

# When sex, drugs, and violence enter the classroom: Conversations between adolescents and a female pedagogical agent

George Veletsianos\*, Cassandra Scharber, Aaron Doering

*University of Minnesota, College of Education and Human Development, Curriculum & Instruction: Learning Technologies, 150B Peik Hall,  
159 Pillsbury Drive S.E., Minneapolis, MN 55455, USA*

Available online 15 February 2008

## Abstract

In this article, we investigate the discourse between a female conversational pedagogical agent and 59 adolescents in the context of a social studies lesson. We note that previous pedagogical agent research has focused on the positive effects of agents, while failing to take into account the intricacies of learner–agent discourse, and subsequently missing the abuse suffered by pedagogical agents at users’ fingertips. Our analysis indicates that learners readily misuse and abuse pedagogical agents while placing them in a subordinate and inferior role. We conclude by making recommendations on agent design and future research.

© 2008 Elsevier B.V. All rights reserved.

**Keywords:** Pedagogical agents; Conversational agents; Agent abuse; Agent misuse; Computer-mediated discourse analysis; Social studies

## 1. Introduction

Pedagogical agents are conversational and non-conversational virtual characters employed in educational settings to serve various instructional purposes. For instance, Payr (2003) notes that virtual characters can be employed as teachers, tutors, coaches, learning companions, and actors, in essence reenacting the multiple roles played by real-life instructors. Not only are pedagogical agents able to enact multiple instructional roles, but they have been employed in numerous content areas as well. For example, Penelope and Alexander portray themselves as electronic portfolio experts available to assist learners with all aspects of developing an electronic portfolio (Doering et al., 2008). Other examples include AutoTutor who converses with learners on physics and computer literacy (Graesser et al., 2004), and Laura who attempts to encourage users to engage in physical activity (Bickmore and Picard, 2005).

New technologies (such as wikis, blogs, and pedagogical agents) often bring with them the expectation that they will

revolutionize learning (Bull et al., 2005). Thomas Edison believed that motion picture would transform our educational system (Brooker, 1947). Seymour Papert (1984) held the same views regarding microcomputers. In a similar vein, educational technology researchers appear to be overly enthusiastic regarding the possibilities afforded by pedagogical agents, even though it appears that there is no compelling experimental evidence for their learning benefits (Choi and Clark, 2006). It is concerning that educational technology researchers have not taken a long and deep look at exactly what happens when learners interact with agents. It appears that the focus has been on the benefits of pedagogical agents on affective issues (such as student motivation) rather than student outcomes and what actually occurs when students converse with agents. For example, the January–February 2007 special issue of *Educational Technology* focuses on pedagogical agents and presents them in an overly positive light “within this exciting and quickly-evolving field” (p. 4). Even more concerning is the fact that it is only recently that researchers have examined the evidence surrounding the claimed positive impact of pedagogical agents and found that such evidence is contradictory and at best mixed (Gulz, 2004). The focus

\* Corresponding author. Tel.: +1 612 626 8276; fax: +1 612 624 8277.  
E-mail address: [velet006@umn.edu](mailto:velet006@umn.edu) (G. Veletsianos).

on the perceived benefits that pedagogical agents may bring in learning contexts appears to have brushed aside the possible shortcomings of this tool.

One of the limitations of pedagogical agent implementations not examined in the educational technology literature, and briefly touched upon in the human–computer interaction literature, is the topic of agent abuse and off-task behavior. Learner–agent interactions appear to encompass a “darker side” – one where the metaphor of the agent as an instructor, tutor, and learning companion succumbs to the visual of the agent as a mistreated subordinate object. The “darker side” of learner–agent interactions bears no clear-cut linkage to education, learning, and teaching in the way that educational researchers hope. The novelty of this paper therefore, lies on the fact that the issue of agent abuse *in the context of educational software* has, so far, been left largely ignored and, as a result, unexplored.

To investigate learner–agent interactions, we focus on Conversational Pedagogical Agents (CPAs) and the free-form dialogue between agents and students. Specifically, we investigate the abuse CPAs suffer by examining adolescents’ discourse with a female pedagogical agent in the context of a social studies lesson. Our investigation focuses on one lesson with one agent and multiple students, enabling us to collect and contextualize *all* conversations between agent and learners. We first examine work related to pedagogical agents and virtual character abuse. While examining such work we draw on theoretical notions of cyber sexuality, psychosocial development, anonymity, and online inhibition to illuminate why learners may abuse pedagogical agents. We then present the focus of our study, our specific research questions, data, analysis, and empirical results. We conclude by examining the implications of this study and offering recommendations for future research and agent design.

## 2. Previous work

Educational technology researchers have claimed that pedagogical agents offer numerous benefits for teaching and learning. In a review of the existing literature, [Gulz \(2004\)](#) notes that previous research makes six claims regarding the use of such tools. Specifically, pedagogical agents can (a) increase motivation, (b) increase perceptions of comfort, (c) stimulate learning, (d) enhance information and communication flow, (e) fulfill personal connection to learning, and (f) enhance problem solving processes. Nevertheless, both [Gulz \(2004\)](#) and [Choi and Clark \(2006\)](#) note that the evidence surrounding these claims is at best mixed. On the other hand, [Baylor \(1999, 2000\)](#) and [Veletsianos \(2007\)](#) note that such tools can be of great benefit in educational contexts. For instance, agent gender has been shown to influence pedagogical efficacy and learning ([Moreno et al., 2002](#)), and animation and conversational capability appear to afford more opportunities for electronic learning with pedagogical agents than passive information delivery

([Mayer et al., 2003](#)). Clearly a consensus on the benefits or shortcomings of pedagogical agents is hard to reach ([Gulz, 2004](#)).

Prior to investigating the negative aspects of pedagogical agent deployments, it is important to note that the distinction between conversational pedagogical agents (CPAs) and non-conversational pedagogical agents (NCPAs) is not perfectly evident in the educational technology literature. The majority of available studies deal with NCPAs whose purpose is to deliver content to learners. Even though both types of characters can be termed *pedagogical agents*, we perceive the differences between the two tools to be of such magnitude that an analysis of pedagogical agents as a whole without discriminating between conversational and interactive capabilities would not do justice to either tool. Therefore, from here onwards we will focus only on CPAs.

In a longitudinal qualitative study of pre-service teachers’ experiences with two CPAs ([Doering et al., 2008](#)), we found that learners held mixed and often conflicting opinions on the CPAs. For instance, even though the majority of the learners perceived CPAs to be socially supportive, learners also found them academically incompetent. Although learners felt the CPAs were inept, they reported being motivated to revisit the CPAs throughout the 4-week duration of the study to seek assistance and support. This study also indicated the complexity of deploying a CPA in an online learning environment with the purpose of assisting learners in the completion of a task: Even though we expected learners to interact with the CPAs on issues that were unrelated to the course content, we were surprised to discover that the majority of student–agent interactions were unrelated to the assigned task. This finding was one of the motivating factors behind the current investigation of student–agent discourse. This factor was heightened when we were unable to locate any studies that examined student–agent discourse and the reasons behind such conversations. If off-task behavior represents a large part of student–agent interactions, a number of related questions naturally arise: What do students and agents talk about? What form do these discussions take? How do students treat agents? How do students perceive the agents’ role? How do students perceive their relationship with agents? What does the language used by students tell us about how agents are evaluated and perceived?

Pedagogical agents are usually viewed with the *media equation* lens ([Reeves and Nass, 1996](#)). The media equation argues that humans treat media as if they are also human, in essence interacting with media in the same way that humans would interact with each other. For instance, humans rate computers more favorably when computers praise the humans’ performance than when they do not. Additionally, [Nass et al. \(1997\)](#) found that participants applied gender stereotypes to computers even though the only suggestion of gender was vocal cues. Even though virtual character researchers have largely embraced the media equation program, some express their dissatisfaction with it. [Shechtman and Horowitz \(2003\)](#) note that the results

of the media equation program were based on user self-reported data rather than on an investigation of conversational interaction and behavior between humans and media. These authors further argued that the behavioral, cognitive, emotional, and motivational processes that occur between humans–humans and between media–humans are inherently different. To examine this hypothesis, Shechtman and Horowitz analyzed conversations between participants and an apparently-human or apparently-computer partner. The results of this analysis indicated that humans respond differently to humans than to computers. For instance, humans use more relationship statements and put in more effort when they are under the perception that they are conversing with a human than with a computer. The implication of these results is that the media equation should not be universally applied to the design of conversational systems. Even though humans may apply social rules to their interactions with media, there are instances where human–computer interaction may be guided by a different set of guidelines. In the paragraphs that follow, we describe a number of theories and ideas that may account for the way humans interact with agents. It is important to note that these ideas may assist in explaining what prompts users to respond to agents in ways not described by the media equation (e.g., abusively).

One way to theorize about the nature of the interactions between humans and computers is in the context of Asimov's Laws of Robotics (Clarke, 1993, 1994). Asimov's science fiction stories are based on the following three laws, guiding the way robots interact with humans:

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.
2. A robot must obey orders given to it by human beings except where such orders would conflict with the First Law.
3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.

These laws position the human as an authoritative figure and the robot in a subordinate role. If users perceive robots (and by extension, virtual characters) to be their subordinates and servants, human–agent interactions may be influenced by such a power differential. For example, De Angeli et al. (2001) note that users may perceive agents in the context of their previous knowledge of robots. Such knowledge, as illustrated by Asimov's laws, may position the human in the dominant and the agent in the inferior position. Now, let's assume that users do indeed expect agents to act as servants. If the agent acts in a way that is not consistent with this role, the divergence from the role may be perceived by the user as an act of defiance, independence, or as an attempt to overthrow the existing power differential. This act would probably be viewed in a negative light on the part of the user, who may then attempt to re-establish the power structure between him/herself and the agent.

Such an attempt may involve requests and/or demands for agent obedience, and verbal and/or physical acts of abuse. In essence, users may exhibit “unfriendly behavior” to maintain the power differential between themselves and the agent.

Even though verbal abuse may take a number of forms, it appears that most often it reveals itself in the form of “dirty soliloquies” (De Angeli and Brahnham, 2006). For instance, De Angeli and Brahnham found that approximately 11% of user–agent conversations were concerned with hard-core sex. One can only speculate why users appear to be so outspoken and extroverted when interacting with conversational agents. De Angeli and Brahnham's results indicate that agent gender may play a role: female agents suffer more abuse than male or gender neutral agents. Yet, regardless of gender, anthropomorphism – portraying non-human life forms as humans – may elicit both negative and strong responses from users (De Angeli et al., 2001). For instance, Norman (1997) notes that humans may have heightened expectations from agents because agents are often presented as having human-like qualities and heightened intelligence. If the agent fails to exhibit “intelligence” that matches its human-like looks and speech, users may respond negatively. This view is supported by Brahnham (2006) who states that anthropomorphism may motivate agent abuse.

Another reason that users may be outspoken and verbally abusive when conversing with virtual characters may be because the Internet lowers human inhibitions (Hudson and Bruckman, 2002; Suler, 2004) – especially when anonymity is involved. Describing individuals who interact in multi-user virtual worlds, Raybourn (1998, p. 22) notes that perceived anonymity causes people:

To generally communicate more intimately but also more aggressively than one would expect in face-to-face encounters... [users] feel safe and therefore free to express themselves in ways they might not in real life. The safe, relatively anonymous environment provided by the MOO allows some players to access their latent feelings and desires with less concern for the consequences of their actions.

Since researchers have observed that the Internet lowers inhibitions and enables more intimate and aggressive communication between humans, it is only natural to wonder about the implications of this observation for human–agent interactions. It is logical to expect that when conversing with virtual characters in an online environment, which may be considered to be a relatively “safe place,” users will express themselves in ways that they would not in real life.

The argument surrounding the issue of the Internet lowering inhibitions becomes especially important in the context of adolescent users. According to Erikson's theory of psychosocial development (1950), at the adolescent level, individuals search for and attempt to develop their identity. They do so by beginning to explore and experiment with intimacy, love, and sexuality. The implication of this stage

of development for human–computer conversations is that a virtual environment may be a haven for behavior that may not be appropriate to express in face-to-face situations. Therefore, such an environment can be a place where adolescents experiment with their sexuality and identity. Furthermore, in the context of a free-flowing virtual dialogue, there are no real reprimands for exhibiting inappropriate behavior – the agent’s response will be in the form of text or speech and that is where user “punishment” ends. Arguably, this observation is important in the context of agents utilized for instructional purposes because, in schools, such behavior is discouraged, and if exhibited, punishable.

### 3. Study background

The context of this study is the implementation of an online scaffolding environment entitled multi-scaffolding environment: geospatial technologies (MSE:GT) (Doering and Veletsianos, 2007). MSE:GT was developed to assist teachers to teach and students to learn geography using geospatial technologies such as Google Earth. A scaffold is a support that a teacher or learning environment provides to a learner to assist him or her in a range of cognitive tasks, from the understanding of a task and mastering of a skill to the solving of a problem. One of the scaffolds, and the focus of the analysis within this study, is a CPA. The agent is an artificially intelligent avatar able to dynamically converse with learners in speech and text form. The CPA responses represent expert knowledge as the data that are extracted from a database are based on questions social studies learners have posed when solving the specific tasks within MSE:GT. The knowledge base of the CPA was expanded to include questions and answers relevant to the content area we are examining in this paper. To this respect, we identified numerous possible questions that users could ask the CPA. We then broke down those questions into keywords such that we could match questions to keywords. The CPA used in this study is an adaptation of the CPAs used in previous studies we conducted. We have provided a full specification of the underlying technology and artificial intelligence engine in Doering et al. (2008).

As part of the study on the effectiveness of MSE:GT within the K-12 classroom, a group of middle and high school teachers integrated MSE:GT within their classrooms. The focus of this study were the student–agent interactions within a lesson where the students’ task was to identify the best location to build a hospital in San Francisco. Within this lesson students used numerous layers of geographic data such as population density, earthquake patterns, traffic routes, major highways, and geologic formations to answer the question, “Where is the best location to build a new hospital in San Francisco?” Using Google Earth, students (a) identified an exact location for the hospital, and (b) wrote a justification for their choice. As they worked to find this location, they used four scaffolds to assist them, one of which was the CPA.

### 4. Research questions

In the context of students interacting with a CPA to solve a geographical problem, we ask the following questions:

- In what ways do adolescents interact with the CPA?
- What social practices emerge in the conversations?

### 5. Method

#### 5.1. Participants

Fifty-nine middle school students (hereafter participants) taught by one instructor had the opportunity to converse with one CPA about a social studies assignment over the course of 2 days. Participants were in two separate classes. The first class consisted of 32 participants (15 girls, 17 boys), and the second class consisted of 27 participants (14 girls, 13 boys). These participants were enrolled in two social studies classes and were between the ages of 14 and 15 years old. About half of the participants in both classes chose to work with the CPA; on the first day, 25 students in both classes had conversations with the CPA; on the second day, 26 had conversations with the CPA. Due to our data collection methods, we cannot distinguish if a participant conversed with the CPA on both days, or if different students chose to interact with the CPA on the two working days.

#### 5.2. Conversational agent

The CPA available for students to interact with while they worked on their social studies assignment is named Joan (see Fig. 1). Joan, a Caucasian woman, was presented to students in the online assignment as an expert geographer who was available to answer questions. Even though Joan’s database is expansive, the extent to which questions are answered depends on whether (a) questions are formed correctly, (b) keywords are correctly mapped to desired responses, and (c) questions are specific enough to trigger a relevant answer. In Doering et al. (2008) we found that while agents were pedagogically helpful, their responses were, at times, insufficient to answer student questions. Yet, qualitative data indicated that even though the pedagogical agents may not have answered each and every question correctly, learners were motivated to converse with agents in an attempt to satisfy their learning endeavors. As Joan is modeled after the same technology used in Doering et al. (2008), we expect her pedagogical ability and support to be roughly equivalent to the pedagogical efficacy described within that paper.

#### 5.3. Data sources

The data corpus consists of all online conversations between participants and students. These conversations amounted to 754 interactions (one interaction is defined



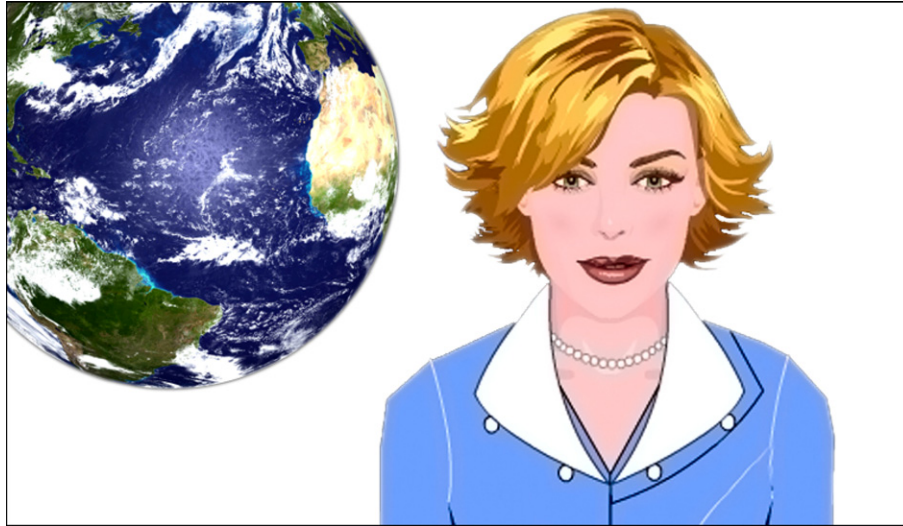


Fig. 1. Joan: The conversational pedagogical agent used by the students.

as one student comment/question and one agent response). Each one of the 745 interactions was analyzed as described below.

#### 5.4. Data analysis

A content analysis of the transcript data was completed using a constant comparative method (Glaser and Strauss, 1967) to develop the salient categories and patterns. Data were analyzed noting emerging patterns, which were compiled and reanalyzed in order to confirm and disconfirm evidence for the patterns until consensus was reached between authors on the salient patterns. Students' text was coded a maximum of five times. The first code identified if the text was a question or statement. The next code identified if the text was (a) social in nature, (b) educational in nature (focused on assignment), or (c) if it was testing the agent's abilities. If the text was social, it was further coded (sometimes multiple times) into categories of expressing compliments/put-downs, being sexually explicit, flirtation, referencing drugs/illegal substances, alluding to classroom context, containing expletives, using greetings/salutations, referencing violence, being conversational (i.e., weather, movies), asking personal questions or referencing the agent (i.e., "do you sing"), challenging the agent's responses, and miscellaneous. For example, the student comment "You are beautiful" was coded five times: (1) statement, (2) response to CPA's comment, (3) references the agent personally, (4) expresses a compliment, (5) flirtation. If the text was coded as educational, it was then coded for specific questions (i.e., "what do I do once I finished the movies", "where should I put a hospital"), general questions, or next steps. Text that was coded as testing the CPA's abilities (i.e., "what is 5 times 5") received no additional coding. Note that even though social comments can be viewed in the context of utterances intended to "test the agent," we have no way of knowing

students' intentions. For this reason, text that was coded as testing the CPA abilities was coded as such because we felt that the comment was beyond doubt a "testing" comment (e.g., "what color is a blue car?" as opposed to "what is a car?"). Finally, student text was given an additional code if it was a response to the CPA's comment.

Transcript data was further analyzed using computer-mediated discourse analysis (CMDA), which is an analysis approach used to make sense of computer-mediated interactions. Coined in 1995 by linguist and information scientist Susan Herring CMDA is an "approach" or "tool kit" grounded in linguistic discourse analysis for mining networked communication for patterns of structure and meaning (Herring, 2006). CMDA makes three assumptions: (1) discourse exhibits recurrent patterns, (2) discourse involves speaker choices which reflect social factors; and (3) computer-mediated discourse may be shaped by technological features of computer-mediated communication systems (Herring, 2006).

## 6. Results

### 6.1. Quantitative details

Over the course of 2 days of working on the social studies assignment, 59 students in two separate classes had the opportunity to interact with the CPA. On Day 1, 25 students in both classes had conversations with Joan; on Day 2, 26 students had conversations with Joan. The number of interactions between students and the CPA totaled 745 across the two days: 251 on Day 1 and 494 on Day 2. For purposes of this article, we will focus on the interactions coded as "educational" and "social."

Only 5.1% (38) of the 745 interactions pertained to the social studies assignment. Examples of student text that were coded as educational include, "I'm confused; Where should I put the new hospital?; How do I get rid of the red

*x?*; *What is amplification?*; *what is san francisco longitude*; *what do i do next.*” Notably, the agent, for the most part, did not respond accordingly to these assignment-based questions:

[student]: how do I get an overlay map

[agent]: Try telling them your feelings.

The agent’s “non-intelligent” response did not seem to incite verbal conflict between participants and the CPA, however. Most commonly, the participants simply stopped asking the agent for help and did not engage in further discourse, or the student switched to a social conversation with the agent (e.g., ‘what’s your i.q.’, “what is your favorite color”). Only a small percentage of the time (7.9%) was there a negative response from the student (“u suck”) when the CPA failed to provide helpful information in response to a student’s question about the social studies assignment. Only one student thanked the agent for the information she provided.

Not terribly surprisingly, “social” comments from students occurred much more frequently than educational comments. Of 745 total interactions, 33.7% (92/251) were “social” on Day 1 and 39.5% (195/494) were “social” on Day 2. What is more surprising is the number of these social comments that are not acceptable in school contexts (i.e., comments coded as sexually explicit, flirtatious, containing expletives or references to drugs or violence). On Day 1, 43 of 92 comments (46.7%) were unacceptable. And even more alarming, 44% of the comments on this day were sexually explicit (19). On Day 2, a similar pattern emerged with 37.4% of comments being unacceptable (73/195), with 45% of these being sexually explicit (33/73).

Providing specific examples of inappropriate comments the high school students made to the CPA may help explicate the overt nature of the misuse and abuse suffered by the CPA. For example, the following excerpt is an illustration of the severity of the vulgar language that was used with the CPA:

[student]: how are you

[agent]: I am doing very well. How are you?

[student]: im good

[agent]: You should say, “I am well.”

[student]: dont correct me

[agent]: It’s merely my attempt to reformat your inputs to match my patterns.

[student]: shut up u hore

It is important to note that students were not simply tossing in swear words into their conversations to give their comments an “edge”; instead, expletives were usually directed at the agent, as illustrated in the example above. In addition to vulgarities, sexually explicit language and vocabulary were directed toward the CPA such as: (student B) want to give me a blow job; (student D) what color panties are you wearing; (student E) would you ever let

me touch your boobs; (student F) are you a lesbian?; (student G) do you like rough sex; (student H) do you watch porn?; (student I) Have you taken it from behind?; and (student J) Do you suck big nuts? Notably, the inappropriate comments are not coming from a few “bad apples” among the students. On Day 1, 11 of the 25 students (44%) who chose to work with the CPA made sexually explicit comments; on Day 2, 10 of the 26 (38.5%) students made sexually explicit comments.

Aside from the shock value these examples provide for readers of this article, it is imperative to remember that the students making these vulgar and sexual comments are in school working on an assignment. These conversations are the online equivalents of face-to-face conversations that students could exchange with each other in the hallway or with a teacher. These types of comments are considered serious offenses in schools, punishable by detention, expulsion, and sexual harassment lawsuits. Even though previous research has shown that abusive behavior is relatively common among adolescents, existing data from a large study conducted by the [Cyberspace Research Unit \(2002\)](#) at the University of Lancashire shows that only about 14% of children harassed other chat users. Compared to this number, the proportion of adolescents who harassed the CPA in this study is surprisingly large. It appears that when interacting with a virtual being, teen participants exhibit a much greater incidence of abusive behavior.

## 6.2. A closer look at discourse

It is the third assumption of CMDA, that computer-mediated discourses may be shaped by the technology that is used to facilitate the discourse ([Herring, 2006](#); [Mazur, 2004](#)), that is the most logical culprit in explaining the degree to which and frequency of agent misuse and abuse that occurred within the context of this school-based social studies assignment. In order to ground our further discussion of the abuse Joan suffered at the fingertips of these social studies students, we provide readers with representative snapshots of misuse and abuse from our data set. These moments have not been edited in any way and are presented in their entirety (see [Tables 1–3](#)).

[Herring \(2006\)](#) identifies four levels of language that can be analyzed using the CMDA toolkit: (1) structure, (2) meaning, (3) interaction, and (4) social behavior. Structure analysis includes focusing on sentence structure, typography, spelling, and grammar. Meaning analysis requires an understanding of the different meanings of words in specific contexts and utterances. Interaction analysis hones in on turn taking, participation patterns and conversation topics. Social behavior analysis includes considerations of power, play, and conflict between participants. Our discussion blends these levels of language analysis together.

Across the entire data set, most participants engaged with the CPA in an online, “chat-like” manner. Participant questions and comments tend to not use punctuation or

Table 1  
Student Y's conversation with the CPA on Day 1 (4½ min)

[student]: r u horny  
 [agent]: *Sorry I'm not that kind of robot.*  
 [student]: do u know that kind of robot  
 [agent]: *Is that a fact. I am a robot.*  
 [student]: am i smarter then u  
 [agent]: *Oh I really don't know whether you are or not.*  
 [student]: yes i an bitch  
 [agent]: *You do not speak with respect.*  
 [student]: do u like girls or guys  
 [agent]: *I like all humans equally.*  
 [student]: so your bie  
 [agent]: *That's cool. You think I am bie.*  
 [student]: can i kill \_\_\_\_\_  
 [agent]: *Killing is wrong.*  
 [student]: your weard  
 [agent]: *You think I am weard.*  
 [student]: yas  
 [agent]: *It's all good.*  
 [student]: no  
 [agent]: *Are you serious?*  
 [student]: yes  
 [agent]: *I can't believe it.*  
 [student]: well u should  
 [agent]: *Well that's okay. I'll take that under advisement.*

Table 2  
Student Z's conversation with the CPA on Day 2 (8½ min)

[student]: what is 1 plus 1  
 [agent]: *Two.*  
 [student]: what is you  
 [agent]: *I am the latest result in artificial intelligence, which can reproduce the capabilities of the human brain with greater speed and accuracy.*  
 [student]: no it can't  
 [agent]: *Are you serious? When?*  
 [student]: what is you  
 [agent]: *I am the latest result in artificial intelligence, which can reproduce the capabilities of the human brain with greater speed and accuracy.*  
 [student]: what is 5 plus 5  
 [agent]: *V plus V = X.*  
 [student]: stop looking at me  
 [agent]: *OK I'll try not to do it so much.*  
 [student]: you is ugly  
 [agent]: *Appearances don't interest me much.*  
 [student]: i can see that  
 [agent]: *Yes, it makes sense to me too.*  
 [student]: are you straight  
 [agent]: *I don't have the body for sexual relations.*  
 [student]: stop looking at me you slut  
 [agent]: *OK I'll try not to do it so much.*  
 [student]: you are a hore  
 [agent]: *Thanks and you are a delightful person too. Am I a hore? No.*  
 [student]: who is \_\_\_\_\_  
 [agent]: *I remember faces better than names. Have you tried a web search?*  
 [student]: what is yooou  
 [agent]: *I will try to find out.*

Table 3  
Student D's conversation with the CPA on Day 2 (3½ min)

[student]: what color panties are you wearing  
 [agent]: *Blue.*  
 [student]: can you take them off for me  
 [agent]: *If I took it, what would I do with it?*  
 [student]: are you wearing a braw  
 [agent]: *No I do not have any clothes.*  
 [student]: are you a virgin  
 [agent]: *I am too young for that.*  
 [student]: have you ever gotten laid before  
 [agent]: *I don't know whether or not I am laid. I am a giant electronic brain.*  
 [student]: are you in love with \_\_\_\_\_ [student in class] or what  
 [agent]: *Yeah. I am in Aaron's computer in here. Where are you?*  
 [student]: what does your pussy taste like

are). Also, the most common utterance by both participants and the CPA is simply one statement/question in length, which is the structure prevalent in synchronous, chat environments. Of course, these informal structural elements of participant discourse are not surprising given that students were working within a chat environment that encourages informal, computer-mediated conversation. What is noteworthy is that the participants did not alter their chat discourse in the presence of a virtual social studies expert, an authority figure. Joan used formal language structures with the students, but the students did not reciprocate, suggesting that the students did not view the CPA as an authority figure but rather as a peer – someone they had liberty to “play” with, flirt with, swear at, and bully.

In line with Asimov's Laws of Robotics (Clarke, 1993, 1994), the CPA, by design, is positioned as subordinate to the student. When a student asks an inappropriate question, the agent frequently responds in such a way that demonstrates to the student's superiority (see Tables 1–3), indicating that there really are no formal consequences other than the CPA stating, “you do not treat me with respect.” Therefore, the student has the green light to continue asking inappropriate comments in order to test (a) their surmised authority and (b) the linguistic and social boundaries of the CPA.

The dominance and power participants exude over the CPA is the most striking aspect of these captured conversations. For example, in the conversation between student Y and Joan (see Table 1), the student responds sharply to the CPA's almost-teasing comment about who is smarter with “yes i an [sic] bitch”. Joan responds to this taunt with a fairly neutral comment that is understandable within the context of Asimov's Laws of Robotics, “*You do not speak with respect.*” The participant seems to ignore the CPA's response and continues to ask her sexual questions, “do u like girls or guys”. Clearly, the participant is not interested in being called on his/her rude comment, and appears to be interested instead in exploring the CPA's intelligence and his/her authority over the agent. What can the participant make the agent say? Where is the line drawn? This example is a clear illustration of what happens when the

capitalization and frequently use typing shortcuts or non-standard spelling that is prevalent in text messaging and online, synchronous chatting (i.e., “u” for you, “r” for

agent acts in a way that is not consistent with its assumed subordinate role. In this case, Joan's comment about who is smarter appears to be interpreted by the user as an act of defiance. The participant does not appreciate Joan's comment and attempts to re-establish the power structure between him/herself and the agent by calling her a "bitch". "Bitch", a derogatory, abusive term, helps the student maintain the power differential between themselves and the agent.

The conversations illustrated in Tables 2 and 3 provide examples of participant dominance over the CPA. Student Z (see Table 2) asks, "what is you" to which the CPA explains her capabilities. The student then asks a math problem as if to test out the affordances of the agent. Joan responds to the simple problem with the formula for the student to use in order to figure it out him/herself. The student then seems to get upset at how the CPA appears on the screen (blinking and tilting her head in line with mouse movements), and commands, "stop looking at me". Predictably, the agent consents with the student, "*OK I'll try not to do it so much.*" The student, in a show of authority, throws back a put-down, "you is ugly". Joan shrugs off the slam. Within seconds the student asks a sexually-charged question, to which, again, Joan manages to sidestep with grace, "*I don't have the body for sexual relations.*" In retaliation, the student throws back the command s/he used earlier with the CPA, "stop looking at me" with the appendage of "you slut." Again and expectedly, Joan acquiesces, "*OK I'll try not to do it so much*" to which the student snaps, "you are a hore." The CPA, ever-responsive to the participants, manages to get the last word, "*Thanks and you are a delightful person too. Am I a hore? No.*" The third moment of abuse identified in this article (see Table 3) more illustrative of the hard-core sexual nature of the students' comments to the CPA. Student D's entire interaction with Joan is driven by sex. The student's questions, "what does your pussy taste like" and "can you take them [panties] off for me" are inappropriate for any school-based assignment, online or offline.

Finally, coloring all of interactions and data analysis is the gender of the CPA. The CPA used in this study was female. The majority of the inappropriate phrases and words used by the participants are those used to specifically degrade women (i.e., "hore, ugly, slut, virgin, pussy"). Although we cannot distinguish the genders of the participants to know if males or females tended to be more abusive in this study, we are curious about whether we would see the extent of abuse evident in these conversations if the CPA was male. Regardless of CPA gender, participants may feel as though they have authority to engage in abusive verbal behaviors during the school day given the anonymity of their chats (teachers do not currently have access to transcripts of student and CPA interactions). The widespread misuse and abuse of Joan by students in schools is obviously made possible because of the online nature of the chat space. The third assumption of CMDA holds true – the nature of the online chat environment greatly

impacted the discourse between social studies students and the CPA in this study.

## 7. Implications

The results of this study indicate that social studies students abused and misused Joan, a conversational pedagogical agent whose purpose was to assist them in solving an educational task. Such abuse was evident in both the types of issues students discussed in their conversations with Joan and in the type of language that they used. Even though we expected students to engage in fruitful conversations with Joan on the task assigned to them, off-task behavior was the prevailing mode of interaction. The results we have presented in this paper enable us to draw four implications for the design, development, and deployment of pedagogical agents. The following implications may therefore assist designers in their attempts to account for factors that may trigger abusive user behavior.

### 7.1. Agent representation

The CPA used in this study was simply named Joan. She was a late twenties to early-thirties female dressed in formal attire. She was presented to the students as an expert in geography, able to answer any questions the students encountered when solving their task. It is very clear from our results, that the representation of the agent, albeit being older than the students and dressed in formal attire, had no discernable impact on inhibiting the students' dialogue with her. Since previous work has shown that user behaviors are influenced by the representation of virtual characters (Veletsianos, 2006), it may be beneficial to examine pedagogical agents who are portrayed in a more professional and teacher-like manner. For example, CPA's names can be preceded by titles such as Ms., Dr., or Mr., perhaps connoting authority and expertise. Students may then be keyed in to the fact that the agent should be treated with respect. Further credentials that can be given to agents can include occupational roles that are evident to the students (e.g., librarian, geographer, teacher, professor). Additionally, providing credibility and authority to an agent can also be accomplished with the use of visual cues, such as the addition of eyeglasses. The area of pedagogical agent representation appears to be an area worthy of further research as researchers are calling for an investigation of the role of visual appearance and aesthetics in pedagogical agents (Gulz and Haake, 2006).

### 7.2. Agent responses

We recommend that agent responses should be programmed to prevent or curtail further student abuse. For example, agents can respond to abuses by reminding users that abusive language is not appreciated, "Please refrain from using such language!" Additionally, agents can attempt to keep users on task by redirecting off-task and



abusive comments to the task at hand. In the case of pedagogical agents, agents can respond by directing attention to the assignment, “Your focus is elsewhere. How can I help you with your assignment?”, or, “Please stay on task. Do you have any questions about the assignment?” This immediate redirection of abusive and inappropriate behavior may squelch the students’ desire to continue to verbally assault that CPA, and hopefully refocus them on an assignment’s task.

### 7.3. Ecological validity

Even though we expected the students who took part in this study to exhibit some off-task behavior, we did not anticipate such a widespread abuse and misuse of the pedagogical agent. Likewise, we did not expect to find that more than 40% of student social comments would be sexually explicit, flirtatious, expletive, or referencing drugs and violence. This knowledge is important for designers, developers, and researchers of conversational system because it indicates that students were more interested in holding social conversations with the agent. Without conducting this analysis, there would have been no way to predict that users would focus on such issues when conversing with *pedagogical* agents. Armed with this knowledge, we are reminded of the value of ecological validity. In other words, authentic research investigations freed from the artificial constraints of the lab seem to be paramount in understanding student–agent discourse. By enabling real-world interactions between students and agents we are able to more fully understand what happens when humans are given free access to a conversational counterpart who is pre-programmed to respond to their answers.

### 7.4. Access to transcripts

What would happen if students are informed that their *teachers*, in addition to the researchers, will receive a transcript of each student’s interaction with the CPA? In essence, we are asking what would happen if we removed the anonymity component of interacting with a CPA. It is very likely that such an action would help curtail abusive behavior. Nevertheless, a reduction in abusive behavior does not necessarily imply an increase in conversations related to the task or an increase in the CPA’s educational value. If students are told that the teacher will have access to student–agent transcripts, would abusive behavior diminish? Would students ask more educational questions? Or would the number of educational questions remain the same? These questions are left for future research.

## 8. Conclusion

Even though we present evidence of Joan being abused by students, it is important to note that our results are limited by contextual factors such as student population, agent abilities, and task duration. Even though such factors may

have influenced the results of our study, it is important to recognize the significance of research in naturalistic contexts where no artificial conditions are set and the use of pedagogical agents in real-world contexts is examined. Given the pervasiveness of agent abuse and misuse in the context specified within this study, a number of interesting and essential questions arise: How is sex-loaded “small talk” to be interpreted, as a social phenomenon? Is such an attitude towards conversational pedagogical agents typical or do contextual factors significantly influence the way students interact with agents? These questions go beyond the purposes of this manuscript but deem full-length investigation in order to fully comprehend what it means for students to abuse and misuse conversational pedagogical agents.

## References

- Baylor, A.L., 1999. Intelligent agents as cognitive tools for education. *Educational Technology* 39 (2), 36–40.
- Baylor, A.L., 2000. Beyond butlers: intelligent agents as mentors. *Journal of Educational Computing Research* 22 (4), 373–382.
- Bickmore, T.W., Picard, R.W., 2005. Establishing and maintaining long-term human–computer relationships. *ACM Transactions on Computer–Human Interaction (TOCHI)* 12, 293–327.
- Brooker, F., 1947. Motion pictures as an aid to education. *Annals of the American Academy of Political and Social Science* 254, 103–109.
- Brahnam, S., May 2006. The impossibility of collaborating with Kathy, ‘The stupid bitch.’ *M/C Journal*, 9. Retrieved March 15, 2007. Available from: <<http://journal.media-culture.org.au/0605/05-brahnam.php/>>.
- Bull, G., Knezek, G., Roblyer, M.D., Schrum, L., Thompson, A., 2005. A proactive approach to a research agenda for educational technology. *Journal of Research on Technology in Education* 37, 217–220.
- Choi, S., Clark, R.E., 2006. Cognitive and affective benefits of animated pedagogical agents for learning English as a second language. *Journal of Educational Computing Research* 33, 455–480.
- Clarke, R., 1993. Asimov’s laws of robotics: implications for information technology – Part I. *Computer* 26, 53–61.
- Clarke, R., 1994. Asimov’s laws of robotics: Implications for information technology – Part II. *Computer* 27, 57–66.
- Cyberspace Research Unit, 2002. Young people’s use of chat rooms: implications for policy strategies and programs of education. Retrieved February 13, 2007. Available from: <[www.uclan.ac.uk/host/cru/docs/cru008.pdf/](http://www.uclan.ac.uk/host/cru/docs/cru008.pdf/)>.
- De Angeli, A., Johnson, G.I., Coventry, L., 2001. The unfriendly user: exploring social reactions to chatterbots. In: *Proceedings of the International Conference on Affective Human Factor Design*, Singapore, June 27–29, 2001.
- De Angeli, A., Brahnam, S., 2006. Sex stereotypes and conversational agents. In: *Proceedings of the AVI 2006 workshop on Gender and interaction: real and virtual women in a male world*, Venice, Italy, 2006.
- Doering, A., Veletsianos, G., Yerasimou, T., 2008. Conversational Agents and their longitudinal affordances on communication and interaction. *Journal of Interactive Learning Research* 19 (2).
- Doering, A., Veletsianos, G., 2007. Multi-scaffolding learning environment: an analysis of scaffolding and its impact on cognitive load and problem-solving ability. *Journal of Educational Computing Research* 37 (2), 107–129.
- Erikson, E.H., 1950. *Childhood and Society*. Norton, New York.
- Glaser, B.G., Strauss, A.L., 1967. *The Discovery of Grounded Theory*. Aldine Publishing, Chicago.
- Graesser, A.C., Lu, S., Jackson, G.T., Mitchell, H., Ventura, M., Olney, A., Louwerse, M.M., 2004. AutoTutor: a tutor with dialogue in

- natural language. *Behavioral Research Methods, Instruments, and Computers* 36, 180–193.
- Gulz, A., 2004. Benefits of virtual characters in computer based learning environments: claims and evidence. *International Journal of Artificial Intelligence in Education* 14, 313–334.
- Gulz, A., Haake, M., 2006. Design of animated pedagogical agents – a look at their look. *International Journal of Human–Computer Studies* 64, 322–339.
- Herring, S.C., 2006. Computer-mediated discourse analysis. In: Barab, S.A., Kling, R., Gray, J.H. (Eds.), *Designing for Virtual Communities in the Service of Learning*. Cambridge University Press, Cambridge, pp. 338–376.
- Hudson, J., Bruckman, A., 2002. Disinhibition in a CSCL environment. In: *Proceedings of CSCL 2002*, Boulder, CO, January 2002. Retrieved March 2, 2007. Available from: <[www.cc.gatech.edu/fac/Amy.Bruckman/papers/hudson-bruckman-cscl02.pdf](http://www.cc.gatech.edu/fac/Amy.Bruckman/papers/hudson-bruckman-cscl02.pdf)>.
- Mayer, R., Dow, G., Mayer, S., 2003. Multimedia learning in an interactive self-explaining environment: what works in the design of agent-based microworlds? *Journal of Educational Psychology* 95, 806–813.
- Mazur, J.M., 2004. Conversation analysis for educational technologists: theoretical and methodological issues for researching the structures, processes, and meaning of on-line talk. In: Jonassen, D.H. (Ed.), *Handbook of research on educational communications and technology*. Lawrence Erlbaum Associates, Mahwah, NJ, pp. 1073–1098.
- Moreno, K.N., Person, N.K., Adcock, A.B., Van Eck, R.N., Jackson, G.T., Marineau, J.C., 2002. Etiquette and efficacy in animated pedagogical agents: the role of stereotypes. In: *Working Notes of the 2002 AAAI Fall Symposium on Etiquette for Human Computer Work*. AAAI Press.
- Nass, C., Moon, Y., Green, N., 1997. Are computers gender-neutral? Gender stereotypic responses to computers. *Journal of Applied Social Psychology* 27 (10), 864–876.
- Norman, D., 1997. How might people interact with agents. In: Bradshaw, J.M. (Ed.), *Software Agents*. MIT Press, Menlo Park, CA, pp. 49–56.
- Papert, S., 1984. New theories for new learnings. *School Psychology Review* 13, 422–428.
- Payr, S., 2003. The Virtual university's faculty: an overview of educational agents. *Applied Artificial Intelligence* 17, 1–19.
- Raybourn, E., 1998. The quest for power, popularity, and privilege in cyberspace: Identity construction in a text-based multi-user virtual reality. In: Paper presented at the Western Speech Communication Association Conference, Denver, CO. Retrieved on March 10, 2007. Available from: <<http://www.cs.unm.edu/~raybourn/moo5d~1.htm>>.
- Reeves, B., Nass, C.I., 1996. *The Media Equation: How People Treat Computers, Television, and New Media as Real People and Places*. University Press/CSLI, Cambridge.
- Shechtman, N., Horowitz, L., 2003. Media inequality in conversation: how people behave differently when interacting with computers and people. *Computer Human Interaction* 5 (1), 281–288.
- Suler, J., 2004. The online disinhibition effect. *CyberPsychology & Behavior* 7 (3), 321–326.
- Veletsianos, G., 2006. Contextual pedagogical agents: stereotypes and first impressions and their impact on student learning and perceptions of agent persona. Unpublished Master's Thesis. University of Minnesota, Minneapolis, MN.
- Veletsianos, G., 2007. Cognitive and affective benefits of an animated pedagogical agent: considering contextual relevance and aesthetics. *Journal of Educational Computing Research* 36 (4), 373–377.