallet, M., Posel, N., & Fleiszer, D. (2008). Learning While Having Fun: The Use of Video Gaming to Teach Geriatric House Calls to Medical Students. *Journal of the American Geriatrics Society, 56*, 1328-1332.

Entertainment Software Association. (2010). Essential Facts About the Computer and Video Game Industry.

de Freitas, S. I. (2006). Using games and simulations for supporting learning. *Learning Media and Technology*, 31(4) 383-58.

Graesser, A. C., Person, N., Harter, D., & the Tutoring Research Group (2001). Teaching Tactics and Dialog in AutoTutor. *International Journal of Artificial Intelligence in Education, 12,* 257-279.

Mayer, R. E. (2001). Multimedia learning. New York: Cambridge University Press.

Millis, K., Forsyth, C., Butler, H., Wallace, P., Graeser, A., & Halpern, D. (in press). Operation ARIES!: A Serious Game for Teaching Scientific Inquiry.

Shute, V. J. (2008). Focus on Formative Feedback *Review of Educational Research, 78*,153-189.

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.

Whitton, N. (2010). Learning with Digital Games: A Practical Guide to Engaging Students in Higher Education. New York: N.Y: Routledge.

Table 1

List of Flaws

Poor or missing comparison group
No random assignment
Dependent variable could be more sensitive, accurate, or precise
Dependent variable is not scored objectively
Dependent variable is not valid
Subject bias
Morality or attrition
Small sample size
Experimenter bias
Premature generalization of results
Confuse correlation with causation

Note. The following list displays all of the possible flaws that will be used in the Case Studies.

Table 2

Example Case Study

Can A Pill Really Enhance Your Memory?

No one can remember everything. Many people forget things everyday- where they put their keys, scheduling a doctor's appointment, calling a friend back. But, maybe researchers are one step closer to helping people to solve their memory problems with a new pill.

Dr. Andrew Phillips, a neurologist from Bluegrass State University, thinks there is definitely a place in today's society for such a pill. He says that he believes "a memory pill could be very effective in helping people remember activities and plans they make on a daily basis." He goes on to say that if a memory pill is shown to be effective, "there would be widespread interest from the public and a huge market for the pill."

In a recent study, Phillips tested the effectiveness of one such memory pill. He recruited 49 adults who responded to an advertisement asking for volunteers who would like to improve their memory. The participants initially completed an interview as a health check. All interviews were conducted by research assistants who were unaware of the goals of the study. During the interview, participants were given 3 lists of 20 common, well known objects to remember. They were immediately asked to recall the lists, and they were given feedback on each item as to whether or not it was correct. Their scores were recorded.

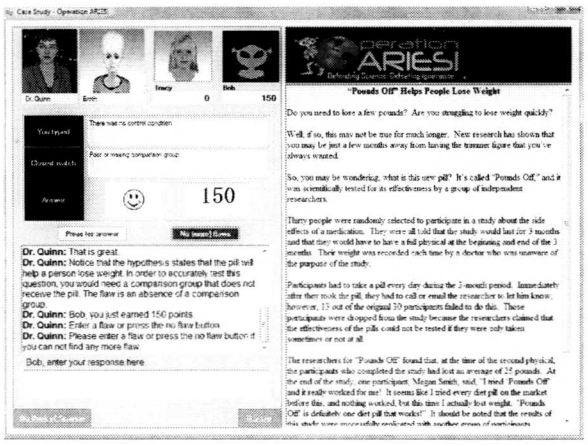
Although they were not made aware of the type of pill they received, only the 24 lowest scoring participants (those who seemed to need the most help) were given the memory pill once a day for 3 weeks. They 25 higher scorers received an identical looking placebo (sugar pill) and were given the same instructions. To ensure that they really did take the pill each day, all participants were required to receive the pills directly from the researcher everyday.

After 3 weeks, the participants returned to the laboratory to complete another interview similar to the first one. Again, they were asked to recall 3 lists of 20 items each.

When Phillips calculated the scores on the recall tests for pre- and post treatments, he found the participants who received the memory pill recalled significantly more items in the second interview than they did in the first, but the placebo group showed no such improvements.

Of the results, Dr Phillips said the following: "These findings are beyond exciting! It seems that the memory pill was effective in improving participants' memory capabilities. I think we need to get this pill out on the market right away!"

Note. The following is an example of a case. Participants will read and analyze each case and identify any flaw in the research they see. The subsequent flaws are present: no random assignment, subject bias, and premature generalization of results.



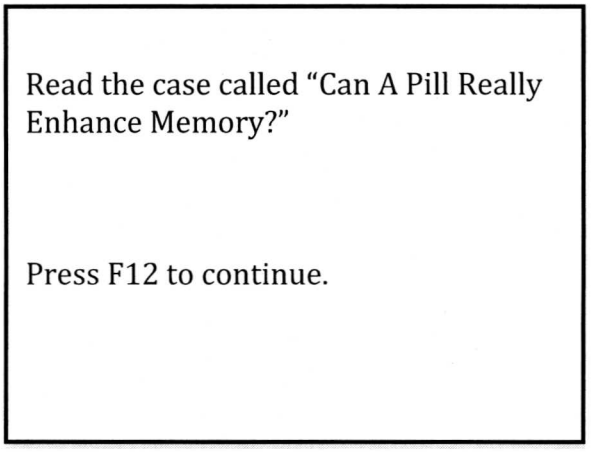


Figure 1. Participants who are in the game condition played the Case Studies module of *Operation ARIES!* (left), participants in the nongame condition used the computerized version programmed in E-Prime (right).

Table 3

Survey Results

	Game				No-Game			
	Pre-Test		Post-Test		Pre-Test		Post-Test	
Measure	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Engagement	3.56	0.89	3.56	0.96	3.62	0.72	3.56	0.89
Enjoyment	3.00	1.36	2.75	1.13	2.70	0.68	2.50	0.63
Motivation	3.63	1.15	4.00	0.89	3.75	0.93	3.69	0.70
Frustration	3.13	1.26	2.88	1.36	1.94	0.77	2.00	0.97
Interest	3.12	1.31	3.38	1.02	2.50	0.82	2.81	0.83
Challenge	3.43	0.89	3.19	0.66	3.13	0.96	3.19	0.83
Choice	2.69	1.14	2.50	0.81	2.75	1.13	2.63	1.36
Game Experience	---	---	3.00	1.31	---	---	2.75	1.23
Game Enjoyment	---	---	3.25	1.18	---	---	2.87	1.20
Learned	---	---	4.00	0.51	---	---	3.89	0.72
Recommended	---	---	3.69	1.14	---	---	3.50	1.14

Appendix A

Smoke and Lung Development

Has your child experienced asthma, decreased lung function, or respiratory tract function? One culprit may be second-hand smoke. As many as 50% of children under the age of five are exposed to second hand smoke. According to a recent poll, 22% of parents are concerned about the ill effects of second hand smoke on their children.

And because of a recent study, we should all become concerned.

Dr. William Kidd, who heads the for-profit organization "Center for Protecting Our Kids from Smoke," recently led an experiment to study the contributions of second hand tobacco smoke to lung diseases.

"We were interested in verifying the deadly effects of smoke exposure on lung health," explained Dr. Kidd. "Our research team decided to expose a group of mice to environmental tobacco smoke and then examine the level of exercise intensity that the animals engaged in while running on a running wheel every day for one week."

At the start of the study, three adult mice were exposed to tobacco smoke for one hour per day, five days a week, for a total of three weeks. Three younger mice were placed in the control group. These mice were not exposed to the smoke. "I did not want to expose the younger mice to the second hand smoke because that would be unethical to hurt their developing lungs," explained Kidd. "The younger mice were fine for the control group though."

After this three week period, a running wheel was placed in each animal's home cage. Based on several hourly observations per day, the exercise intensity for each animal was determined by Dr. Kidd and his research assistants. Kidd found a significant difference in the exercise intensities of the two groups while running in the wheel.

"The group exposed to environmental tobacco smoke engaged in significantly less intense running in the running wheel than the control group," said Kidd in a recent telephone interview.

Kidd concluded that second hand smoke does have harmful effects on lung health, and he is planning to replicate the findings later this year.

Can A Pill Really Enhance Your Memory??

No one can remember everything. Many people forget things everyday—where they put their keys, scheduling a doctor's appointment, calling a friend back. But, maybe researchers are one step closer to helping people to solve their memory problems with a new memory pill.

Dr. Andrew Phillips, a neurologist from Bluegrass State University, thinks there is definitely a place in today's society for such a pill. He says that he believes "a memory pill could be very effective in helping people remember activities and plans they make on a daily basis." He goes on to say that if a memory pill is shown to be effective, "there would be widespread interest from the public and a huge market for the pill."

In a recent study, Phillips tested the effectiveness of one such memory pill. He recruited 49 adults who responded to an advertisement for a memory enhancement study. The participants initially completed an interview as a health check. All interviews were conducted by research assistants who were unaware of the goals of the study. During the interview, participants were given 3 lists of 20 common, well-known objects to remember. They were immediately asked to recall the lists, and their scores were recorded.

Although they were not made aware of the type of pill they received, only the 24 lowest scoring participants (those who seemed to need the most help) were given the memory pill once a day for 3 weeks. The 25 higher scorers received an identical looking placebo (sugar pill) and were given the same instructions. To ensure that they really did take the pill each day, all participants were required to receive the pills directly from the researcher every day.

After 3 weeks, the participants returned to the laboratory to complete another interview similar to the first one. Again, they were asked to recall 3 lists of 20 items each.

When Phillips calculated the scores on the recall tests for pre- and post-treatment, he found that the participants who received the memory pill recalled significantly more items in the second interview than they did in the first, but the placebo group showed no such improvements.

Of the results, Dr. Phillips said the following: "These findings are beyond exciting! It seems that the memory pill was effective in improving participants' memory capabilities. I think we need to get this pill out on the market right away!"

Attack of the Plastic

Bisphenol A (BPA) is a compound found in many common plastics, such as the plastics that come in contact with food. In the U.S., the current acceptable daily exposure limit for BPA in humans is 50μg/kg. This is because higher doses have been linked to negative outcomes for users. For years, scientists believed that BPA was tied to increases in hormones, such as estrogen and testosterone; however, fairly recent research has called this into question.

In an attempt to clarify the connection between BPA and hormones, a consumer protection organization investigated the effects of BPA exposure on testosterone levels. Dr. Anne Maygill, conducted the experiment. Maygill was not affiliated with the plastics industry or the consumer protection organization. She was also blind to conditions while conducting the experiment and analyzing the data.

Maygill studied the testosterone levels of forty rats that were exposed to BPA-containing materials that had been left in their cages. Rats were exposed to the BPA-containing plastics for various amounts of time, ranging from a day to months. Five rats had actually died because of high exposure. Testosterone level was defined on a 3-point scale: high, medium and low, based on blood samples.

The resulting data showed no correlation between testosterone levels and the amount of BPA exposure. On the basis of this finding, Maygill asserts that BPA does not cause increases or decreases in testosterone; however, follow-up studies are underway to examine this issue further.

Save the Produce!

Dear Editor,

Here is a little useful information for you and your readers.

It seems that consumers are told to eat sufficient servings of fresh fruits and vegetables... but shopping for produce is getting more and more complicated. Distinctions are made between organic produce, non-organic yet not genetically modified produce, genetically modified food, locally grown versus imported food, etc. Unless you grow your produce in your own backyard, you have likely felt frustrated by these confusing choices.

Now we can add one more new kid on the produce block: irradiation, also known as ionized radiation. It has become common practice for food companies to irradiate produce, as well as various other food products. Irradiation creates free radicals that destroy the microorganisms in foods that bring about food spoilage. In other words, it slows the rate at which the food on the shelves in stores goes bad. Untreated produce goes bad quite quickly; irradiated produce can sit on the shelves for a week or even longer and still look freshly picked.

Although there are benefits to using this technique, one potential problem is that the free radicals created during radiation combine with chemicals in food which generate various compounds, some of which are toxic, and can lead to cell death.

Until now, no systematic studies have looked into the long-term effects of such by-products on consumers' health, but the potential harm has been sufficient to upset health enthusiasts. To examine the potential side-effects, a team of chemists fed one of two batches of corn to two different colonies of corn-fed only laboratory mice (something like 60 mice in total) for two weeks. The corn was brought in fresh daily from two different farms…one that irradiates their produce and another that doesn't. The irradiated corn was fed to one of the mice colonies (about half the mice), and the non-irradiated corn was fed to the other colony.

After two weeks, the researchers compared blood samples from the mice fed with irradiated corn to that of the other mice to see if their immune systems were different. Of course they only compared blood samples from mice that were still alive after two weeks. I guess 15 mice from the irradiated group and 2 from the non-irradiated group died. Anyway, after applying each blood sample to a small glass tray, a trained specialist used a microscope to assess the amount of iron in the blood, which is needed for transportation of oxygen in the blood.

The researcher found that mice in the non-irradiated group had significantly more iron in their blood. The results suggest that irradiated produce may pose a consumer health hazard at least for mice; however, the researchers are beginning a second study to verify this finding. So, you just may want to be a little more careful when you're out picking up fruits and veggies on the way home tonight.

Happy and safe shopping!

Let's Dance

Stephanie Webber, a dance instructor at a local college, is always looking for new ways to help dance students improve their techniques. One way she thought to do this was to show a dancing video that she strongly believed would help improve dance ability. The video showed dance performances from the popular TV show "Dancing with the Stars."

Before using the video with her own class, she decided to test out its effectiveness with randomly chosen shoppers at a local mall. She solicited shoppers by asking if they would be interested in improving their dancing skills.

When her study began, she measured each of the participant's dancing ability using a test in which she had them dance alone while holding a broom. All dancers were videotaped, and their movements were coded using an objective scoring technique that has been validated in a number of previous studies. Webber, who has been trained in the technique, did the scoring herself.

Ten participants at the mall completed the broom test at their own pace and then watched the video. After they watched the video, they took the test again, and once again, they were videotaped and the tapes were scored.

Webber found that the participants' dancing ability was significantly better after they watched the video than it had been before. In order to be certain that the video was truly effective, she repeated the study with another 10 dance students and found the same result. Webber feels confident that watching the dancing video does improve dancing skills and plans to use the findings in an advertisement for her dance studio.

Bacteria with a bite!

Bacteria are not usually thought of as being intelligent, but they can adapt to hostile environments. And, this is what makes some bacteria absolutely terrifying.

Normally dangerous bacteria are commonly treated with antibiotics. However, history shows that over time, bacteria strands develop resistance to the antibiotics designed to destroy them. Simply put: antibiotics are becoming wimpier with the passing of time.

In an effort to curb this problem, Dr. Renee Smiles headed a team of researchers to test the rate at which bacteria would develop resistance to a new antibiotic that is being developing. Their goal was to determine whether this new antibiotic would show a slowing down of bacterial resistance.

Several small samples of a lab culture of bacteria were collected from patients who reported influenza symptoms. The samples were placed on an agar medium (a substance that helps bacteria grow) in separate, small Petri dishes labeled according to how long the sample would be left for observation. Pads containing the test antibiotic were placed on the bacteria colonies in the different dishes.

After various numbers of reproductive cycles ranging from three to ten, the coloring of the area where the bacteria had been present on the pads in each dish was scored by a researcher. To do this, Dr. Smiles used a microscope to determine a coloring score based on a scale of 0 (color free) to 2 (high color saturation). Based on research on the particular strain of bacteria, the color test is a valid assessment of growth in which no color indicates an absence of bacteria.

Smiles found that there was a significant relationship between the length of time that the sample was left in the dish before being scored and the resistance of the bacteria colonies as indicated by high color scores. That is, the color scores for the bacteria were higher for those bacteria that were left the longest.

As such, Dr. Smiles announced that "the length of exposure to the bacteria does not affect the usefulness of the new antibiotic early on in an illness. But, if the illness persists beyond several bacterial reproductive cycles, the antibiotic would not be effective." Because of this finding, the drug company has determined that the antibiotic should be pulled from further testing and not made available to the medical community for use.

Appendix B

Flaws	Lung Development	Memory Enhancement	Plastic
Poor or missing comparison group	____	____	____
No random assignment	____	____	____
DV could be more sensitive, accurate, or precise	____	____	____
DV is not scored objectively	____	____	____
DV is not valid	____	____	____
Subject bias	____	____	____
Mortality or attrition	____	____	____
Small sample size	____	____	____
Poor sample selection	____	____	____
Experimenter bias	____	____	____
Premature generalization of results	____	____	____
Confuse correlation with causation	____	____	____

Flaws	Save the Produce	Lets Dance	Bacteria
Poor or missing comparison group	____	____	____
No random assignment	____	____	____
DV could be more sensitive, accurate, or precise	____	____	____
DV is not scored objectively	____	____	____
DV is not valid	____	____	____
Subject bias	____	____	____
Mortality or attrition	____	____	____
Small sample size	____	____	____
Poor sample selection	____	____	____
Experimenter bias	____	____	____
Premature generalization of results	____	____	____
Confuse correlation with causation	____	____	____

Appendix C

Pre-Questionnaire

Z-ID_____

1) How engaged were you during the tasks?

Not engaged Engaged

1 2 3 4 5

2) How much did you enjoy what you were doing?

Not engaged Engaged

1 2 3 4 5

3) How motivated were you in answering correctly?

Not engaged Engaged

1 2 3 4 5

4) How frustrated were you?

Not engaged Engaged

1 2 3 4 5

5) How interesting was this to you?

Not engaged Engaged

1 2 3 4 5

6) How challenging was it?

Not engaged Engaged

1 2 3 4 5

7) How much choice did you have in what you were doing?

Not engaged Engaged

1 2 3 4 5

Post-Questionnaire

Z-ID_____

1) How engaged were you during the tasks?

Not engaged Engaged

1 2 3 4 5

2) How much did you enjoy what you were doing?

Not engaged Engaged

1 2 3 4 5

3) How motivated were you in answering correctly?

Not engaged Engaged

1 2 3 4 5

4) How frustrated were you?

Not engaged Engaged

1 2 3 4 5

5) How interesting was this to you?

Not engaged Engaged

1 2 3 4 5

6) How challenging was it?

Not engaged Engaged

1 2 3 4 5

7) How much choice did you have in what you were doing?

Not engaged Engaged

1 2 3 4 5

8) Rate your gaming experience.

Not engaged Engaged

1 2 3 4 5

9) How much do you enjoy playing video games?

Not engaged Engaged

1 2 3 4 5

10) How much did you learn?

Not engaged Engaged

1 2 3 4 5

11) Would you recommend this as a homework activity?

Not engaged Engaged

1 2 3 4 5

12) Any suggestions?