

Authors: Judy Robertson, Beth Cross, Hamish MacLeod and Peter Wiemer-Hastings
Correspondence address:
Judy Robertson, Division of Informatics, University of Edinburgh, 2 Buccleuch Place,
Edinburgh, EH8 9LW, UK.
Email: judy@inf.ed.ac.uk

Published in International Journal of Ai in Education 2004.

Children's interactions with animated agents in an intelligent tutoring system

JUDY ROBERTSON, BETH CROSS, HAMISH MACLEOD AND PETER WIEMER-HASTINGS

*Division of Informatics, University of Edinburgh, 2 Buccleuch Place,
Edinburgh, EH8 9LW, UK.
Email: judy@cogsci.ed.ac.uk*

Although animated pedagogical agents are frequently found in intelligent tutoring systems and interactive learning environments, their effect on users' attitudes and learning requires further investigation. This paper reports findings of a field study designed to investigate the impact of animated pedagogical agents on primary school children's attitudes to and interactions with the StoryStation system. Fifty-nine pupils used either a version of StoryStation with an animated agent interface or an equivalent one with a normal graphical user interface to write a story. Analysis of questionnaire data indicated that pupils who used the agent version rated StoryStation more highly than those who used the non-agent version. This effect was more pronounced for girls. Analysis of program use revealed that girls were more likely to interact with the agent version, while boys were more likely to interact with the non-agent version.

INTRODUCTION

Animated pedagogical agents are increasingly common interface components in intelligent tutoring systems and interactive learning environments. After pioneering empirical work with the animated agent "Herman the Bug", Lester et al. (1997) recommended that "designers of interactive learning environment should give serious consideration to including an animated pedagogical agent in their software" (p12) because of their positive effects on pupils' motivation and learning. The range of potential benefits of animated pedagogical agents in various tutoring domains were well documented in (Johnston et al., 2000). These suggested advantages have encouraged more designers of learning environments to include animated pedagogical agents or learning companions in their software. For example in the recent Intelligent Tutoring System Conference (Cerri et al., 2002), three of the invited speakers discussed issues relating to agent interactions with humans or designs for pedagogical agent architectures. In addition, several papers were presented on social agents, collaboration with agents and technologies for implementing agents. For the purposes of this paper, the term "agent" is used inter-changeably with "animated pedagogical agent" to refer to a life-like character. We do not refer to agents in the sense of software architectures, or to encompass features such as adaptivity or intelligence.

Currently there is a lack of empirical evidence supporting the positive impact of agents on students' motivation or learning (Dehn and van Mulken, 2000). This study presents quantitative and qualitative data from a field study of 59 primary school children which assessed the effects of a group of animated pedagogical agents on pupils' attitudes to and interactions with the StoryStation tutoring system. The study reported here did not investigate the impact of the tutoring system on children's writing development; this work is still in progress.

Previous research into users' relationships with computers and animated agents is described in Section 2. The StoryStation project, its design process and educational philosophy is presented in Section 3, with a description of the interface and animated agents. Section 4 outlines the methodology used in the field study, the results of which are presented in Section 5. Discussion and possible interpretations of the results informed by interview data and previous research are offered in Section 6. The paper concludes with suggestions for further work.

PREVIOUS ANIMATED AGENT RESEARCH

The long term goal of the StoryStation project is to improve children's writing abilities using a supportive software writing environment. In order to determine a viable method for doing this, we must first ascertain whether or not animated agents can be acceptable and beneficial to pupils. Previous research described in this section suggests that agents have positive effects on pupils' motivation and problem solving abilities, but this is not yet well established. Furthermore, it is likely that users in different target user groups will have different attitudes towards agents, an issue which needs further research. This section describes previous related research into interactions between users and computers generally, before focusing on the potential advantages of animated agents in particular. Previous empirical studies of the effect of animated agents in learning systems on users' attitudes and learning are summarised.

In a series of studies of the ways in which humans interact with computers, Reeves and Nass (1996) have gathered evidence to support their "computers as social actors" theory. Their findings suggest that humans interact with computers using social rules appropriate to interactions with other humans. For example, they found that even expert users obeyed norms for politeness when evaluating their interactions with a computer. The users were more likely to be honest and accurate when filling out a form to rate the performance of a piece of software when they filled out the form on another computer, rather than the computer which ran the original bit of software. Although the users were not aware that they answered to "spare the feelings" of the first computer, they nevertheless were inclined to rate the software more highly when evaluating it on the machine with which they interacted. This illustrates the point that even expert users interact with computers as if they were social actors.

Given that software with the most simple of textual interfaces can induce this effect, it is reasonable to suppose that animated pedagogical agents which are designed to capitalise on peoples' anthropomorphic tendencies can have a positive effect on users' motivation and behaviour with software.

Johnson et al. (2000) discuss the possibilities for increased social interaction offered by animated agents in interactive environments. Users interacting face to face with animated agents have the advantage of increased bandwidth of communication with the learning environment. Agents can express emotion through voice, facial expressions and body language, thus increasing the social presence experienced by the user. Engaging with agents as social actors is hypothesized to be motivating to users and therefore have a positive effect on students' learning (as discussed in the section describing studies conducted by Moreno et al. (2001)). Furthermore, an agent interface is particularly appropriate for training applications where the agent can demonstrate complex tasks in a simulated activity, navigate users around a virtual environment, or focus the users' attention on salient aspects of the task through gestures. Johnson et al. (2000) review a number of learning applications which have successfully incorporated agent interfaces. An interesting learning environment is Design-A-Plant (Lester et al. 1997), because it was designed for 12 year old children – a similar user group to the StoryStation target audience.

Design-A-Plant is a learning environment designed to teach botanical anatomy and physiology which incorporates an animated pedagogical agent named Herman the Bug. In a study with 100 middle school children, Lester et al. (1997) investigated the effects of five versions of Herman the Bug on pupils' motivation and learning. These five "clones" were as follows: a muted version offering no advice at all; a task specific verbal version offering instruction to the user on how to complete a particular task; a principle based verbal version offering high level domain advice; a principle based verbal and animated version offering high level advice both with voice and accompanying animations; and a fully expressive version offering task specific and principled advice with verbal instruction and accompanying animation. Analysis of questionnaire data reporting the children's attitudes to the agents showed that the users responded positively to all clone types, including the mute version. There was a significant interaction between clone type and question. Users of the fully expressive clone rated the utility of the agent's advice and encouragement particularly highly. Comparison of pupils' pre-test and post-test scores showed significant increases in scores, and that the magnitude of the effect was dependent on the clone type. The fully expressive, principled verbal and

animated, and principle based verbal versions obtained the greatest magnitude of score increases. The researchers conclude that designers of learning environments should seriously consider including animated pedagogical agents in their software because of their positive motivational effects, even if the agent offers no advice.

Since this study was published, animated pedagogical agents have been included in various educational software packages with a range of target user groups. The "T'riffic Tales" story writing environment for young children incorporates an empathic cartoon style agent named Louisa (Brna, Cooper and Razmerita, 2001). This agent is intended to represent a slightly older peer, who is available to provide assistance and affective support as the pupil writes. There are as yet no published results on the effectiveness of Louisa. Ryokai, Vaualle, and Cassell (2003) describe an animated agent named Sam who is designed as a story listening partner for young storytellers. The agent is humanoid in appearance and was designed to look child like and gender neutral. Sam and either a pair of children or a single child take it turns to tell a story situated in a toy castle. The stories which Sam "tells" are pre-recorded stories which are relevant to the story told in the previous turn. Results of a study in which 28 five year old girls interacted with Sam illustrated that it functions as a virtual peer in two ways. Firstly, it acts as a more linguistically advanced conversational partner, which can model the decontextualised language necessary for storytelling. Children who collaborated with Sam created stories which contained the more linguistically advanced features used by the virtual peer, such as more quoted speech and temporal and spatial references, than children who interacted with a child of the same age. Secondly, there is some indication that Sam could be useful in encouraging children to coach and criticise it on its storytelling behaviour. The authors intend to further investigate how changes to Sam's appearance will alter the children's perceptions of it as a virtual friend.

The "Teachable Agents" project at Vanderbilt investigates the use of social agents in intelligent learning environments. This project explores the concept that people can learn from teaching others. Biswas et al. (2001) describe a Teachable Agent environment in which pupils can teach an agent named Billy how to solve time/speed/distance problems. Billy is represented in a comic book style, and appears as a character in an animated story embedded in the environment. The designers of this environment intend that teaching this computer agent will help pupils to gain a deeper understanding of the subject matter, and increase their motivation to learn.

Animated pedagogical agents have been used in a number of educational domains, and for target user groups ranging from pre-schoolers to college students. There appears to be a design assumption that agents will be appropriate, and this is seldom tested independently of other aspects of the learning environment. However, further research is still required to confirm that animated pedagogical agents are more effective than traditional graphical user interfaces in terms of either students' motivation or learning. In addition, it is likely that different subsets of these user groups are likely to prefer different styles of interfaces. One would expect that interfaces which are suitable for four year olds might not be so appealing to fourteen or twenty-four year olds. Dehn and van Mulken (2000) reviewed empirical studies of the impact of animated agents on users' experiences with the software, their behaviour while using it and their performance on the task with which the agent is intended to assist. The authors reported that there are few empirical studies which address these issues, and that the results are inconclusive. They concluded that the literature to date "does not provide evidence for a so-called persona effect, that is, a general advantage of an interface with an animated agent over one without an animated agent" (Dehn and van Mulken; 2000: p.17). Furthermore, they comment that the methodological validity of some studies is questionable. For example, Lester et. al (1997) concluded that the presence of the animated agent in an intelligent tutoring system improved the learners' problem solving skills. Dehn and van Mulken pointed out that these conclusions are suspect because there was no control condition that provided the same advice without an animated agent. There are also some problems with the questionnaire used to assess users' motivation when using the software. The muted agent had higher mean scores than the more expressive agents on the questions "Was Herman the Bug's advice useful to you?" and "Did you believe the advice you got from Herman the Bug?". As the muted version of the software gave no advice at all, these means suggest that the validity of the attitude scores may be questionable. In addition, some of

the questions use structure and vocabulary which may have been difficult for some members of their 12 year old user group to understand e.g. “As you progressed in the educational program, did Herman the Bug become more helpful?”. This potential difficulty may have been compounded by the fact that the researchers were not on hand to answer queries when the children filled out their forms.

Given the lack of evidence from the small number of studies, some of which are confounded, Dehn and van Mulken called for further methodologically sound studies in this area. A series of such studies was conducted by Moreno et al. (2001).

Moreno et al. (2001) conducted the studies using the Design-a-Plant environment with the agent Herman the Bug (Lester et al., 1997), to test two competing hypotheses about the effect of agents on users’ learning. The *constructivist hypothesis* is that users who interact with animated pedagogical agents invest socially and emotionally in the agent and consequently find the learning task personally meaningful. When users become motivated to learn with the agents, they try hard to understand the material. This results in the formation of a coherent mental model of the domain, which in turn leads to improved problem solving skills in the domain (Mayer and Wittrock, 1996). The *interference hypothesis* is an argument from cognitive resources. Entertaining material which does not contain information pertinent to the lesson may distract the user from the learning task. Extraneous sounds or animations may divert the users’ attention from the important learning material, by overloading working memory with non-essential items. The constructivist hypothesis is supported by arguments that animated agents are social actors with whom users can engage (Reeves and Nass, 1996) , while the interference hypothesis is supported by Moreno and Mayers’ earlier findings (Moreno and Mayer, 2000) that extraneous music and sounds in a multi-media presentation hindered users’ recall and understanding of the educational material.

Moreno et al. (2001) measured the effect of animated agents on users’ attitudes, retention of material, and problem solving transfer. Two experiments with college age and 12-13 year old users showed that both groups preferred to use the version of Design-A-Plant with Herman the Bug, rather than equivalent software which presented the same information without the agent. In addition, there is evidence that users learned more deeply (as measured by problem solving transfer tasks) when using the agent version. These results support the constructivist hypothesis rather than the interference hypothesis – the presence of a social actor in the environment motivated the users and encouraged them to work harder to understand the material. This also suggests that a computer environment with an animated agent functions more like a social actor than a computer environment with a normal graphical user interface.

Moreno and colleagues’ next three studies investigated aspects of the animated agent environment which could potentially impact on the users’ motivation and learning: user participation, the visual presence of the agent, and the modality in which the agent’s advice was presented. In a comparison of an agent version of the software in which users participated in the design of a plant to an agent version in which users did *not* participate in designing a plant, users who participated were better able to recall facts about plants and solve difficult problem solving tasks. Attitudes were not significantly different between those who participated and those who did not. A study which aimed to identify the relative contributions of voice and visual appearance to the social agency effect of animated agents suggested that voice is an important factor in creating social presence, while the visual appearance of the agent has less of an impact. College students interacted with Design-A-Plant in one of four conditions: a version with a visual animated agent and narration of information through speech; a version with a visual animated agent and information presented via on-screen text; a version with no visual image of the agent but with information presented by narrated speech; and a version with no visual image of the agent and information presented via on-screen text. Students interacting with versions of the Design-A-Plant environment in which the agent’s image was deleted throughout displayed no significant differences in attitude, information retention or problem solving skills to their counterparts who used versions with the visual agent. This result supports neither the constructivist nor interference hypotheses, as the visual image of the agent neither increases nor detracts from users’ motivation or learning. However, there was an effect for voice – users who interacted with a version in which information was narrated did better on retention tests,

performed better on transfer tasks and rated the program more highly in attitude tests. This results supports the constructivist hypothesis, indicating that voice is an important factor in inducing social presence. These findings were repeated in a second experiment in which the visual images of the animated agent Herman the Bug were replaced with a video of a human, and Herman's voice was replaced with a recording of a human. As the users did not learn better or become more motivated by the visual images of the human, but were more motivated and learned better when listening to the human voice, this study confirms that voice is an important component of social presence. Even although the users who interacted with the human agent were presented with socially plausible gestures and eye-contact of another person, it made no significant difference to their attitude or behaviour compared to those interacting with a normal graphical user interface.

Note that these results were from studies with college students; it may be the case that visual image and voice have more or less of an impact on younger users' learning experience.

STORYSTATION

Background

StoryStation is an intelligent tutoring system designed to assist children with story writing (Robertson and Wiemer-Hastings, 2002), based on the prototype system Select-a-Kibitzer (Wiemer-Hastings and Graesser, 2000). It is aimed at pupils who have reached a basic competency with the mechanics of writing, but require help to improve specific story writing skills. The target age group is 10-12 years old. StoryStation was developed by a design team consisting of eight 10-12 year old pupils, two teachers and a researcher. Pupils and teachers were consulted throughout the design process, and were involved in requirements gathering, critical evaluation, design, animation, and pilot testing activities (see Robertson, 2002). The software design was also informed by the aims and objectives of the Scottish National English Language Curriculum, which Scottish schools are required to teach.

Writing a story is a complicated task, which many children find difficult. It has been characterized as a problem solving process (Flower and Hayes, 1980), as a creative design process (Sharples, 1998), and as "the negotiated construction of meaning" (Flower, 1994). In Sharples' model of writing as design, three main activities take place during the composition process. These are: *planning*, *composing*, and *revising*. The writer swaps between these activities at different stages of writing, and the proportions of time spent in each activity and the sequence of activities varies between writers. While composing and revising, the writer must consider many constraints at different levels, from high level concerns such as a potential reader's reaction, to low level surface features such as spelling. The problem for learner writers is that the process of satisfying these multiple constraints can be overwhelming. Flower (1994) likens the constraints to multiple, possibly conflicting "voices" which lead the writer in different directions. Learner writers must develop the meta-cognitive skills to switch their attention between the different voices at appropriate points in the composition process. StoryStation provides help with several aspects of writing, but the pupil is in control of deciding which help is appropriate for them at any given time.

Asking for help requires the pupil to switch attention from the composition phase of the writing process to the revising phase. This switch between phases in the problem solving process represents a meta-level skill necessary for good writing. Teachers involved in the StoryStation project have mentioned that it is difficult to persuade lower ability children to edit their work; once they have written the story, they consider themselves to be finished. Sharples (1998) explains this sort of observation with theory that children under 10 years old lack the cognitive skills to monitor and control their own language use; older children at earlier stages of writing development may suffer from the same difficulty. StoryStation is designed to encourage pupils to review and revise their work by providing positive feedback.

In addition to finding writing difficult, pupils often find it daunting and demoralising – they struggle with writing apprehension (Robertson, 2001). Madigan, Linton and Johnson (1996) describe writing apprehension as anxiety, self-deprecating thoughts and concern about how written work will be received. For sufferers of writing apprehension, writing is “an unpleasant, unrewarding activity that they actively avoid (Madigan, Linton, and Johnson, 1996 ; p 295). In the UK, there is concern among educators that boys in particular suffer from low motivation towards writing tasks, and that they under-perform on written assessments. The most recent data from the UK Department of Education and Skills shows that only 60% of eleven year old pupils met government performance targets for standard writing tests (DfES, 2003). The breakdown of the results by gender reveals that 52% of boys performed adequately in comparison to 69% of girls. Concern about boys’ underachievement in writing lead the UK Office for Standards in Education to study the methods of writing instruction used in schools where boys do well in order to extract general guidelines for raising boys’ motivation and achievement (OfStEd, 2003). They characterize schools where boys do well in writing with a list of factors including: prompt, detailed feedback specifying both what was done well and what could be improved; a balance between teacher provided support and pupil independence; staged writing tasks incorporating time for feedback and review; and a school ethos which encourages boys to be proud of their written work. The StoryStation learning environment was designed to promote these features of good writing instruction, as described below. For this reason, it is anticipated that StoryStation may be particularly suitable for motivating boys and so the quantitative analysis reported in the Results section studies this issue.

Given the evidence that pupils have negative attitudes towards writing, it is particularly important to provide a motivating, positive writing environment for pupils which encourages them to engage in the writing process on more than one occasion. StoryStation was designed to provide constructive feedback, and to adapt feedback to users’ skills over time. This positive feedback is realised in different ways in the two versions of StoryStation. In the animated agent version, praise is given by animated characters, while it is provided in a solely textual form in the non-agent interface. Previous research by Moreno et al. (2001) suggests that the presence of a social actor in the interface in the form of an animated agent motivates students and encourages them to work harder to understand the material. For this reason, we wanted to evaluate the effect of advice presented by a social actor on pupils in our target user group and in the domain of writing.

The software is based on a philosophy of positive reinforcement which has been successfully used by an experienced teacher in the design team, and is consistent with the guidelines identified for good writing instruction by a recent government report (OfStEd, 2003). Pupils are praised when they display specific writing skills, and are encouraged to use these techniques in future stories. The expert teacher’s classroom experience was that acknowledgement and recognition of pupils’ achievements motivated them and encouraged them to take pride in and responsibility for the further development of their writing skills.

Software features

The support in StoryStation is of two types: tools and assessment features. The tools – a dictionary, a thesaurus, word banks and a word count facility- are designed to help children while they are planning or composing their stories. Pupils are often encouraged by their teachers to use thesauruses or list of “good words” to make their stories more interesting. StoryStation offers similar support in a digital form. The dictionary and thesaurus tools are implemented using WordNet, a semantic network developed at Princeton University¹. The word bank facility is a simple database of words related to story topics, editable by a teacher. For example, StoryStation currently has word bank entries for two particular stories and a more general bank of words to help children write spooky stories.

The assessment features are intended for the revision phase of the writing process. StoryStation can currently provide feedback on spelling, vocabulary usage and characterization

¹ Available from <http://www.cogsci.princeton.edu/~wn/doc.shtml>

skills. The vocabulary skills it can assess are: use of adjectives; use of connecting words (such as “and”, “but” or “because”); and use of unusual words. These skills were identified as important by the teachers who took part in the design process. Adjectives and adverbs in the text are identified using the Qtag part of speech tagger². Target connecting words, as specified by the teachers on the design team are identified by simple text search. At present there is no further grammar checking, although this is an obvious extension project. Unusual words are identified using a statistical algorithm which uses general word frequency information from the British National Corpus, and word frequency information for particular ability levels from a corpus of children’s stories.

The story writing skill of portraying characters is also considered to be an important part of the curriculum. StoryStation can assess the pupil’s descriptions of the character’s feelings, appearance, personality and dialogue between characters. It is based on a story analysis scheme described in Robertson (2001) which was derived from studying the characterization techniques used by both children’s authors and children at various stages of writing development.

The feedback is generated by a teaching rules module based on comparisons of linguistic features of a pupil’s story to her previous stories, or to the average performance of other pupils of the same ability level. The teaching rules specify which sort of feedback should be output by the user interface (e.g. praise or suggest improvement), although the actual textual content of the feedback is stored in a separate file which is editable by teachers. The student model keeps track of the pupil’s current performance by storing the quantitative linguistic measures generated by the assessment algorithms for spelling, vocabulary and characterisation. If the pupils has asked for feedback on the skill before while writing the current story, the feedback on her current use of that skill will be based on a comparison to the student model entry for the session. If not, the student models for previous sessions will be checked to see if there is any data on the pupil’s use of that skill in previous stories. If this information is not available, a student model which represents the average performance for pupils of that ability level will be consulted. These average values were derived from the analysis of a corpus of 400 children’s stories which were grouped by National Curriculum scores as assessed by teachers.

Interface

There are two versions of StoryStation. In the first version, each tool and assessment features is presented via a unique animated agent. In the second version, exactly the same feedback is presented in a text box on a more traditional graphical user interface. Both versions were designed by pupils in the design team. Figure 1 shows the main window in the agent version. The tools and assessment features can be accessed by clicking on the buttons along the right hand side and bottom of the screen. In this version, the buttons represent doors to the agents’ “houses”. For example, if the pupil clicks on the baby alien’s door the baby alien will appear along with the thesaurus window. Figure 2 shows the non-agent version of this interface. Help can be accessed by clicking on coloured buttons. All tools and assessment features have the same interface as the agent version, except that neither the agents nor speech synthesis are present.

² Available from <http://web.bham.ac.uk/O.Mason/software/tagger/>



Figure 1 – the main window in the agent version of StoryStation

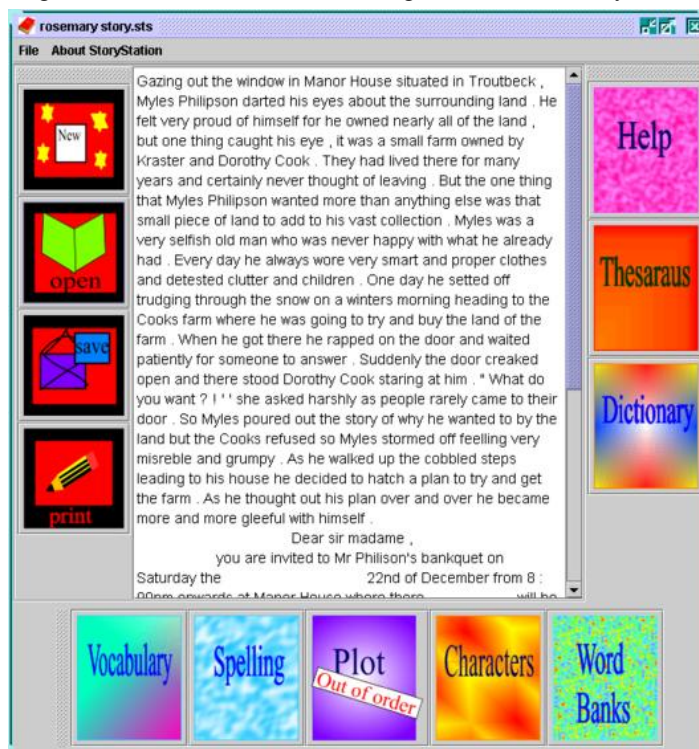


Figure 2 – the main window in the non-agent version of StoryStation

An example feedback window in the agent version in Figure 3. This agent is named Eye-Eye. His task is to assess the pupil's character descriptions. In this example, the pupil has requested assessment on her description of characters' feelings. Figure 4 shows the same feedback in the non-agent version. Note that in the agent version, the pupil can listen to a speech synthesizer reading out the text which is presented in the speech bubble. The same information is also present in the feedback text box. In contrast, the non-agent user sees only the text in the feedback text box. In both versions, the good example of "feelings" words have been highlighted blue in the story text window.

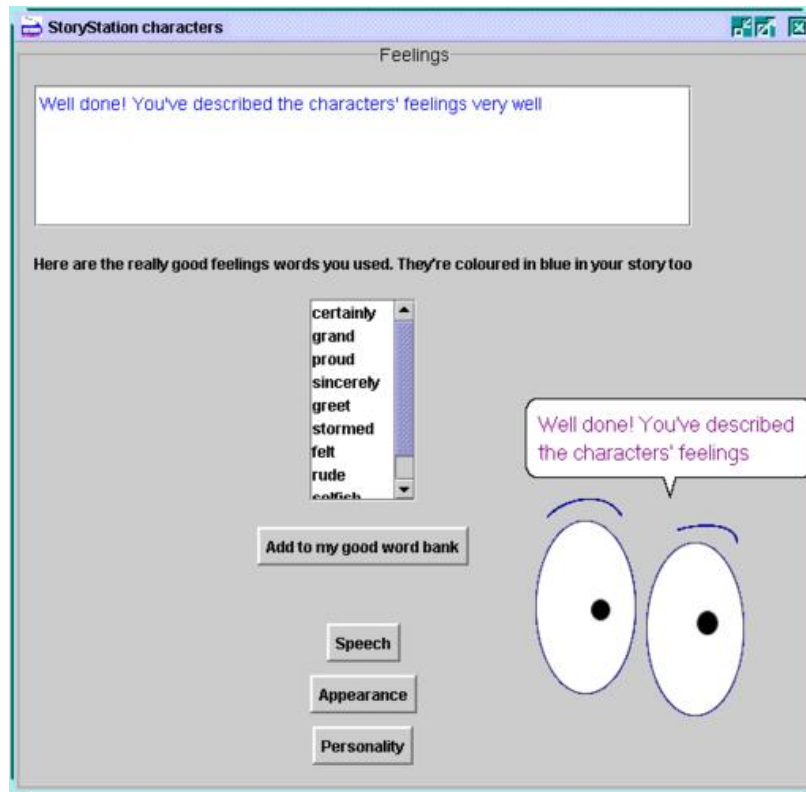


Figure 3- Feedback on descriptions of character's feelings in the agent version



Figure 4- Feedback on descriptions of character's feelings in the non-agent version

The agents used in the software were designed and animated by the children in the design team. We assumed that children of the same age as the target user group were more likely to be able to design agents which would appeal to their peers than the adults on the project. Further discussion of the advantages of involving children in the design process can be found in Robertson (2002). The children created them in the graphics package XaraX, before the researcher converted them to Microsoft Agents. Figure 5 shows the agents and their functions. Each agent was created by a different pupil and allocated to a function according to the pupils'

choice. No attempt was made to match the function to the appearance of the agent, as the functions are too abstract to map neatly to pictorial representations. Each of the agents has a “show” animation which is played when the agent first appears in screen, and an unobtrusive “idle” animation such as blinking. The agents which represent assessment features also have different animations which are played alongside positive and negative feedback from StoryStation. These are intended to convey pleasure or mild displeasure (it is important that the agents do not offend the pupils when they make mistakes). Note that some agents’ “moods” are easier to identify than others e.g. Jim’s facial expression can look happy or sad whereas Whiskers relies on body movements to convey his feelings. Eye-Eye can display a range of emotions by manipulating the position of his eyebrows. The children also selected voices for their agents by selecting accents, gender and pitch in the Microsoft Agent Character Editor. Three of the characters used the female voice offered by the synthesizer (Whiskers, Rabbit, and, surprisingly, Jim). The other five characters used the male voice, although the alterations of pitch and speed made them sound less obviously male.

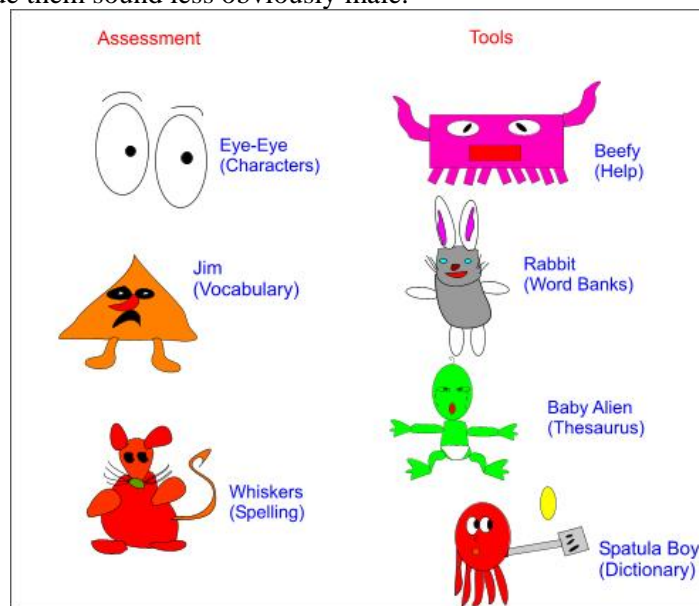


Figure 5 The StoryStation agents (not actual size)

METHOD

A field study was conducted to evaluate pupils’ attitudes to and interactions with the agent and non-agent versions of the StoryStation system. The purpose of this was to investigate the impact of animated pedagogical agents on a target user group of 10-12year olds in the domain of story writing. If the findings from previous studies about the motivational benefit of animated agents were replicated this would be of interest to educators, as low motivation is a problem in writing instruction. As boys are considered to be particularly at risk from negative attitudes to writing, and therefore low attainment in writing assessment, it was of interest to discover motivating factors in the software which influenced boys in particular. Previous empirical studies on animated pedagogical agents have not addressed gender differences in user attitudes to agents. This study was intended to establish whether:

1: Agent users have different attitudes to StoryStation than users of the non-agent version. Does the presence of animated agents influence the pupils’ enjoyment of the task, or their inclination to use the system again?

2: Agent users interact with StoryStation in a different way to non-agent users. Does the presence of animated agents have an impact on how often users request feedback from the system, or how often they use the software tools?

3: Male and female users have different attitudes to, and interactions with, StoryStation. Do boys find agents generally any more or less motivating than girls? Do they tend to interact with them more or less frequently?

This study was not intended to investigate the effects of the animated agents on the development of pupils' writing skills; further longer term study is required for this. A between-subjects design was preferred over a within-subjects design for two main reasons. From a practical perspective, in running a field study in a busy school, it is necessary to co-operate with teachers' existing classroom timetables as far as possible. It suited the teachers involved in the study better to release all the children from class work on one occasion, than to release them on two separate occasions. Secondly, from a more theoretical standpoint, we believed that there were potential ordering effects which could not be easily counterbalanced. We were interested in pupils' attitudes towards the agents and previous empirical results suggested that people respond to agents in a social way. However, their responses to the agents in StoryStation might depend on whether they had previously encountered advice from StoryStation before in a non-textual form. They might be less inclined to view the agents as social actors capable of dispensing advice if they knew that the program could give the same advice without agents. The between subjects design was intended to make it clearer that any effects were due to manipulation rather than to prior exposure to the software.

Participants

Participants in the study were 59 ten – twelve year old pupils at a middle class suburban state funded primary school. Thirty one of these pupils were boys, twenty eight were girls. All pupils spoke English as their first language. The school has a good reputation in the area, particularly for writing instruction. The pupils were drawn from three classes, one of which had two teachers who shared the job. Three of the teachers were female, and the teacher of the oldest class was male. Pupils took part in the study if they returned a parental consent form, and if time tabling constraints allowed. One girl did not complete the session because she chose to practice for the school concert instead.

The pupils regularly use iMac computers as part of their school work; there is a ratio of 80 computers to a school population of 400. Pupils usually use computers to type out a fair copy of their finished stories after they have been approved by the teacher. The open plan layout facilitates easy access to computing equipment under the supervision of teachers or classroom assistants. Furthermore, all the pupils have a computer at home. Most pupils used PCs (rather than macs) at home for a variety of tasks including story writing, game playing, web searching, emailing and sending instant messages.

Procedure

The field study took place in a small computer room next to the open plan computer labs. The pupils used StoryStation on PC laptops provided by the research project. Three pupils took part in the study at a time, each using a separate computer.

The pupils were initially introduced to the project by the researchers who explained that they were going to use a program which was designed to help them with their writing. They were that the researchers would like to know their opinion of the software and how it could be improved. It was emphasized that it was not a test of their writing abilities, but an opportunity for them to share their opinions about the program. The researchers did not mention who created the software.

The pupils then listened to a ten minute recording of one of the researchers (who is also a children's storyteller) telling a version of the legend "The Screaming Skulls of Calgarth". After the story, the researcher checked that they understood the story and explained that their task was to write their own version of it. It was emphasized that they could improve the story and change the details if they wanted to.

The researcher demonstrated the appropriate version of StoryStation and checked that the pupils understood the functionality. Then the pupils used StoryStation to write their story. Time

on task varied from 35 minutes to 85 minutes; this unfortunate variation could not be avoided due to time tabling constraints. If the pupils asked for help, the researchers first suggested that they use StoryStation to answer the question. If the pupil still had difficulty, the researcher provided assistance. This assistance was mostly clarifying instructions, reminding pupils of the story plot, or generally re-assuring them. Three pupils with learning difficulties received further help and typing assistance as was normal classroom practice, but their data were removed from the analysis. The researchers also provided general encouragement and praise to the pupils.

After using StoryStation, the pupils filled in a visual analogue questionnaire to assess their attitudes to the software. One of the researchers explained how to fill it in, and tested their understanding of the scale with an example question unrelated to the software. She then read out the questions, pausing to allow the children to fill in their answers. She did not look at the pupils' answers, nor did the children appear to copy each other. The pupils were encouraged to ask for clarification if they did not understand a question, although none did. This procedure was intended to minimize confusion and consequently misrepresentative answers.

Finally, the pupils took part in group based semi-structured interviews. The pupils discussed their opinions of the agents (if they used the agent version of the software), and contrasted their classroom experiences of writing to their experience with StoryStation. The researchers thanked the pupils before they returned to the classroom. Pupils received a print out of their story the next day.

Measures

Attitude

The pupils' attitudes to StoryStation were measured using a visual analogue questionnaire. The statements in the scale covered issues relating to the pupils' enjoyment of using the software, their perceptions of its effect on their writing and its usability. These questions evaluate the extent to which StoryStation motivated pupils to write stories. As discussed in Section 3, pupils can find writing a demoralising task which they seek to avoid. The questions were intended to establish whether pupils found using StoryStation a negative experience, and whether it encouraged them to use it again. The questionnaire consists of the following ten statements.

1. I enjoyed using StoryStation
2. I think StoryStation made my writing worse
3. I think I would like to use StoryStation again
4. Using StoryStation help me to write better
5. I found StoryStation confusing
6. StoryStation makes writing stories easier
7. I think I need someone to help me use StoryStation
8. I think the StoryStation advice was useful
9. A teacher is more helpful than StoryStation
10. StoryStation is boring

Pupils were asked to place a cross on a line which started at "No" and finished at "Yes", to show how much they agreed with each sentence. The statements were carefully constructed to avoid awkward negatives, or complex language. The scale was tested and refined through a StoryStation pilot study with eight pupils from another school.

Interactions with the Tutoring System

Interactions with StoryStation were automatically recorded in log files. The system recorded every request for help from the pupils, as well as the feedback provided by StoryStation. The requests for help were categorized into two main types for analysis purposes: tool usage and requests for feedback. Help requests categorized in the tool usage type were: thesaurus lookups, dictionary lookups, word bank lookups and word counts. Requests for

feedback were interactions with the spell checker, requests for assessment of unusual words, describing words, joining words, character feelings, appearance, personality and dialogue. In order to compensate for the variation in time on task, the raw scores for total interactions, requests for feedback and instances of tool interactions were normalised by the time on task. The measures used in analysis were total interactions per 10 minutes; feedback requests per 10 minutes and tool interactions per 10 minutes.

RESULTS

Attitudes

The results of a two way multivariate analysis of variance with software version and gender as the independent factors, are discussed below. The dependent measures were the visual analogue scale statements reported in the previous section. There were 32 users of the agent version (19 boys and 13 girls), and 22 users of the non-agent version (10 boys and 12 girls) included in the analysis. Data from four pupils who did not answer all of the questions were removed. In addition, after testing for outliers, the data from a boy in the non-agent version whose answers were inconsistent were removed. The descriptive statistics of the measurements of the visual analogue responses in centimetres are shown in Table 2. Note that the extreme “no” end of the scale was measured at 0cm, and the extreme “yes” end of the scale is at 8cm.

It can be seen that the pupils enjoyed using StoryStation, and did not find it boring. They found the StoryStation advice useful, indicated that it made it easier to write stories, and that their stories were better. On usability issues, they did not find StoryStation confusing, did not think that they needed someone to help them use it, and would like to use it again. Pupils did not appear to think that a teacher is more helpful than StoryStation.

Table 1. Descriptive statistics of visual analogue scale responses

	Means				Standard Deviation			
	<i>Agent</i>		<i>Non-agent</i>		<i>Agent</i>		<i>Non-Agent</i>	
	Girls	Boys	Girls	Boys	Girls	Boys	Girls	Boys
1. I enjoyed using StoryStation	7.79	7.2	7.02	7.19	.51	1.01	1.59	.93
2. I think StoryStation made my writing worse	.15	.59	.48	.68	.13	.9	.45	.95
3. I think I would like to use StoryStation again	7.87	7.46	6.83	6.86	.48	1.01	1.73	1.46
4. Using StoryStation help me to write better	5.70	5.80	5.33	6.2	.66	2.35	2.53	1.90
5. I found StoryStation confusing	.338	.80	1.78	1.17	.48	1.04	2.34	1.92
6. StoryStation makes writing stories easier	7.62	6.07	6.81	6.27	1.3	2.1	1.57	2.53
7. I think I need someone	1.18	1.44	.625	.85	1.66	1.63	.68	1.1

to help me use StoryStation

8. I think the StoryStation advice was useful	7.14	6.86	6.23	6.78	.2	1.29	2.3	1.47
9. A teacher is more helpful than StoryStation	1.89	2.61	2.22	2.88	2.23	2.56	1.75	2.33
10. StoryStation is boring	.9	.69	.96	.71	2.21	1.34	1.25	1.01

There was a significant effect for software version on the statement “I think I would like to use StoryStation again” ($F = 5.79$; $p = .02$). Users of the agent version indicated more strongly that they wanted to use the software again.

There was a significant effect for software version on the statement “I found StoryStation confusing” ($F = 4.51$; $p = .04$). Users of the non-agent version indicated more strongly that they found StoryStation confusing.

There was a significant effect for software version on the statement “I think the StoryStation advice was useful” ($F = 5.07$; $p = .03$). Users of the agent version indicated more strongly that they found the advice useful than non-agent users. However, this main effect must be interpreted with caution, as there is a significant software version by gender interaction for this statement ($F = 3.89$; $p = .05$). Girls are more likely to find the advice of the agent version more useful than the non-agent version, as shown in Figure 6.

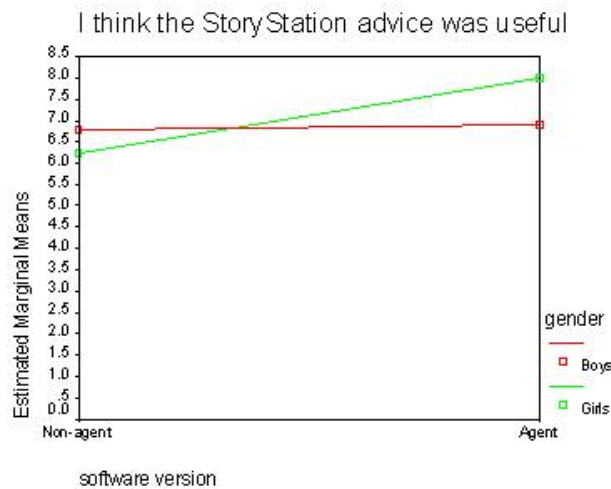


Figure 6 . Interaction between software version and gender in responses to “I think I would like to use StoryStation again”.

There is a significant interaction for gender and software version for the statement “Using StoryStation helped me to write better” ($F = 3.89$; $p = .05$). As can be seen in Figure 7, girls are more likely to believe more strongly that the agent version of StoryStation helped them to write better. Boys are more likely to believe more strongly that the non-agent version helped them to write better.

feedback
per 10
minutes
Tool usage
per 10
minutes

2.43	1.45	1.69	1.41	1.74	.84	1.03	.58
------	------	------	------	------	-----	------	-----

Software version and gender had no significant effect on the total interactions ($F = .127$; $p = .724$ and $F = 1.61$; $p = .211$ respectively). There was a significant interaction between software version and gender ($F = 5.92$; $p = .019$).

There were no significant effects for software version or gender on the requests for feedback ($F = 1.63$; $p = .207$ and $F = .01$; $p = .92$ respectively). The interaction between software version and gender approached significance ($F = 3.59$; $p = .06$).

There were no significant effects for software version or gender on the instances of tool usage ($F = .76$; $p = .39$ and $F = 3.39$; $p = .07$ respectively). There was no significant interaction between software version and gender ($F = 1.80$; $p = .19$).

Note that the covariate, ability level, had no significant effect on total interactions, requests for feedback or instances of tool usage.

Figure 8 shows the differences between groups for total interactions with StoryStation graphically. It can be seen that the girls interacted with StoryStation more frequently when using the agent version, while the boys interacted more frequently when using the non-agent version.

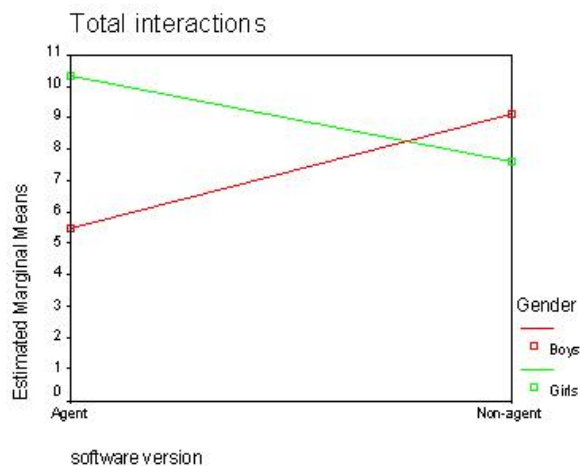


Figure 8. Total interactions with StoryStation by software version and gender

Figure 9 shows the interaction between software version and gender on request for feedback. It shows that boys using the non-agent version were more likely to ask for feedback than girls in the non-agent group. It also shows that girls using the agent version were more likely to ask for feedback than boys using the agent version.

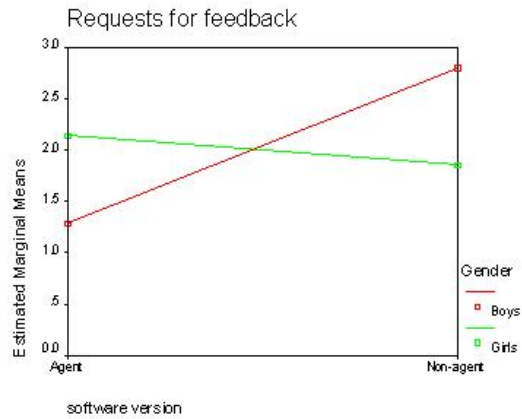


Figure 9. Requests for feedback by software version and gender.

Figure 10 shows instances of tool usage by gender and software version. It can be seen that girls are more likely to use the StoryStation tools than boys, particularly when using the agent version of the software.

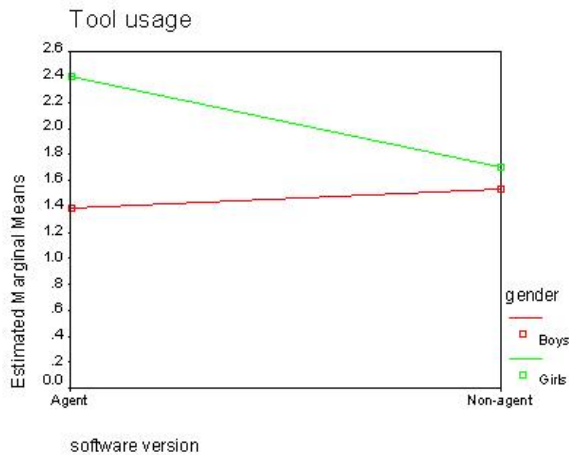


Figure 10. Instances of tool usage by software version and gender

In summary, there was a significant interaction between software version and gender for the total interactions between the pupils and StoryStation. Girls were more likely to interact when using the agent version of StoryStation, while the boys were more likely to interact when using the non-agent version.

Results Summary

Evidence from the visual analogue scale responses suggests that users preferred the agent version of the software. They were more likely to want to use the system again and were less likely to find it confusing.

However, there is no evidence that use of the agent version results in more interaction with the tutoring system for all users. There is some evidence that there is an interaction between

software version and gender, suggesting that girls are likely to interact more when using the agent version, while boys are more likely to interact with the non-agent version. This is consistent with software version and gender interactions in the responses to the visual analogue scale which show that girls are more likely to rate advice from the agent version highly. It is also consistent with the finding that girls are more likely to believe more strongly that the agent version helped them to write better, whereas the boys are more likely to believe more strongly that the non-agent version helped them to write better.

DISCUSSION

Pupils were more likely to want to use the agent version of StoryStation again. This result suggests that pupils have a preference for the agent version of StoryStation. However, it is interesting to note that there is no significant difference between the levels of enjoyment expressed by agent and non-agent users. Both groups indicated that they enjoyed using StoryStation very much. This is possibly due to general youthful enthusiasm in both groups, or excitement at the opportunity to compose (rather than simply type out) a story on the computer. It would seem that the results about the perceived usefulness of the StoryStation advice, and the desire to use it again, point to a more subtle effect from the presence of the agents than simple enjoyment. A possible explanation of the effect is that the agent users believe they formed relationships with the agents which made them more eager to engage with the program again. These relationships possibly add some emotional meaning to the StoryStation feedback which leads the children to place more value on the advice and, in the case of girls, rate its impact on their writing skills more highly.

It is interesting to note that it was the non-agent users who were more likely to find StoryStation confusing, even although the agent version suffered from more technical problems and crashes. This suggests that technical problems in the system did not have an impact on the agent users' understanding of the software. There are at least two possible reasons why the agent users found StoryStation less confusing. Firstly, there are more visual cues in the interface to help them to remember where to find the help features. The agent users can associate each help feature with a unique animated character. Recall of these characters is reinforced by a drawing on the appropriate help button. In contrast, the non-agent users have only different colours to distinguish between help buttons. Secondly, it may be the case that the pupils find the idea of receiving advice from a character less confusing than receiving advice from a computer. They are used to hearing advice from their teachers; the experience of getting advice from an animated "person" is perhaps more similar to previous experiences.

The attitude findings reported above are generally consistent with those reported in Moreno et al (2001). In a study of seventh grade students' attitudes to agents, Moreno and colleagues found that students who used the agent version of the software enjoyed it more, although they were no more likely to want to use it again than non-agent users. The results about pupils' interactions with the tutoring system are not comparable with Moreno et al.'s results about transfer and retention of problem solving skills; further study is required to assess the impact of agents on pupils' learning with StoryStation. No gender effects are reported in this publication, so it is not possible to compare the gender trends discovered in the present study.

The gender trends uncovered in this study warrant further investigation: why do boys and girls have different attitudes and interaction patterns with agents? It is worth noting that these gender trends may be related to other habits or behaviours associated with each sex, rather than attributes of each sex per se. For example, it is possible that the boys in this study were more likely to be experienced computer users than the girls, and so that the difference in attitudes to the agents were in fact related to novice versus expert perception of animated characters. In fact, all the pupils used computers at school and at home and neither gender appeared to be more familiar with animated agents. Two other possible explanations arise from analysis of the semi-structured interview data.

One explanation is that boys did not like the particular agents used in this study as much as girls. During the interviews, some boys noted that they would prefer cartoon style agents such

as Bart Simpson or characters from the Beano. Others remarked that the agents looked as if they were “for kids”, and commented that their younger brothers or sisters might like to use them. Analysis of the questionnaire data by age group and gender suggests that the oldest age group of boys (12 year olds) responded least well to the agents. When asked to comment on this, a twelve year old boy from the StoryStation design team said “Well, you wouldn’t want your friends to see you playing with the helpers on screen. They’d laugh at you” It is not clear whether it is embarrassing to be seen playing with any agent, or whether the particular StoryStation agents are a problem. Perhaps human-like cartoon characters would have been more suitable for older boys than the childish looking StoryStation agents. Further study is required to investigate this possibility.

Another possible explanation comes from boys’ existing classroom relationships. During the semi-structured interviews boys commented that they were embarrassed to ask their teacher for help, especially if they had forgotten her previous advice. Some of them mentioned that the teacher might shout at them if they asked for help. StoryStation offers privacy for users – no-one need know how many times the user has asked for help, and it does not chastise them for forgetting advice. Educationalists have noted that for some boys, it is important not to be seen to try and fail in public: they would rather that their failures were attributed to lack of effort than lack of aptitude (Jackson, 2002). It is possible that the non-agent version of StoryStation offers more privacy than the agent version. In the agent version, there are 8 personas to witness the pupil’s mistakes and judge his work. The computer without animated personas and synthesized voices, is less of a social actor and therefore possibly more acceptable to those who are demotivated by scrutiny of their efforts.

Boys also switched more fluently to talking about technical aspects of the software. They seemed more motivated to come to grips with technical features and less interested in suspending disbelief to immerse themselves in a pretend world. They seemed to want to immerse themselves in was a technical world, rather than a make believe world where animated characters are real. In this view, the agents could have been seen as a distraction and impediment.

Although these explanations begin to suggest why boys using the non-agent version had higher motivation and more interaction with the tutoring system, more research is required. A follow up study exploring pupils’ preferences for the visual style of agents may reveal whether boys are less motivated by agents in general, or the StoryStation agents in particular. It is important to note that the non-agent condition was a positive experience for boys, and given their willingness to interact with the tutoring system in this condition, it is likely that it would have a positive effect on their story writing skills. This is an important finding because of educators’ concern with boys’ under-achievement in literacy, particularly in the area of writing (see Background section). Therefore, a tutoring system which appeals to boys and encourages them to develop writing skills offers a new approach to raising literacy standards in this problem area.

In view of Moreno et al’s (2001) findings regarding the importance of auditory feedback, it would be interesting to systematically explore this modality in StoryStation. Users in this study usually chose not to wear their headphones to listen to the agents’ voices after the first couple of minutes. In fact, users tended only to wear the headphone when there was a distracting level of background noise. It is possible that voices are less important to users when they present feedback rather than new instructional material. As some of the pupils mentioned that they would have preferred recorded human voices to match the agents’ appearances rather than synthesized robotic voices, voice quality may also have a strong effect on this age group.

A high priority for future research with StoryStation is to investigate the impact of the agent and non-agent versions on the pupils’ story writing skills. This will be accomplished by a longitudinal pre-test post-test design at a local primary school over a period of eight weeks. It would be interesting to confirm whether the positive effects of agent on botany problem solving transfer tests, reported by Moreno et al., (2001) would also apply to the domain of writing. As cognitive theories of writing (Flower and Hayes, 1980) describe writing as a problem solving process, it is reasonable to hypothesise that agents would also have a positive effect on pupils’ writing skills. However, given the interaction between software version and gender discovered

in pupils' interactions with the tutoring system, it is probable that agents will have a more positive effect on girls' writing skills.

CONCLUSIONS

This paper presented the findings of a field study conducted to evaluate the impact of agent and non-agent versions of the StoryStation software on 10 -12 year old pupils. In confirmation of findings of previous studies (Moreno et al., 2001; Lester et al., 1997), users rated the agent version of the system more highly. In particular, they were more likely to want to use it again, rated the advice more highly and found it less confusing. Girls were more likely to rate the advice of agents highly, and to think it improved their writing more than boys.

Software version had no significant effects upon the users' interactions with StoryStation . However, there was a significant interaction for gender and software version; girls were more likely to interact with the agent version, while boys were more likely to interact with the non-agent version. This is consistent with the gender differences in users' attitudes as reported in the attitude questionnaires.

This results of this study raise questions about varying patterns in different learners' interactions with agents. Further research is required to gain an understanding of the ways in which agents can be used positively within the specific context of learners' cognitive skills and social roles.

ACKNOWLEDGEMENTS

StoryStation was funded by an EPSRC grant. The authors would like to thank the pupils and teachers at Blackhall, and Sinclairtown primary schools, in Scotland. Morag Donaldson, Joe Beck, Jack Mostow and Senga Munro also provided assistance.

REFERENCES

- Biswas, G., Katzlberger, T., Bransford, J., Schwartz, D., & The Teachable Agents Group at Vanderbilt (TAG-V). (2001). Extending intelligent learning environments with teachable agents to enhance learning. In J. Moore, C. Redfield & W. Johnson (Eds.), *Artificial Intelligence in Education - AI-ED in the Wired and Wireless Future* (pp. 389-397). Amsterdam: IOS Press.
- P.Brna, B.Cooper, L.Razmerita, "Marching to the wrong distant drum: Pedagogic Agents, emotion and student modeling", Workshop on Attitude, Personality and Emotions in User-Adapted Interaction in conjunction with User Modeling 2001, Sonthofen, Germany, July 2001.
- Cerri, S. , Gouarderes, G. and Paraguacu, F. (2002) *Intelligent Tutoring Systems: 6th International Conference, ITS 2002*. Biarritz, France and San Sebastian, Spain, June 2002 Proceedings.
- Dehn, D. and van Mulken, S. (2000). "The impact of animated interface agents: a review of empirical research". *International Journal of Human-Computer Studies* 52 1-22.
- DfES (2003). "Autumn Package: 2003: Key Stage 2 National Summary Results". Department for Education and Skills Pupil Performance Data. <http://www.standards.dfes.gov.uk/performance/ap/>. Retrieved on 23/10/03.
- Duran-Huard, R. and Hayes-Roth, B. (1996) *Children's Collaborative Playcrafting*. Knowledge System Laboratory Report No. KSL 96-17. Department of Computer Science, Stanford University.
- Flower, L. (1994). *The construction of negotiated meaning: A social cognitive theory of writing*. Southern Illinois University Press, Carbondale, EL.
- Flower, L. and Hayes, J. (1980). The dynamics of composing: making plans and juggling constraints. In Gregg, L.W. and Steinberg, E. R., Editors, *Cognitive Processes in Writing*. Lawrence Erlbaum Associates.
- Jackson, C. (2002) "Laddishness as a Self-Worth Protection Strategy", *Gender and Education*, 14(1): 37-52.

- Johnson, L., Rickel, J. and 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., Bhoga,R. (1997). "The Persona Effect: Affective Impact of Animated Pedagogical Agents" in *Proceedings of CHI 97, Atlanta*.
- Madigan, R. , Linton, P. and Johnson, S. (1996). *The Paradox of Writing Apprehension*. In Levy, C. and Ransdell, S. (Editors) *The Science of Writing: Theories, Methods, Individual Differences and Applications*. Laurence Erlbaum Associates, New Jersey.
- Mayer, R.E., Wittrock, M.C.(1996). Problem-Solving Transfer. In *Handbook of Educational Psychology*. New York: Simon &Schuster Macmillan. P. 49.
- Moreno, R. , Mayer, R., and Lester, J. (2001) "The Case for Social Agency in Computer-Based Teaching: Do Students Learn More Deeply When They Interact With Animated Pedagogical Agents?" *Cognition and Instruction* 19(2).177 -213.
- Moreno, R. & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology*, 97, 117-125.
- OfStEd (2003). "Yes he can: Schools where boys write well". Office for standards in Education. HMI 505.
- Sharples, M. (1998). *How We Write: Writing as Creative Design*. Routledge, London.
- Reeves, B. and Nass, C. (1996) *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Cambridge University Press, September 1996.
- Robertson, J. (2002) "Experiences of Child Centred Design in the StoryStation Project" In Bekker, M., Markopoulos, P. and Kersten-Tsikalkina, M. (Ed.s). *Proceedings of Workshop on Interaction Design and Children*, Eindhoven.
- Robertson, J. (2001). *The Effectiveness of A Virtual Role-play Environment as Story Preparation Activity*. Unpublished PhD thesis, University of Edinburgh.
- Robertson, J. and Wiemer-Hastings, P. (2002). "Feedback on children's stories via multiple interface agents" In the proceedings of the International Conference on Intelligent Tutoring Systems, 2002, Biarritz, France.
- Ryokai, K., Vaucelle, C., Cassell, J. (2003) "Virtual Peers as Partners in Storytelling and Literacy Learning" *Journal of Computer Assisted Learning* 19(2), 195- 208.
- Wiemer-Hastings, P and Graesser, A. (2000) "Select-a-kibitzer: A computer tool that gives meaningful feedback on student compositions". *Interactive Learning Environments*, pp 149-169.