Jl. of Educational Multimedia and Hypermedia (2005) **14**(1), 47-59

Effect of a Socratic Animated Agent on Student Performance in a Computer-Simulated Disassembly Process

ROBERTO PEREZ AND HOWARD SOLOMON
Florida State University
USA
rgp6722@mailer.fsu.edu
hms3683@garnet.acns.fsu.edu

The purpose of this study was to examine the effect of a Socratic animated agent on user performance in a computer-based CD player disassembly simulation. Vocational Education students were randomly assigned to one of two versions of the simulation. One version offered text-based feedback in the form of statements; the other one featured an animated agent providing hints and clues. Results showed that performance (measured by number of wrong tool choices) did not significantly differ between groups. Similarly, both groups stopped making wrong tool choices at approximately the same point in the disassembly process. The short exposure to the treatment may account for these results. Suggested modifications that could be incorporated in future studies are proposed in light of this factor.

Computer-based instruction (CBI) has significantly evolved in recent years, moving from an initial, basic drill-and-practice approach, to complex applications that incorporate simulations and animations. One such application is animated agents, which have been claimed to allow for a more natural, human-like interactivity between user and computer (Baylor, 2002; Johnson, Rickel, & Lester, 2000).

Animated agents are characters, or conversational interfaces, that mediate between the user and the computer program, to offer different forms of assistance, guidance, and feedback in the use of that program. These agents range from animated shapes (e.g., the Clippit desktop assistant in Microsoft Windows 95 and beyond) to human-like entities, including cartoon-like shapes and characters, talking animals, and a variety of others.

A number of studies have been carried out on animated agents as used in CBI, to learn about their effect on users. Dehn and Van Mulken (2000), in their review of the most salient experimental studies conducted on the use of animated agents, compared existing literature by classifying studies according to the kind of agents used (e.g., physical appearance, voice, etc.), the variables examined (both on the cognitive and the affective domains), and the results obtained. This comparison of current literature shows that agents can be engaging (Koda & Maes, 1996), but also create an uncomfortable feeling (Sproull, Subramani, Kiesler, Walker, & Waters, 1996); they can be helpful (Van Mulken, André, & Müller, 1998), but to some extent distracting (Koda & Maes); and users can interact with them (Takeuchi & Naito, 1995), or they can ignore agents completely (Lester, Stone, Converse, Kahler, & Barlow, 1997). Dehn and Van Mulken concluded that, given the contradictory results shown by current literature, it appears that the effectiveness of animated agents is directly affected by the combined elements of physical features of the agent, target audience, domain in which agent is used, and type of task to perform.

Among the studies described by current literature, only a few of them focus on how the pedagogic style of the agent may affect user performance and attitude. In one of those studies, Baylor (2002) compared two agents that advised students through the development of an instructional plan. One of the agents represented a constructivist approach (emphasized learner-centeredness, process over product, etc.), and the other displayed an instructivist style (followed a traditional, systematic method). Students who worked with the constructivist agent reported a significant change in their perspectives about instructional planning, and reflected an underlying constructivist pedagogy in their plans. Students who worked with the instructivist agent reported a lower disposition towards instructional planning. Other studies have examined different agent personal characteristics (which affect their perceived pedagogical style). Hietala and Niemirepo (1998), for example, compared four agents with different genders and different skill levels. Users tended to prefer one or the other depending on their own personality type (introvert or extrovert), as well as the difficulty of the task. In view of this research, the original Dehn and Van Mulken's list of combined elements that affect agent effectiveness could be expanded to include the agent's pedagogical style. This component has been shown to play an important role in determining agent effectiveness, and it should be taken into consideration when examining the interaction between animated agent and user.

One pedagogical approach that has proven effective in face-to-face instruction is generally referred to as Socratic Questioning. This approach, in its original form, is based on conversation-oriented questioning, inquiry, exploration, and discovery process, with a high degree of interaction and involvement between the learner and the Socratic questioner. A number of variations of this method can be used to maintain a rich, probing, intellectually challenging discussion, where "discussion" is more loosely defined and operationalized: instead of deep, involved, sometimes intricate interactions like the ones that may take place in a human-human Socratic Questioning approach, agent-human "discussions" could be more structured and controlled, while still challenging students intellectually.

Stevens & Collins (1977) laid the foundations for the use of Socratic tutoring in CBI environments by establishing one of the first structures for computer-mediated Socratic questioning. They developed the WHY system, a Socratic tutor that used a number of dialogue structures to help students develop reasoning skills such as formulating and testing hypotheses. This work led to advanced intelligent systems such as the CirCSIM-Tutor by Zhou et al. (1999). This text-based system used a number of hinting strategies in a conversation-based approach to guide users through the right path of reasoning.

Still, imperfect as intelligent systems may currently be, the implementation of a Socratic system within a context of controlled, predictable, expected interactions is worth studying. This system would capture the essence of the Socratic approach, guiding the user toward successful performance through the use of hints delivered in the form of questions.

The very nature of a Socratic approach, which is the conversational interaction between two human beings, can be created to a certain extent by making use of a conversational interface, that is, an animated agent. This agent would help counteract the unnatural feeling of simulating a conversation in a CBI environment, given that the computer is far from being perceived as having "human-like" characteristics. The choice of a Socratic-like questioning approach as the pedagogical style of an animated agent gives rise to a relationship that seems to make technology and pedagogy fit together well.

The purpose of the present study was to examine the effect of an animated agent using a Socratic-like questioning approach on student performance during a set of disassembly tasks performed in a CBI simulation program. To complete the disassembly task, students had to repeatedly select from an assortment of tools that would enable them to remove the next item in the disassembly sequence. Corrective feedback, in one of two forms, was provided by the system after each wrong choice of tool. This experiment tried to determine whether the type of feedback interface (text vs. animated agent) and the style used (corrective feedback as statements vs. Socratic tutoring as questioning) would combine to positively influence user performance.

It was hypothesized that, (a) students using the Socratic animated agent would make a total number of wrong tool choices that would be lower than those students using a textual, statement-based feedback system, and (b) those students using the Socratic animated agent would stop making wrong tool selections earlier in the disassembly process. Both hypotheses were supported by Moreno's (2001) studies on the positive effects of animated agents on motivation, Hietala and Niemirepo's (1998) findings on frequency of interaction between users and agents, Lester et al.'s (1997) results on the positive effect of agents on user performance, as well as Zhou et al.'s (1999) positive experiences on the use of a Socratic tutoring approach in a CBI environment.

METHODS

Participants

Participants in this study were 27 post-secondary vocational education students in a medium-size technical college in North Florida, enrolled in Electronics Technology and Computer Repair classes. The population consisted of both teenage and adult students, all of them males. This distribution was influenced by the nature of computer-related classes and programs of study, where female attendance is scarce. Ethnicity and SES were not considered for this study. Students volunteered to participate in the study during their regular class hours. An almost equal number of participants was assigned to each condition (13 to the Textual Feedback group, 14 to the Socratic Agent Feedback).

Materials

A CBI program adapted from the Panasonic Virtual CD Mechanism (2000) was used for this study. The program simulated a CD-Rom player disassembly process, and it was designed to show users the right disassembly sequence for the Panasonic CD-ROM player. As Figure 1 shows, the working area featured an instructions box, where instructions were provided for each of the 18 disassembly steps; an onscreen tool selection panel with eight available tools, where the user would choose the right tool for the right task; a Velcro strip, where the different disassembled parts were attached; and a drop box, where removed screws were kept. The combination of these elements simulated the environment in which a computer technician would work.

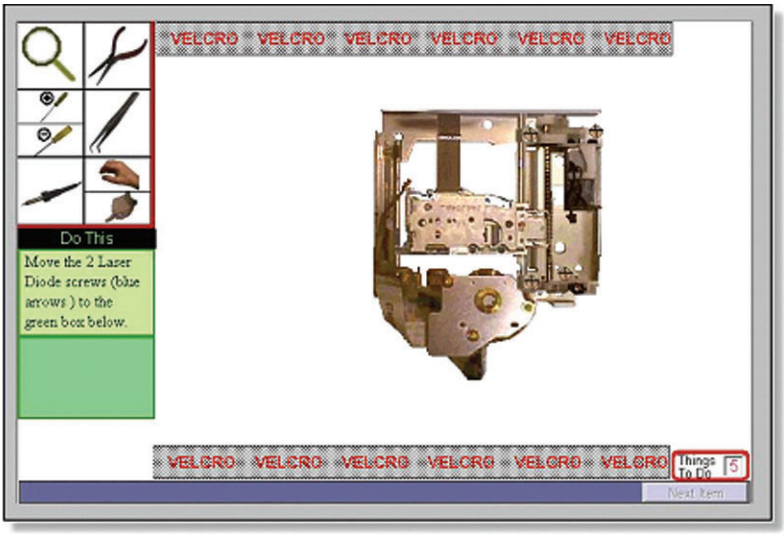

Figure 1. The simulated workspace for the disassembly process

During the study, participants were allowed to move to the next disassembly step only after the current step had been performed successfully. The activities in this CBI program were designed to take between 20 and 45 minutes.

The program was rewritten for this study in two versions, one with text-based feedback and the other with an animated Socratic agent providing feedback. In both versions, each step of the disassembly process could not proceed until the right tool had been chosen. The agent selected for this study was Microsoft's *Genie*, shown in Figure 2. This agent used Microsoft technology to display a number of behaviors on the screen (appearing, disappearing, making different facial expressions) in response to a user's actions.

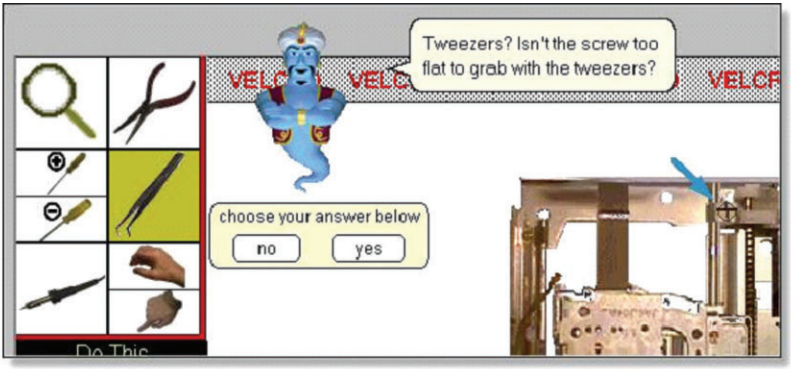

Figure 2. Genie responds to a wrong tool choice and initiates a conversational exchange. The agent would not be visible until a wrong tool was chosen

There were no body gestures (pointing, moving to a specific area, etc.) performed by the agent, since its role was to engage the user in a specific train of thought, not to instruct him/her on specific procedural tasks. Similarly, facial expressions were limited to "surprise" and "smile," just enough to convey a friendly and more natural communication setting.

Independent Variables

The independent variable for this study was the type of feedback offered to provide help and guidance each time the wrong tool was selected. Group 1 received textual feedback (TF) in the form of a statement each time a wrong tool choice was made. This feedback, in a text box next to the tools, informed the user that he/she needed to make another selection. The statements were of the type: "The soldering iron is not the right tool for this task. Please select another tool." Group 2 received feedback from a Socratic agent (SAF), who would appear in a relevant area of the screen and use a computer simulated voice to speak a number of questions and hints, in order to make the user think and reflect about the choice they had just made. The questions were of the type: "A soldering iron? Now, do you need to apply heat to perform this task? If not, what kind of tool would you need instead?" The kind of reasoning the agent elicited was intended to help the student make the right choice in that, and subsequent, similar situations.

Dependent Variables

The dependent measures for the study were the *total number of wrong tool choices*, and the *disassembly step at which wrong tool choice stopped*. These were straightforward measures, tracked as users interacted with each version of the CBI program. Each wrong tool selected at each disassembly step was tallied. Several wrong tools could actually be chosen at any given disassembly step. A high number of wrong tool choices made indicated a low level of user performance (and thus, low level of CBI version effectiveness), and vice versa. "High" and "low" were determined by comparing both groups with each other, and classifying them as "higher" and "lower," and not by comparing the tallied scores to predetermined standard. For the second DV, each disassembly step was monitored to detect at what point no further wrong tools were selected. Mistakes stopping in the early steps of the disassembly process indicated a higher level of user performance, which could be attributed to a higher level of CBI version effectiveness.

Procedures

Participants interacted with their assigned version of the Panasonic Virtual CD Mechanism (2000) during their Electronics Technology or Computer Repair classes. The classes themselves, which for the most part consisted of independent study activities, were held in the main classroom. The research activities were held in computer labs adjacent to the classroom. Students were individually contacted and offered the opportunity to participate in the study as independent study activities were taking place. Those who volunteered were randomly assigned to one of the two versions of the CBI, and taken to one of the two independent computer labs where the computers used for the study were set up. The CBI program gave participants a brief introduction to our goals (i.e., to field-test the disassembly simulation program), although those students were not aware that different kinds of feedback (textual vs. Socratic tutoring agent) would be offered. Once they completed their disassembly computer simulation, students returned to their independent study activities. The last screen of the CBI program thanked them for their participation in this experience. The average duration of the disassembly simulation process was 20 minutes. All study-related activities were supervised by members of our research team, with the permission of the instructors of each of the classes.

RESULTS

Wrong Tool Choice

Means and standard deviations regarding the total number of wrong tool choice across groups are presented in Table 1. As shown by a *t*-test, there was no significant difference between the total number of wrong tool choices made by the groups that received textual feedback (M = 7.2, SD = 4.7), and the group that received feedback from the Socratic agent (M = 6.9, SD = 5.2).

Table 1

Means and Standard Deviations of Wrong Tool Choice across groups.

	Performance		
Treatment	**M**	**SD**	**N**
Textual feedback	7.2	4.7	13
Socratic Agent feedback	6.9	5.2	14

A visual examination of raw scores revealed no outliers, and an approximately normal distribution of scores indicated the normality assumption had not been compromised. The independence assumption was robust enough given the random assignment of students to one treatment or another. At the same time, a Levene's test supported the fact that the assumption of equal variances had not been violated, $F(1,25)=.27$, $p=0.37$.

A t-test was used to compare means scores across groups. The results of this test showed a significance value of .85, which indicated there was no statistically significant difference between the scores in both groups. These results did not support our hypothesis that the group with the Socratic agent would make fewer total mistakes in the tool selection task than the group with the textual feedback.

Last Step Where Mistakes Were Made

Table 2 shows means and standard deviations for the last recorded step in the disassembly process where users last made a mistake. This data indicates that the group receiving feedback from the animated Socratic agent (M

= 11.4, *SD* = 5.1) stopped making mistakes at the same point in the disassembly process as the group that received textual feedback (*M* = 11.4, *SD* = 3.8).

Table 2

Means and Standard Deviations of Last Step where a Mistake was Made.

	Performance		
Treatment	**M**	**SD**	**N**
Textual feedback	11.4	3.8	13
Socratic Agent feedback	11.4	5.1	14

An examination of this data showed an approximately normal curve and no presence of outliers, which supports the robustness of the normality assumption. The robustness of the independence assumption was supported by random assignment of students to treatments. The robustness of the homogeneity of variances assumption was supported by a Levene's test, $F(1,25)=1.3$, $p=0.44$.

The *t*-test used to compare means across groups for this variable revealed a significance value of .98, which was high enough to determine that there was no statistically significant difference between the results from both groups. This does not support our second hypothesis that the group with the Socratic agent feedback would stop making mistakes earlier in the disassembly process.

DISCUSSION

The purpose of this study was to examine the effect of a Socratic animated agent on student performance during a computer-based simulation of a disassembly process. Our first hypothesis that students in the Socratic agent group would make fewer wrong tool choices during the disassembly process was not supported by the results from this study. This is actually in line with previous research as cited by Dehn and Van Mulken (2000) that mainly showed no effect on the cognitive domain resulting from the use of animated agents. In our experiment, our intention was to add a new variable to be examined in the agent-human relationship, which was the Socratic tutoring approach, and the hypothesis was formulated based on this new component and its positive results as a pedagogical approach in face-to-face environments.

The lack of statistical difference between groups regarding user performance may be attributable to the fact that students were already experienced in the use of tools and the kind of care that has to be taken during any disassembly process of any electronic components. Thus, the feedback provided by either the agent or the text box may have added little to their previous knowledge or experience with manual disassembly tasks. In this connection, the fact that this disassembly was done on a computer, where behaviors are different from a "hands-on," real environment, may have been the main factor for the mistakes made. For example, in the computer environment a click on a screw was used to remove it, while in a real situation the user would have had to grab a screwdriver, align it with the screw head, and rotate his/her hand to remove the screw. These new "codes," or behaviors, may have affected both groups equally, accounting for the number of mistakes the program recorded. Thus, this would have made the feedback regarding tool choice irrelevant, since feedback could have been more pertinent on the specific computer-simulation "codes" (i.e., drag-and-drop, single click, double click, etc.). The software involved was originally designed to train technicians about the order of part disassembly and reassembly. So the ability to choose the proper tool for a removal or installation procedure was a precondition of the use of the software. That there was no significant difference measured in what should have been a precondition may be expected.

Our second hypothesis that students interacting with the Socratic agent would stop making mistakes earlier in the disassembly process than the group receiving text feedback was not supported by the data obtained. One of the main reasons for this result could be the short time that students spent with each treatment. An average time on task of 20 minutes was recorded for this study, which means students did not have much exposure to either of the two versions of the CBI program so that they would, for instance, develop the kind of analytical skills the Socratic agent was intended to encourage through its questioning and intellectual challenging approach. This brief exposure to one feedback method or the other was enough to help students move on with the program and complete the task, but not to affect their cognitive skills so that one of the groups (the one with the Socratic agent) would develop better reasoning skills to apply to subsequent disassembly steps, so that mistakes would stop earlier in the disassembly process.

In spite of the no significant difference results regarding user performance, informal debriefings with students have revealed that the agent had an effect on the affective domain (most students liked it). The effect of the animated agent on the affective domain has not been considered in relation to this study. Observations of participants at the time of the study who indicated that participation in the study came to be regarded as "fun," remain anecdotal, although they appear to indicate that animated agents may be used as a tool to improve user attitude. Likeability and appeal are desirable features to have in any kind of educational intervention, which can lead to higher levels of motivation and user satisfaction.

Future studies in this area can be improved by allowing enough exposure to the treatment (i.e., the Socratic animated agent), so that learners have a chance of developing the kind of intellectual reasoning skills the agent is intended to trigger on the part of the students. The procedural nature of the task being learned allows questioning by an agent to occur a maximum of nineteen times. While the agent is modeling a way of thinking about tool choices, very few of the agent's thought processes are exposed to the learner when they are only initiated as the result of a learner error. Merrill (1985) noted that corrective feedback studies rarely expose the effectiveness of an instructional method since learners spend comparatively little of their learning time in corrective feedback situations.

At the same time, more rigorous results could be yielded by dividing up the treatment used in this research into two separate ones. Thus, one of the studies could look at the nature and content of feedback itself, for example, a questioning kind of feedback versus a direct statement telling the user what to do, both implemented in the same format. The second study could focus on the format in which the message is presented, comparing the effect on the user of a plain text format versus an animated agent format. Given enough subjects and resources, these two independent studies could also be carried out at as 2x2 factorial design, with *message content* as one independent variable with two levels (statement; questioning) and *message format* as a second independent variable with two levels (text; animated agent). A cross-analysis of this data would provide information to identify unique effects and to determine how each factor influences student proficiency in the disassembly process.

Similarly, generalizability of the results could be improved by carrying out follow-up studies where users are asked to perform different tasks under similar conditions. The tasks could be of the same procedural nature (disassembly of a piece of electronic equipment), or they could be of a cognitive type (e.g., a problem-solving activity). Results then could be compared to analyze whether there are inherent elements in a given type of task that make a specific treatment (e.g., plain text, animated agent, etc.) more effective for some kinds of tasks than others.

The implications of this study for developers of educational software that makes use of an animated agent involve the importance of considering the pedagogical approach on which the animated agent is based, and the applicability of a discussion-based, Socratic tutoring approach. Although this study has provided anecdotal evidence that shows animated agents are an interesting tool that can make a significant contribution to user motivation in CBT, further studies are needed to identify what pedagogical principles and theories of learning and cognition make animated agents effective and helpful to the users. That type of research is essential to build a knowledge base from which developers and educators alike can draw, in order to design and implement appropriate interventions making use of animated agents that would improve the users' learning experience.

References

Baylor, A. (2002). Expanding preservice teachers' metacognitive awareness of instructional planning through pedagogical agents. *ETR&D, 12*(2), 5-22.

Dehn, D.M., & Van Mulken, S. (2000). The impact of animated interface agents: A review of empirical research. *International Journal of Human-Computer Studies, 52*(1), 1-22.

Hietala, P., & Niemirepo, T. (1998). The competence of learning companion agents. *International Journal of Artificial Intelligence in Education*, *9*(3-4), 178-192.

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education, 11*, 47-78.

Koda, T., & Maes, P. (1996). *Agents with faces: The effect of personification*. Paper presented at the 5th IEEE International Workshop on Robot and Human Communication (RO-MAN'96), Tsukuba, Japan.

Lester, J., Stone, B., Converse, S., Kahler, S., & Barlow, S. (1997). Animated pedagogical agents and problem-solving effectiveness: A large-scale empirical investigation. In B. Du Boulay & R. Mizoguchi (Eds.), *Proceedings of the 8th World Conference on Artificial Intelligence in Education*, (pp. 23-30). Kobe, Japan: IOS Press.

Merrill, J. (1985). Levels of questioning and forms of feedback: Instructional factors in courseware design. *Journal of Computer-Based Instruction*, *14*(4), 18-22.

Moreno, R. (2001). Cognitive and motivational consequences of adapting an agent metaphor in multimedia learning: Do the benefits outweigh the costs? In W. A. Lawrence-Fowler & J. Hasebrook (Eds.), *Proceedings of WebNet 2001 World Conference*, (pp. 873-878). Norfolk, VA: Association for the Advancement of Computing in Education.

Panasonic Virtual CD Mechanism [Computer Software] (2000). Secaucus, NJ: Panasonic Services Co.

Sproull, L., Subramani, M., Kiesler, S., Walker, J., & Waters, K. (1996). When the interface is a face. *Human-Computer Interaction, 11*, 97-124.

Stevens, A., & Collins, A. (1977). The goal structure of a Socratic tutor. *Proceedings of the National ACM Conference*, *USA*, (pp. 256-263).

Takeuchi, A., Naito, T. (1995). Situated facial displays: Towards social interaction. In I. Katz, R. Mack, L. Marks, M.B. Rosson, & J. Nielsen (Eds.), *Human factors in computing systems: CHI'95 conference proceedings*, (pp.450-455). New York: ACM Press.

Van Mulken, S., André, E., & Müller, J. (1998). The persona effect: How substantial is it? In H. Johnson, L. Nigay, & C. Roast (Eds.), *People and computers XIII: Proceedings of HCT98*, (pp. 53-66). Berlin: Springer.

Zhou, Y., Freedman, R., Glass, M., Michael, J., Rovick, A., & Evens, M. (1999). Delivering hints in a dialogue-based intelligent tutoring system. *Proceedings of the Sixteenth National Conference on Artificial Intelligence*, *USA*, (pp. 128-134).