Evaluation of a Virtual Training Environment for Aggression De-escalation

Tibor Bosse[1], Charlotte Gerritsen[2], and Jeroen de Man[1]

[1]Department of Computer Science
VU University Amsterdam
Amsterdam, the Netherlands
t.bosse@vu.nl, j.de.man@vu.nl

[2]Netherlands Institute for the Study of
Crime and Law Enforcement
Amsterdam, the Netherlands
cgerritsen@nscr.nl

KEYWORDS
simulation-based training, aggression de-escalation, dialogue system, evaluation.

ABSTRACT

Public transport employees are confronted with aggressive behavior on a regular basis. As such encounters can have serious consequences, employees need to be well prepared, so that they know how to deal with incidents of aggression. The current paper describes an ongoing endeavor that is aimed at the development and evaluation of a simulation-based training environment for public transport employees, by which they can practice their verbal aggression de-escalation skills during face-to-face conversations. A prototype of the training environment is presented, as well as an experiment to evaluate the environment in several steps. The results indicate that the prototype is evaluated positively with respect to user satisfaction, whereas there is room for improvement with respect to learning effectiveness.

INTRODUCTION

People working in the public sector often have to deal with aggressive behavior. According to a national safety investigation in the Netherlands in 2011, almost 60% of the employees is confronted with unwanted behavior on a daily basis (Abraham et al., 2011). This behavior can include verbal or physical aggressive behavior, but also sexual assault or discrimination.

The municipal public transport operator in Amsterdam (GVB) is one of the organizations of which the employees have to face aggressive behavior on a regular basis. In 2014, the GVB reported 443 incidents of aggressive behavior against employees (GVB, 2012). Only a small part of this number relates to physical incidents, but verbal forms of aggression can be perceived as unwanted as well. Typical examples of incidents are situations where travelers insult a bus driver, or intimidate a tram conductor to get a free ride. Such confrontations may have a range of serious consequences for employees, including reduced work pleasure, decreased work performance, sick leave, various mental health symptoms and even post-traumatic stress disorder (PTSD).

To better prepare them for these incidents, companies like the GVB offer their employees resilience training. Such training is typically performed in a group setting based on role-play, where employees learn to communicate with aggressive clients in a de-escalating manner. Although this form of training has shown to be successful, it is quite expensive with respect to both money and time. Furthermore, the training is not always easy to control or repeat systematically.

As a complementary approach, we propose the use of simulation-based training of aggression de-escalation. This is in line with a number of recent initiatives that show promising results regarding the possibility to train social and communicative skills based on simulated environments involving virtual humans (Bruijnes et al., 2015; Hays et al., 2012; Kim et al., 2009; Vaassen and Wauters, 2012). The main idea of the current system is that public transport employees can practice their aggression de-escalation skills by engaging in conversations with aggressive virtual travelers. By designing the scenarios in such a way that the virtual characters calm down if they are being approached correctly, but become more aggressive if they are being treated inappropriately, trainees will receive immediate feedback on their performance. By using such a system, employees have the ability to practice their aggression de-escalation skills in a cost-effective, personalized and systematic manner.

In this paper, a prototype of such a training environment is presented, which has been developed in collaboration with the public transport company GVB. In addition, an experiment is described that has been performed to evaluate different aspects of the training environment.

LEARNING GOALS

To design an effective training tool, a first question to be asked is what should be the learning goals of the system. For the current context, these learning goals are similar to the ones used in the real world training of the public transport company, and are related to the development of *emotional intelligence*: employees should be able to recognize the emotional state of the (virtual) conversation partner, and choose the communication style that suits this emotional state.

More specifically, when it comes to aggressive behavior, it is important that employees learn to recognize the nature of the aggression. Here, two main categories can be distinguished: aggression can be either *emotional* (or *reactive*) or *instrumental* (or *proactive*) (Dodge, 1990). One of the key

differences between these two types is the absence or presence of anger (Miller and Lyna, 2006).

In case of emotional aggression, the aggressive behavior typically is caused by an angry reaction to a negative event that frustrates a person's desires, cf. the *frustration-aggression hypothesis* (Berkowitz, 1978). Such a person is likely to be angry with respect to whatever stopped him from achieving his goal. By a carry-over effect, the anger can be transferred to new situations as well (Angie et al., 2011). Examples in the public transport domain are people getting angry because the tram is late while they have to attend an important meeting, or because they want to enter the tram while carrying food or drinks that are not allowed. When dealing with an emotional aggressor, *supportive* behavior from the de-escalator is required, for example by ignoring the conflict-seeking behavior, calmly making contact with the aggressor, actively listening to what he has to say, showing empathy, and suggesting solutions to his problems.

In contrast, in case of instrumental aggression, the aggressive behavior is only used 'instrumentally', to achieve a certain predetermined goal. Such behavior is not a direct response to a negative event and is less strongly related to heavy emotions. A well-known example of this type of aggression in the domain of public transport involves someone who wants to travel without paying for his ticket. This type of aggression often starts with an attempt to persuade the conversation partner, e.g. "Oh, I forgot my wallet, can I just come along for two stops?" or "Hi honey, I don't have to pay for a short ride, do I?". Often, in case the employee does not give the aggressor what he wants, the aggressive behavior will reveal itself through more threatening remarks like "I know where you live", "I will be back tomorrow with my friends", or "I will be waiting for you at the end of your shift".

A possible basis for this behavior can be found in the social learning theory, which states that if a person has used aggression to achieve a goal in the past, and if this behavior was successful, then by operant conditioning (s)he will be likely to follow the same behavioral pattern in the future. So, the behavior is learned through positive reinforcement (Bandura, 1963). Hence, to de-escalate instrumental aggressive behavior, a *directive* type of intervention is assumed to be most effective. It is necessary to show the aggressor that there is a limit to how far he can pursue his aggressive behavior, and to make him aware of the consequences of this behavior.

To conclude, the presented training environment will be centered around two main learning goals, namely 1) *recognizing* the type of aggression of the conversation partner (i.e., emotional or instrumental), and 2) selecting the appropriate *communication style* towards the conversation partner (i.e., supportive or directive).

To assess the type of aggression, employees need to carefully observe the verbal and non-verbal behavior of the aggressive individual. In general, reactive aggressors will show more arousal (e.g., flushed face, emotional speech) than proactive aggressors. Also, the context should be taken into account (e.g., someone who just finds out that he lost his ticket will be more emotional that someone who knew this all along, and just tries to intimidate the tram driver to ride for free).

TRAINING ENVIRONMENT

In Bosse et al. (2014), a global overview is presented of the simulation-based training environment that is being developed within the current project. The environment consists of two main components, namely a *virtual reality environment* and a *training agent*. The virtual reality environment has the form of a 3D graphical environment that simulates a particular context in the real world (e.g., the interior of a tram including travelers),[1] with which the user can interact based on a dialogue system. The training agent is an intelligent virtual tutor that monitors the behavior of the trainee and generates personalized support. Two types of support are used, namely run-time modifications of the scenario to adjust its difficulty level to the trainee's performance (scaffolding) (Bosse et al., 2015), and personalized feedback on the trainee's performance in terms of after-session hints (Bosse and Provoost, 2015). To evaluate the overall training environment in a systematic manner, the current paper focuses exclusively on the virtual reality component (and the underlying dialogue system); evaluation of the training agent is left for future research.

The virtual reality environment is based on the InterACT software,[2] developed by the company IC3D Media.[3] InterACT is a software platform that has been specifically designed for simulation-based training of interpersonal skills. Unlike most existing software, it focuses on smaller situations, with high realism and detailed interactions with virtual characters. True-to-life animations and photo-realistic characters are used to immerse the player in the game. An example screenshot of a training scenario for the public transport domain is shown in Figure 1. In this example, the user plays the role of a tram conductor that has the task of calming down an aggressive virtual traveler.

To enable users to engage in a conversation with an emotional conversational agent (ECA), a dialogue system based on conversation trees is used. The system assumes that a dialogue consists of a sequence of spoken sentences that follow a turn-taking protocol. That is, first the ECA says something (e.g. "I forgot my public transport card. You probably don't mind if I ride for free?"). After that, the user can respond, followed by a response from the ECA, and so on. In InterACT, these dialogues are represented by conversation trees, where vertices are either atomic ECA behaviours or decision nodes (enabling the user to determine a response), and the edges are transitions between nodes.

[1] Although the focus of this paper is on public transport, in principle the approach can be applied to any domain involving aggressive behavior in face-to-face conversations.
[2] http://www.interact-training.nl/.
[3] http://ic3dmedia.com/.

The atomic ECA behaviors consist of pre-generated fragments of speech, synchronised with facial expressions and possibly extended with gestures. Scenario developers can generate their own fragments using a motion sensing input device such as the Microsoft Kinect camera and a commercial software package FaceShift.[4] As the recorded fragments are independent from a particular avatar, they can be projected on arbitrary characters.

Each decision node is implemented as a multiple choice menu. Via such a menu, the user has the ability to choose between multiple sentences. Hence, the emphasis of the current system is on the verbal aspects of aggression de-escalation. In the system used for the current study, three options are available with every decision node. These options have been created in such a way that one of them is clearly *supportive*, another one is clearly *directive*, and the third option is neutral. Here, the supportive and directive option relate to the communication styles explained earlier. Figure 1 illustrates how these three options can be instantiated in terms of concrete sentences (in this case: A=neutral, B=directive, C=supportive).

For the current evaluation study, a number of scenarios have been developed, in collaboration with (and approved by) domain experts of the public transport company. To be precise, the scenarios address 9 different situations in which a conflict may arise, such as 'traveler is not allowed to take hot coffee on board' and 'tram arrives 10 minutes late'. Moreover, for each scenario three variants have been written: two variants in which the virtual character shows emotional aggression, and one in which it shows instrumental aggression.

The contents of the scenarios (i.e., the conversation fragments) have been recorded with the help of professional trainers of the public transport company. Each of the 9x3 scenarios has been recorded with a female trainer and with a male trainer, with a specific focus on showing emotional behaviors. Hence, in total a set of 54 scenarios (9x3x2) has been created. The scenarios have been set up in such a way that if the user takes the appropriate communication style, the character calms down and conflict is resolved; however, if the user takes an inappropriate communication style, the situation will escalate. On average, a scenario lasts about 3 to 4 interactions (i.e., both the user and the virtual character speak 3-4 sentences before the scenario ends).

METHOD

This section describes the experiment that was conducted to investigate the impact of the virtual training environment on the user experience as well as the performance of potential end-users from the public transport domain.

Participants

Initially, 30 people were selected to participate in the experiment. All participants were employees of the public transport company (in particular: professional tram conductors and tram drivers). Among these participants, initially 15 were allocated to the training group and 15 to the control group, based on their availability. However, after this allocation had been made, 6 participants withdrew from the study. This resulted in a training group of 14 participants and a control group of 10 participants. Within the training group, 8 participants were male and 6 were female. The average age in this group was 42,7 (σ = 13.1). Within the control group, 5 participants were male and 5 were female. The average age in this group was 48,3 (σ = 9.9).

Experimental Design and Procedure

For the experiment, a pre-test post-test design has been used, where the pre-test and the post-test were separated by a period of 4 weeks. At the start of the pre-test, all participants (in both groups) filled out an informed consent form and

Figure 1: Example screenshot of a training scenario

[4] http://www.faceshift.com/.

provided their personal data. This, as well as all other data gathered in this experiment, was collected anonymously. After that, they made the pre-test, which had the form of a written exam that had been developed in advance by instructors of the public transport company. The exam was composed of 7 multiple choice questions with 4 options each and 5 open questions, which were designed to be representative for the learning goals of the training environment. All closed questions consisted of a particular context description, similar (but not identical) to the ones used in the virtual training (e.g., "a traveler enters the tram and shouts to you that he refuses to pay for his ticket because your tram is much too late"), followed by four alternative responses of which the participant should select the most appropriate one. The open questions were more general, but also related to the learning goals (e.g., "how can you recognize emotional aggression of a traveler?").The post-test was also made by all participants. This test also had the form of a written exam; it had the exact same structure as the pre-test, only the contents of the scenarios and questions were slightly modified to prevent a learning effect (e.g., by changing some properties of the main character, or by rephrasing the multiple choice answers).

In the period between the pre-test and the post-test, the training group performed 4 training sessions, in which they worked with the software for about 30 minutes. More details about these sessions is provided in the next sub-section. The control group did not participate in these training sessions.[5]

After the last training session, the participants in the training group filled out a usability questionnaire. This questionnaire consisted of 13 statements about which the participants had to express their opinion on a 5-point Likert scale. The questionnaire was inspired by Witmer and Singer (1998), and included statements about issues such as user experience, presence, and perceived effectiveness. In the end, the statements were grouped into 4 categories, namely *content*, *interaction*, *emotional*, and *effect*, to obtain an average score on these aspects. The *content* category contained statements about the perceived realism of the scenarios and the characters (e.g., 'the virtual characters showed believable behavior'). The *interaction* category contained statements about how natural it was to interact with the characters (e.g., 'I felt that my answers had an influence in the behavior of the virtual characters'). The *emotional* category addressed the perceived sense of presence of the participants (e.g., 'during training I felt engaged in the scenarios'). Finally, the *effect* category contained statements asking the participants for their opinion about the effectiveness of the training (e.g., 'I think this type of training is a useful addition to real world training').

Training Sessions

All training sessions were executed in a computer room at the public transport company. At the start of a session, participants received a document with instructions about how to work with the training software. They were instructed to solve each virtual scenario to the best of their ability by identifying the type of aggression they observed during the conversation and by selecting the appropriate response in the multiple choice menu. After having read the instructions, they could start the training software.

Upon launching the software, the start menu shown in Figure 2 was displayed. In the upper part of the menu, participants had to input their personal ID and gender. Below that, they could select which scenarios they wanted to run. As can be seen, there were 4 training sessions (corresponding the 4 weeks of the training), each of which consisted of 10 scenarios. The sets of scenarios were chosen in such a way that they were representative for the types of incidents encountered on the job (for instance, they contained more male aggressors than female aggressors, and more cases of emotional aggression than instrumental aggression). All 40 scenarios were slightly different from each other; hence, no scenario was presented more than once. The order in which the scenarios were offered was determined randomly.

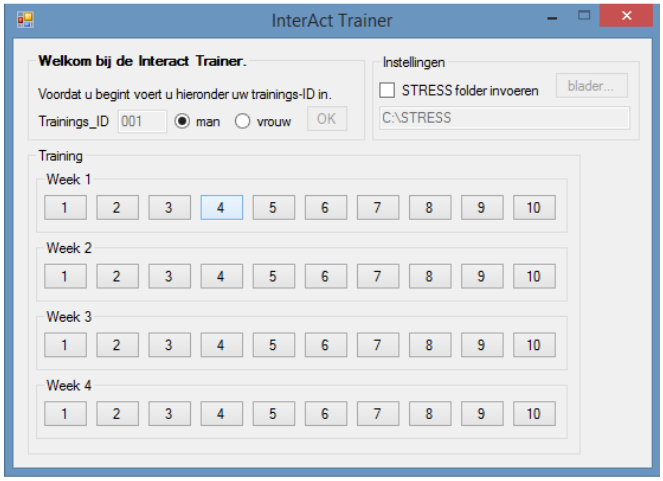

Figure 2: Start menu of the training software (in Dutch)

At the end of each scenario, participants had to indicate whether they thought the aggressive behavior shown by the virtual character was emotional or instrumental. All choices they made in the multiple choice menu were logged, as well as the time it took them to play a scenario.

Variables

The variables that were measured during the study were selected in such a way that they could roughly be related to the training evaluation model by Kirkpatrick (1994). This model distinguishes four levels on which training programs can be evaluated, namely *satisfaction* ('did the participants enjoy/appreciate the training?'), *learning* ('was there an increase in knowledge/skills during training?'), *impact* ('did the participants change their behavior on the job as a result of the training?'), and *results* ('did the training positively affect the organization?'). In the current study, the emphasis is on the first two levels (satisfaction and learning), where the evaluation of *learning* can be further divided into two sub-questions, namely 'did the participants' performance within the training environment improve over time?' and 'did the training result in an increased performance in a different environment that involves the same skills?'. Below,

[5] Note, however, that the participants in both groups did continue their regular work activities in the meantime.

we will refer to these two aspects of learning by *learning during training* and *transfer of learning*, respectively.

Based on this categorization, we can relate the different levels of evaluation to measurable variables in the following way. To evaluate *satisfaction*, the results of the usability questionnaires filled out by the training group (i.e., the answers given to the Likert questions) were analyzed.

To evaluate *learning during training*, the behavior of the participants of the training group during the training sessions was analyzed. In particular, we measured their performance in terms of *identification* (i.e., how well are they able to recognize the type of aggression of the virtual characters?) and *response* (i.e., how well are they able to provide the appropriate responses to the aggressive behavior). As both measures were applied to emotional as well as instrumental aggression separately, this resulted in 4 scores (2x2) to evaluate learning during training. By observing the change of these scores over the four weeks of the experiment, we could evaluate whether the participants improved over time.

To evaluate *transfer of learning*, the written exams made during the pre- and post-test were used. Here, by comparing the scores for the pre-test with the scores for the post-test (in a within-subjects analysis), we could investigate whether the participants' knowledge had improved. Additionally, by comparing the improvement of the training group to that of the control group (in a between-subjects analysis), we could investigate whether the training had an added value over the regular work activities. In this analysis, the independent variable was the condition (i.e., training or no training), and the dependent variable was the change in score between the pre- and post-test. The scores for the pre- and post-tests were obtained by having an instructor of the public transport company grade all exams.

Finally, note that besides for evaluation, the performance of the participants in the different tests (the pen-and-paper exams and the simulation-based training) could be used for assessment purposes as well. That is, by observing the behavior of their employees during the study, the public transport company could gain more insight in how they act in various situations that are representative for real world incidents.

RESULTS

In the following sections, the results obtained during this evaluation study are presented as described above. That is, the first part shows the *satisfaction* of the participants using the training software, while the section thereafter present their *learning during training*. The final section shows the results on the pen-and-paper exams with regard to the *transfer of learning*.

Satisfaction

The experimental group completed a questionnaire asking about their opinion on the training sessions. The answers to these questions are grouped in four categories as explained above; interaction, content, emotional and effect. The scores (on a scale from -2 up to 2) are shown in Figure 3.

The first category, *content*, contained questions regarding the scenarios and virtual characters. With an average score of 0.5 the results were mainly positive, however there were critical remarks as can be seen by the rather larger standard deviation of 0.66. Similar results are found for the second category, *interaction,* and are, with an average score of 0.4 and a standard deviation of 0.71, again mainly positive with some negatives.The worst results are found on the category asking about the *emotional* aspects of the training. This entailed questions about their personal involvement in the scenario or whether they got frightened by the aggression of the virtual characters. With an average score of -0.3 the results do not look promising, however again the standard deviation is rather large (0.73) indicating some positive results as well.The last category contained questions to their personal belief whether such a training has an *effect*. For example, if they think they improved in their interaction with travelers or if they believe such a training is useful addition to the current role-play scenarios. Overall, responses to these questions were positive (average 0.7), with almost no negative scores across the participants (standard deviation 0.52).

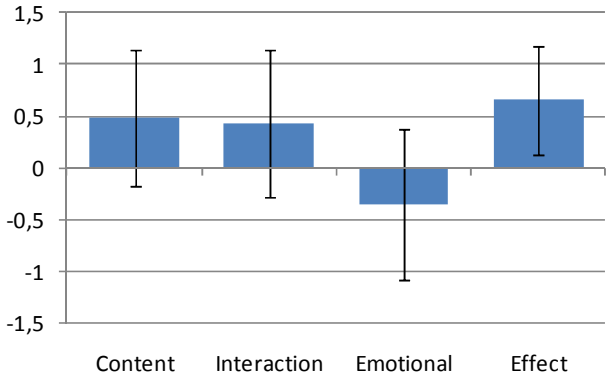

Figure 3: Average scores on satisfaction

Learning during Training

The experimental group underwent 4 weekly training sessions, each of which consisted of 10 scenarios. For each scenario they had to identify the type of aggression as well as respond correctly to de-escalate the situation. Figure 4 shows both the percentage of correctly identified aggression types as well as the correct responses split into instrumental (i) and emotional (e) aggression per week. Higher scores are better, where scores of 0.5 for identification and 0.34 for response would be expected with random answers.

Firstly, none of the measurements show an increase over time, indicating there is no real increase in performance over these 4 weeks. But, when taken a closer look, it can be seen that participants were able to identify emotional aggression correctly approximately half the time and subsequently responded well half of the time. However, instrumental aggression was identified correctly more often, while the response on these situations was the worst of all. This is confirmed by paired t-tests as well; the difference between the identification of and response to instrumental aggression is significant ($t(46) = 7.37,\ p < 0.001$), while for emotional aggression this is not the case ($t(46) = 1.45,\ p = 0.153$). Furthermore, instrumental aggression was identified

correctly more often ($t(46) = 5.87$, $p < 0.001$), while the response to emotional aggression was significantly better ($t(46) = 3.45$, $p = 0.001$).

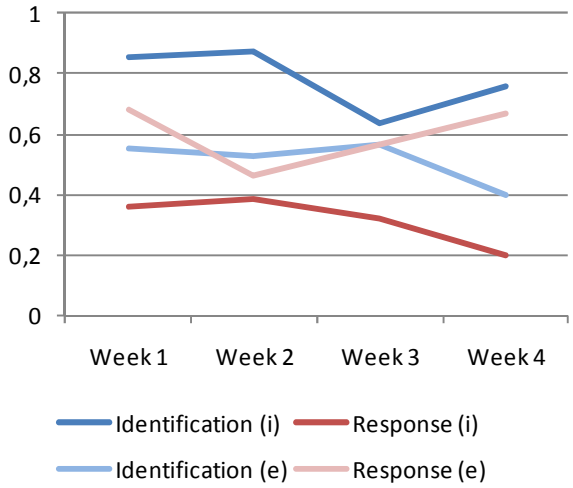

Figure 4: Average scores for learning during training

Transfer of Learning

The results of the pen-and-paper tests of both the experimental and control group are shown in Figure 5. The total score is subdivided in a score for the open questions (blue color, max. 10 points) and multiple choice questions (red color, max. 7 points). The error bars represent the standard deviation of the total score.

It can be seen that both groups performed better on the post-test as confirmed by a t-test with $t(23) = 2.64$, $p = 0.014$ for the experimental group and $t(18) = 3.31$, $p = 0.004$ for the control group. On closer examination, it turns out that there is no significant change in the score on the multiple choice questions, but the higher scores are due to better answers on the open questions. The important question is whether the experimental group experienced a greater increase than the control group, which unfortunately does not show in this data ($t(21) = 0.06$, $p = 0.950$).

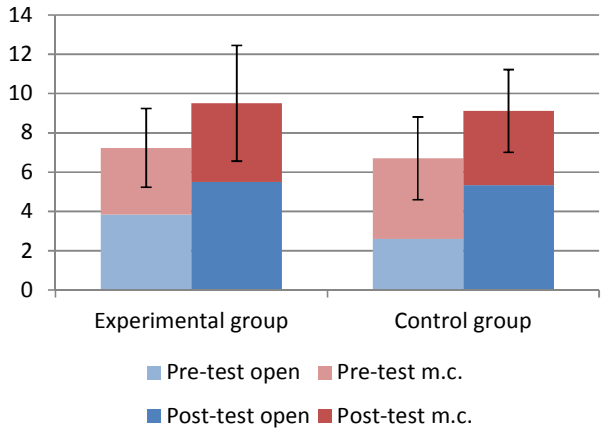

Figure 5: Average results for transfer of learning

DISCUSSION

Firstly, did the virtual training help improve the participants in the experimental group more than those that did not use the training? Unfortunately not, but nevertheless this research helps us in understanding why. Considering the difference between the pre- and post-test, the improvement in both groups was mainly due to a better score on the open questions. As they did the same tests, this improvement could be due to the second test being a bit easier, in which case there would be no 'real' improvement in either group. Another potential explanation could be that merely being part of the experiment already made the participants reflect on the topic of aggression de-escalation during the 4 weeks of the experiment (even if not all of them participated in the training sessions). This might explain why both groups obtained a better score in the open questions.

On the other hand, participants did not seem capable of translating this increased understanding of aggression de-escalation (as measured with the open questions) to correct decisions on how to act in concrete situations (as measured with the closed questions). This is in line with the results retrieved from the training sessions, which did not show any improvement in performance during the scenarios.

Then, what do these results learn us? Throughout the training, participants identified instrumental aggression quite well, but did not respond accordingly. As the correct response for instrumental aggression is very direct, it might be that they preferred a more 'polite' or 'friendly' answer. Another explanation for the lack of improvement might be that each of the participants already has a set way of responding and has difficulty in changing this 'default' approach. This option is backed by the data as well; by looking at the scores of those participants with more than 2 years of experience in comparison with the others, there is no difference in test scores as well (pre-test $t(20) = -0.72$, $p = 0.480$; post-test $t(20) = -1.09$, $p = 0.288$).

From this, the conclusion might be drawn that on your own, it is difficult to learn correct responses for the different types of aggression. Then, it would be important to provide timely feedback such that a trainee understands the mistake and is able to improve on it. Providing such feedback was not yet implemented in this training software, but is being developed (Bosse and Provoost, 2014).

It should be considered as well that there might be a more fundamental problem, such as a flaw in the pedagogical approach or simply a lack of motivation from the participants. However, from the subjective evaluation, it is clear that participants do see the usefulness of such a training and already experience a belief of improvement due to it. Nonetheless, it is important to consider methods to improve the emotional involvement of trainees during the various scenarios as this was shown to be insufficient and could potentially affect the learning as well. Improvements can be made by for example changing from a standard desktop screen to a head-mounted display or increasing the intensity of the aggression shown by the virtual agents. All in all, we believe the potential of such a virtual training is supported by these results.

CONCLUSION

The current paper introduced a prototype of a simulation-based training environment that enables public transport employees to practice their verbal aggression de-escalation skills during face-to-face conversations. The design of the system is centered around two learning goals, namely the ability to recognize the type of aggression (emotional or instrumental) and the ability to select the most appropriate communication style for the observed aggression type (supportive or directive).

The prototype was evaluated by means of an experiment in which 24 employees of the public transport company of Amsterdam participated. The results indicate that with respect to user satisfaction, participants were moderately positive about the content of the virtual scenarios and the mechanisms to interact with the characters. Also, they were very positive about the potential of the system as an effective learning tool. The only category for which their opinion was below neutral involved their perceived sense of (emotional) engagement and presence.

Regarding the performance during training, no significant improvement was found, which might be explained by the fact that this particular task is difficult to learn without specific feedback. In line with these results, also no transfer of learning was found to a similar task on paper (in particular, to the closed questions, where participants had to indicate how they would behave in fictional scenarios). In contrast, participants in the training group did show an improved performance regarding the open questions of this paper task, but this improvement was not significantly larger than that of the control group.

Finally, an interesting side effect was that the training environment also proved useful as an assessment tool. For instance, it allowed us to conclude that the current group of participants is significantly better in identifying proactive aggression than reactive aggression, but at the same time has significantly more difficulties in dealing with proactive aggression than with reactive aggression.

Inspired by the current findings, our future research will concentrate on two main aspects. First, we will try to incorporate additional elements in the training with the aim to enhance users' emotional engagement and presence. Examples of such elements are more extreme aggressive behavior of the virtual characters (e.g., louder voice volume, more threatening facial expressions and utterances), the use of immersive technology like head-mounted displays, and the use of mechanisms to introduce a 'simulated threat', e.g., based on air blast devices or electric surges. Secondly, we will integrate our previously developed modules for learner feedback (Bosse and Provoost, 2014) within the system, to explore whether this has a positive impact on learning effectiveness.

ACKNOWLEDGEMENTS

This research was supported by funding from the National Initiative Brain and Cognition, coordinated by the Netherlands Organization for Scientific Research (NWO), under grant agreement No. 056-25-013. The authors wish to thank Maria ter Beek, Ernst van der Horst, Aly Lubberink, Jermaine Ravenberg, Caroline de Ridder, Petra ter Weeme, and all participants in the study for their contribution to the project.

REFERENCES

Abraham, M., Flight, S., and Roorda, W. 2011. "Agressie en geweld tegen werknemers met een publieke taak". Research for the program 'Veilige Publieke Taak 2007 - 2009 – 2011'. Amsterdam: DSP.

Angie, A.D., Connelly, S., Waples, E.P., and Kligyte, V. 2011. "The influence of discrete emotions on judgment and decision-making: A meta-analytic review". Cognition & Emotion, 25(8), 1393-1422.

Bandura, A. 1963. "Social learning and personality development". New York: Holt, Rinehart, and Winston.

Berkowitz, L. (1978). "Whatever Happened to the Frustration-Aggression Hypothesis?" American Behavioral Scientist, 21, 691-708.

Bosse, T., Gerritsen, C., and Man, J. de. 2014. "Agent-Based Simulation as a Tool for the Design of a Virtual Training Environment". In: Proceedings of the 14th International Conference on Intelligent Agent Technology, IAT'14. IEEE Computer Society Press, pp. 40-47.

Bosse, T., Gerritsen, C., Man, J. de., and Tolmeijer, S. 2015. "Adaptive Training for Aggression De-escalation". In: C.J. Headleand et al. (eds.), Proceedings of ALIA 2014, CCIS 519, Springer Verlag, pp. 80-93.

Bosse, T. and Provoost, S. 2014. "Towards Aggression De-escalation Training with Virtual Agents: A Computational Model". In: Proceedings of the 6th International Conference on Human-Computer Interaction, HCI'14. Springer Verlag, pp. 375-387.

Bosse, T. and Provoost, S. 2015. "Integrating Conversation Trees and Cognitive Models within an ECA for Aggression De-escalation Training". In: Proceedings of PRIMA 2015, Springer Verlag, in press.

Bruijnes, M., Linssen, J.M., op den Akker, H.J.A., Theune, M., Wapperom, S., Broekema, C., and Heylen, D.K.J. 2015. "Social Behaviour in Police Interviews: Relating Data to Theories". In: Conflict and Multimodal Communication, Springer Verlag, pp. 317-347.

Dodge, K.A. 1990. "The structure and function of reactive and proactive aggression". In D. Pepler and H. Rubin, (eds.), The development and treatment of childhood aggression (pp. 201-218). Hillsdale, NJ: Erlbaum.

GVB Amsterdam. 2012. "Incidenten overzicht 2002 tot heden". Technical Report.

Hays, M., Campbell, J., Trimmer, M., Poore, J., Webb, A., Stark, C., and King, T. 2012. "Can Role-Play with Virtual Humans Teach Interpersonal Skills?". In Interservice/Industry Training, Simulation and Education Conference (I/ITSEC).

Kim, J., Hill, R.W., Durlach, P., Lane, H.C., Forbell, E., Core, C., Marsella, S. Pynadath, D. and Hart, J. 2009. "BiLAT: A game-based environment for practicing negotiation in a cultural context". International Journal of Artificial Intelligence in Education, vol. 19, issue 3, pp. 289-308.

Kirkpatrick, D.L. and Kirkpatrick, J.D. 1994. "Evaluating Training Programs". Berrett-Koehler Publishers.

Miller, J.D. and Lyna, D.R. 2006. "Reactive and proactive aggression: Similarities and differences". Personality and Individual Differences, 41(8), 1469-1480.

Vaassen, F. and Wauters, J. 2012. "deLearyous: Training interpersonal communication skills using unconstrained text input". In: Proceedings of ECGBL. pp. 505–513.

Witmer, B.G. and Singer, M.J. 1998. "Measuring presence in virtual environments: a presence questionnaire". Presence: Teleoperators and Virtual Environments 7, pp. 225-240.