Evaluating the Persona Effect of an Interface Agent in an Intelligent Tutoring System

Maria Moundridou, Maria Virvou

Department of Informatics, University of Piraeus

Abstract. This paper describes the evaluation of the persona effect of a speech-driven anthropomorphic agent that has been embodied in the interface of an Intelligent Tutoring System (ITS). This agent is responsible for guiding the student in the environment and communicating the system's feedback messages. The agent was evaluated during an experiment in terms of the effect that it could have on students' learning, behaviour and experience. The participants in the experiment were divided into two groups: half of them worked with a version of the ITS which embodied the agent and the rest worked with an agent-less version. The results from this study confirm the hypothesis that a pedagogical agent incorporated in an ITS can enhance students' learning experience. On the other hand, the hypothesis that the presence of the agent improves short-term learning effects was rejected based on our data.

1 Introduction

The perpetual, main challenges that researchers in the computer assisted learning area face are how to guarantee effective learning, increase students' motivation and engagement and generally enhance the learning experience. A promising approach for these challenging aims concerns the incorporation of animated pedagogical agents in computer assisted learning environments.

Animated pedagogical agents are characters embodied in the learning environment that exhibit life-like behaviour through speech, facial expressions, gestures and body movements. They are presented as three-dimensional graphics, real video or two-dimensional cartoon-style drawings. Their responsibilities range from acknowledging a student's action to providing help in a problem-solving situation. Several educational systems exist which embody animated pedagogical agents. Among them we cite Adele (Shaw et al., 1999), Herman the Bug (Lester et al., 1999) and PPP Persona (Rist et al., 1997). The reader is also referred to

(Johnson et al., 2000), in which a number of animated pedagogical agents representing the current state of the art are described and discussed.

Adele (Agent for Distance Learning: Light Edition) operates in a Web-based distributed simulation environment, where she guides and assesses students as they work through clinical cases. Adele consists of two components: the animated persona which is a Java applet that runs in a separate window, and the reasoning engine which monitors the student's actions and decides how Adele should respond to each of these actions. Herman the Bug is an animated pedagogical agent inhabiting Design-A-Plant, a learning environment for the domain of botanical anatomy and physiology. Herman observes students' actions as they build plants that can thrive in a given set of environmental conditions and provides explanations and hints. In the process of explaining and hinting, Herman performs various actions, such as walking, flying, swimming, teleporting, etc. PPP Persona is an animated interface agent that presents multimedia material to the user. While the user views the presentation the agent can comment on particular parts and highlight them through pointing gestures. The repertoire of the persona's presentation gestures includes gestures expressing approval or disapproval, warning or recommendation, etc.

Despite the existence of quite a few educational systems that incorporate animated agents, Dehn & van Mulken (2000) point out that empirical investigations concerning the effect of animated agents on learning are still small in number and differ with regard to the measured effects. They came to this conclusion after they conducted a comprehensive and systematic overview of the empirical studies performed so far. Indeed, examples of such studies show a diversity of results. Walker et al. (1994) investigated subjects' responses to a synthesised talking head displayed on a computer screen in the context of a questionnaire study. Their findings showed that compared to subjects who answered questions presented via text display on a screen, subjects who answered the same questions spoken by a talking head spent more time, made fewer mistakes, and wrote more comments. The study of Lester et al. (1997) with different versions of Herman the Bug revealed that the presence of a lifelike character in an interactive learning environment can have a strong positive effect on student's perception of the learning experience. Van Mulken et al. (1998) performed an empirical study to examine the effect of PPP Persona both on subjective and objective measures. The results of their study indicate that the presence of the agent has neither a positive nor a negative effect on comprehension and recall performance. However, even the mere presence of the Persona causes presentations and tests to be experienced by users as more entertaining and less difficult.

The diversity of results among empirical studies conducted so far and their small number signifies that the evaluation of the persona effect is an open research problem that needs to be addressed further. In this

paper we will report on an educational system called WEAR which embodies an animated speaking character in its interface and we will then discuss the evaluation study that we performed in order to evaluate the effect of the interface agent.

2 Description of the System and its Interface

WEAR (WEb-based authoring tool for Algebra Related domains) is a Web-based authoring tool for the construction of Intelligent Tutoring Systems (Virvou & Moundridou, 2000). WEAR incorporates knowledge about the construction of exercises and a mechanism for student error diagnosis that is applicable to many domains that can be "described" by algebraic equations. For example, such domains could be chemistry, economics, physics or math itself. WEAR also allows the authoring of electronic textbooks by instructors and delivers them over the WWW to learners (Moundridou & Virvou, 2001). These textbooks offer navigation support to students, adapted to their individual needs and knowledge. WEAR associates with each student a level of knowledge according to his/her past performance in solving problems with the tool. The tool suggests each student to try the problems and study the topics corresponding to the associated level of knowledge.

When a student attempts to solve a problem the system provides an environment where the student gives the solution step by step (Figure 1). At first the student is presented with a problem statement like the one shown in Figure 1[1]. The student is requested to write down the equations that are needed to solve the problem and then s/he is requested to mathematically solve the problem. To detect the erroneous answers the system compares the student's solution to its own at every step. The system's solution is generated by WEAR's Problem Solver. The Problem Solver incorporates knowledge about how to solve systems of linear equations correctly and may generate the solution to a problem using information about the specific domain to which the problem belongs (e.g. physics). During the process of solving a problem the student's actions are monitored by the system. In case of an erroneous action the Problem Solver passes the student's answer to the Student modeller, which is then responsible for diagnosing the cause of the error. The errors that are recognised by WEAR's Student modeller are the following:

- *Domain errors*. These include errors that are due to the student's unfamiliarity with the domain being taught. Such errors occur when the student enters an equation different from the one needed for the problem to be solved, or when s/he enters an "almost" correct equation (lacking a variable,

---

[1] This problem is based on a problem we found at http://www.algebratutor.org/big_list.html

using an erroneous relationship between variables, etc.). For example, if a student enters the equation *l=30-2/x* instead of *l=30-2*x*, then the error is attributed to the category of Domain errors and in particular to the sub-category of "erroneous relationship between variables".

- *Mathematical errors*. These include errors that are due to the student's lack of skills in solving mathematical equations. Such errors could be calculation errors, errors in isolating the unknown variable, etc. For example, if a student trying to isolate *x* in the equation *2*x=30-l* enters *x=(30-l)*2* instead of *x=(30-l)/2*, then the error is attributed to the category of Mathematical errors and in particular to the sub-category of "wrong isolation of the unknown variable".

Figure 1 about here.

WEAR's student interface includes an animated speaking face which is responsible for communicating the instructions and any feedback messages to the students (Figure 1). This interface agent is responsible for guiding the student to the environment and providing feedback to her/him while s/he is solving problems. The talking head component of the system uses speech synthesis to automatically produce speech output from text using MBROLA (http://tcts.fpms.ac.be/synthesis/mbrola.html), a freely available speech synthesiser.

3 Evaluation Study

As we have already mentioned, the aim of the experiment we conducted was to examine how the presence of the speech-driven anthropomorphic agent would affect students. Dehn & van Mulken (2000) argue that the possible effects of animated interface agents on users can and should be classified into: (i) effects on the user's subjective experience of the system, (ii) effects on the user's behaviour while interacting with the system and (iii) effects on the outcome of the interaction as indicated by performance data. In this evaluation we follow this classification and we examine the effect of the anthropomorphic agent with respect to experience, behaviour and performance of students. In the remainder of this section we will first describe the experimental setting and then report on the results along the three dimensions of effect.

3.1 Experimental Setting

The participants in the experiment were 48 college students from the University of Piraeus. Twenty-six of them were studying Informatics and the rest Economics. The students were randomly divided into two equal-sized groups: the "Agent" group (group A) and the "Non-Agent" group (group NA). Both groups participated in the experiment in the same way. The difference between them was the version of the ITS they interacted with: the version used by group A embodied the talking head, whereas group NA worked with an agent-less version, that passed textual messages to students. To be able to explore the agent's effect on students' learning, the content of the messages that both versions of the ITS passed to students was completely identical. Such messages concerned feedback messages, guidance to the environment etc. To generate the ITS that we used as a testbed, we fed WEAR with several quite simple math problems. An example of such a problem is the following: "Scott starts jogging from point X to point Y. A half-hour later his friend Garret who jogs 1 mile per hour slower than twice Scott's rate starts from the same point and follows the same path. If Garret overtakes Scott in 2 hours, how many miles will Garret have covered?" Most of these problems can be expressed by a system of linear equations. Solving the system of equations produces the solution to the problem. This is no new knowledge for college students. However, not all of them seem to know how to deal with this kind of math problems.

The experiment started with a pre-test, consisting of five problems. The students were asked to work on the pre-test using paper and pencil. For each student, the time s/he needed and the score s/he achieved in the pre-test was recorded. After completing the pre-test, students worked with the system to solve problems of the same level of difficulty. At this stage the students had the chance to practice and consolidate their knowledge using the system's facilities. Students were allowed to work with the system for as long as they would like and to solve as many problems as they wished. While working with the system, the students' actions were logged. The log files provided information concerning time (time spent on each problem, response times, total time spent working with the system) and performance (grade achieved, mistakes per problem). They also provided other useful information such as the number of times a student dropped a problem s/he was attempting to solve, or the number of times a student requested to see the solution of the problem before or after solving it correctly by himself/herself. The post-test that followed consisted again of five similar problems that students were asked to solve using paper and pencil and being timed again. Finally, students were asked to complete a questionnaire concerning their experience from interacting with the system.

The collected data from pre- and post-tests, log files and questionnaires were used to compute the mean values for each group and the associated p-values. The null hypothesis of no difference between the means of the two groups, H0: $\mu 1 - \mu 2 = 0$, was tested against the alternative of different means, HA: $\mu 1 - \mu 2 \neq 0$. To decide upon rejecting or not the null hypothesis we subjected our data to two-tailed t-tests and checked whether the calculated p-value was greater or less than a specific level of significance α. In our test we fixed this level at 5%. If the calculated p-value was less than .05, we rejected the null hypothesis and concluded that the difference between means of the two groups was statistically significant whereas larger values indicated that the difference between the means of the two groups was not statistically significant.

3.2 Students' Behaviour while Interacting with the System

In general, in the literature of computer aided learning, reaction times and/or error rates in a user's responses are often used as indicators of the degree of attention that the user shows. However, these indicators are often interpreted in a very different way. Sometimes, these indicators are even interpreted in a contradictory way. For example, in the study of Sproull et al. (1996) two interfaces were compared: one that was using textual output and another that was using a talking face in addition to the textual output. The authors measured the time subjects needed to complete some psychological test items and the number of items they skipped. The longer response times that were produced by subjects in the face condition were interpreted by the authors as an indication of a higher degree of users' attentiveness. Similarly, the fact that subjects in the face condition left more items unanswered than did subjects in the text condition was also assumed to reflect a higher degree of attentiveness. In another study (Takeuchi and Naito, 1995), the extra time that subjects in the face condition needed for their reaction in a card matching game was thought as an indication of a lower degree of attentiveness, since the authors argue that the facial display distracted users from the game.

As it can be seen from the above mentioned studies, there seems to be no definite interpretation of the students' behaviour concerning the degree of their attention to the system. For this reason, we conducted a short empirical study on this matter by consulting 15 human experts who were mainly classroom instructors. These experts were asked to specify the criteria that they would use to consider a student as attentive. In particular, the experts were given a list of criteria, which were either used in previous studies or were a combination of criteria and they were asked to select from them the criterion that they would use to measure the attentiveness of students. The experts almost unanimously favoured the following criterion: A student **A** is more attentive than a student **B** if **A** spends more time with the system but not "too long", given that both

students deal with the same number of tasks. The experts provided us also with a threshold value over which we should consider the time students spent as "too long".

Based on the above mentioned experts' criterion and the collected data from our experiment, the conclusion we reached was that the presence of the anthropomorphic agent neither increased students' attention nor it distracted them from their tasks. In particular, group A spent on average 19.96 minutes working with the system while group NA spent 18.25 minutes. The mean time each group spent, as well as the time spent by each individual, was below the 30 minutes threshold that we were suggested to use; so we assumed that neither group seemed to be non-attentive. Having solved almost the same number of problems, we noticed a slight numerical difference between the mean times of the two groups. However, this difference was not statistically significant ($t(46)=0.97$; $p=0.336$) and thus we concluded that students' attention is not affected by the presence of the agent.

3.3 Students' Subjective Experience of the System

Students' responses to a questionnaire were used to measure their attitude towards the system. The corresponding questions that were asked were the following:

1. Did you enjoy working with the system?

2. Was the system easy to use?

3. Were the problems that you were asked to solve with the system difficult?

4. Did the system help you to improve your problem-solving skills?

5. How useful was the system compared to a human tutor?

6. What did you like most in the system?

7. Are there any comments that you would like to make?

Students' answers to questions 1-5 were scored on a five point Likert scale (e.g. in question 2 the possible answers were: 1-very difficult, 2-difficult, 3-average, 4-easy, 5-very easy) and were taken into account to produce numerical results which are illustrated in Table 1. The results (Table 1) are generally congruent with findings from similar studies: In all of these questions the ratings that group A gave are numerically higher than the ratings of group NA. Furthermore, the differences between the means of the two groups' ratings are statistically significant in most of the questions.

Table 1 about here

In conclusion, based on the students' subjective experience of the system (as summarised in the above table) we could say that the presence of the agent adds value to a system, since users working with the agent version found the system more enjoyable and easier to use than did those working with the agent-less version. In addition, the agent seems to affect the students' attitude towards the tasks that they should perform and the knowledge they should acquire eventually. Students working with the agent found the problems that they should solve less difficult than did students working without the agent. In addition, students of the agent group also felt that the system was more useful in helping them improve their problem-solving skills than did students of the other group. However, in the quite critical question of comparing the usefulness of the system to that of a human tutor students were more restrained in their ratings irrespective of the group they belonged to. This finding was quite surprising for us, since we expected that the presence of the speech-driven anthropomorphic agent could give students the impression of a real tutor and thus affect their ratings in that particular question.

Questions 6 and 7 were asked to students so that they could express more freely what they thought about the system. The answers to these questions were not taken into account in any numerical results but they showed what the students felt about the agent, what they would expect from it and how it could be improved. The students' answers to questions 6 and 7 were quite encouraging because there were no negative comments about the students' overall impression concerning the agent. However, a lot of students made comments on the agent's voice. This is probably an indication that the most important aspect of the agent is the speech rather than the animation.

Examples of students' comments include the following answers:

- "I liked the talking face but I got nervous with his voice"
- "I liked the simulated instructor. He was amusing. A real instructor is more strict and I may make errors due to my anxiety"
- "The face that was talking was nice but I would prefer a female voice"

3.4 Learning Outcomes

To test whether the presence of the agent improves the learning outcomes, we studied and compared data collected from the pre-tests and post-tests and from the interaction with the system (Table 2). Some fundamental findings for our evaluation were the following:

- Both groups scored similarly in the pre-test and needed about the same time to complete it. This was an indication that both groups were equally capable of dealing with the tasks they had to perform.

- Both groups improved their time and scores in the post-test. Based on this, we could assume that the tutoring system (irrespective of the existence of the agent) achieved its goal to improve the students' performance.

Table 2 about here

However, the pointed improvement in time and grade differed between the two groups: In average, group A spent 19% less time to complete the post-test and achieved 15% higher grade than the pre-test. In contrast to group A, the improvement of group NA's average time and grade was 14% and 8% respectively. Although numerically higher, the improvement both in time and grade is not statistically significant between groups (time improvement: $t(46)=-0.86$; $p=0.394$, grade improvement: $t(46)=1.22$; $p=0.227$). As for the students' performance when solving problems with the system, group A achieved 6.5% higher grade than group NA but this was again not statistically significant ($t(46)=1.69$; $p=0.098$). In conclusion concerning the learning outcomes, we should note that although the experiment results showed some trends favouring the agent version of the system, none of these were statistically significant and thus we cannot claim that the presence or absence of the animated agent affects the learning outcomes.

4 Discussion and Conclusions

The conclusions of previous related empirical studies in the field of computer assisted learning, are rather controversial as to the effect of animated agents on students' performance, behaviour and experience. Therefore, it has become an open research question whether the effect of animated agents is beneficial to students or not. This paper's contribution on this topic is the fact that it confirms some of the positive and/or non-positive effects that have been reported previously and highlights some assets of the use of agents that may be investigated further.

In our case study there were mainly two advantages that were induced by the presence of the agent. The first advantage concerned the enjoyment that students felt when they interacted with a system that embodied a speaking animated interface agent. The presence of this advantage was not surprising since it was the most common finding among previous studies, e.g. (Walker et al., 1994; Lester et al., 1997). The other advantage we came up with in our study was that students working with the agent version of the system found the problems that they should solve less difficult than students working without the agent did, despite the fact

that the performance of both groups of students was similar. The importance of this finding is great concerning the motivation of students to work with the system; it shows that students working with the agent version are more motivated than students working with the agent-less version.

However, there was an important but not positive result of our case study. This was the fact that the presence of the interface agent did not manage to improve significantly the short-term learning outcomes. Furthermore, the students' attentiveness to the system was not promoted by the talking face. Although this is not an encouraging finding with respect to the educational benefits of such interfaces, there have also been other empirical studies that have shown similar results, such as (van Mulken et al., 1998).

In conclusion of our investigation that explored the short-term effects of interacting with an interface agent, we have found that the agent version increases the motivation of students but does not necessarily promote the learning outcomes. This may not seem very encouraging at a first glance but we believe that in fact the use of animated agents does promote the learning process significantly. Indeed the confirmed increase of students' motivation is an important asset for computer aided learning systems, which may show positive results in learning outcomes in the long run.

As a matter of fact, it is within our future plans to examine the long-term effects of the interactions with the animated agent. In this examination we will investigate how the increase of students' motivation may be connected to long-term effects in the learning outcomes. In addition we intend to use WEAR to produce ITSs in other domains as well and run more experiments involving students from various age groups. The experiments are going to be similar to the one discussed here but we will have the opportunity to evaluate the effect of the speaking agent along the two dimensions of the domain taught and the age of students.

References

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

Johnson, W. L., Rickel, J. & Lester, J. (2000) Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education*, **11**, 47-78.

Lester, J., Converse, S., Kahler, S., Barlow, S., Stone B. & Bhogal R. (1997) The persona effect: affective impact of animated pedagogical agents. In *Human Factors in Computing Systems: CHI'97 Conference Proceedings* (ed. S. Pemberton), 359-366. ACM Press, New York.

Lester, J., Stone, B. & Stelling, G. (1999) Lifelike Pedagogical Agents for Mixed-Initiative Problem Solving in Constructivist Learning Environments. *User Modeling and User-Adapted Interaction*, **9**, 1-2, 1-44.

Moundridou, M. & Virvou, M. (2001). Authoring and Delivering Adaptive Web-Based Textbooks Using WEAR. In *Proceedings of ICALT-2001, IEEE International Conference on Advanced Learning Technologies*, to appear.

Rist, T., André, E. & Müller, J. (1997) Adding Animated Presentation Agents to the Interface. In *Proceedings of the 1997 International Conference on Intelligent User Interfaces* (eds. J. Moore, E. Edmonds & A. Puerta), 79-86. ACM Press, New York.

Shaw, E., Ganeshan, R., Johnson, W. L. & Millar, D. (1999) Building a Case for Agent-Assisted Learning as a Catalyst for Curriculum Reform in Medical Education. In *Proceedings of the 9th World Conference on Artificial Intelligence in Education –AIED'99* (eds. S. Lajoie & M. Vivet), 509-516. Frontiers in Artificial Intelligence and Applications, Vol. 50, IOS Press, Amsterdam.

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

Takeuchi, A. & Naito, T. (1995) Situated Facial Displays: Towards Social Interaction. In *Human Factors in Computing Systems: CHI'95 Conference Proceedings* (eds. I. Katz, R. Mack, L. Marks, M.B. Rosson & J. Nielsen), 450-455. ACM Press, New York.

van Mulken, S., André, E. & Müller, J. (1998) The Persona Effect: How Substantial Is It? In *People and Computers XIII: Proceedings of HCI'98* (eds. H. Johnson, L. Nigay & C. Roast), 53-66. Springer, Berlin.

Virvou, M. & Moundridou, M. (2000). A Web-Based Authoring Tool for Algebra-Related Intelligent Tutoring Systems. *Educational Technology & Society*, **3**, 2, 61-70.

Virvou, M., Sgouros, N., Moundridou, M. & Manargias, D. (2000) Using a speech-driven, anthropomorphic agent in the interface of a WWW educational application. In *Proceedings of ED-MEDIA 2000, World Conference on Educational Multimedia, Hypermedia & Telecommunications* (eds. J. Bourdeau & R. Heller), 1724-1726. AACE, Charlottesville VA.

Walker, J.H., Sproull, L. & Subramani, R. (1994) Using a Human Face in an Interface. In *Human Factors in Computing Systems: CHI'94 Conference Proceedings* (eds. B. Adelson, S. Dumais & J. Olson), 85-91. ACM Press, New York.

Table 1: Mean responses for subjective experience of the system and t-test results for the difference between A and NA group means.

Questions	Scale for answers	Agent group (n=24)		Non-Agent group (n=24)		Mean difference t(df); p
		Mean	Standard Deviation	Mean	Standard Deviation	
1. Did you enjoy working with the system?	1..5, where 1: not at all and 5: very much	4.67	0.64	2.92	1.10	t(46)=6.74; p=0.000
2. Was the system easy to use?	1..5, where 1: very difficult and 5: very easy	4.54	0.59	3.54	0.78	t(46)=5.02; p=0.000
3. Were the problems that you were asked to solve with the system difficult?	1..5, where 1: very difficult and 5: very easy	4.54	0.66	3.17	0.87	t(46)=6.18; p=0.000
4. Did the system help you to improve your problem-solving skills?	1..5, where 1: not at all and 5: very much	4.04	0.75	3.50	0.88	t(46)=2.29; p=0.027
5. How useful was the system compared to a human tutor?	1..5, where 1: useless and 5: very useful	3.54	0.72	3.38	1.01	t(46)=0.66; p=0.515

Table 2: Means of time and grade from the pre- and post-test and from working with the system; t-test results for the difference between A and NA group means.

Variable	Agent group (n=24)		Non-Agent group (n=24)		Mean difference t(df); p
	Mean	Standard Deviation	Mean	Standard Deviation	
Pre-test time (in minutes)	12.50	3.31	12.96	2.84	t(46)=-0.51; p=0.609
Pre-test grade (scale from 0 to 5)	4.38	0.77	4.46	0.59	t(46)=-0.42; p=0.675
Post-test time (in minutes)	9.83	1.81	10.75	2.05	t(46)=-1.64; p=0.107
Post-test grade (scale from 0 to 5)	4.88	0.34	4.75	0.44	t(46)=1.10; p=0.277
Rate of change in time from pre to post-test	-0.19	0.15	-0.14	0.22	t(46)=-0.86; p=0.394
Rate of change in grade from pre to post-test	0.15	0.25	0.08	0.15	t(46)=1.22; p=0.227
Time spent working with the system (in minutes)	19.96	6.60	18.25	5.53	t(46)=0.97; p=0.336
Grade achieved with the system (scale from 0 to 5)	4.75	0.44	4.46	0.72	t(46)=1.69; p=0.098

Figure 1: Solving a problem with the Agent version of the system.