

A serious game to improve the verbal resilience against doorstep scams *

L.M. van der Lubbe¹, C. Gerritsen¹, T. Bosse²

¹ *Vrije Universiteit, Amsterdam, The Netherlands*

² *Radboud Universiteit, Nijmegen, The Netherlands*

l.m.vander.lubbe@vu.nl

Abstract

Doorstep scams, scams in which con artists tell convincing but fraudulent stories in order to enter the house of a victim and/or steal personal belongings or information, have a high impact on victims. These victims are often elderly people. Existing campaigns mostly focus on the behavioural aspects that help to prevent a doorstep scam from happening, but verbal skills also contribute to the resilience of possible victims. In this research, a serious game is designed and evaluated for this purpose. The results of the evaluation are positive. The serious game is specifically seen as a valuable addition to existing training meetings.

Keywords: Serious Games, Virtual Agents, Voice Analysis, Doorstep Scams, Elderly Users

1 Introduction

Doorstep scams are scams in which a con artist has a convincing, but fraudulent, story with the purpose of coming into your house and/or stealing money or gaining personal information for fraudulent purposes. Doorstep scams frequently happen; numerous news reports about different stories exist. Elderly people are at a high risk of becoming victims of such scams. Since doorstep scams often have high emotional and economical impacts, various campaigns try to educate people on this topic in order to prevent doorstep scams from happening (e.g. ‘Spot it, Stop it’ [1] or ‘scam awareness month’ [2]). Such campaigns focus on behavioural aspects of the prevention of doorstep scams, such as opening a door with a chain on to prevent intrusion. However verbal skills (such as refusal assertiveness) are also important to be more resilient against doorstep scams. There is no known large scale prevention campaign that focuses on the conversation that takes place within a doorstep scam and the verbal skills needed to prevent these scams from happening. On a smaller scale, human actors are sometimes used to play a doorstep scam scenario with a group of people within, for example, larger group meetings about safety.

Doorstep scams are acknowledged by the Dutch Ministry of Safety and Justice as high impact crimes. Because of this, and because existing campaigns are insufficient, they funded research towards a virtual doorstep scam resilience training. The current research is executed

*Earlier, shorter versions of this paper are published as: *L. M. van der Lubbe, T. Bosse, and C. Gerritsen, "Design of an Agent-Based Learning Environment for High-Risk Doorstep Scam Victims," Cham, 2018, pp. 335-347: Springer International Publishing.* and *L. M. van der Lubbe, C. Gerritsen, D. Formolo, M. Otte, and T. Bosse, "A Serious Game for Training Verbal Resilience to Doorstep Scams," Cham, 2018, pp. 110-120: Springer International Publishing.*



in collaboration with a large Dutch organization for elderly called KBO-PCOB [3]. This research describes a virtual, simulation-based training, in the form of a serious game. This serious game can be used to improve the verbal resilience of the players, in order to minimize doorstep scams from having negative outcomes. As the high-risk victims of such scams are elderly people, the intervention is aimed at this specific target group and therefore the design necessities of this specific target group are taken into account.

For the design of the system, theories from the domain of serious game research are used. We studied the hierarchy of players' needs [4]. This hierarchy, based on psychology, has different layers with needs of players that need to be successively fulfilled to increase the motivation of players. By using this hierarchy, we aimed to fulfil all those needs in order to increase the motivation of players. Moreover, literature about design principles for elderly users [5] is studied and used in the design of the app, to ensure that the app is accessible for the aimed target group. Such design principles are for example about recommended font size and contrast levels.

As mentioned, human actors can be used in small scale training sessions. An alternative to human actors can be virtual agents. Virtual agents are autonomous entities in a virtual environment that look and behave like living organisms (for example human or animal characters), that can interact with other human or virtual agent(s) in their environment [6]. Research in other domains (such as [7, 8]) has shown that the low costs and the degree of control of training with virtual agents are important advantages over training with human actors. Other advantages are for example that the training is available at any time and for a larger audience. When game elements are added to a virtual training, the virtual training becomes a serious game: 'a game in which education (in its various forms) is the primary goal, rather than entertainment' [9]. Previous research has shown that virtual trainings or serious games are effective learning resources [10–13]. Moreover, virtual trainings can be repeated easily, and as many times as the trainee wants.

The objective of this paper is to present the serious game that we have developed. This serious game aims to improve the players' verbal resilience against doorstep scams, by both teaching them more about what to say and how to use their voice in an assertive way. To achieve this goal, the serious game uses virtual agents in the role of scammer in interactive scenarios of doorstep scams and voice analysis to measure the assertiveness in the voice of a player. Moreover, this paper presents a first evaluation of the system, with the goal to present insights in the perceived usability of the system by users from the target group. This evaluation gives insights in the interaction between the user and the system, as well as how the target group thinks that the system can be used after publishing. Moreover, the evaluation showed that using principles from theory in the design process of a serious game, is actually helpful. For example, the evaluation shows that all participants were capable of playing the serious game without assistance.

This paper starts with a background section explaining more about the domain of doorstep scams and explores which virtual trainings and serious games already exist for skills that are relevant in the prevention of doorstep scams. Based on this research, Section 3 describes the design and implementation of a serious game to increase the verbal resilience against doorstep scams. In Section 4, a preliminary evaluation of this serious game and the results of the evaluation are described. In Section 5 some side projects that explore future work with (similar) serious games are described. Finally, Section 6 concludes the paper with a summary of the system and some concluding remarks.

2 Background

To understand the domain of doorstep scams in the Netherlands, a field study was conducted to understand the content and progress of doorstep scams. For this field study the following sources were used: a focus group meeting with the partner KBO-PCOB, various conversations with domain experts, (news) articles and reports.

This section begins with a description of the findings of the field study in Section 2.1. Next, different skills, relevant for the resilience against doorstep scams, and existing (virtual) trainings and serious games for these skills are discussed in Section 2.2. Finally, the topic of serious gaming is introduced in Section 2.3.

2.1 Doorstep scam scenarios

Doorstep scams often happen at the front door, but can also happen on the street or via phone. During the field study, we identified three locations where doorstep scams often happen: at the front door, on the street, and on the phone. For each of these locations, we collected various frequently happening scenarios. An enumeration of these scenarios can be found in Appendix A.

The fact that doorstep scams are a serious problem can be derived from the number of campaigns and news articles that can be found on the subject. However, it is hard to find statistics about the scope of the problem. This is due to the fact that there are multiple criminal activities associated with doorstep scams, which makes the registration inconclusive. Furthermore, often victims do not report a doorstep scam [14], due to shame for example.

2.2 Virtual training for verbal resilience

There is no previous work found in the domain of virtual trainings for doorstep scams. We therefore investigated which (virtual) training exist for skills that are relevant for doorstep scam resilience and which virtual training applications exist for the target group of elderly people.

2.2.1 Assertiveness training

Being assertive means that you are behaving confident and that you dare to say what you think or believe [15]. This is a form of verbal resilience. Winship & Kelley [16] used a verbal response model to train assertiveness. Participants that were trained using this model showed an increase in their assertive behaviour. Research furthermore showed that assertiveness training within a group can also be effective [17]. It is also shown that verbal modelling and therapist coaching can increase the refusal of unreasonable requests [18, 19]. In other research [20] video feedback was found to be an effective approach for assertiveness group training.

Saying no Saying no, or refusal skills, are part of assertiveness trainings, applied in various domains among which are smoking behaviour, drugs usage, shoplifting and rape prevention. There are various ways in which somebody can say no: you can simply say no, make aversive statements, give a reason for not accepting an offer, change the subject or walk away [21]. For adolescents the most often used strategies are simply saying no to drugs [22, 23], alcohol [24, 25], or smoking [26–28], and giving a reason for not accepting the offer [24, 25]. In order to learn students to resist direct and indirect pressures to engage in negative behaviours, an effective prevention program should both show different verbal strategies as well as the need to be assertive when refusing an offer [22, 23, 27]. While nonverbal assertive skills can be

used for different types of situations, practicing verbal strategies for specific situations is also needed [21].

Assertiveness is not only determined by the content of the message, but also by nonverbal aspects and the speech characteristics [29]. In this research, we focus on speech characteristics next to the content of the messages. Speech characteristics of assertiveness are for example speaking firmly or authoritative [30], duration of a reply [31, 32], medium latency of the response [29, 31], the loud volume of the voice [29, 31, 33], and the medium fluency [29, 33]. Another research points out that both the volume and the speed of the response are cues of confidence [34]. The Behavioural Assertiveness Test - Revised (BAT-R) [35] can assess assertive behaviour, via role play scenes. It studies different aspects in the behaviour of a participant: nonverbal aspects such as the number of smiles or the duration of eye contact, speech characteristics such as the characteristics mentioned before, and content aspects such as praise, appreciation and requests for new behaviour. However, the validation for this test is mixed [31]. One remark made is that for example the volume of voice is not unique for assertiveness. Within our research saying no, although often implicitly, is an important aspect of the assertive behaviour taught to the player. As suggested within the research of Nichols et al. [21] it is important to give verbal strategies for specific situations, this is done in this virtual training.

2.2.2 Social skill training with virtual agents

Although some of the above-mentioned trainings are virtual, there are no examples mentioned yet of training programs for assertiveness with a virtual agent. However, virtual agents, sometimes in combination with virtual reality devices, are used for many different types of training programs, among which are social skills [36, 37]. An example is the *automated social skills trainer* (ASST) [38], a training in which human-agent interaction takes place via user speech and language interaction. The proposed training focuses mostly on communication, previous research has been done in the same field, however no comparable virtual trainings are found. *deLearyous* [13] is a serious game used to train interpersonal communication skills. Players learn how to use Leary’s Rose in their advantage. The communication with the agent is in this case performed by unconstrained written language input. *Communicate!* [12] aims to train interpersonal communication skills. This game is used to train communication in a consultation setting. The interaction with the agent in this game is via multiple choice.

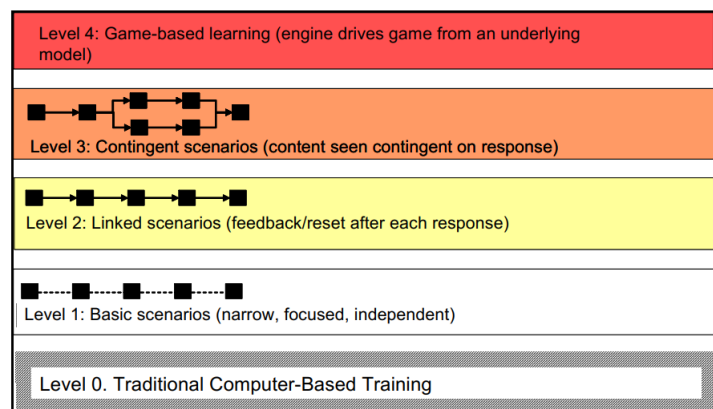


Figure 1: Levels of engagement for computer-based trainings [39]

2.3 Serious games

When certain game elements are added to virtual trainings, they are considered as serious games. Goh et al. [40] give an overview of several strengths of serious games. One of these strengths is the fact that serious games offer covert learning, in combination with an already existing positive attitude towards games this makes serious games an easier accepted platform for learning. Furthermore, serious games are seen as fun and motivational since players want to achieve goals within the game. Another important strength of serious games is the feeling of control players have; they can practice a scenario as many times as they want to master a skill and they always have the option to shut down the tablet, giving them a feeling of safety. Greitzer [39] defines five levels of engagement for computer-based trainings (see Figure 1). These levels show different approaches to computer-based trainings, in which the level of engagement says something about the active role of the trainee and the complexity of the scenarios.

Siang et al. [4] created, based on the hierarchy of needs from Maslow [41], a hierarchy of players' needs (see Figure 2). This hierarchy is based on psychological theories and shows how psychology can be used in game design. The needs need to be addressed from the bottom to top successively. So, the rules need needs to be fulfilled before the safety need. With the use of this hierarchy, game designers can steer the motivation of players.

2.3.1 Serious games for elderly users

Various types of serious games targeting elderly users exist to address different difficulties the target group faces. An often used type of game is the exergame, a game in which the player has to perform some sort of physical activity. Mostly these games are used to address problems that have to do with physical activity of the elderly users and related problems, such as balance and postural control. Often used technologies for such games are Nintendo Wii Fit sometimes together with the Balance Board (e.g. [10, 42]) or XBox Kinect (such as [42]). Exergames are also used for seniors with subsyndromal depression [43].

Serious games for elderly users can be used to address problems with cognitive abilities. So called brain training games, such as Brain Age [11], improve the elderly players' attention and memory skills. While these games are often designed especially to serve as a serious game, the research of Whitlock et al. [44] found that a non-serious game, World of Warcraft, improved the attention and spatial orientation of the elderly players as well.

Serious games for elderly do not only address physical or cognitive abilities, but it can

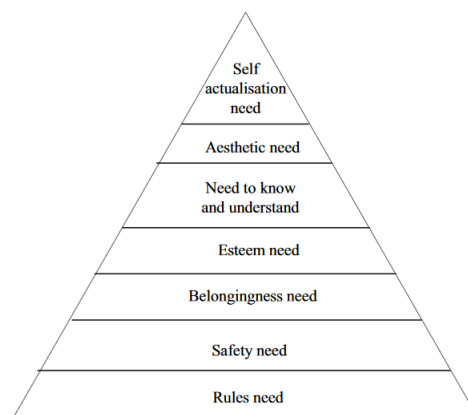


Figure 2: *Hierarchy of players' need [4]*

also be used to enhance the users' social contacts, for example intergenerational interactions [45]. SilverGame [5] is a platform consisting of different activities to promote social activities, such as singing, dancing or performing fitness activities together with the use of video communication. The activities also serve as entertainment and promote exercise.

3 Design and implementation of serious game

This section describes the design and implementation of the virtual training that is developed during this research. An overview of the system is described in Section 3.1. In the next section, Section 3.2, the voice analysis module is explained. Section 3.3, describes the components of the training scenarios in more detail. First the scenarios are introduced (Section 3.3.1), followed by a section about the virtual agents and the environment (Section 3.3.2). Next, more about how the scores are calculated is described in Section 3.3.3. Finally, Section 3.4 describes the game layer that is built around the virtual training.

3.1 Overview of the system

The platform for which the virtual training is developed is a tablet. Since this is an often used platform by elderly people, KBO-PCOB posed this requirement.

Besides choosing a suitable platform for the target group, it is also important to keep the target group in mind when designing the game. Tailoring a serious game to the elderly target group reduces the risks of factors that reduce the efficiency of the game for this target group, such as information overload [10]. Not only the content and game play of the game must be tailored to the target group, also specific design principles [5] are used to adjust the game design to the target group. This includes high contrasts and a larger font type.

As explained earlier, serious games have some strengths, such as being fun and motivational. The virtual training will therefore be designed in the form of a serious game. Moreover, it was explained earlier that for assertiveness, the behaviour (way of using the voice) is also important. Hence, the application consists of three components: the voice analysis module, the training scenarios and the game layer.

The application starts with a main menu with four different options: scenarios, scores, explanation, and credits. On the scores pages the player can find their top ten scores and the average number of stars received for each scenario. It is also possible to reset all the scores.

For the application, six different scenarios were written, based on the field study (see Section 2.1). To ensure that the scenarios are credible, they have been evaluated with various domain experts provided by the KBO-PCOB, for example in a session with professional actors. Based on these evaluations, changes were made to the scenarios, which resulted in the six scenarios mentioned in Table 1.

Players can choose from these six scenarios, which can be played from a first person

Table 1: Scenarios used in the training

	Location	Short description	Goal con artist
1	Front door	Energy meter check	Enter the house
2	Front door	Package delivery	Enter the house with package
3	Phone	Fraudulent back activity	Gain personal bank information
4	Phone	Lottery won	Gain personal information
5	Street	Money for charity	Collect money for non-existing charity
6	Street	Sell bracelet for charity	Sell overpriced/Fake bracelets

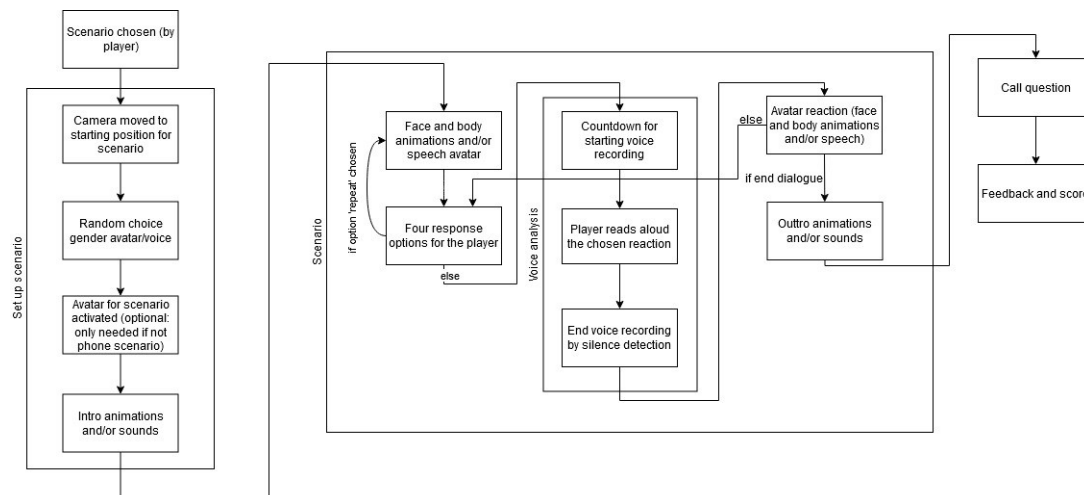


Figure 3: Flowchart of app flow

perspective. Moreover, the player can choose to play using voice analysis or not, in order to train the assertiveness of his/her voice. The voice analysis measures the assertiveness of the players' voice, and influences the scenario based on this measurement. The voice analysis is only available when playing with a network connection and is, if connected, by default turned on.

Figure 3 shows the flow of the application when playing a scenario. For each type of scenario (door, street or phone) a different intro is used. In the case of the door scenario the doorbell rings, the door is opened and the camera moves a bit forward. For the street scenarios the camera and an animated dog, placed close to the camera to represent the players' dog, move towards the avatar. In case of the phone scenarios a ring tone is played after which the screen of the phone placed in the environment changes, representing an incoming call. The end scenes of a scenario can either be closing the door (door scenarios), walking away (street scenarios), or a hung up sound (phone scenario).

A scenario always starts with the virtual opponent. After the turn of the virtual opponent the player generally has four possible responses to choose from. One of these responses is to repeat the last turn of the agent (this response is only available if the virtual opponent has said something in the previous turn). This response is added to accommodate the target user group; in case they did not hear the agent correct. The other three responses influence the progress of the scenario.

When the scenario has ended the player is asked whether or not he/she would call the emergency number in such a situation. This question was added as during the field study it was found that victims often do not report a doorstep scam. Moreover, many people do not know that it is allowed to call the emergency number in case of a (suspected) doorstep scam. To motivate them to do so they are offered an opportunity to file a report in the application at the end of the scenario and they will receive feedback on this decision. After this the player will receive feedback on their choices during the scenario and see his/her score. For all feedback a 'read aloud' option is available. Section 3.3.3 discusses how the feedback and score are established.

3.2 Voice analysis module

As stated in Section 2.2 being assertive is not only represented in the content of your message, but also in the way the message is communicated. The application addresses both: the con-

tent is represented in the different choices provided by the interactive conversation, the voice analysis module addresses the influence of how the message is communicated. The technique used for this module is based on the research of Formolo and Bosse [46], the technical details of the module are beyond the scope of this article.

The voice analysis module is based on the Interpersonal Stances theory. This concept stems from social psychology, and can be defined as ‘the ways in which speakers and writers linguistically demonstrate their commitment to or attitudes about a person or proposition’ [47]. The module classifies the voice in 2 types of attitudes: Dominant (normally referred to as Above) and Submissive (Below). A modified version of the openSmile toolkit extracts the voice features, while an Support Vector Machine (SVM) algorithm classifies the extracted features into the categories. The SVM model was built using 4-fold cross validation over a dataset with 681 sentences of four people instructed in how to act into both categories. Details about the algorithm and the SVM tuning are described in [48]. The final accuracy of the module is 86.56%.

When voice analysis is performed, the recorded voice of the player is analysed by the algorithm. The application detects silence to determine when the recording ends. The ambient noise is measured while people are using the application. Silence is defined as a period of 3 seconds in which the volume is 20 decibels above the ambient noise volume maximum. The recording stops after 10 seconds or if silence is detected. The module runs on a server, it receives the recording of the user’s voice and returns the classification status. In case of communication failure, the client application ignores the voice information and continues the dialogue without the voice analysis. The output of the module is the confidence percentage between the 2 categories (submissive or dominant), resulting in an assertiveness score.

The next section explains how the voice analysis is used within the training scenarios.

3.3 Training scenarios

The training scenarios consist of different components that together form interactive training scenarios with tailored feedback. These components are described in the next sections.

3.3.1 Scenarios

The interactive conversations are the scripts of doorstep scams, represented in a conversation tree. Within these conversation trees, see an example in Figure 13 in Appendix B, vertices represent either atomic agent behaviours or decision nodes (which enable the user to select a response), whereas edges represent transitions between nodes. They are turn-based, always starting and ending with the virtual agent. The blue rectangles show the avatar’s dialogue. Each round the player is offered three choices as response to the agent (round rectangles in Figure 13). In general, these choices are of a good, moderate and bad level, influencing the conversation in the same way (represented in the colours of the round rectangles in Figure 13). Good in this context means assertive, bad means submissive. For some reactions of the player there are two outgoing arrows with conditions. These conditions are used by the voice analysis module and indicate what the influence of the assertiveness of the players’ voice is on the progress of the scenario. In general, the outcomes of a conversation have the same mapping: good means no scam happened, and bad means that a scam happened. The red rectangle in Figure 13 shows a negative end state with the dialogue of the avatar.

When the module is turned on the player will not only be asked to make a choice within the scenario, the player is also asked to say this choice aloud and record this (within the application). The progress of the interactive scenario is then no longer only determined by the choices made, but also by the level of assertiveness measured in the voice of the player. When the voice analysis module is not used, each reaction of the player directly leads to a

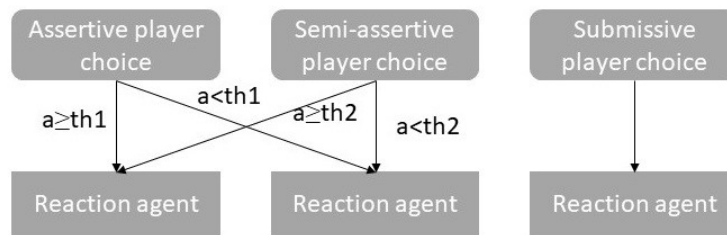


Figure 4: Flow scenario taking voice analysis into account

reaction of the virtual agent. However, when the module is turned on in some cases one choice of the player can lead to two different reactions by the agent, determined by the level of assertiveness.

Figure 4 shows the flow of a choice moment when voice analysis is taken into account. When no voice analysis is performed the flow chart does not have thresholds and diagonal lines. The voice analysis module returns an assertiveness score (a). A high assertiveness score means a very assertive reaction. When a player has a high assertiveness score this will lead to the reaction normally given to an assertive player choice, although this might not be the most assertive choice when looking at the content. This also works the other way around. Threshold1 (th1) is lower than threshold2 (th2), since a very assertive player choice needs a little bit less assertive voice to have an assertive impact on the virtual agent.

Based on the information obtained during the field study a general outline is made for all the scenarios:

1. Background: getting to know the background of the story the con artist is telling.
2. Identity: getting to know about the identity of the con artist and his/her relation to the story that has been told.
3. Alternative: finding out alternatives to the suggested behaviour by the con artist.

In each scenario these three aspects can be found in this order. It depends on the specific scenario how these aspects are addressed. Following the same outline in each scenario gives the players guidance for other (real life) doorstep scams. Furthermore, it was discussed during the field research that it is important for people to get familiar with scenarios instead of only hearing about the theory of what to do to prevent/during a doorstep scam. It was mentioned that the fact that people have to read and speak certain sentences, such as ‘can I see your identification card?’, people might feel less of a burden to say this sentences in other situations as well. This supports our choice for scenarios that are somewhat similar in outline, so that these elements can be practised repeatedly and are easier to adopt for use in real world situations.

Figure 5 shows a screenshot of the application when the player has reached a choice moment, during the package delivery scenario at the door. At the bottom of the screen the four possible reactions are visible, the bottom one is the repeat option, the other three options are randomly ordered. In the upper right corner, the gauge for the voice analysis is showing the players last voice analysis score.

3.3.2 Virtual avatars and environment

A three-dimensional environment is created using the game engine Unity [49], using various assets from the Asset Store [50]. Scripts for the functioning of the system are mostly programmed in C#. The environment features a small part of a residential area. Within this residential area there is a decorated street, that is used for the scenarios that take place on the

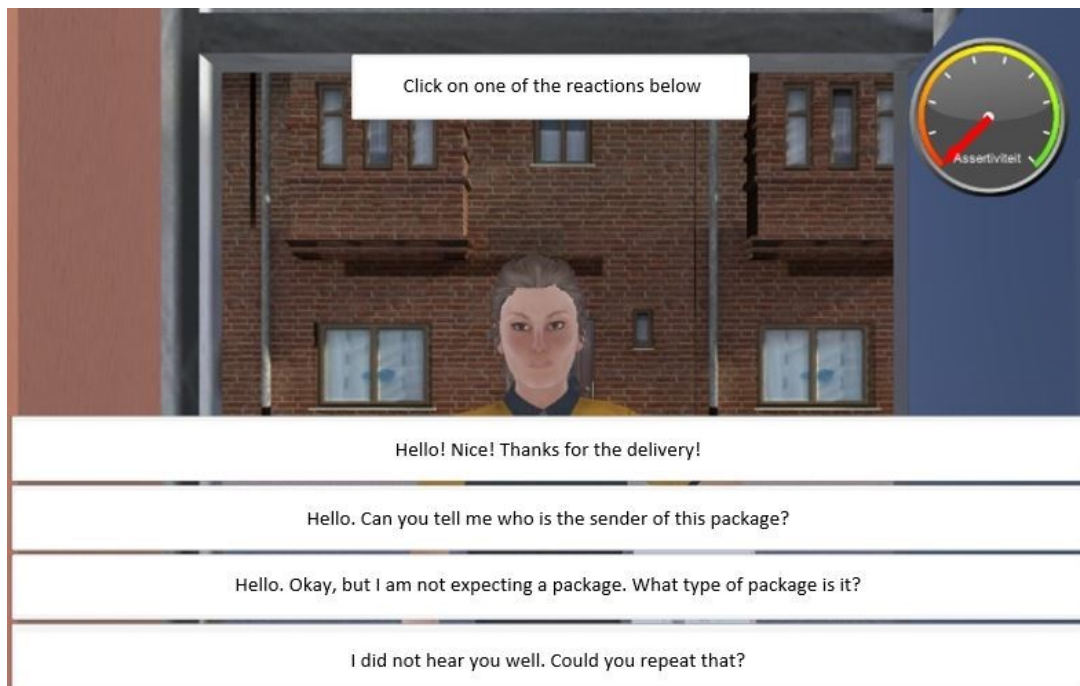


Figure 5: (Translated) screenshot of a choice moment

street. Furthermore, one of the houses within the environment is partly furnished to feature the scenarios taking place at the front door and phone scenarios that take place in the living room. Figure 6 shows the different viewpoints within the environment. Players cannot move around in the virtual environment.

The virtual agents used within the scenarios, playing the role of doorstep scam artist, are modelled and animated using iClone [51]. For each scenario two similar agents are created, one female and one male. The outfits of the virtual agents are either recreated working uniforms, used to increase the credibility, or modern clothes. Figure 7 shows the eight avatars that are created for the door and street scenarios.

For the face animations, Facial Motion Capture [52] software, with a plug-in for iClone, is used. During two recording sessions a male and a female actor their face expressions were recorded while also recording the voice for the different dialogues. Their voices are also used for the scenarios that are not using a virtual agent (phone scenarios). After recording, the facial animations needed some further editing because they were insufficiently realistic.

The flow of the application is programmed in a generic way, so that it is easy to add or change the content of the scenarios. The scenario specific content (avatars, dialogues, responses, tips, and feedback) are therefore stored in a database. By linking the code to the



Figure 6: Screenshots of different environments



Figure 7: Avatars created for door and street scenarios

database the specific content is shown within the application. An overview of all the databases and the data stored in these databases can be found in Table 2.

3.3.3 Feedback & scores

After each scenario the player will receive feedback. This feedback consists of a general reaction on the outcome achieved (whether or not you became a victim of a doorstep scam) and some general tips for the specific scenario. This is the feedback that is shown to the player by default. This feedback is stored in the feedback database (see Table 2).

Moreover, there is extended feedback. This feedback includes a paragraph that is specific for the scenario and is therefore independent of the outcome reached by the player. Furthermore, it includes feedback on the choice the player made for the emergency call question. There are three different feedback texts available for this. The first is for players who chose to call the emergency number, the second is for people who did not choose to call the emergency number but did have a negative outcome, and the last is for people who did not call the emergency number but did also not become a victim in the scenario.

The most tailored feedback are the tips. In each scenario a number of Boolean variables are defined, that are by default true. If a specific dialogue that is linked (which is stored in the dialogue database) to a variable is played during the scenario, this variable is set to false. For the variables that are still true at the end of the scenario a tip is given to the player. The tips are showed in such an order that they follow the progress of the scenario. An example is a variable about asking the identification, this is linked to the dialogue where the virtual opponent shows its identification. If this dialogue is not played during the scenario, the player receives feedback about this afterwards.

The tone of the feedback is positive. The goal of the feedback is to inform and educate players, not to punish them for wrong behaviour. This positive tone can be found in the general formulation of the sentences as well as in the fact that the feedback suggests other types of behaviour instead of telling the player what not to do. Since the general outline of the different

Table 2: *Overview of all used databases*

Database	Description
Avatar	This database contains the names of all the available avatars and links them to the right scene (location) and scenario number.
Dialogue	This database contains all the turns of the virtual opponent. Storing: <ul style="list-style-type: none"> - Scene number, scenario number and the dialogue ID - Text of the dialogue - The tip variable IDs that is turned false in this dialogue (if applicable, more details on this in Section) - The type of end state (if applicable) - The body animation of the virtual avatar (if applicable)
Responses	This database contains the responses that are linked to the dialogues. Storing the following: <ul style="list-style-type: none"> - Scene number, scenario number and the ID of the dialogue that the response is corresponding with - Text of the response - The ID of the default next dialogue - The threshold for the speech analysis score (if applicable) - The ID of the next dialogue when the speech analysis score is below the defined threshold, and the ID of the next dialogue when the score is above the threshold (if applicable) - The action (animation) linked to the response (if applicable) - The type of response (2=good, 1=average, 0=bad) - A boolean if there is speech analysis for this response (default = true)
Tip	This database contains the tip variable IDs (corresponding with the dialogue table), their name and text.
Feedback	This database contains the different types of outcomes. The IDs correspond with the type of end state defined in the dialogue table. Furthermore, it contains the name and the feedback text of the outcomes.

scenarios is comparable, players can use the feedback for different scenarios. Furthermore, since the scenarios are comparable to real world doorstep scams, the feedback is also reusable in different possible real world doorstep scams.

Besides feedback the player also receives a score at the end of each scenario. The higher the score of the player, the better the performance during the training was. The highest score is 105. The score is calculated using the average score for the choices, the average score for the voice, a score for the result of the scenario, and a bonus (of 5 points) if the emergency number is called.

Next to a score a player also gets a number of stars (0-5) for a scenario. A player will receive one star if he/she earned more than 11 points during the game, two stars for more than 33 points, three stars for more than 55 points, four stars for more than 77 points, and the maximal number of five stars is achieved when more than 100 points are earned during the game.

3.4 Game layer

In Section 2.3 the levels of engagement of computer-based training programs were discussed. The described training aims at level three engagement, in which a limited amount of branching within the interactive scenario follows the choices of the player.

The game layer of the application is built around the actual training scenarios with the purpose of motivating players to (repeatedly) do the training. The scores and stars, discussed in Section 3.3.3 are one of the measures that are taken to do so. Moreover, a high score table in the app encourages players to improve their local scores. Siang et al. (2003) distinguished the different player needs. Those needs are addressed in the following way:

- Rules need: The players will receive an explanation about the rules of the game at the beginning, as well as instructions while they are playing.
- Safety need: The player will receive feedback at the end of each training scenario. The positive tone of the feedback will give them the feeling that they can handle the scenarios. Furthermore, when the player falls for a doorstep scam within the training this will not affect the player in real life.
- Belongingness need: The feedback that the players receive after each scenario helps them to improve their resilience against doorstep scams as they can use the feedback for other scenarios as well. The feedback will guide them to achieving better outcomes and achieve the goal of the game: to prevent a doorstep scam.
- Esteem need: The players have control over the scenarios since they can make choices influencing the progress of the scenario, as well as the option to quit a scenario. Furthermore, by the feedback received from the game the players will be encouraged, which will boost their esteem.
- Need to know and understand: By repeatedly training different scenarios players will improve their resilience skills, which gives them the ability to train with even more advanced scenarios that can be unlocked at a certain skill level.
- Aesthetic need: Several state of the art game development techniques have been used to build a 3D-environment with credible virtual agents.
- Self-actualisation need: (Advanced) players can play scenarios multiple times to test the reaction the virtual agent gives on different reactions, allowing them to test different reaction strategies.

4 Evaluation

Different consultations of domain experts have taken place during the development of the training. Once a first prototype was finished, this was evaluated during a focus group session. The results of this session are discussed in Section 4.1. Section 4.2 describes the setup and results of the final evaluation of the virtual training.

4.1 Focus group evaluation of prototype

During the design of the training, the partner organisation and its experts were repeatedly consulted. When a first prototype version was finished, a focus group with five elderly safety advisors of the KBO-PCOB (two females, three male) was organized to evaluate this version of the system. At the beginning of the focus group, a general instruction about the application was given. Next, participants could use the application in a private session for 30 minutes. After this, a group discussion was held to talk about the findings of the participants. This discussion was guided by open questions.

The overall reaction on the prototype was positive. Added value was especially seen in using the training scenarios in larger settings in which different users can interact with and

help each other. Small remarks were made, that have resulted in some bugfixes and small textual changes for the final version of the application. Moreover, a bar was added to show players the remaining time for the voice recording.

4.2 Final evaluation

The final evaluation was in the form of two focus groups: one with eight elderly safety advisors and one with 11 elderly potential end users. This second group had a position as tablet coach for elderly for KBO-PCOB, which means that they are familiar with using a tablet. The elderly participants have ages between 68 and 84, with an average age of 74. Only 3 (16%) participants were female. Both sessions were organized on the same day, the 17th of July 2018, at the headquarters of the KBO-PCOB in Utrecht.

The focus groups started with a general explanation of the project, followed by 30 minutes of individual interaction with the system. After this, the participants provided us with qualitative feedback in a group discussion.

Next to this, participants filled in a questionnaire before the group discussion, which consisted of two parts. In the first part, some background information of the participants was asked, as well as some information about their experience with doorstep scams. The second part was based on the questionnaire used in a previous study of a virtual training [53]. The questionnaire consisted of 23 statements, about which the participants give their opinion using a 7-point Likert scale. The questions could be divided into four different categories:

- Content: five statements about the experienced realism of the scenarios and virtual avatars.
- Interaction: six statements about the believability of the interaction with the virtual avatars.
- Engagement: four statements about the engagement of the players with the scenario.
- Effect: eight statements about the opinion about the effect of the application.

All statements were identical to the ones used in [47], however three statements (Statement 14, 15, 16) were merged into one, as this was more suitable for this context. Moreover, five statements (Statements 10, 12, 16, 18, 22) were added to evaluate specific elements of the training, such as the voice analysis. A complete list of all the statements used within this evaluation can be found in Appendix C.

Moreover, participants were given the opportunity to write down remarks after each question if they wanted to further elaborate on their rating further. Next to that there was some room for remarks at the end of the questionnaire.

4.2.1 Results

Overall, the participants were positive about the application. In their opinion, the training was fun to do, and useful as well. Moreover, it was easy to use the application. The voice analysis clearly had added value for the participants. Again, some remarks were made to further improve the application. Minor changes, such as increasing font and button size and some minor textual changes, were made based on this before publishing the application.

To analyse the results from the questionnaire, the answers were coded on a scale from -3 up until 3, numerically representing the answers on the 7-point Likert Scale that is used. If a question was not answered, these values are left blank.

Figure 8 shows the grouped results per category of the questionnaire. This figure shows that, in general, the participants (n=19) are average to strongly enthusiastic about the system.

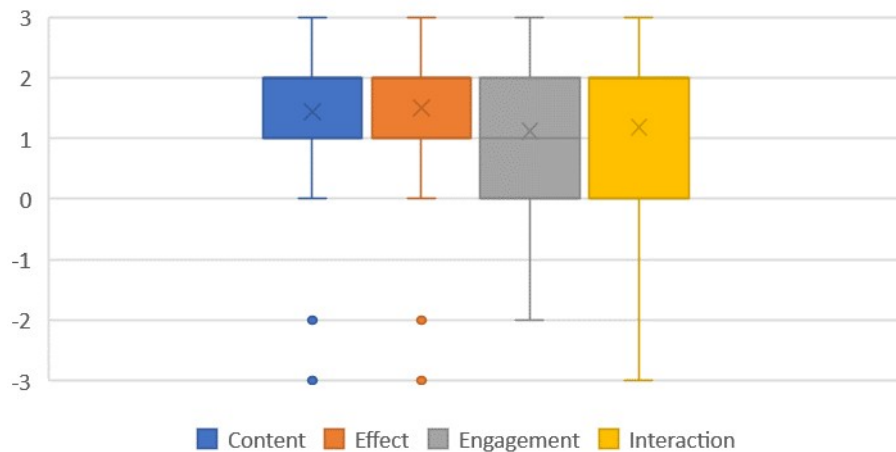


Figure 8: Boxplot for each topic in final evaluation

Mainly, people were positive about the added value of this medium in addition to the existing trainings. Moreover, the scenarios were found to be recognizable, and the interaction with the avatars was reasonable believable. The questions about engagement are rated a bit lower, but still above neutral. The highest average is found for the topic about effects, however the differences with the other topics are small.

The average score for each individual statement are also studied. This shows that all statements, but one (-0,25), are above neutral. This question was about the accuracy of the voice analysis. This question was only answered by seven participants, and of these participants two participants answered negatively to this question, one participant answered positively, and four answers were neutral.

The highest average score was given to the statement about the added value of the virtual training to role-plays with human actors (average score of 2,05).

Finally, the average scores for all statements in each topic are studied. For *content*, the averages can be found in Figure 9. This figure shows the previously discussed statement with the negative average. All other statements are evaluated with an average above one. The realism of the scenarios and the usefulness of the feedback are evaluated the highest. For the realism it was noted by a participant that in real doorstep scams, the scammers are more pertinacious, they want to hear arguments and to have a discussion if you counteract them.

For *effect*, the average scores are shown in Figure 10. One of the statements from this topic, is the statement that has the highest average of all statements, about the added value of the training to role-plays. This is the only average score of this topic as well as of the whole questionnaire that is evaluated with an average above 2. Another statement that is evaluated higher than other statements in this topic is about whether people liked to do the training. One of the participants noted as a remark to this question that he was surprised that it was such a learning moment for him. The lowest average score is for the statement about whether the voice analysis teaches participants about how they can use their voice in an assertive way.

For *engagement*, the average scores are shown in Figure 11. One statement has an average that differs more than 1 point from the other averages of this topic. This is a statement about whether participants felt personally addressed by the avatars. All other statements have an average above 1 and the highest average for the feeling of immersion. One of the participants noted that it felt like there was eye contact with the avatar.

Figure 8 showed that the average score for engagement was the lowest of the 4 categories. However, from Figure 11 it becomes clear that this low average is caused by the average score for the statement about whether participants felt personally addressed by the avatars. When

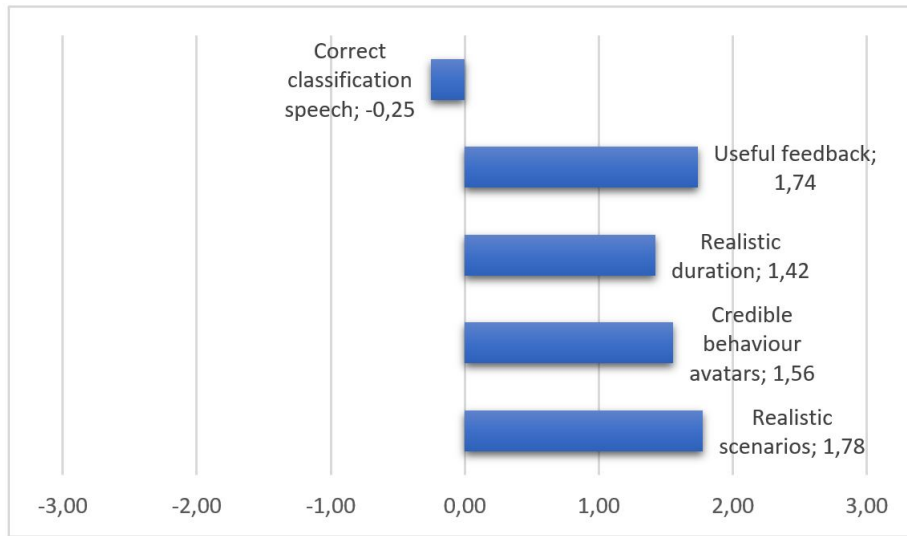


Figure 9: Average scores for content topic



Figure 10: Average scores for effect topic

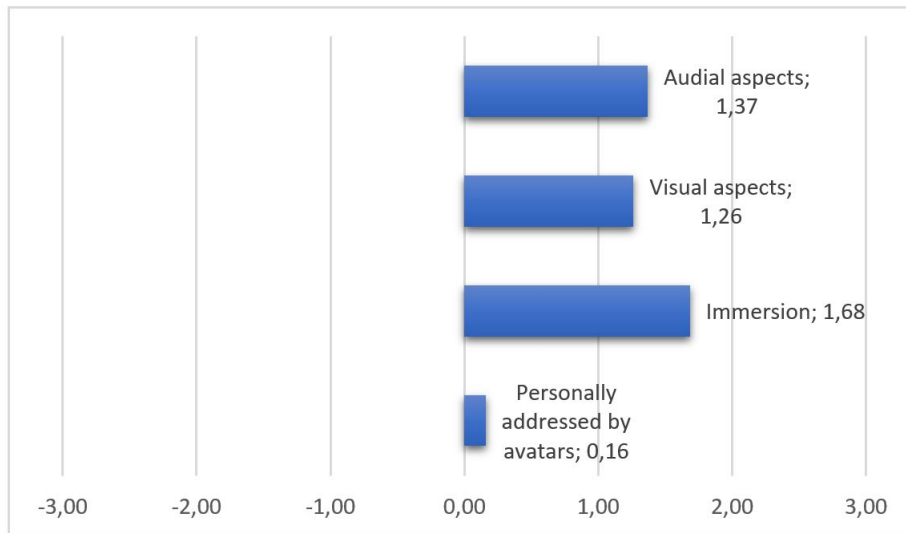


Figure 11: Average scores for engagement topic

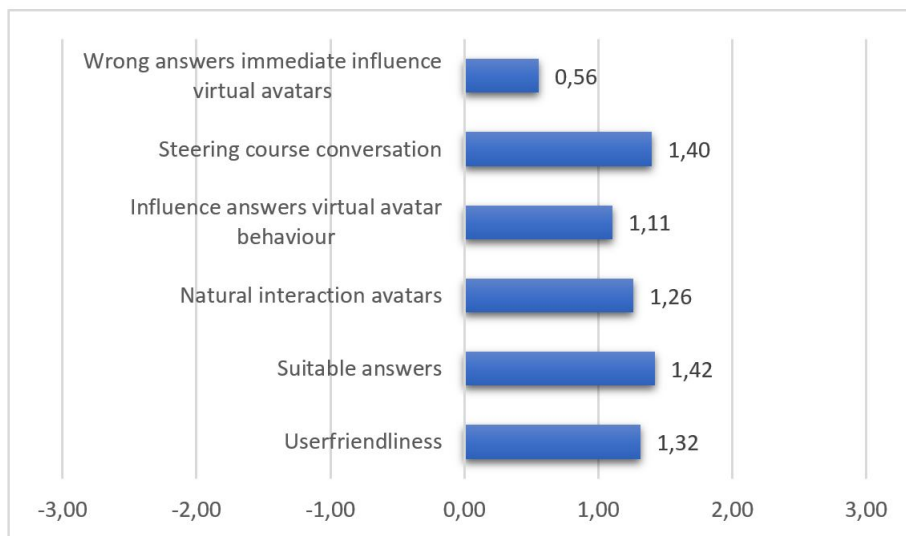


Figure 12: Average scores for interaction topic

this statement is left out of the average, it becomes much higher (1,44) which is the same as the average of the content topic.

Finally, for *interaction*, the averages are shown in Figure 12. All statements are evaluated with an average above 1, besides one statement. That statement is about whether wrong answers influenced the virtual avatars' behaviour. Another statement about the influence of users' choices on the behaviour of avatars was evaluated with a 1,11. It was noted that the virtual opponents could be more pressing. One participant reported to have noticed frustration after a resilient answer, the virtual avatar seemed frustrated. Whether or not a participant noted differences in the behaviour of the avatar might be influenced by the type of answers given. Whether the interaction with the avatars felt natural was rated with a 1,26. A participant suggested that using real actors would be better. The highest average in this topic is for whether there was always a suitable answer in the multiple choice menu. A similar average is given to whether participants felt that they could steer the course of the conversation.

In the group discussions during the final evaluation, the participants also noted that they

see an added value of the app in prevention meetings with groups of elderly. Moreover, it was noted that the explanation of how the app works, as well as the answers, are clear for the user. Points for improvement that were noted had to do with making some components in the app clearer (for example fonts or button sizes). Moreover, some participants would like to see real people instead of avatars. Some remarks were made about the content of the scenarios, among which was a recommendation to use the name of the player in a scenario to make it more personal.

4.2.2 Discussion of the results

The accuracy of the voice analysis module is not evaluated well, while during the training of the module the accuracy that was measured was high. This can be caused by different aspects. The performance can be lower in real-life situations compared to the situations of the voice samples used for training and testing the algorithm, for example due to background noise. Furthermore, we encountered some technical difficulties with the voice analysis, which might also have influenced the perception of the participants. Regarding the voice analysis, we learned that more testing and fine-tuning is needed to ensure a good accuracy in real-life settings. Moreover, from the evaluation it is clear that participants do not feel that they learned about an assertive use of voice. This might be improved by increasing the accuracy of the voice analysis as well.

Another point that was negatively evaluated was the statement about whether players felt personally addressed by the avatars. However, no additional feedback on this statement was given to explain this negative evaluation. Moreover, the other statements about engagement are evaluated higher, so it is unclear what causes this negative evaluation. One of the remarks that was made during the group discussion was that it might be more personal to use the name of the player in the scenarios. Another remark that has been made is that it would be better to use real actors instead of virtual characters. However, using actors would make the development more time consuming. In creating such serious games, the costs of development have to be considered in comparison to the believably and expected outcome. For future research, it is valuable to further study if the appearance of the virtual agents can be altered in such a way that the target groups finds them more realistic, and to study if this changes how players feel addressed.

5 Future work: extending the serious game with more game elements

The developed serious game is a fully functional, publically available tablet application. However, it is just a first version and it offers many opportunities for extension. As the game layer of this version only has limited game elements, different students worked on side projects to explore how other game elements could be integrated in this serious game as well¹. This section describes an overview of their work, and the implications of their work for future extensions of the serious game, and in general for a virtual training in the form of a serious game.

5.1 Adding a quiz

In order to increase the learning value of the training as a whole, the first project explored the potential of adding a quiz to the training scenarios [54]. It is relevant to explore this, as the

¹All students created their own simplified version of the virtual training to modify for their projects.

final evaluation of the serious game showed that the score for the statement about the learning value of the scenario was only 1,26.

The student compared his simplified version of a training scenario (control condition) with a version in which he added a quiz at the end (experimental condition). The quiz consisted of the following four components:

- True or false questions;
- Drag & drop questions in which three statements had to be placed in the right order;
- Matching questions in which a statement has to be matched with a scenario;
- Multiple choice questions.

In the evaluation of this project, 32 elderly participants took part. To evaluate the potential of adding a quiz, both objective and subjective measures were used. The objective measures concerned knowledge questions about the prevention of doorstep scams. Subjective measures concerned the perceived learning results, clarity, entertainment, motivation, challenge, and user-friendliness.

This work did not find an added value in the learning results of adding a quiz to a scenario game, but the research did indicate that the attitude of participants towards the version with the quiz was different. For perceived entertainment, motivation, and challenge, significant differences were found for the difference in average score between the two groups, in favour of the experimental condition. This indicates that adding an extra element might not directly contribute to better learning results, however it does not harm but in fact even improve the perceived entertainment, motivation, and challenge. This in turn could lead to more use of such a virtual training, which might increase the learning results on the long term. However, as the research only had a limited number of participants, that interacted with a simplified version for a limited amount of time, more research is needed to further explore this.

5.2 Adding different reward systems

Another approach to increase the learning outcome of the virtual training can be to add more reward systems to the serious game. However, while the project described in 5.1 was directly oriented on the knowledge (by adding a knowledge quiz), this student project aims to make the whole virtual training more gamified, and therefore more fun and motivation [55].

The student created three different designs for reward systems, based on literature research towards reward systems (for elderly) and learning and motivational theories [56–64], that could be added to the virtual training. These designs were discussed during one of the sessions at the KBO-PCOB in the form of a focus group (with six elderly participants that also took part in our focus group). The following three designs were evaluated:

1. A performance graph, showing the results from the last three attempts of the player. This is a reward system that aims for the individual user. In the performance graph the user sees a grade (0-10), showing the performance of the user in the scenario. Moreover, based on this grade, the player earns a trophy which can have three levels (bronze, silver or gold). While this gives feedback at the end of the scenario, another reward system within this prototype gives feedback during the scenario. That system is a bar that shows the level of openness/closeness of the player, indicating how close the player is to preventing or become the victim of a doorstep scam. The bar fluctuates as the player makes choices in the scenario.

2. The trophies from the first reward system are reused in this reward system. If a player earns a golden trophy, this gives the player an attempt in an edited version of hangman. The player can collect different words by playing games of hangman in between the scenarios. Next to this, in this version there are also goals to achieve for the player, which awards badges to the player.
3. In the third reward system a leaderboard is introduced, showing different values (high score, total score, number of prevented losses) to compare the users. During the scenario a bonus score is shown, that fluctuates with the choices made by the player.

During the focus group meeting, people were positive about the bar as it provides feedback during the scenario (prototype 1). However, remarks were made to increase the clarity of this mechanism. The trophies were also evaluated positively (prototype 1). The reactions to the performance graph (prototype 1) were more mixed, which was also the case for the hangman game (prototype 2). While participants liked it as it would be fun to play, they were worried that it might be too distracting. The badges (prototype 2) were seen as something fun, but the participants indicate to not have a desire to achieve them. Finally, the idea of a leaderboard (prototype 3) was liked, but the fluctuating bonus score during the scenario, introduced in prototype 3, was not.

Based on this feedback, the student decided to use the trophies and the hangman game, as well as the bar during the scenario, for the final prototype. Taking into account the remarks made during the focus group. In the control condition only the trophies were present, while in the experimental condition had all the reward systems of the final version implemented. In the final evaluation 54 elderly participants took part, 27 in each group. Objective data was collected, being the choices made in each scenario. Subjective data, both in the form of open and closed questions, was collected to give some background information to explain the objective data. No significant effects were found for an increase in the learning results for the experimental condition. In the subjective data it became clear that participants often did not pay attention to the bar during the scenarios and that people already had knowledge about doorstep scam prevention and might therefore not have learned enough to measure a difference. The results for the hangman component were mixed.

This study is an interesting example of how conventional game elements can be incorporated in new virtual training games as reward mechanism. Although the results do not show an added learning value, the idea of it is not rejected. Furthermore, feedback during the scenario was seen as a valuable addition to the serious game. Moreover, it shows that it is important to test if game elements are clear to users, as in the current design the bar missed its goal as it was overseen by many users.

5.3 *Increasing motivation*

While participants of the final evaluation of the serious game indicated that they like playing the game, the score for the statement about whether participants wanted to use the system in the future was lower. This project [65] aims to increase this score, by adding a reward mechanism to stimulate repeated play.

Similar to the research setup described in Section 5.2, three different prototypes of motivation mechanisms, based on literature research [57, 66–71], were created and discussed during a focus group at the KBO-PCOB:

- A reward garden, shown in the menu of the game. New flowers to extend the garden appear when a scenario is completed, and is independent of the outcome of the scenario. This is used to stimulate repeated training.

- A score system in which the score is based on the choices made in the scenario.
- A dog mascot (cartoon style) was added to the scenarios to give immediate feedback, both in facial expression as well as textual, after a choice of the player.

The participants of the focus group were enthusiastic about the score system and especially enthusiastic about the dog mascot. They suggested adding a high score functionality to the game. While they were enthusiastic about the dog mascot, they were also worried that the immediate feedback would take the player out of the flow of the game.

The reward garden was received with less enthusiasm; it was unclear to the participants when new flowers appeared. Moreover, there would be a need to remove flowers at a certain moment, or have a maximum number of flowers as there is only a limited amount of space for these. The participants had mixed reactions on this, some were not enthusiastic about removing flowers, while others were neutral about the concept of gaining flowers.

For the final evaluation of this study, two versions of virtual training were created. In the control condition, a version was used that includes a score and high score system. For the experimental group this version was extended with the dog mascot. However, instead of giving immediate feedback, the dog mascot now only gives feedback at the beginning and end of a scenario, to address the concerns about the interruption of the flow of the app.

In the final evaluation, 62 participants, equally divided over the two groups, played the game and filled in a questionnaire afterwards. This questionnaire contained both questions about the background of the participants, experience with doorstep scams, and questions related to the experience with the game they played.

This research explored how motivational mechanisms around the scenario game can be used to increase the motivation of players to continue to play the game. Although it did not find any significant results, the focus group did point out that feedback during the game in the form of a mascot was received with enthusiasm. However, there were concerns about the interruption of the flow of the scenario due to this. The final evaluation showed that reducing it to only providing feedback at the beginning and end of a scenario did not stand out enough for players to notice it and to have an effect in this study. In the research described in Section 5.1, feedback during the scenario was also preferred by participants. Both studies also showed that to meet this desire a balance needs to be found between feedback and scenario flow. In future research, more extended attempts to do so can be researched.

5.4 Possibilities for future work

In the three described side projects, extensions for the virtual training are studied, focusing on the learning outcomes, rewards, and motivation. Although the studies only had a limited number of participants, the results point out some interesting things. Participants expressed a preference for feedback during the scenario when explicitly asked, but once implemented, this did not have an effect and was not even noticed by the players. However, when making additional components in the virtual training more outstanding, this might harm the flow of the app or make the interface (too) complex. This underlines the importance of (repeated) testing with possible end users to ensure that components are incorporated in the system in an optimal way, being noticeable without harming the flow of the app.

There are many other possibilities to extend the existing game. The game could be extended by increasing the number of scenarios or adding more game elements, such as social features. It could also be considered to replace the way to select one of the multiple choice reactions (touch) by voice control, which could simplify the flow of the system as no additional step is needed for the voice analysis. However, it needs to be studied whether this is an accessible way for elderly users to interact with the system.

It is also possible to use the key principles, the scenarios and the voice analysis, for a serious game on a different platform, such as a head-mounted display for a virtual reality (VR) version of the game. Currently, such technology is not commonly used among the aimed target group. Nevertheless, it is interesting to see if such a technology could increase the engagement of players. Moreover, it would be possible to also include a physical threat that comes with some types of doorstep scams. VR-technology allows for such things to be included, however the safety of the elderly target group needs to be guaranteed.

6 Conclusion

In this paper the design, implementation and evaluation of a serious game to improve the verbal resilience of elderly users against doorstep scams is discussed. In collaboration with KBO-PCOB, different scenarios for this serious game were developed. The serious game aims to increase verbal resilience in two ways: via the content of the message and via the assertiveness of the voice of the user. For the design of the serious game, principles from theory were used. The hierarchy of players' needs was used to ensure that players are sufficiently motivated by the game. Moreover, design principles for an elderly target group are consulted. This made the game accessible for the target group, in the final evaluation all participants were able to use the application without assistance. During the final evaluation, 19 participants filled in a questionnaire about their experience with the virtual training. Overall, the evaluation of the system was positive, specifically the potential effect of the system was evaluated high, in particular the added value of the training to existing trainings/campaigns. Overall, the engagement category scored the lowest, but still positive.

The side projects that are introduced in Section 5 showed how different components, serving their own purpose, could be added to the virtual training to further enhance it. The research showed that there is a difficult balance between sticking close to the reality with realistic and uninterrupted scenarios versus giving feedback and helping players to make the right choices.

The long-term effect of the virtual training on people's performance is not studied in this research. This has different reasons, among which the fact that studying the effect of this doorstep prevention training is ethically challenging. It is always hard to measure potential situations, situations that might happen or not. To study the effect of this application, it is possible to study changes in the registered number of doorstep scams before and after the publication of the application. However, it remains difficult to draw conclusions on the effect of the application. There might be other reasons why the number decreased e.g. prevention by police, use of special locks on the door. It is, however, possible to test the knowledge of people before and after working with the application and again measure the effects after a longer period. Another possibility, that was beyond the scope of this project, is to test the reaction of participants in staged situations with the help of actors. However, the current study sheds light on a first important step that has been taken, it showed the feasibility of using a serious game for verbal resilience for the specific target group and it also gives more insight in how game elements can be used in this context and for this target group.

One of the important findings of this research is that participants saw a great added value of the training in incorporating it in existing training sessions. Moreover, this serious game shows how multiple choice interaction can be enriched with voice analysis. The application ('Trucs tegen babbeltrucs') is published and freely available in the Google Play Store and Apple's app store. Moreover, the application is used by the KBO-PCOB for prevention training sessions.

Acknowledgements

This project has been done in collaboration with the KBO-PCOB and has been funded by the Dutch ministry of Safety and Justice.

The development of the virtual training has been done with the help of Daniel Formolo and Marco Otte.

The original work described in Section 5.1, 5.2, and 5.3 is individually performed by students from the Vrije Universiteit Amsterdam; Celino Toussaint, Marvin Lau, and Dennis Walter. For this paper, their work is summarized by the authors.

References

- [1] “Doorstep scams - scams and fraud - age uk,” <https://www.ageuk.org.uk/information-advice/money-legal/scams-fraud/doorstep-scams/>, accessed: 2020-01-30.
- [2] “Scams awareness month - citizens advice,” www.citizensadvice.org.uk/about-us/campaigns/current_campaigns/scams-awareness-month/, accessed: 2020-01-30.
- [3] “Samen sterker voor senioren - kbo-pboc,” www.kbo-pcob.nl/, accessed: 2020-01-30.
- [4] A. C. Siang and R. Radha Krishna, “Theories of learning: a computer game perspective,” in *Fifth International Symposium on Multimedia Software Engineering, 2003. Proceedings.*, Conference Proceedings. doi: 10.1109/MMSE.2003.1254447 pp. 239–245.
- [5] S. Srikwan and M. Jakobsson, “Using cartoons to teach internet security,” *Cryptologia*, vol. 32, no. 2, pp. 137–154, 2008. doi: 10.1080/01611190701743724
- [6] S. Vosinakis and T. Panayiotopoulos, “Simhuman: A platform for real-time virtual agents with planning capabilities,” ser. Intelligent Virtual Agents. Springer Berlin Heidelberg, Conference Proceedings. doi: 10.1007/3-540-44812-8_17 pp. 210–223.
- [7] T. Bosse and C. Gerritsen, “Towards serious gaming for communication training - a pilot study with police academy students,” ser. Intelligent Technologies for Interactive Entertainment. Springer International Publishing. doi: 10.1007/978-3-319-49616-0_2 pp. 13–22.
- [8] T. Bosse, C. Gerritsen, and J. de Man, “An intelligent system for aggression de-escalation training,” in *Proceedings of the Twenty-second European Conference on Artificial Intelligence*. IOS Press, 2016. doi: 10.3233/978-1-61499-672-9-1805 pp. 1805–1811.
- [9] D. R. Michael and S. L. Chen, *Serious Games: Games That Educate, Train, and Inform*. Muska & Lipman/Premier-Trade, 2005. ISBN 1592006221
- [10] E. I. Konstantinidis, A. S. Billis, C. A. Mouzakidis, V. I. Zilidou, P. E. Antoniou, and P. D. Bamidis, “Design, implementation, and wide pilot deployment of fitforall: An easy to use exergaming platform improving physical fitness and life quality of senior citizens,” *IEEE Journal of Biomedical and Health Informatics*, vol. 20, no. 1, pp. 189–200, 2016. doi: 10.1109/JBHI.2014.2378814
- [11] R. Nouchi, Y. Taki, H. Takeuchi, H. Hashizume, Y. Akitsuki, Y. Shigemune, A. Sekiguchi, Y. Kotozaki, T. Tsukiura, Y. Yomogida, and R. Kawashima, “Brain training game improves executive functions and processing speed in the elderly: A randomized controlled trial,” *PLOS ONE*, vol. 7, no. 1, p. e29676, 2012. doi: 10.1371/journal.pone.0029676
- [12] J. Jeuring, F. Grosfeld, B. Heeren, M. Hulsbergen, R. IJntema, V. Jonker, N. Mastenbroek, M. van der Smagt, F. Wijmans, M. Wolters, and H. van Zeijts, “Communicate! - a serious game for communication skills -,” ser. Design for Teaching and Learning in a Networked World. Springer International Publishing, Conference Proceedings. doi: 10.1007/978-3-319-24258-3_49 pp. 513–517.

- [13] F. Vaassen, J. Wauters, F. Van Broeckhoven, M. Van Overveldt, W. Daelemans, and K. Eneman, “delearyous: Training interpersonal communication skills using unconstrained text input,” *Proc. of ECGBL*, pp. 505–513, 2012.
- [14] P. van der Meer, “Onderzoeksrapport veiligheid 2017 (version 1.0),” *Technical Report commissioned by KBO-PCOB*, 2017.
- [15] “Assertive - meaning in the cambridge dictionary,” www.dictionary.cambridge.org/dictionary/english/assertive, accessed: 2020-01-30.
- [16] B. J. Winship and J. D. Kelley, “A verbal response model of assertiveness,” *Journal of Counseling Psychology*, vol. 23, no. 3, pp. 215–220, 1976. doi: 10.1037/0022-0167.23.3.215
- [17] S. A. Rathus, “An experimental investigation of assertive training in a group setting,” *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 3, no. 2, pp. 81–86, 1972. doi: [https://doi.org/10.1016/0005-7916\(72\)90003-1](https://doi.org/10.1016/0005-7916(72)90003-1). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/0005791672900031>
- [18] R. M. McFall and D. B. Lillesand, “Behavior rehearsal with modeling and coaching in assertion training,” *Journal of Abnormal Psychology*, vol. 77, no. 3, pp. 313–323, 1971. doi: 10.1037/h0031010
- [19] R. M. McFall and A. R. Marston, “An experimental investigation of behavior rehearsal in assertive training,” *Journal of Abnormal Psychology*, vol. 76, no. 2, pp. 295–303, 1970. doi: 10.1037/h0030112
- [20] J. P. Galassi, M. D. Galassi, and M. C. Litz, “Assertive training in groups using video feedback,” *Journal of Counseling Psychology*, vol. 21, no. 5, pp. 390–394, 1974. doi: 10.1037/h0037094
- [21] T. R. Nichols, J. A. Graber, J. Brooks-Gunn, and G. J. Botvin, “Ways to say no: Refusal skill strategies among urban adolescents,” *American Journal of Health Behavior*, vol. 30, no. 3, pp. 227–236, 2006. doi: 10.5993/AJHB.30.3.1. [Online]. Available: <https://www.ingentaconnect.com/content/png/ajhb/2006/00000030/00000003/art00001>
- [22] M. A. Miller-Day, J. Alberts, M. L. Hecht, M. R. Trost, and R. L. Krizek, *Adolescent relationships and drug use*. Psychology Press, 2014. ISBN 1135663483
- [23] P. Englander-Golden, J. Elconin, and V. Satir, “Assertive/leveling communication and empathy in adolescent drug abuse prevention,” *Journal of Primary Prevention*, vol. 6, no. 4, pp. 231–243, 1986. doi: 10.1007/bf01330264
- [24] J. T. Shope, L. A. Copeland, R. Maharg, T. E. Dielman, and A. T. Butchart, “Assessment of adolescent refusal skills in an alcohol misuse prevention study,” *Health Education Quarterly*, vol. 20, no. 3, pp. 373–390, 1993. doi: 10.1177/109019819302000308
- [25] L. A. Rohrbach, J. W. Graham, W. B. Hansen, B. R. Flay, and C. A. Johnson, “Evaluation of resistance skills training using multitrait-multimethod role play skill assessments,” *Health Education Research*, vol. 2, no. 4, pp. 401–407, 1987. doi: 10.1093/her/2.4.401
- [26] A. Charlton, K.-E. Minagawa, and D. While, “Saying “no” to cigarettes: a reappraisal of adolescent refusal skills,” *Journal of Adolescence*, vol. 22, no. 5, pp. 695–707, 1999. doi: <https://doi.org/10.1006/jado.1999.0264>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0140197199902646>
- [27] G. E. Turner, C. Burciaga, S. Sussman, E. Klein-Selski, S. Craig, C. W. Dent, H. R. C. Mason, D. Burton, and B. Flay, “Which lesson components mediate refusal assertion skill improvement in school-based adolescent tobacco use prevention?” *International Journal of the Addictions*, vol. 28, no. 8, pp. 749–766, 1993. doi: 10.3109/10826089309062171
- [28] S. Sussman, A. W. Stacy, C. W. Dent, C. Burciaga, D. Burton, and B. R. Flay, “Refusal assertion versus conversational skill role-play competence: Relevance to preven-

- tion of tobacco use,” *Statistics in Medicine*, vol. 12, no. 3-4, pp. 365–376, 1993. doi: 10.1002/sim.4780120318
- [29] R. M. Eisler, P. M. Miller, and M. Hersen, “Components of assertive behavior,” *Journal of Clinical Psychology*, vol. 29, no. 3, pp. 295–299, 1973. doi: 10.1002/1097-4679(197307)29:3<295::aid-jclp2270290305>3.0.co;2-9
- [30] T. R. Nichols, S. Birmel, J. A. Graber, J. Brooks-Gunn, and G. J. Botvin, “Refusal skill ability: An examination of adolescent perceptions of effectiveness,” *The Journal of Primary Prevention*, vol. 31, no. 3, pp. 127–137, 2010. doi: 10.1007/s10935-010-0213-9
- [31] A. S. Bellack, M. Hersen, and S. M. Turner, “Role-play tests for assessing social skills: Are they valid?” *Behavior Therapy*, vol. 9, no. 3, pp. 448–461, 1978. doi: [https://doi.org/10.1016/S0005-7894\(78\)80089-6](https://doi.org/10.1016/S0005-7894(78)80089-6). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0005789478800896>
- [32] R. M. Eisler, M. Hersen, and P. M. Miller, “Effects of modeling on components of assertive behavior,” *Journal of Behavior Therapy and Experimental Psychiatry*, vol. 4, no. 1, pp. 1–6, 1973. doi: [https://doi.org/10.1016/0005-7916\(73\)90030-X](https://doi.org/10.1016/0005-7916(73)90030-X). [Online]. Available: <http://www.sciencedirect.com/science/article/pii/000579167390030X>
- [33] C. L. Ridgeway, “Nonverbal behavior, dominance, and the basis of status in task groups,” *American Sociological Review*, vol. 52, no. 5, pp. 683–694, 1987. doi: 10.2307/2095603. [Online]. Available: www.jstor.org/stable/2095603
- [34] C. E. Kimble and S. D. Seidel, “Vocal signs of confidence,” *Journal of Nonverbal Behavior*, vol. 15, no. 2, pp. 99–105, 1991. doi: 10.1007/bf00998265
- [35] R. M. Eisler, M. Hersen, P. M. Miller, and E. B. Blanchard, “Situational determinants of assertive behaviors,” *Journal of Consulting and Clinical Psychology*, vol. 43, no. 3, pp. 330–340, 1975. doi: 10.1037/h0076717
- [36] K. Bosman, T. Bosse, and D. Formolo, “Virtual agents for professional social skills training: An overview of the state-of-the-art,” ser. Intelligent Technologies for Interactive Entertainment. Springer International Publishing, Conference Proceedings. doi: https://doi.org/10.1007/978-3-030-16447-8_8 pp. 75–84.
- [37] M. C. Howard and M. B. Gutworth, “A meta-analysis of virtual reality training programs for social skill development,” *Computers & Education*, vol. 144, p. 103707, 2020. doi: 10.1016/j.compedu.2019.103707
- [38] H. Tanaka, S. Sakriani, G. Neubig, T. Toda, H. Negoro, H. Iwasaka, and S. Nakamura, “Teaching social communication skills through human-agent interaction,” *ACM Trans. Interact. Intell. Syst.*, vol. 6, no. 2, pp. 1–26, 2016. doi: 10.1145/2937757
- [39] F. L. Greitzer, “Ingredients of effective and engaging online learning,” in *Inter-Lab 2005 Conference*, Conference Proceedings.
- [40] D. H. Goh, R. P. Ang, and H. C. Tan, “Strategies for designing effective psychotherapeutic gaming interventions for children and adolescents,” *Computers in Human Behavior*, vol. 24, no. 5, pp. 2217–2235, 2008. doi: <https://doi.org/10.1016/j.chb.2007.10.007>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0747563207001616>
- [41] A. H. Maslow, ““motivation and personality”, 3 sub edition,” *New York*, 1987.
- [42] D. Schoene, S. R. Lord, K. Delbaere, C. Severino, T. A. Davies, and S. T. Smith, “A randomized controlled pilot study of home-based step training in older people using videogame technology,” *PLOS ONE*, vol. 8, no. 3, 2013. doi: 10.1371/journal.pone.0057734
- [43] L. M. Taylor, R. Maddison, L. A. Pfaeffli, J. C. Rawstorn, N. Gant, and N. M. Kerse, “Activity and energy expenditure in older people playing active video games,” *Archives of Physical Medicine and Rehabilitation*, vol. 93, no. 12, pp.

- 2281–2286, 2012. doi: <https://doi.org/10.1016/j.apmr.2012.03.034>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0003999312002584>
- [44] S. T. Chen and I. Chiang, “Effects of improvement on selective attention: Developing appropriate somatosensory video game interventions for institutional-dwelling elderly with disabilities,” *Turkish Online Journal of Educational Technology-TOJET*, vol. 11, no. 4, pp. 409–417, 2012.
- [45] J. Senger, T. Wälisch, M. John, H. Wang, A. N. Belbachir, B. Kohn, A. Smurawski, R.-Z. Lübben, and G. Jones, *Serious Gaming: Enhancing the Quality of Life among the Elderly through Play with the Multimedia Platform SilverGame*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 317–331. ISBN 978-3-642-27491-6
- [46] D. Formolo and T. Bosse, “Human vs. computer performance in voice-based recognition of interpersonal stance,” ser. Human-Computer Interaction. User Interface Design, Development and Multimodality. Springer International Publishing, Conference Proceedings. doi: 10.1007/978-3-319-58071-5_51 pp. 672–686.
- [47] T. Gales, “Identifying interpersonal stance in threatening discourse: An appraisal analysis,” *Discourse Studies*, vol. 13, no. 1, pp. 27–46, 2011. doi: 10.1177/1461445610387735
- [48] D. Formolo and T. Bosse, “Towards interactive agents that infer emotions from voice and context information,” *EAI Endorsed Transactions on Creative Technologies*, vol. 4, no. 10, 2017. doi: 10.4108/eai.4-9-2017.153054
- [49] “Unity real-time development platform,” www.unity3d.com, accessed: 2020-01-30.
- [50] “Unity asset store,” <https://assetstore.unity.com/>, accessed: 2020-01-30.
- [51] “3d animations software - iclone - reallusion,” www.reallusion.com/iclone/, accessed: 2020-01-30.
- [52] “Faceware technologies, inc.” www.facewaretech.com, accessed: 2020-01-30.
- [53] C. Gerritsen and T. Bosse, “Virtual communication training for the law enforcement domain,” *EAI Endorsed Transactions on Creative Technologies*, vol. 5, no. 14, 2018. doi: 10.4108/eai.16-1-2018.153642
- [54] C. Toussaint, “Gamificatie: Het effect van het toevoegen van een quiz achter een serious game simulatie over babbeltrucs voor ouderen,” 2018.
- [55] H. Lau, “The effect of reward systems in the gamification of doorstep scams simulation training for elderly,” 2018.
- [56] B. De Schutter and V. Vanden Abeele, “Meaningful play in elderly life,” in *Annual Meeting of the International Communication Association, Date: 2008/01/01-2008/01/01, Location: Montreal, Quebec, Canada*, Conference Proceedings.
- [57] E. Deci and R. Ryan, “Intrinsic motivation and self-determination in human behavior: Springer science business media,” 1985.
- [58] R. N. Landers, K. N. Bauer, R. C. Callan, and M. B. Armstrong, *Psychological Theory and the Gamification of Learning*. Cham: Springer International Publishing, 2015, pp. 165–186. ISBN 978-3-319-10208-5
- [59] E. A. Locke, K. N. Shaw, L. M. Saari, and G. P. Latham, “Goal setting and task performance: 1969-1980,” *Psychological Bulletin*, vol. 90, no. 1, pp. 125–152, 1981. doi: 10.1037/0033-2909.90.1.125
- [60] A. Nagle, P. Wolf, R. Riener, and D. Novak, “The use of player-centered positive reinforcement to schedule in-game rewards increases enjoyment and performance in a serious game,” *International Journal of Serious Games*, vol. 1, no. 4, pp. 35–47, 2014. doi: 10.17083/ijsg.v1i4.47
- [61] G. Richter, D. R. Raban, and S. Rafaeli, *Studying Gamification: The Effect of Rewards and Incentives on Motivation*. Cham: Springer International Publishing, 2015, pp. 21–46. ISBN 978-3-319-10208-5

- [62] M. Sailer, J. U. Hense, S. K. Mayr, and H. Mandl, “How gamification motivates: An experimental study of the effects of specific game design elements on psychological need satisfaction,” *Computers in Human Behavior*, vol. 69, pp. 371–380, 2017. doi: <https://doi.org/10.1016/j.chb.2016.12.033>. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S074756321630855X>
- [63] O. Selcen and Ş. Sercan, *Gaining Reward vs. Avoiding Loss: When Does Gamification Stop Being Fun?* Hershey, PA, USA: IGI Global, 2016, pp. 48–71. ISBN 9781466686519. [Online]. Available: <http://services.igi-global.com/resolvedoi/resolve.aspx?doi=10.4018/978-1-4666-8651-9.ch003>
- [64] V. H. Vroom, *Work and motivation*, ser. Work and motivation. Oxford, England: Wiley, 1964.
- [65] D. Walter, “Increasing motivation in elderly to continue playing serious games: A case study of doorstep scams prevention,” 2018.
- [66] M. Csikszentmihalyi, S. Khosla, and J. Nakamura, “Flow at work,” *The Wiley Blackwell Handbook of the Psychology of Positivity and Strengths-Based Approaches at Work*, pp. 99–109, 2016. doi: 10.1002/9781118977620.ch7
- [67] I. J. Deary and G. Der, “Reaction time, age, and cognitive ability: Longitudinal findings from age 16 to 63 years in representative population samples,” *Aging, Neuropsychology, and Cognition*, vol. 12, no. 2, pp. 187–215. doi: 10.1080/13825580590969235
- [68] J. Olds and P. Milner, “Positive reinforcement produced by electrical stimulation of septal area and other regions of rat brain,” *Journal of Comparative and Physiological Psychology*, vol. 47, no. 6, pp. 419–427, 1954. doi: 10.1037/h0058775
- [69] I. Pairoa and P. Arunrangsiwed, “The effect of brand mascots on consumers’ purchasing behaviors,” *World Academy of Science, Engineering and Technology, International Journal of Social, Behavioral, Educational, Economic, Business and Industrial Engineering*, vol. 10, no. 5, pp. 1612–1615, 2016.
- [70] D. Thompson, T. Baranowski, R. Buday, J. Baranowski, V. Thompson, R. Jago, and M. J. Griffith, “Serious video games for health: How behavioral science guided the development of a serious video game,” *Simulation Gaming*, vol. 41, no. 4, pp. 587–606, 2008. doi: 10.1177/1046878108328087
- [71] K. Van Calster, W. Lens, and J. R. Nuttin, “Affective attitude toward the personal future: Impact on motivation in high school boys,” *The American Journal of Psychology*, vol. 100, no. 1, pp. 1–13, 1987. doi: 10.2307/1422639. [Online]. Available: www.jstor.org/stable/1422639

A *Extended explanation scenarios*

- Stories at the door:
 - The electricity, gas, or water, needs to be checked by the con artist, therefore he/she must enter your home.
 - The con artist has a delivery (package, flowers) that he/she wants to give to the victim. Either this is an excuse to enter the house or it is an excuse to seduce the victim to do a small payment (e.g. delivery costs), often with the purpose to steal more money from the victim than just the payment.
 - A con artist in the role of handyman wants to do some job for the victim, however he/she asks (a large amount of) money that needs to be paid right away. The job will not be done (correctly), but the money is taken.

- Stories on the phone:
 - The con artist calls with a story about fraudulent payments made, pretending to be the bank. He/she will ask private information, such as their debit/credit card PIN.
 - The con artist tells the victim that he/she has won a price, but in order to claim the price the victim needs to give information or make a small payment.
 - The con artist calls pretending to do a survey and asks the victim to give personal information. This information is later used to perform a financial scam or identity theft.

- Stories on the street:
 - The con artist claims to be collecting money or selling something for charity. Either the money is not used for charity or the con artist uses this to be able to easily pickpocket the target.
 - The con artist sells a newspaper or magazine, however the victim ends up with a long and expensive subscription.
 - The con artist sells some goods at the street, either the prize is too high (or the quality too low) or the good turns out to be completely worthless.

B Full flowchart (translated) example scenario

C Questionnaire statements translated from Dutch

1. The trainings software was user friendly.
2. The multiple choices menus always contained an answer I agreed with.
3. I felt like I got better in handling the scenarios correctly.
4. I believe the events in the scenarios were realistic.
5. I believe the virtual characters acted credibly.
6. Interacting with the virtual characters felt natural.
7. I felt like my answers influenced the behaviour of the virtual characters.
8. I felt capable steering the course of the conversations.
9. When I said something wrong I directly noticed it from the behaviour of the virtual characters.
10. The duration of the scenarios was realistic.
11. I felt personally addressed by the virtual characters.
12. The feedback at the end of each scenario was useful.

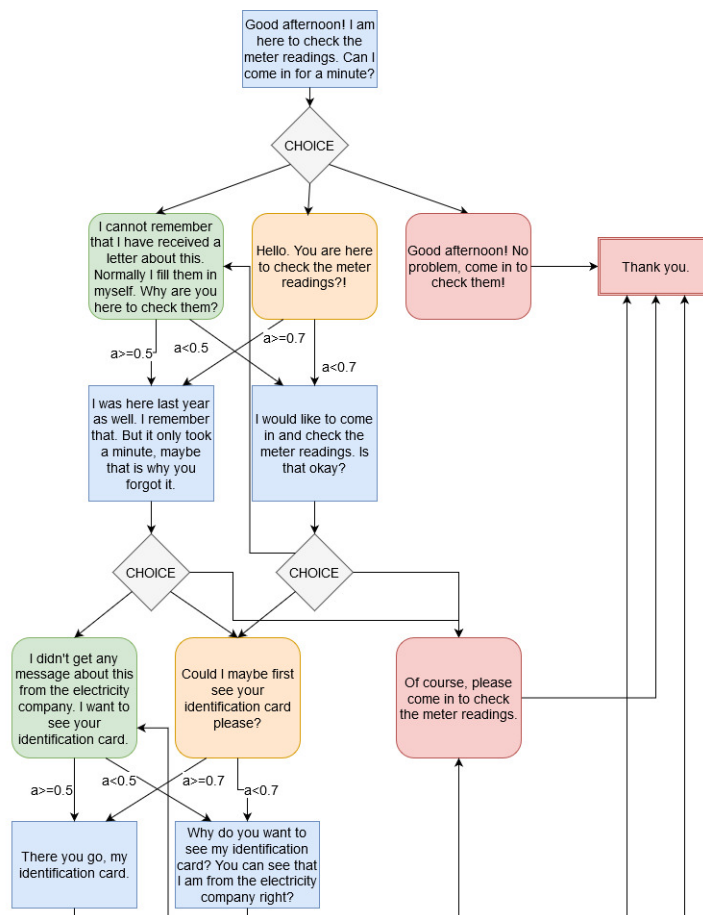


Figure 13: A (translated) scenario flowchart (part I)

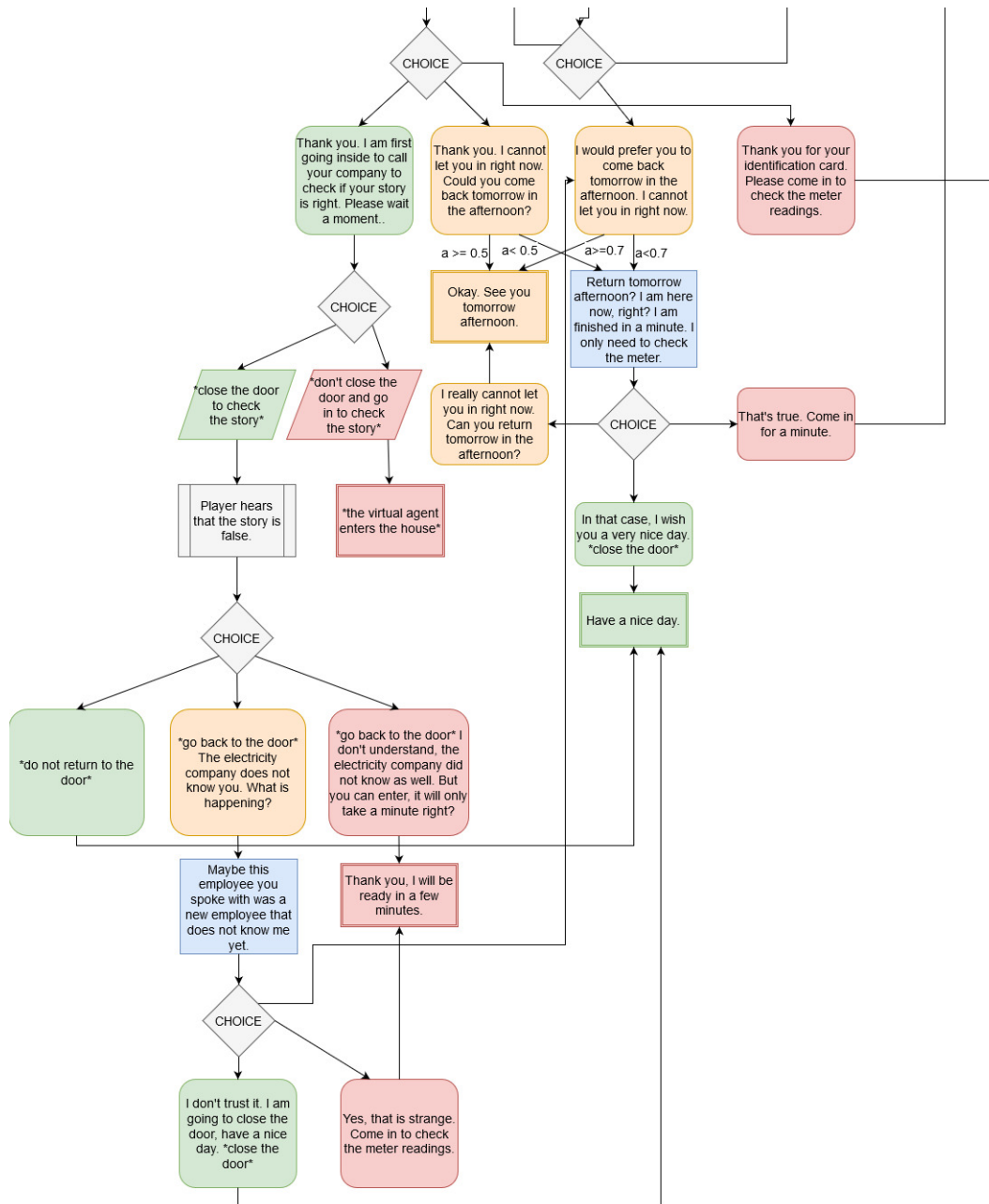


Figure 13: (part II) A (translated) scenario flowchart

13. During the training I felt immersed in the scenarios.
14. The visual aspects of the scenarios made me feel like the scenarios were real.
15. The audial aspects of the scenarios made me feel like the scenarios were real.
16. For people that tried the voice analysis only: I felt that the voice analysis classified my voice correctly.
17. This training made me think about the what to do and what not to do during a doorstep scam.
18. After following this training I am better capable to determine the right way of handling such situations.

19. For people that tried the voice analysis only: This training taught me how to use my voice in an assertive way during a doorstep scam.
20. I believe this training can be a useful addition to role-play training.
21. I liked doing the training.
22. The resilience training met my expectations.
23. I would like to continue doing this type of training in the future.