frontiers in Computer Science

ORIGINAL RESEARCH
published: 31 July 2020
doi: 10.3389/fcomp.2020.00028

# A Technical and Conceptual Framework for Serious Role-Playing Games in the Area of Social Skill Training

*Julia Othlinghaus-Wulhorst* and H. Ulrich Hoppe*

*COLLIDE Group, Department of Computer Science and Applied Cognitive Science, Faculty of Engineering, University of Duisburg-Essen, Duisburg, Germany*

OPEN ACCESS

***Edited by:***
*Carlos Vaz De Carvalho, Polytechnic Institute of Porto, Portugal*

***Reviewed by:***
*Patricia Martin-Rodilla, University of A Coruña, Spain*
*António Fernando Coelho, University of Porto, Portugal*
*Ramon Fabregat Gesa, University of Girona, Spain*

****Correspondence:***
*Julia Othlinghaus-Wulhorst*
*othlinghaus@collide.info*

***Specialty section:***
*This article was submitted to Human-Media Interaction, a section of the journal Frontiers in Computer Science*

***Received:*** *28 December 2019*
***Accepted:*** *24 June 2020*
***Published:*** *31 July 2020*

***Citation:***
*Othlinghaus-Wulhorst J and Hoppe HU (2020) A Technical and Conceptual Framework for Serious Role-Playing Games in the Area of Social Skill Training. Front. Comput. Sci. 2:28. doi: 10.3389/fcomp.2020.00028*

Virtual role-playing games can provide an authentic experience of situated learning and allow for trying out different problem-solving and communication strategies without consequences in the real world. This is of particular interest and benefit for the training of social skills. This article presents a conceptual and technical framework for serious role-playing games for the training of specific social skills in virtual 2D learning environments involving chatbots in dialog-centric settings. It summarizes different use cases and evaluation results from prior studies. From the design perspective, several distinctive conceptual features characterize our framework: (1) chat-like interaction with an AI-controlled chatbot, (2) separate phases of immersion and reflection to facilitate a change of perspective that is considered conducive for learning, (3) the learning process is emphasized by means of adaptive feedback based on individual analyses. We propose a system architecture that is based on three components: (1) AI-controlled chatbots that adapt to the player's behavior, (2) a multi-agent blackboard system as the backbone in order to keep components independent and optimize performance due to parallel processing, and (3) intelligent support for an automated evaluation of the player's performance and feedback generation. The training scenarios presented and discussed in this article include workplace-oriented conflict management, patient-centered medical interviews, and customer complaint management. First evaluation studies indicate that the scenarios may be well-suited for real training situations. Due to its flexible architecture, our framework and approach can easily be tailored to different settings and use cases and thus serve as a basis for future research focusing on the adaptation to other contexts and systems. On the basis of these developments, we elaborate important design dimensions, reflect and discuss general issues and major challenges, summarize and contrast different approaches and strategies, and identify opportunities for serious role-playing games in the area of social skills training.

**Keywords: virtual role play, intelligent support, serious games, social skills, chatbots**

## INTRODUCTION

In recent years, serious games have been established as an efficient medium in education and professional training (Michael and Chen, 2006; Marr, 2010). The serious gaming approach attempts to use the appeal of digital games not only for entertainment purposes but also to convey "serious" content and to train practice-oriented skills (Ritterfeld et al., 2009). The combination of the serious gaming approach with role play scenarios is particularly promising. Role play enables learners to explore new situations and train how to act and react in these situations (Martens et al., 2008). Virtual role-playing games provide mobile, safe, and continuable environments, whereas traditional role plays can be time-consuming, costly, and difficult to administer (Totty, 2005). In addition, they lack repeatability. One general problem in the evaluation of role play experiences for educational purposes is the effort involved in analyzing and reflecting on the actual role play following the enactment. Traditional scenarios typically rely on video recording and, if applicable, note-taking. However, virtual learning environments enable structured recordings with integrated indexing, navigation instruments, search functions, and cross-references between different media and data sources. In addition, computer-supported analyses can help to evaluate and track the learners' performance. This is an important aspect, since without feedback and post-role-play reflection, the transfer to real word situations cannot be ensured (Lim et al., 2009). An additional important advantage of serious role-playing games in contrast to other virtual learning activities and environments is the motivational component, which may lead to intense and passionate involvement of learners (Susi et al., 2007).

Based on a series of different instances of role-playing games for the training of specific social skills, this article presents the underlying conceptual and technical framework that facilitated the implementation of the different applications. This framework is characterized by using scripted chatbots as training cases in a dialogic setting. A multi-agent architecture supports both the actual dialogic processing as well as the evaluation of the dialogs and the generation of adaptive feedback. Conceptual and technical aspects of this framework are described in chapters Framework: Conceptual Approach and Framework: Technical Approach, following up on a discussion of related work in this area (chapter Related Work). Chapter Case Studies assembles several case studies conducted with different instances of virtual role-playing environments based on the framework, reporting on experience and evaluation results. Chapter Dimensions of the Design of Serious Role-Playing Games for the Training of Social Skills combines this specific experience with general issues in the design of serious role-playing games to devise a set design dimensions in the sense of important aspects to be considered in the design, description and comparison of serious role-playing games.

## RELATED WORK

### Serious Role-Playing Games for the Training of Social Skills

Serious games can be defined as "any form of interactive computer-based game software of one or multiple players to be used on any platform and hat has been developed with the intention to be more than entertainment" (Ritterfeld et al., 2009) and with an explicit focus on education. Games of this category are supposed to convey specific knowledge or train certain skills by using the attractiveness of entertainment games (Susi et al., 2007). Serious games can generally cover many different subject areas, but their application is mainly found in healthcare, education, and training, including military or employee training in companies (Marr, 2010). Serious games are widely accepted as an important and efficient medium with respect to education, training, and behavioral change (Michael and Chen, 2006). They are recognized to have several benefits: Serious games facilitate learning experiences while not having negative or harmful impacts (Ritterfeld et al., 2009). Games in general not only have a positive effect on the development of the player but can also be conducive to many different skills. Among others, Mitchell and Savill-Smith suggest that such target competences can be related to cognitive, social, analytical, and strategic aspects (Mitchell and Savill-Smith, 2005). Squire and Jenkins also made a comparable assessment (Squire and Jenkins, 2003). Further advantages include the reduction of costs and time associated with the use of serious games. They make it possible to recreate situations or working conditions that would otherwise not be possible in the real world (Corti, 2006; Susi et al., 2007). Serious games intend to facilitate deep and sustained learning (Gee, 2007) and prove to be more effective than traditional pedagogy and other educational technologies (Prensky, 2000; Ritterfeld et al., 2009).

Michael and Chen differentiate between games that educate and games that train (Michael and Chen, 2006). Games that educate should convey knowledge, facts and processes in a playful way, thereby contributing to education, while games that train are intended to improve the learners' skills in virtual environments or simulations. Our work is focused on the second category, more specifically on serious games for the training of social skills based on role play. Social skills can be seen as a sub-category of soft skills. The term *soft skills* refers to a broad concept that describes a set of personal attributes or traits expressing how persons know and manage both themselves and their relationship with other people (Dell'Aquila et al., 2017). While no universal definition of the term "soft skills" is available, Dell'Aquila et al. combine several different approaches to the following definition (Dell'Aquila et al., 2017): "Soft skills are not domain or practice specific; experientially based; both self and people orientated; goal-related behaviors; inextricably complementary to hard technical knowledge and skills enabling completion of activities and accomplishment of results; and crucial for effective leadership performance." Social skills refer

to soft skills related to interaction with other people. It describes "the ability to interact with others in a given social context in specific ways that are societally acceptable or valued and at the same time personally beneficial, mutually beneficial, or beneficial primarily to others" (Combs and Slaby, 1977) and includes, e.g., communication, cooperation, assertion, responsibility, empathy, engagement, and self-control (Gresham and Elliott, 2008). Role play is a great instrument to train interaction with other people. Assuming roles provides the opportunity to train to act and re-act in new situations. It facilitates the creation of knowledge and meaning through concrete experiences (Lim et al., 2009). Also, the observation of role play can lead to conclusions about own behavior (Martens et al., 2008). The integration of role play in a serious gaming context seems to be particularly promising, as this combination (a) incorporates a highly motivational character and (b) creates opportunities for exploration and experimentation in a protective environment without any consequences in the real world. In addition, virtual role plays may be much more effective than conventional approaches in settings where the social component is a crucial factor (Lim et al., 2009).

Several serious role-playing games for the training of social skills are available. They can be assigned to three main categories of relevant social skills: (1) leadership skills, (2) communication skills, and (3) conflict management. Examples for serious role-playing games for training leadership skills are *Virtual Leader* (Knode and Knode, 2011), *TeamUp* (Bezuijen, 2012), and *Learn to Lead* (Di Ferdinando et al., 2011). *Virtual Leader* is a simulation game in which students practice leadership styles and approaches within a 3D environment using avatars and intelligent agents in order to create a preferably realistic environment (Knode and Knode, 2011). Players participate in virtual business meetings with animated characters and are required to make a series of decisions in five scenarios with increasing complexity. *TeamUp* is a collaborative game for the training of teamwork and leadership skills, developed at the TU Delft (Bezuijen, 2012). In this game, four players need to work together to overcome several challenges, each designed to cover a specific element of effective teamwork. In *Learn to Lead*, the players have to lead a simulated team of employees (e.g., workers in a bank, a post-office, or a local government office) that is competing against other teams (Di Ferdinando et al., 2011). In this game, the players have two main objectives: First, they need to ensure that the company is running efficiently and productively. Second, they need to ensure that their teams develop in the desired manner. The Productive Leadership Game is a simulation game that is supposed to foster leadership competencies to improve team-based and organizational productivity (Kesti et al., 2017). A recapitulatory overview of serious role-playing games for training leadership skills can be seen in **Table 1**.

There are various examples for serious role-playing games aiming at the training of communication skills: *ENACT* (Marocco et al., 2015) is an online game for the standardized psychometric assessment and training of negotiation skills based on Rahim's model of conflict handling styles (Rahim and Bonoma, 1979). In this game, players assume different characters to negotiate with computer-controlled virtual 3D agents in various scenarios representing everyday life situations. They can always choose one of four possible pre-defined sentences to communicate with the agents. In *DREAD-ED*, players become part of a crisis management team that is dealing with an emergency situation (Haferkamp et al., 2011). The game is organized into a series of timed rounds, separated by phases in which a tutor can provide feedback to the players. Bosse et al. developed a game targeted at police academy students that focuses on decision-making aspects in critical situations like the so-called "door scene" in which a police officer has been informed about an incoming emergency call and is supposed to find out if it is indeed a case of domestic violence or not (Bosse and Gerritsen, 2016). The players interact with virtual characters in a realistic 3D environment by using a relatively simple interaction paradigm based on multiple choice and dialog trees. In the game *deLearyous*, players assume the role of a manager who just announced that the parking facilities of the company are no longer free and needs to deal with the reaction of an employee (Vaassen and Wauters, 2012) by using unconstrained written natural language input. The design of the virtual character representing the employee is based on a framework for interpersonal communication called Leary's Rose (Leary, 1957). *JUST-TALK* is a serious game to train law enforcement personnel for encounters with persons showing symptoms of serious mental illness (Hubal et al., 2003). The players interact with these computer-controlled characters using spoken natural language. They are supposed to look for indications of particular forms of mental illness so that they can adapt their approach in an appropriate way und thus defuse the situation. In *POINTER*, a game developed for interview training targeted at police officers, the players assume the role of a police officer interacting with a subject in the context of a police interview (Linssen et al., 2014). The subject here is a virtual agent who is not cooperating during the interview. The players' task is to interact with the subject in a way that makes it cooperate in order to gather information from them. *ELECT BiLAT* is a simulation game in which soldiers practice bilateral engagements within a cultural context (Lane and Hays, 2008). The recruits are supposed to conduct meetings and negotiations with local leaders. *Maritime City* is a game targeted at social workers. It aims at training the ability to read emotional states of persons and improving communication skills in verbal and non-verbal forms (Flynn et al., 2011). In this game, players are asked to investigate a disturbance at a house where a woman is living with her two children and need to investigate a range of approaches for each part of the scenario. *TARDIS* is a scenario-based serious game simulation platform that supports social training and coaching in the context of job interviews (Gebhard et al., 2018). It is specifically intended to be used by young people and job-inclusion associations to explore, practice, and improve their skills in a diverse range of possible interview situations by interacting with virtual agents acting as recruiters. *Communicate!* is a serious role-playing game designed to support practicing interpersonal communication between health care professionals such as doctors, pharmacists, or psychologists and a patient or client (Jeuring et al., 2015). In the scenarios included in the game, the players find themselves in a consultation with a virtual character during which they can choose between

**TABLE 1 |** Serious games for the training of leadership skills (overview).

| Game | Author | Use of AI | Mode | Learning objective | Underlying framework/model/theory |
|---|---|---|---|---|---|
| Virtual Leader | Knode and Knode (2011) | Yes | Singleplayer | Leadership styles | – |
| TeamUp | Bezuijen (2012) | No | Multiplayer | Teamwork, leadership skills | – |
| Learn to Lead | Di Ferdinando et al. (2011) | No | Singleplayer | Leadership skills | Full-range theory |
| Productive Leadership Game | Kesti et al. (2017) | No | Singleplayer | Leadership competencies | Human capital production function |

various options. They receive immediate feedback through the utterances and emotions of the conversational partner. The game *SALVE* (Augello et al., 2016) is using AI-controlled chatbots participating in medical consultations and is based on the Social Practice Theory (Schatzki, 1996). In contrast, Even et al. developed a serious game primarily targeting schizophrenia patients to support rehabilitation programs for social skills (Even et al., 2016). This approach is combining role play with problem-solving exercises on which remediation therapies rely. A recapitulatory overview of serious role-playing games for the training of communication skills can be seen in **Table 2**.

Conflict management is an important social skill that has been the subject of serious role-playing games in the past. *Choices and Voices,* for example, is an interactive simulation game for preventing violent extremism. In it, players explore and discuss issues and influences leading to tension and disruption in communities (Memarzia and Star, 2011). In this game, players face several moral dilemmas in which their decisions determine the outcome of the game (for themselves, their family, and their friends). This is supposed to show the significant consequences real life decisions can have. The storytelling game *Façade* asks players to resolve a conflict between a married couple. Through communication with the conflicted parties, they are to investigate the causes of their issues and provide counseling (Mateas and Stern, 2003). The emphasis here is on believable characters, natural language conversation, and a dynamic storyline. In *Office Brawl* the player assumes the role of a mediator, who is moderating a conflict between two parties in a workplace-oriented setting, using AI-controlled virtual characters (Glock et al., 2011). As a project manager in the game, the player needs to handle an argument between two members of a team. *FearNOT!* is a virtual drama for anti-bullying education targeted at children (Aylett et al., 2005). In this game, the bullying behavior of one of the characters is leading to dramatic episodes. The victim is seeking advice of the player who can interact with this character by using free text input. It is supposed to allow children to explore what happens in bullying situations in which they take responsibility for what happens to a victim without feeling victimized themselves. The game *LOITER* lets prospective police officers enact street interventions with loitering juveniles (Linssen et al., 2014) and aims to improve their social awareness. Here, players can experiment with different ways of interacting with the juveniles. *Self City* is a serious game developed for emotionally impaired adolescents, which is supposed to help them develop skills such as process-oriented thinking and conflict resolution (Van Dijk et al., 2008). In this game, players can walk around online in a virtual city. On their way to the cinema, they experience challenging social situations and learn how to deal with them. Players are accompanied by a daemon that provides advice in conflict situations and suggests alternative actions. The *Junior Detective Computer Game* has been developed as part of a multi-component social skills intervention for children with Asperger syndrome (Beaumont and Sofronoff, 2008). Here, players take the role of a trainee at the Detective Academy and are taught how to recognize complex emotions in computer-animated and human characters. They need to complete several missions, such as dealing with bullying, playing with others, and trying out new things. A recapitulatory overview of serious role-playing games for training conflict management can be seen in **Table 3**.

## Frameworks for the Design of Serious Games

There is a number of existing models and frameworks for the general design of serious games, which describe fundamental components of such systems and support formal approaches to game design. A very general approach is the so-called *MDA (Mechanics—Dynamics—Aesthetics) framework* (Hunicke et al., 2004). It proposes three different perspectives for understanding and designing games: *Mechanics* refer to the actual implementation of the game. They describe its particular components (actions, behaviors and control mechanisms) at the level of data representation and algorithms. *Dynamics* relate to the overarching design goals and run-time behavior of the mechanics acting on player inputs and each other's output over time. *Aesthetics* refers to the resulting game experience. They describe the desirable emotional responses evoked in players, when interacting with the game system. Although the MDA framework is widely accepted and practically employed, it has weaknesses and limitations (Walk et al., 2017): It focuses too much on game mechanics, neglecting many design aspects of games, including an over-arching narrative. Therefore, it is not really suitable for all types of games, including particularly gamified content or any type of experience-oriented design.

Another approach toward serious game design is the *Four-Dimensional Framework* suggested by De Freitas and Oliver (2006). It postulates four main dimensions of learning processes to be considered in the design process of serious games: the *context* in which learning takes place (e.g., classroom-based or outdoors, access to equipment, technical support), the *learner specification* (e.g., learner profile, pathways, learning background), the *mode of representation* (e.g., level of fidelity,

**TABLE 2 |** Serious games for the training of communication skills (overview).

| Game | Author | Use of AI | Mode | Learning objective | Underlying framework/model/theory |
|---|---|---|---|---|---|
| ENACT | Marocco et al. (2015) | No | Singleplayer | Negotiation skills | Model of conflict handling styles |
| DREAD-ED | Haferkamp et al. (2011) | No | Multiplayer | Disaster Communication | Theories of crisis and emergency risk management |
| The "Door Scene" | Bosse and Gerritsen (2016) | No | Singleplayer | Communication skills (police domain) | Education program of police academy students |
| deLearyous | Vaassen and Wauters (2012) | Yes | Singleplayer | Communication skills (workplace) | Interpersonal circumplex (Leary's Rose) |
| JUST-TALK | Hubal et al. (2003) | Yes | Singleplayer | Communication skills (law enforcement) | – |
| POINTER | Linssen et al. (2014) | Yes | Singleplayer | Communication skills (police domain) | Cognitive model for social interaction |
| ELECT BiLAT | Lane and Hays (2008) | Yes | Singleplayer | Cultural social conventions (military domain) | – |
| Maritime City | Flynn et al. (2011) | Yes | Singleplayer | Communication skills (social work domain) | – |
| TARDIS | Gebhard et al. (2018) | No | Singleplayer | Communication skills (job interview) | – |
| Communicate! | Jeuring et al. (2015) | No | Singleplayer | Communication skills (health care domain) | – |
| SALVE | Augello et al. (2016) | Yes | Singleplayer | Communication skills (healthcare domain) | Social practice model |
| Serious game for schizophrenia patients | Even et al. (2016) | No | Singleplayer | Communication skills (emotion recognition) | Social skills programs for schizophrenia |

**TABLE 3 |** Serious games for the training of conflict management (overview).

| Game | Author | Use of AI | Mode | Learning objective | Underlying framework/model/theory |
|---|---|---|---|---|---|
| Choices and Voices | Memarzia and Star (2011) | No | Singleplayer | Prevention of violent extremism | National curriculum |
| Façade | Mateas and Stern (2003) | Yes | Singleplayer | Conflict resolution | – |
| Office Brawl | Glock et al. (2011) | Yes | Singleplayer | Mediation | – |
| LOITER | Linssen et al. (2014) | Yes | Singleplayer | Street interventions | Cognitive model for social interaction, Virtual Storyteller (VST) |
| Self City | Van Dijk et al. (2008) | No | Singleplayer | Process-oriented thinking, conflict resolution | Dialogical self-theory |
| Junior Detective Computer Game | Beaumont and Sofronoff (2008) | No | Singleplayer | Bullying, conflict resolution | Social skills programs for individuals with Asperger syndrome |

interactivity, and immersion used in the game), and *pedagogic considerations* (e.g., learning models, approaches for learning support). Like the MDA framework, this framework is a high-level model, meaning that it specifies a limited number of generic concepts that can or should be taken into consideration when designing or evaluating serious games, but only on a very general level with no concrete design or evaluation guidelines (Mayer et al., 2014).

This also applies to the *RETAIN (Relevance Embedding Translation Adaptation Immersion & Naturalization) model* by Gunter et al. (2006). This model was developed to support game development and to assess whether a serious game is appropriate for educational purposes, how well the academic or pedagogical content is immersed and embedded in the game's narrative and how knowledge transfer is promoted. *Relevance* means that the information students learn in the game should be relevant to the game world as well as to the players' targeted objectives. *Embedding* should be done in a way that learning objectives and fantasy are tightly coupled. *Transfer* refers to how well players can recognize and apply newly learned information outside the game environment. *Adaptation* means that players apply their learned knowledge to create new scenarios that apply literacy skills in a new domain. *Immersion* should be facilitated by the game environment and the ability to create customizable social presence. *Naturalization* means that players should be encouraged to gradually use their own skills to gain the

knowledge necessary for success in other problems and subject areas (Kenny and Gunter, 2011).

The *Triadic Game Evaluation* (TGE) (Harteveld, 2011) approach stresses three different perspectives for the design and evaluation of serious games: reality, meaning and play. The *reality* component determines the game subject, variables and definitions. It could be represented by players from the real world or a representation of the real world inside the game. Evaluation criteria in regards to this component include fidelity, realism, and validity. The *meaning* component of the framework considers how a meaningful effect beyond the game experience can be achieved and incorporates aspects such as communication, learning, rhetoric, and opinions. Evaluation criteria include reflection, transfer, and relevance. The *play* component refers to the fact that games are primarily highly interactive and engaging tools that immerse players into a fictitious situation, and is related to game elements like actors, rules, resources, challenges, and competition. Evaluation criteria for this component are engagement, fun, and immersion. The TGE framework claims that games need to be designed equally along these three components (Kortmann and Harteveld, 2009). In contrast to the aforementioned models, this framework comes with a concrete agile development model that describes different software engineering phases and decision moments in the creation process. However, specific design and implementation guidelines are not included.

In summary, the various promising approaches to training social skills by means of role-playing games are still defined on a very general level. Our aim is to provide a comprehensive conceptual and technical framework for the concrete design and implementation of serious role-playing games for the training of social skills in dialog-centric settings with virtual characters through which we would support more efficient and effective design and implementation of such game environments.

# FRAMEWORK: CONCEPTUAL APPROACH

From the design perspective, several distinctive conceptual features characterize our framework: (1) chat-like interaction with an AI-controlled chatbot, (2) separate phases of immersion (role-playing) and reflection to facilitate a change of perspective that is considered conducive for learning, (3) the learning process is emphasized by means of adaptive feedback based on individual analyses.

## Chatbots in Virtual Role-Playing Environments

Chatbots are computer programs (conversational agents) that communicate with users in natural language. Their purpose is to simulate a human conversation via text or voice interactions. Originally, chatbots were developed for entertainment purposes. However, especially in today's world, in which the possibilities of computer use are becoming more and more diverse, the use of chatbots can be extended to many other areas. Chatbots are found in daily life now, such as personal assistants (like Google Assistant, Amazon Alexa, or Apple's Siri), search engines, customer service and support, and healthcare coaching (Winkler and Söllner, 2018). They can be used in a variety of domains including business, e-commerce, entertainment, medicine, and others (Kerly et al., 2006; Shawar and Atwell, 2007).

Chatbots can also be used successfully for learning. Past studies even show that chatbots present feasible means to improve learners' results (Kerly et al., 2006). They have been used for a variety of purposes including medical education and therapy, language learning, as well as receiving feedback and strengthen motivation and self-efficacy (Winkler and Söllner, 2018). Chatbots have also been used in serious role-playing games, as shown in the examples in chapter Related Work. The use of chatbots in serious role-playing games has several advantages. First, having a chatbot interact with the player instead of a human ensures a certain level of standardization that could never be achieved in a setting with human actors. Second, scenarios including a chatbot are repeatable, independent of time and place, and no additional resources are needed. An important part of chatbots is the creation of dialogs. A chatbot can only be as good as its knowledge base used for answer generation (Abdul-Kader and Woods, 2015). The problem of the "classic" chatbots is that they do not allow to store the course of the conversation and have no real understanding of the answers. However, a realistic and responsive behavior of chatbots is important to increase the players' engagement and contribute to the immersive nature of role plays. To achieve this, our approach proposes several technical workarounds that will be explained in detail in chapter Multi-Agent Architecture.

## Immersion and Reflection

The educational impact of serious role-playing games highly draws on the "willing suspension of disbelief" by the players who commit to the role they are supposed to play (Lim et al., 2009). Thus, this kind of system intends to create a certain degree if immersion. Janet H. Murray defines the term immersion as follows: "A stirring narrative in any medium can be experienced as a virtual reality because our brains are programmed to tune into stories with an intensity that can obliterate the world around us… The experience of being transported to an elaborately simulated place is pleasurable in itself, regardless of the fantasy content. Immersion is a metaphorical term derived from the physical experience of being submerged in water. We seek the same feeling from a psychologically immersive experience that we do from a plunge in the ocean or swimming pool: the sensation of being surrounded by a completely other reality, as different as water is from air, that takes over all of our attention, our whole perceptual apparatus" (Murray, 2017). When players identify themselves with the character they are assuming in the game and are immersed, their motivation to proceed and succeed in the game increases (Annetta, 2010). This intrinsic way of motivating learners is something conventional instruction modes do not have (Yee, 2006). Players become immersed in a game because they find it satisfying, and through this intrinsic motivation, they get more engaged in the learning task (Annetta, 2010).

In terms of experience-based, authentic learning, it seems reasonable to carry out the enactment in an immersive situation. However, there is reason to believe that the immersion tends

to impede the critical self-reflection that is important for the learning process (Malzahn et al., 2010). Reflection is a successful tool to improve the learning process (Jonassen et al., 1993), and it is needed to ensure the transfer to real-life situations (Lim et al., 2009). During the reflection process, people recapture, rethink, and evaluate their experiences to develop new understandings and appreciations (Boud et al., 1985). It is to be expected that the amount of reactive attention required for immersion impedes the players' ability to distance themselves from the role, which in turn interferes with self-reflection. Thus, the requirement of role distance in phases of reflection suggests that the mode should be changed to help the learner step out of his role and adopt a different perspective. Based on this assumption, we decided to separate the actual role-playing game from the reflection session in our framework.

## Adaptive Feedback

As stated above, an important challenge for serious role-playing games is shaping the narrative experience and the pedagogical outcomes that generally depend on post-role-play reflection and feedback (Lim et al., 2009). Feedback on the performance of the player(s) during the role-playing session is necessary to ensure the transfer to real-life settings. It is supposed to help learners to improve their performance by providing information about the correctness of their actions (Shute, 2008). Johnson et al. identified four feedback characteristics: (1) the *type* of feedback (e.g., outcome-based or process-based feedback), (2) the *timing* of feedback after an action (i.e., immediate or delayed feedback), (3) the *modality* in which the feedback is presented (e.g., spoken or text-based feedback), and (4) *adaptation* to learner characteristics (e.g., in regards to prior knowledge or spatial ability) (Johnson et al., 2017).

Our framework relies on adaptive feedback based on an automated, individual performance analysis. We differentiate between three types of feedback: The first one is implicit feedback during the role-playing session through the reactions of the chatbot (*ingame feedback*). These reactions can be non-verbal (e.g., facial expressions) or verbal. Real-life situations are simulated through both types of reactions to the players' actions. The second one is a general summary of the analysis results (*aftergame feedback*). Players should receive an overall feedback on their performance during the role play that summarizes the most important aspects (positive and negative). The third type is direct and specific feedback on single incidents during the role play that can be provided through prompts in a replay of the conversation. A replay offers several advantages: The whole conversation can be shown again step by step and augmented with individual feedback at certain points, commenting on specific actions of the player. Also, it provides the possibility to navigate between the different phases of a conversation, pause the replay, or jump to the next feedback marker. As a result, it is much more flexible and searchable than, e.g., a video of a conventional role play.

# FRAMEWORK: TECHNICAL APPROACH

In our approach, the technical implementation of such systems entails three main challenges: (1) dialog modeling of the chatbot, (2) implementing a multi-agent system as the backbone in order to keep components independent and optimize performance, and (3) performance analysis and feedback generation. The following section will present our approach toward each of these aspects in detail.

## Dialog Modeling

In our framework, the *Artificial Intelligence Markup Language* (AIML) is used for the implementation of the chatbots' conversational logic. It is a common XML-based solution for passive AI-controlled chatbots, which comes with an easy syntax and a small number of control structures (Wallace, 2004). AIML relies on a simple pattern matching. It consists of categories, each containing a pattern and a template. If the user input is matching a pattern, the template defines the answer or action to be given. Recursion and wildcards allow for many different inputs matching one single pattern, while the ability to store a context and the use of variables and conditions allow a complex and sophisticated chatbot design.

Although AIML has a long history and is a common solution for chatbots used in educational contexts, it has certain limitations. One problem is the passive nature of AIML. An AIML chatbot only reacts to an input it receives, it cannot take the initiative. This behavior can be bypassed by using external triggers to make the bot become active when required in certain situations. Another problem is that an AIML chatbot (as is true for all artificial natural language processing) cannot truly grasp the sense of what has been said. The AIML chatbot only checks the user input against predefined patterns; if there is no match, it can at most output some default statements (which need to be predefined as well). To solve this problem, our framework proposes the use of sentence openers in dialog-centric role play scenarios. This means that players always have to select a sentence opener from a predefined set and supplement it with free text input to compose a message.

This approach has several advantages: First, a sentence opener already defines the general gist of a message (e.g., affirmation, rejection, proposal, inquiry). As a result, it is at least possible to provide a default answer that is tailored to the selected sentence opener even if the free text input following the opener does not match a predefined pattern. Furthermore, if each phase of the chat conversation has unique sentence openers, the chatbot always has some kind of context information. Second, the use of sentence openers reduces the complexity of the dialog scripts dramatically because the possible starting points of all input sentences are already known. Third, sentence openers provide support to the players and help them phrase their messages. In addition, sentence openers improve the overall atmosphere of the simulated conversation and make it seem more realistic and natural. Last, sentence openers (in contrast to fully predefined text messages) still allow for free text input that can be analyzed in detail and influence the course of the game.

## Multi-Agent Architecture

Our technical framework is based on a uniform multi-agent system architecture with a blackboard as the communication and integration mechanism. The blackboard is realized through a so-called *tuple space*. The components (agents) in this system are loosely coupled, i.e., they do not communicate with each other directly but only via entries on a central tuple space server (Gelernter, 1985). These entries have a simple tuple structure that contains primitive data types (integers, characters, booleans) and strings. According to the original concept of Gelernter, there are only a few generic operations (read, write, take, wait-to-take, etc.) to interact with such a blackboard. In contrast to a pure database solution, however, there are active trigger mechanisms such as notifications. The *SQLSpaces* developed in the COLLIDE group itself serve as a specific implementation basis in our framework (Weinbrenner, 2012). While the server itself is implemented in Java, the system framework of SQLSpaces provides clients for the agent programming in various programming languages. SQLSpaces also facilitates the logging of relevant data of each gaming session, which can later be used for analysis and comparison.

The overall system consists of a user interface and various agents, each of which is responsible for one task in either dialog analysis, feedback creation, or game control. The user interface in the three implemented training scenarios described in this article have been implemented as a web application using HTML, CSS, and JavaScript (2D frontend). Previous implementations were based on OpenSimulator3 (3D frontend), but since there were no specific advantages of 3D environments over 2D environments, we decided to go ahead with a 2D approach (Malzahn et al., 2010). As described above, the client (user interface) and all agents are writing and reading tuples from the tuple space server without communicating with each other directly, which results in a loosely coupled and adaptive system. That means, agents can easily be adapted, added, or removed depending on the actual application scenario.

The agents can be divided into three groups, depending on their functionality. *Pervasive agents* are overarching agents, which are crucial in connecting the individual game components. The *register agent,* for example, is managing the log-in of the player (or players in a collaborative scenario). When a new client is logging in, the register agent receives a request via the tuple space (callback) and starts a new gaming session. The *silence agent* reacts if a player has been inactive for a certain amount of time, in which case the agent is triggered and sends an internal message to which the chat bot responds. After the fourth internal message from the silence agent, the conversation ends. *Pre-processing agents* are used to pre-process the player's input before the answer to it is generated in order to provide the best possible answer. This pre-processing is mainly used to overcome the limited capabilities of AIML: Analyzing certain aspects separately helps to prioritize specific behaviors, i.e., make sure that the chatbot is reacting adequately to rude or aggressive behavior. In addition, this procedure reduces the structure of the AIML scripts and supports the feedback creation. Each of the implemented scenarios uses different pre-processing agents depending on the context. All pre-processing agents analyze the player's input regarding one specific aspect. **Figure 1** shows the basic architecture.

## Performance Analysis and Feedback Generation

Both performance analysis and feedback generation always depend on the context and the learning objectives of the serious role-playing game. As described above, our architecture is using analysis agents, each of which is responsible for the evaluation of one specific aspect of the player's communication behavior. They are divided into pre-processing agents and regular agents.

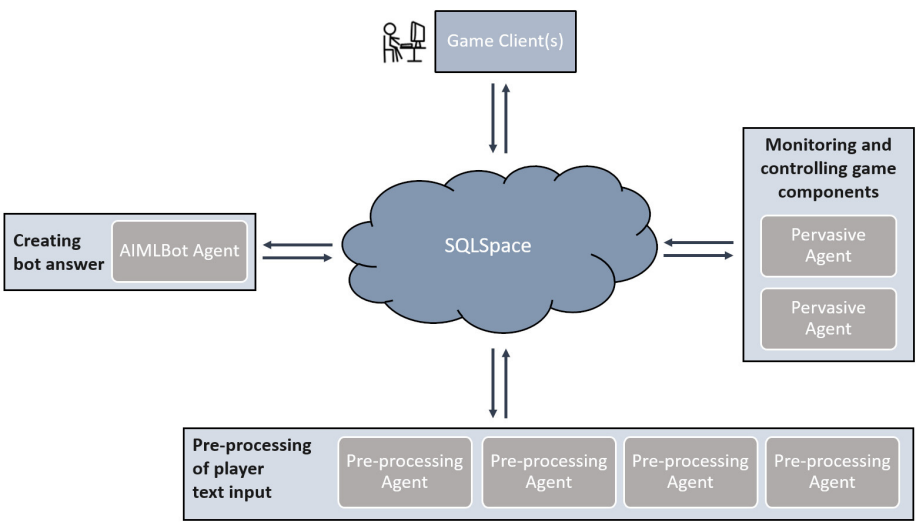


**FIGURE 1 |** Basic system architecture.

Pre-processing is necessary for generating a suitable chatbot response. For example, if a player acts aggressive or rude, the chatbot should react to this behavior regardless of any other information the player's message contains. The results of the pre-processing are collected, and if an immediate reaction to a specific behavior is required, the text input is modified. If, for instance, a swearword has been detected in a player's message, the complete input string is replaced by a specific trigger ("swearword"), causing the bot to react appropriately. The same applies to other behaviors. In case the pre-processing agents do not find anything that needs an immediate reaction of the chatbot, the bot receives the original text input. Simultaneously, all other analysis agents evaluate the message and add their feedback to it in the form of feedback tags (e.g., #praise#, #interruption#, or #criticize#). These feedback tags mark any situations in which the player is supposed to receive feedback during the replay that is taking place in the reflection phase following the role play session. The tags are filtered out during the chat session; the players do not get to see them during the game, but they play an important role in the feedback generation.

# CASE STUDIES

Based on the framework described above, the research group COLLIDE at the University of Duisburg-Essen has conducted various case studies with different instances of virtual role-play environments. The training scenarios include workplace-oriented conflict management, patient-centered medical interviews, and customer complaint management.

## Case Study: Conflict Management

*ColCoMa* (Collaborative Conflict Management) is a collaborative serious game for the training of conflict management strategies in an organizational context within a role-playing scenario, developed at the COLLIDE group in 2012. It involves two players in a conversation with an AI-controlled chatbot acting as a mediator in a 2D virtual environment. The following description of the approach and game design is based on the work of Emmerich et al. (2012).

### Approach

In ColCoMa, two players have a conversation about a fictitious conflict, moderated by an AIML chatbot in the role of a mediator. The main goal of the players is to resolve the conflict by showing constructive and appropriate behavior during the conversation. Each player is assigned a predefined role in this fictitious scenario: As a member of the computer support hotline of a big software company, Mr. Meier is conscientiously taking much time for his customers. Mrs. Schmidt is his supervisor. She is dissatisfied with Mr. Meier's way of working. She notices that he takes too much time for the customers and therefore does not work efficiently in her eyes. Mr. Meier does not agree with her, and the situation escalates after a negative appraisal of Mr. Meier's performance. In order to support immediate understanding of the situation and empathy with the assigned role, the scenario is kept as simple and comprehensible as possible and focuses on the main conflict as well as the person's feelings.

### Game Design

The players are introduced to the game and the scenario through a cartoon-like picture story that is told from their respective role's perspective and is supposed to result in conflicting points of view. The conversation itself takes place in a chat window where graphical representations of the mediator and the other player's character are shown to create the association of sitting opposite each other. The dialog partners can communicate via simple text messages. Facial animations can be evoked via common emoticons. The interface also includes a notepad with hints as well as a help section that offers additional information on the game controls and the fictitious scenario if needed. **Figure 2** is showing the basic user interface.

The conversation is divided into five phases according to Proksch (2010): (1) framing phase, (2) topic collection, (3) working on the conflict, (4) looking for a solution, and (5) contract. The framing phase represents the starting point of the mediation talk and is important for establishing certain rules for the conflicted parties and their behavior toward each other. The actual conflict is not yet the focus. Instead, the participants state their personal hopes and mediation goals and reflect on their own point of view as well as the opponent's position. In the second phase, both parties are supposed to name relevant topics they would like to put on the agenda during the mediation talk, like performance review, working conditions, the participants' perspective in the company as well as their behavior toward each other. The mediator chatbot recognizes the topics based on a list of keywords and phrases. In order to be able to advance in the game, the two players need to name three topics; otherwise the mediator terminates the conversation due to a lack of contribution. If only two topics are volunteered, the mediator will suggest a third one. The mediation talk itself takes place in the third phase. The main task during this phase is to discuss the selected topics in detail. Both players are given the opportunity to explain why a topic is important to them, what changes they would like to see in regard to the specific topic, and what they themselves can contribute to realize these changes. They are also given the opportunity to comment on whether the other party's perception is correct and to rectify their position if this is not the case. The aim of the fourth phase is to find solutions for the different topics that are acceptable for both parties. Finally, in phase five, they are supposed to agree to adhere to the solutions they came up with and enter into a contract.

The mediation talk is followed by a reflection phase in which both players receive feedback on their performance in order to help them reflect on their behavior. At the start of this phase, players get the opportunity to directly exchange feedback with each other in a free chat without the mediator. After this free chat, each of them receives an overall feedback on the own performance during the mediation talk. Finally, the players take part in a replay session of the whole chat conversation, but this time augmented with individual feedback commenting on especially positive and negative contributions of the players. A change of the graphical interface during the replay is supposed to reinforce role distance, which is assumed to be conducive for learning (see chapter Immersion and Reflection).

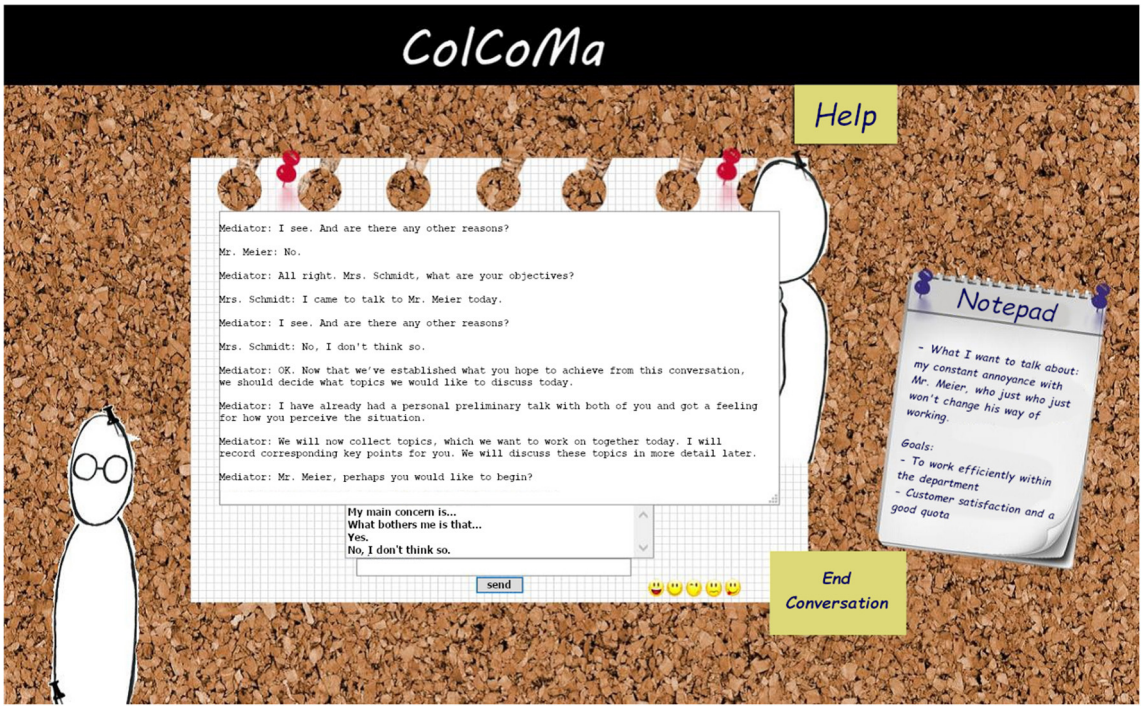


**FIGURE 2 |** ColCoMa chat interface.

The performance analysis and assessment is based on general rules that conflicting parties have to adhere to during a mediation talk, such as not being aggressive or rude, not being reproachful, and not impairing the opponent's autonomy (Stauss and Seidel, 2010). Instead, the participants are supposed to have an open and constructive attitude, name topics and issues in a concrete way, and help the other party understand their perspective. The evaluation of the players' performance during the mediation talk is done by several analysis agents, each responsible for one specific behavioral aspect, e.g., rudeness (by comparing the players' input to a list of swearwords and defamations), aggression (e.g., by checking for multiple exclamation marks or use of all-caps spelling), emotion-showing (e.g., use of emoticons), or the use of I- and you-statements (by counting the amount of words referring to the speaker and those referring to the dialog partner). Some of the analysis results are used just for the overall feedback that is provided to players after the conversation.

## Evaluation Results

In 2018, an eye-tracking study has been conducted in collaboration with the Dortmund University of Applied Sciences and Arts (Othlinghaus-Wulhorst et al., 2018). The results of this study will be summarized and discussed in this section.

Apart from getting feedback on the prototype, the main goal of the study was to investigate the question if there is a correspondence between gaze synchronicity of the two players and the quality of collaboration. Twenty subjects (average 22.8, SD = 2.84, 5 females, 15 males) participated in the study and have been tested in dyads, using two desktop-based eye-trackers to track the players' gaze during the experiment. To investigate the research question, three main hypotheses have been examined: The first hypotheses postulated "a positive relation between the convergence of visual foci of attention (gaze synchronicity) and the successful completion of the game (achievement score)" (Othlinghaus-Wulhorst et al., 2018). In this study, *gaze synchronicity* has been defined as the extent to what the two players have been looking at the same areas of interest in the same time interval during the course of a gaming session. The so-called *achievement score* has been used to measure the success in the game and reflects the players' performance during the mediation talk based on three criteria: (1) automated feedback generated by the system, which summarizes the players' behavior during the game, (2) the successful completion of the topic collection phase (which has been considered a major milestone in the game), and (3) the successful completion of the game, which is achieved when both players sign a contract, which includes the agreements and rules they worked out together with the mediator. Referring to the hypothesis, a highly significant correlation between the gaze synchronicity and the achievement score has been found on the aggregate level (taking overall eye-tracking convergence as a global parameter).

In the second hypothesis it is assumed that "there is a positive relation between the convergence of visual foci of attention (gaze synchronicity) and the quality of collaboration in the chat." (Othlinghaus-Wulhorst et al., 2018). In order to define the *quality of collaboration*, a rating scheme has been developed, which includes five dimensions: (1) argumentation

(players discuss or bring forward justifying arguments), (2) agreement/disagreement (players endorse or dissent from one another), (3) collaborative orientation (players refer to each other, ask questions, provide feedback or refer to topics brought up by the other player), (4) solution orientation (players try to find or propose a solution), and (5) shared awareness/reinforcing shared history (players share common knowledge or explain their situation). Based on this scheme, all chat messages have been analyzed and checked against the five dimensions and assigned a total quality score. Finally, all matches of a gaming session have been added up to a percentage indicating the overall quality of the collaboration for a pair of players. Relating to the hypothesis, a high correlation between the gaze synchronicity and the collaboration quality has been found, especially for the dimension's agreement/disagreement, solution orientation and shared awareness.

The third hypothesis proposes "a dynamic (time-related) congruence between similar eye movements (synchronicity) and the quality of collaboration in the chat" (Othlinghaus-Wulhorst et al., 2018), meaning that there is not only a gaze synchronicity on the aggregate level, but also synchronicity between convergent eye-tracking and chat interaction during the course of the game. This hypothesis could not be verified. It is assumed, that the specific nature of the chat might be a reason for this, as three persons are involved (the two players and the mediator chatbot) and thus the two human actors do not really communicate directly, but only to the mediator. They answer his questions and do not really have the chance to communicate with each other directly, which is resulting in a predefined structure of the chat conversation and rather long time interval between the utterances of the two players.

## Case Study: Patient-Centered Medical Interview

In 2013, a training scenario for medical interviews has been developed at the COLLIDE group. It is supposed to give medical students the opportunity to train doctor-patient conversations autonomously and systematically in the form of role plays with simulated patients. The following description of the approach and game design is based on the work of Behler et al. (2013):

### Approach

This scenario for the training of doctor-patient communication has the basic goal to train the communication strategies between doctor and patient and is tailored to the target group of medical students. Here, the player takes the role of a locum doctor for family medicine whose goal is to uncover all the symptoms of a patient in a given time. To achieve this, they have to use methods of the GOG (*Gesundheitsortientierte Gesprächsführung*, engl. "health-oriented negotiation") (Schwantes and Kampmann, 2007), in order to create a pleasant conversation atmosphere. To successfully master the game, it is necessary to behave in accordance with this concept and to bring in the guidelines in the course of the conversation. Another important learning objective is to build trust and empathize with the patient, as these aspects play a central role in the doctor-patient conversation (Kruse, 2000). Medical diagnosis is not a learning objective in this scenario, so the game can be used independently of progress in medical studies.

### Game Design

At the beginning, the player enters the waiting room, where several patients are already sitting and waiting for their call. The patients represent different scenarios, which differ in the content and level of difficulty. The level of difficulty is determined by the number of symptoms to be identified and the willingness to talk about his or her condition. In the waiting room, the trophies and high scores already achieved by the current player are also displayed. By choosing a patient, the player starts the scenario and enters the doctor's office, where the actual interview with the patient takes place.

In the office, the player communicates with the patient via text input. The player chooses a suitable sentence opener and completes the sentence with free text. The sentence openers are related to GOG phases. In addition to verbal interaction, the player can also use items from the doctor's bag (information leaflets, stethoscope, pills, and syringe) and conduct non-verbal actions like nodding, smiling or touching the patient, which are also important in real interpersonal communication (Ziebarth et al., 2014). The items provide a playful added value. The player has to find out when which item is reasonable to use and receives bonus points for this, but only in combination with appropriate topics—otherwise, points are deducted. **Figure 3** is showing the basic user interface of this scenario.

In order to win the game, i.e., to achieve the highest possible score, the player has to collect points for recognized symptoms as well as points for trust-building and empathic contributions and actions. The accumulated sum of trust and empathy points in the game represents the conversation atmosphere and serves as a threshold value that defines how quickly a symptom is revealed by the patient. The patient reveals symptoms when the player addresses a scenario relevant topic and has reached the corresponding threshold value.

The main conflict is between the limited time available to the player to find the symptoms and the patient who only reveals them under certain conditions. This situation resembles a doctor's real conflict between time pressure and the desire to help patients comprehensively. Each scenario of this game contains a side mission to increase replayability. While the main task includes finding relevant symptoms, a secondary task could be, e.g., to point out the benefits of assisted living to an elderly patient to ensure long-term care. Side missions give more depth to the game as they refer to the social situation of the patient and thus lead to more immersion (McMahon and Ojeda, 2008). Players receive bonus points and trophies for solving side missions. As an additional incentive system, the total number of points is entered to a leaderboard, which all players can see. According to Festinger's theory of social comparison (Festinger, 1954), this motivates players to improve their own abilities, which are represented by the points.

The gaming session is followed by a reflection phase. First, the players are presented with their individual score in the fields trust, empathy and symptoms. Afterwards, they receive detailed feedback in the form of an augmented replay, in


**FIGURE 3 |** The doctor's office (basic game interface).

which the analysis results are presented. In this analysis, the player interaction is, e.g., checked for the use of paraphrases, emotions expressed to the patient, showing choices to the patient, addressing him or her by name and the use of all phases of the GOG. Although the phases do not have to be passed linearly, goal guidance and explanations for example are particularly relevant toward the end of the conversation. Pauses, nods and facial expressions are evaluated as well. In addition, behaviors are taken into account that do not directly lead to an improvement of the score but influence the course of the conversation. For example, the patient reacts verbally to excessive talking of the player and a lack of balance between the doctor and the patient as the subjects of the player's statements. This leads to a loss of time, which increases the central conflict of the game. As in a real situation, the player receives his feedback directly from the patient and can react to it in the process of the conversation.

## Evaluation Results

The prototype has been evaluated in two studies, both performed by the COLLIDE group in cooperation with the Department of Family Medicine of the Charité in Berlin (Ziebarth et al., 2014). The results of these studies will be summarized and discussed hereinafter:

The focus of the studies was the examination of usability and playability, as well as immersion and reflection. The following key questions have been deducted from the global objectives: (1) Does the flow of the game feel natural? (playability), (2) Is the player able to manage the game well? (usability), (3) Is the game immersive? (immersion), (4) Is the reflection phase at the end of the game perceived as helpful? (reflection), (5) Which functions are used?, (6) What are the difficulties in using them?, (7) How is the game perceived by the target group? *Playability* has been evaluated using self-created items relating to the clarity of the goal, structural problems regarding game flow (i.e., the use of sentence openers to create chat messages), "functional" playability (i.e., the extend of feeling understood by the patient), and the complexity of the game (Ziebarth et al., 2014). *Usability* has been assessed based on the following categories of ISO 9241-110: conformity with user expectations, suitability for learning, self-descriptiveness, and error tolerance. The items of the questionnaire were phrased based on the German inventories Isonorm[1] and IsoMetrics[2]. The aspect of *immersion* was measured based on the approach developed by Jennett et al. for measuring immersion in digital games (Jennett et al., 2008). The items selected for the studies address the subjective enjoyment of the game's representation, fun factor, immersion, and emotional involvement (Ziebarth et al., 2014). For the assessment of the *reflection* support, participants were asked what they thought the game is aiming to train, if they viewed the annotated replay, and if they thought about what they could have done differently (Ziebarth et al., 2014).

The first experiment was an observational study with 7 medical students (average 21, SD = 2.582, 6 females, 1 male). Although the results indicate that the idea and approach of the game were assessed quite positively, the observations showed slight problems with the general usage of the game. While the interaction principles have been generally well-understood, a few participants reported problems with expressing themselves using the predefined sentence openers. Also, the free text supplementing the sentence opener was often not understand by the chatbot, because some topics have not been considered in the

[1]https://abeto-online.de/ep/index,id,3314.html
[2]http://www.isometrics.uni-osnabrueck.de/

design. Apart from these limitations, the students liked the game as an alternative for the training of medical interviews before performing them with human patients. The second experiment was an online study with 21 medical students (average 23.05, SD = 4.295, 15 females, 6 males). The online questionnaires ($n$ = 21) as well as the questionnaires completed by the participants on paper during the observation ($n$ = 7) were included in the evaluation of the questionnaire in the subject areas playability, usability, immersion and reflection. The results mostly support the findings gained in the observation study. While most of the participants liked the user interface [M = 4.18 (of 5), SD = 0.819] and understood the goal of the game [M = 3.68 (of 5), SD = 0.905], its suitability for learning [M = 3.82 (of 5), SD = 0.782] and the self-descriptiveness [M = 3.33 (of 5), SD = 0.603] were considered good (above average), and the imaginative immersion [M = 3.29 (of 5), SD = 0.076] as well as the emotional involvement (M = 2.93, SD = 1.086) showed only average values. In addition to the questionnaires, a total of 36 conversation transcripts were evaluated in order to uncover possible weak points in the text recognition module of the system, which was used to fix and further improve the AIML scripts.

## Case Study: Customer Complaint Management

The case study CuCoMaG (*Customer Complaint Management Group reflection*) is a serious role-playing game for the training of customer complaint handling based on theories of consumer psychology and complaint management, originally developed in 2016 at the COLLIDE group in the context of a student master project (Doberstein et al., 2016; Othlinghaus and Hoppe, 2016). It has been re-designed and evaluated in 2019 by Othlinghaus-Wulhorst et al. (2019). The following description of the approach and game design is based on these works.

### Approach

In this game, the player assumes the role of a customer service employee in the fictitious company *LittleOnes*, a producer and seller of personalized clothing for children via an online shop. The player is confronted with a chatbot in the role of a complaining customer, who is reporting a certain problem. The player is communicating with the customer through a simple chat environment. Like in the other scenarios, the player has to select a sentence opener and supplement it with free text in order to formulate a chat message. The chat setting is ideal in this use case, as it simulates everyday work situations for people working in the customer support sector. **Figure 4** is showing the general user interface of this scenario. This game has one distinctive feature that sets it apart from the previous ones presented: It offers explicit support for group reflection (Othlinghaus and Hoppe, 2016). Group reflection enables a collective exchange and thus collaborative learning (Schuster, 2010). The group reflection session, which is supposed to take place subsequent to the role play session, is designed to be guided by a trainer. This trainer is given a special group reflection tool that he can use to arrange an interactive after-action review process (Othlinghaus and Hoppe, 2016). The tool allows him to show and discuss important sequences from chat conversations of several players, review specific actions and aspects of their communication behavior. The data provided by the tool can be used to give feedback to the players and initiate group discussions to help them reflecting their actions and improving their performance.

### Game Design

The game includes three different scenarios. The scenarios differ according to the type of customer used, especially in terms of conversation style (Rahim and Bonoma, 1979) and the problem situation of the customer, and thus in the level of difficulty. The first scenario serves as a base level and tutorial. The customer's problem can be solved quite easily by the player, since the conversation is reduced to the conversation phases in which only information content has to be collected and no pure "soft skill" phases have to be passed through. The conversation therefore only includes the greeting phase, the problem-solution phase and the conclusive phase. The customer in this scenario can be classified as an *integrating* customer according to the model of Rahim and Bonoma (Rahim and Bonoma, 1979), who differentiated five different styles of handling interpersonal conflicts, and is therefore open to reach a solution acceptable for both parties and is showing problem-solving behavior. According to a study conducted by Cho et al. (2002), the customer's problem is the third most common cause of non-public online customer complaints: delivery problems. The aim of this scenario is to help the player becoming acquainted with the user interface and let him walk through the basic milestones of the complaint conversation.

In the second scenario, the level of difficulty is increased. The customer is emotionally aroused because of his problem and must be calmed down. According to the classification of Rahim and Bonoma (1979), this customer is considered a *compromising* customer. The customer's problem is the most common problem within non-public online complaints (Cho et al., 2002): he has (among other things) problems with the customer service. The individualization on the delivered product is wrong and in a previous attempt to complain, the customer did not achieve a satisfactory result because there was a misunderstanding between the customer and the other member of the support staff. This makes the customer also a *follow-up complainant*, as it is the second time that he has contacted the customer service about the same problem (Stauss and Seidel, 2010). The result of the scenario is that after retrieving the database, the player learns that an error in production caused the incorrect individualization. Possible solutions to the problem in this scenario are replacement with the correct product or a refund. The goal of the scenario is to successfully pass through all five phases of a complaint process.

The third scenario is the one with the highest level of difficulty. Unlike the previous scenarios, it is less about processing the information milestones than about showing patience and applying soft skills. This customer can be classified as a *dominating* customer (Rahim and Bonoma, 1979) and a grouser (Stauss and Seidel, 2010). He is only focused on his own needs and shows little or no understanding for the other side. He tries to force a solution that is optimal for him and is looking for a continuation of the conflict. He has problems with the

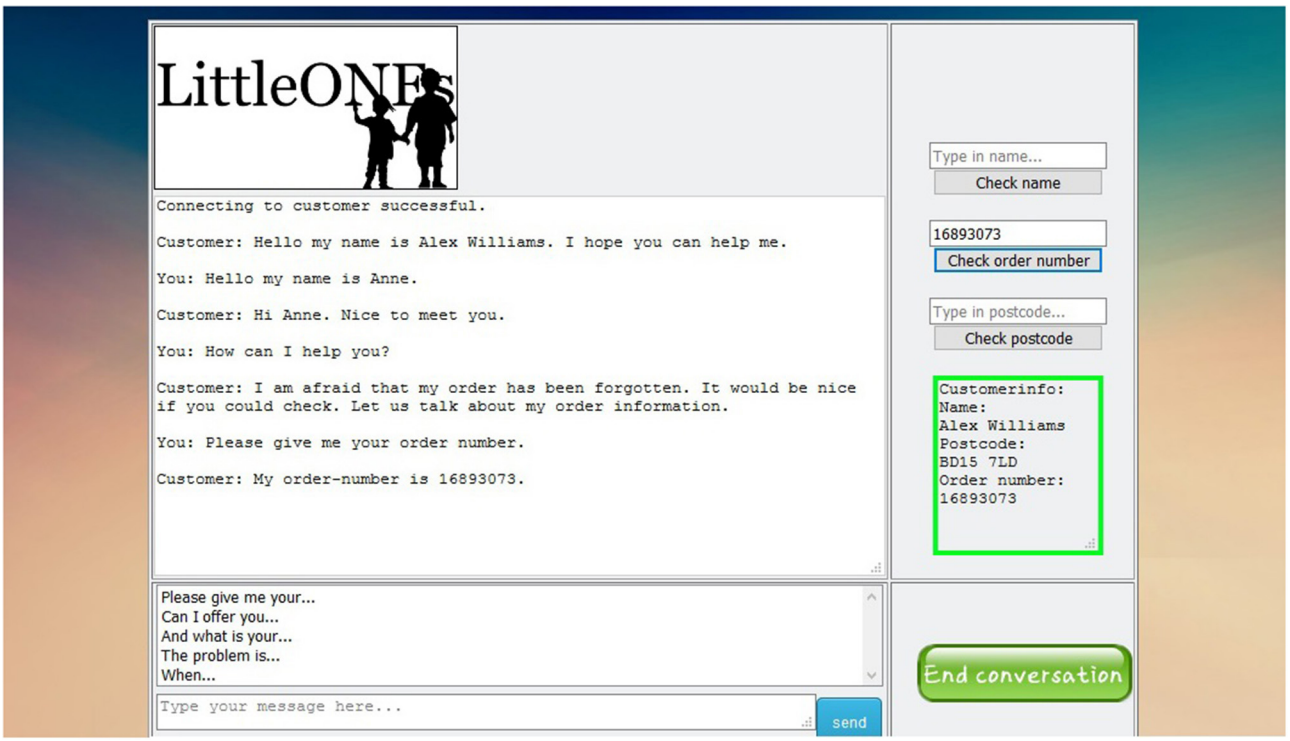


**FIGURE 4 |** Basic Interface of CuCoMaG.

business rules and conditions, which is the second most common problem with non-public online complaints according to Cho et al. (2002). The costumer is not reasonable and reacts abusive. When checking the database, the player learns that the customer is regularly complaining. The player's best result may be to not respond to the customer's provocations and finally end the conversation, since in this scenario, the player is not able to step into the next phase of the complaint process. This is called *active farewell* (Stauss and Seidel, 2010). The goal of this scenario is to deal with extreme situations. In this scenario, the player has to prove his ability to deal with provocations and difficult customers.

## Evaluation Results

The playability, game experience, as well as the perception of the serious game itself and the interaction with the chatbot in particular have been evaluated in a mixed method study combining qualitative and quantitative methods (Othlinghaus-Wulhorst et al., 2019). The results of this study will be summarized and discussed in this section.

To investigate, whether the scenarios are perceived as realistic, if the developed scenarios' chatbots behave as intended, and whether their style of conversation is influencing how the players experience the chatbots, three hypotheses have been formulated (Othlinghaus-Wulhorst et al., 2019)—mainly relying on the game experience and perception of the chatbots (subjective measures) and evaluation of the chat transcripts (objective measures):

1. *"Participants who play the second scenario ("compromising") achieve different results in the game experience questionnaire (GEQ) dimensions tension, negative affect, and challenge than participants who play the third scenario ("dominating")."*
2. *"Participants who play the second scenario ("compromising") achieve different results in the Holtgraves questionnaire dimensions comfortable, thoughtful, polite, responsive, and engaging than participants who play the third scenario ("dominating")."*
3. *"Participants with prior experience/knowledge in complaint management achieve better results than participants without prior experience/knowledge."*

20 subjects (average 26.05, SD = 7.99, 15 females, 5 males) participated in the study. of the subjects indicated that they had prior experience in customer complaint management. All participants of the study played two scenarios, either the first ("integrating customer") and the second ("compromising customer"), or the first and the third ("dominating customer"). The distribution was randomized. Subsequent to the gaming session, the participants were asked to answer several post-experiment questionnaires to collect their experiences and perceptions during the game: (1) The *Game Experience Questionnaire* (GEQ) (Ijsselsteijn et al., 2013), (2) a questionnaire for the evaluation of educational role-playing games (Dell'Aquila et al., 2017), and (3) a questionnaire for measuring the human-like qualities of the chatbots developed by Holtgraves et al. (2007). In addition to the questionnaires and a subsequent

qualitative interview, the chat transcripts have been evaluated in regard to the answer quality of the chatbot. Based on human coding, every answer of the chatbot during the gaming session was assigned to one of three categories: *constructive*, *comprehensive*, and *nonsensical*. In order to estimate which of the predefined sentence openers have been used frequently, rarely or not at all, the frequency of uses for each one was counted.

The first hypothesis could only be partially confirmed. There were only significant differences in the dimension *negative affect*, but not in the dimensions *tension* and *challenge*. The lack of significant results could be possibly caused by methodical conditions. First, the number of participants was rather small. Second, the participants were asked to evaluate the perception of both played scenarios combined. The second hypothesis could be partially confirmed as well. Players who played the second scenario indeed showed significant differences in the dimensions *thoughtful*, *polite*, *responsive*, and *engaging*, but not in the dimension *comfortable*. As predicted, there were no significant differences in the dimensions *human* and *skilled*. This suggests that there is a difference in the style of conversation but not in the quality of the chatbots' implementation. In order to be able to examine hypothesis 3, we needed to define "success in the game." It has been determined by (a) a relative score calculated by the system and (b) the total number of inputs, as it was assumed that fast completion is an indicator for effective complaint management. Unfortunately, this hypothesis could not be tested due to the small sample size ($n = 4$).

The analysis of the chat transcripts revealed that *constructive* chatbot responses were the ones occurring most often, followed by *comprehensible* responses. The number of *nonsensical* responses has been quite low for all scenarios, which underlines the quality of the chatbot scripts. Reponses that were categorized as comprehensible, but not constructive, were default outputs, which were implemented for every sentence opener in case the free text part of the chat message was not understood by the chatbot. This way, the chatbot was able to show that he still understands the general gist or intention of the message. Sentence openers used to obtain information from the customers (e.g., "Tell me…", "Please describe…") were used most frequently, as well as the sentence opener "I am sorry…", which is not surprising, since apologies are almost always suitable in the given situation and clearly associated with polite behavior. In general, the results of the study showed that the idea and approach of the game were rated positively, but the evaluation also revealed problems, e.g., with the use of the sentence openers. It could be validated that the chatbots' style of discussion is influencing the players' perception of them, which emphasizes the successful design of the dialog scripts.

# DIMENSIONS OF THE DESIGN OF SERIOUS ROLE-PLAYING GAMES FOR THE TRAINING OF SOCIAL SKILLS

The previous chapter has assembled examples instances of serious role-playing environments and ensuing empirical studies. In this chapter, we combine and inter-relate this experience with general issues in the design of serious role-playing games to devise and propose a set of design dimensions that constitute important aspects for the conceptualization, description, and comparison of serious role-playing games.

## Learning Context

Many practical considerations have to be taken into account when designing serious role-playing games for the training of social skills. The probably most important is (as it is the case for every educational game) to have a clear educational purpose for using them (Whitton and Hollins, 2008). Digital games have a great motivational potential, but this potential needs to be utilized to convey the pedagogical goals and learning objectives. The goal of the game should be aligned to the learning outcomes as much as possible, otherwise learners may learn something, but it may not be what was intended. Learning objectives and intents need to be translated into concrete mechanical elements of gameplay by mapping learning mechanics and game mechanics onto each other. In games that pursue the goal to impart and train certain skills, learners should be given the opportunity to put these skills into practice in order to facilitate skills acquisition and provide a context in which these skills are useful (Naido et al., 2000).

Furthermore, the setting of the game needs to be appropriate for the learning context (Whitton and Hollins, 2008). As many studies mentioned in chapter Related Work show, role-playing games are an ideal instrument for the assessment and training of soft skills. However, the chosen scenario and storyline need to be appropriate in the given thematic context and should be described adequately for the players, so they are able to develop immediate understanding and empathy with the role they are assigned. The storyline may be fictitious, but the concepts used in it should be real to ensure that a transfer to real-word settings is possible (Pivec, 2009). Also, the desired learning outcome will not be achieved unless the correct game situation is chosen for the selected topic (Salen and Zimmerman, 2004). Another important point is that the educational design must be based on an underlying corpus of background theories. This includes general psychological and pedagogical concepts and guidelines for the design of serious games, as well as theoretical foundations of the learning material itself.

## Technical Architecture and Set-Up

From the technical perspective, major issues regarding the implementation of serious role-playing games are flexibility, reusability, and extensibility/adaptability. Thanks to the use of a multi-agent blackboard architecture, our framework for scenario-based game development can easily be adapted and tailored to different settings and use cases, while the web-based gaming environment ensures easy access and platform independence. To adapt the framework to a new scenario, the following elements are needed: (1) a new GUI including sentence openers to be provided in menu selection, (2) new AIML scripts, and (3) modified or additional agents in the backend. The actual effort of course depends on the expertise of the developer.

The use of chatbots in serious role-playing games entails some major challenges. There are different approaches and technologies for natural language processing, each coming with specific advantages and disadvantages. In our approach, we use AIML as technological basis for the dialog modeling of the chatbots, but of course, there are many more approaches (e.g., data-driven technologies).

## Dialog Models and Degrees of Freedom in Communication

There is a range of possibilities for introducing dialogs with virtual characters (not necessarily chatbots) in digital role-playing games. Brusk and Björk summarize different dialog models in games (Brusk and Björk, 2009): In some games, dialogs are the only way of interacting with the game, meaning that the dialog *is* the gameplay. In other games, dialogs are integrated as separate modes. Either they are taking place concurrently to other actions, or solely with no other activity occurring at the same time. In our scenarios, dialog is indeed the main gameplay element. There may be side tasks, but the focus is on the communication behavior of the players and their interaction with other characters (chatbots and/or other players, depending on the setting of the game).

In dialog-centric role play settings, one major design decision is related to the degrees of freedom in communication the players have. There is a range of communication models from fully predefined single choice inputs to free text composition. The choice is mainly depending on the setting, the narrative structure of the game, and the technical implementation. Using single choice inputs within underlying conversation trees are rather easy to implement, but provide the least freedom in communication. The players always have to select a predefined answer from a given set, and have no possibility to express themselves. Also, the game plot and the structure of the dialog is predefined. There may be decision nodes in the communication tree allowing for different lines of action, but the freedom of choice is very limited.

In our approach, we decided to integrate chatbots as dialog partners for the players. As illustrated in chapter Multi-Agent Architecture, it is very hard for natural language processing artificial intelligence to really grasp the sense of what has been said and a sophisticated chatbot design and implementation is a complex task. Thus, free text input poses a big challenge for developers. The use of sentence openers appears to be a compromise between these two ends of the spectrum. On the one hand, it limits the possible inputs, which reduces the complexity of the AIML scripts immensely and helps the chatbot to understand the general gist of a text input. On the other hand, it still offers the players the possibility to formulate their own inputs and express themselves more freely.

## Feedback and Scaffolding Elements and Mechanisms

As we have shown in chapter Adaptive Feedback, feedback is crucial for ensuring the success of any serious role-playing game. It allows the learners to reflect on what happened during the role play and to analyze the consequences of their actions. In our approach, we differentiate between *ingame* and *aftergame* feedback. Ingame feedback refers to implicit feedback during the role-playing session. We realize this kind of feedback mainly through the reactions of the chatbots. Other feedback mechanisms are conceivable, but they should not corrupt or break the immersion during the role play situation. The balance between keeping the realism and immersion on the one hand and providing information on the status of the conversation as well as the players' performance and progress on the other hand is proposing a major design challenge for this kind of games. In our approach, aftergame feedback is an important point for enabling reflection processes. We consider a combination of an overall summary presented after the role play session and some kind of augmented replay of the dialog particularly helpful and promising.

Another important challenge for research and development in the area of serious role-playing games is to establish intelligent mechanisms for support and guidance (scaffolding). Learners should be provided with appropriate support in order to enable them to master the challenges of the game and achieve the learning goals. Ideally, a serious game should also adapt to the learners' level of knowledge, skills, as well as progress and current performance, as adaptation and personalization are considered key factors for education (Bellotti et al., 2010). Kickmeier-Rust and Albert suggest the introduction of micro-adaptive interventions (Kickmeier-Rust and Albert, 2010). This approach allows for interventions, support, guidance or feedback in a meaningful and personalized way, embedded in the game flow. These adaptive educational mechanisms are supposed to support the learner by hinting or providing appropriate feedback in certain situations, e.g., when misconceptions occur or when the progress is unsatisfactory (Kickmeier-Rust and Albert, 2010). The idea is to provide help to the learners by intelligently monitoring and interpreting their behavior in a non-invasive manner, which we consider a very promising approach. At this point of time, scaffolding, adaptation and personalization are incorporated in our framework only to a limited extend, thus augmenting these dimensions in our approach proposes a significant challenge for future research.

## Relation Between Immersion and Reflection

As described in chapter Immersion and Reflection, one major advantage of games is their motivational and immersive potential. Immersion holds the potential to motivate learners and make them get more engaged in learning task and this potential needs to be used to full capacity in the role play situation. Getting immersed in a game requires some degree of (perceived) realism, because if learners do not perceive a scenario as realistic, they are likely to regard the game experience as irrelevant to their understanding of the real world (Sutcliffe, 2002). Thus, realism is an important characteristic of any successful serious role-playing game design. Ribbens and Malliet identified seven factors of perceived game realism: (1) simulation realism, (2)

freedom of choice, (3) character involvement, (4) perceptual pervasiveness, (5) authenticity regarding subject matter, (6) authenticity regarding characters, and (7) social realism (Ribbens and Malliet, 2010).

A properly designed serious role-playing game also needs to provide support for reflection, allowing the learners to re-think and reflect on their actions. There are approaches claiming that it could be beneficial to have reflection taking place within the game itself without letting the learner step out of the game world by offering reflection activities within the game (Yusoff et al., 2009). However, as we have argued in chapter Immersion and Reflection, there is reason to assume that immersion tends to hinder the critical self-reflection, and based on this assumption, we decided to separate the actual role play phase from the reflection phase in our framework, allowing the learners to step out of the game world and their role and take over a distant perspective during the reflection phase. Following Malzahn et al. (2010), we claim that reflection needs role distance, which is not compatible with a high degree of immersion (although this is desirable during the actual role play). Accordingly, phases of enactment (role play) should be separated from reflection. Reflection phases should enable to take a third-person perspective on the prior experience, which requires an accessible/readable representation of this experience. During this phase, immersion is explicitly undesirable in order to help learners to view their own actions from the perspective of an external observer.

## Collaboration Support

An increasing popularity of multi-user virtual environments and games is causing a growing interest in the use of collaborative technologies for learning scenarios and recent research is indicating the positive effects of collaborative learning (Whitton and Hollins, 2008). Collaborative learning in the context of games describes a learning situation in which more than one learner participates in a learning (game) activity pursuing a common goal (Romero et al., 2012). In collaborative scenarios, learners work together on a common goal, they share and construct a certain level of knowledge, expertise and understanding (Romero et al., 2012). Major pedagogical benefits of bringing collaborative elements in gaming environments are (among others) providing multiple perspectives, creating self-awareness of the learning process, and thus making learning authentic and relevant (Whitton and Hollins, 2008). Serious games can provide a context for solving tasks and learning together with others. Integrating collaborative elements in a serious game may increase the players' motivation and foster the development of cognitive skills (Romero et al., 2012). In addition, collaborative virtual environments allow for a detailed recording of all collaborative interactions and thus may help to get a better understanding of those (Dillenbourg, 1999). Dillenbourg claims that it should be the aim of research to determine under which conditions collaborative learning is efficient (Dillenbourg et al., 1996).

He identifies three main criteria for rich and successful collaborative learning interactions (Dillenbourg, 1999): interactivity, synchronicity and negotiability. *Interactivity* is an integral part of any collaborative situation. It is not the frequency of interactions that defines the degree of interactivity, but the extent to which the interactions influence the other persons' cognitive processes. *Synchronicity* means that persons involved in a collaborative situation wait for messages from others and process them immediately. *Negotiability* relates to the structure of collaborative dialog being more complex than a hierarchical situation. That means one person will not impose her view only based on her authority, but will (to a certain extent) argue for her standpoint, justify, negotiate, and try to convince.

We have provided an example of a collaborative scenario in chapter Case Study: Conflict Management., in which two human players are involved in a mediation talk moderated by an AI-controlled mediator. However, not in all scenarios it is desirable and reasonable to include collaborative elements. It always depends on the context, the scenario and the learning objectives. If a task can be solved by one player, there is no need for collaboration. Thus, the tasks incorporated in the game should be only solvable if players act together and there should be a common goal (Wendel et al., 2013).

# CONCLUSION

In this article, we presented a technical and conceptual framework for serious role-playing games for the training of specific social skills in virtual learning environments involving chatbots in dialog-centric settings. From the design perspective, three distinctive conceptual features characterize our framework: (1) chat-like interaction with an AI-controlled chatbot, (2) phases of immersion (role-playing) and reflection are separated to facilitate a change of perspective that is considered conducive for learning, and (3) the learning process is emphasized by means of adaptive feedback based on individual analyses. The technical conception is based on three main components: (1) AI-controlled chatbots that adapt to the player's behavior, (2) a multi-agent blackboard system as the backbone in order to keep components independent and to optimize performance due to parallel processing; and (3) intelligent support for an automated evaluation of the player's performance and feedback generation.

Different use cases based on this framework have been presented, including scenarios for the training of workplace-oriented conflict management, patient-centered medical interviews, and customer complaint management. First evaluation studies indicate that this approach is assessed positively, the scenarios are perceived as useful and realistic and may qualify for real training situations. Due to the flexible architecture, our framework can easily be tailored to different settings and use cases and thus serve as a basis for future research focusing on the adaptation to other contexts and systems.

Our framework facilitates the building of serious virtual role playing games in that it allows for tailoring and adapting a given component architecture with very limited effort, comprising the provision of a specific GUI with sentence openers, a new set of AIML scripts (chatbots), and (possibly) a modification/extension of the backend agents. The framework provides all the basic mechanisms such as the inter-operability between GUI, chatbots, and agents through a tuple space. The basic architecture is

available as a kind of modifiable prototype. Other than model-driven development (Schmidt, 2006), our approach does not use meta-level descriptions in combination with generators in the overall systems engineering process. Only the AIML-based specification of chatbot behavior could be conceived as a meta-level element. However, this is limited to one of the components and only imported and exploited in our application framework. Although our system architecture and basic mechanisms are predefined, these premises do not preclude the quality of the ensuing application instances. These depend very much on the specification of chatbot scripts as well as on the GUI design. Accordingly, our evaluations have relied on standard instruments to measure game experience and usability as the main human-oriented factors.

Based on our experience, we formulated a set of general dimensions and challenges in the design of serious role-playing games for the training of social skills. In summary, we identified six major aspects: The *learning context* builds the basis of each serious game and relates to its theoretical foundation and the desired learning outcomes. The *technical architecture and set-up* refer to technologies and tools that are used for the technical implementation of such games and the underlying system architecture. *Dialog models and degrees of freedom in communication* relates to the question of how the communication with the non-playing dialog partner(s) is carried out, structured, and controlled, from predefined answers to sentence openers to free text input. *Feedback and scaffolding elements and mechanisms* are essential for the transfer of learning to application in the real world and can be integrated in many different ways. The *relation between immersion and reflection* refers to the question whether phases of immersion and reflection overlap or occur separate from each other. *Collaboration support* relates to the number of (human) players involved in the game and the question whether it enables collaborative learning.

## DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation, to any qualified researcher.

## AUTHOR CONTRIBUTIONS

JO-W has been involved in a guiding role (main responsibility) in the design, implementation, and evaluation of the serious role-playing games ColCoMa and CuCoMaG described in chapter Case Studies. JO-W and HH are (co-)authors of the relevant publications connected to these developments. JO-W wrote the main manuscript with text input from HH. All authors have reviewed and approved the manuscript.

## ACKNOWLEDGMENTS

The authors would like to thank several students and employees of the University of Duisburg-Essen who were involved in the design, development, evaluation, and publication of the mentioned use cases, namely Katharina Emmerich, Katja Neuwald, and Anna Jedich (ColCoMa), Philipp Behler, Ingo Börsting, Heike Choi, Evelyn Fricke, Stefan Liszio, Christian Klöpfel, and Anna Kizina (training scenario for medical interviews), as well as Dorian Doberstein, Nadja Agreiter, Marco Bäumer, Menglu Cui, Shaghayegh Abdollahzadegan, Diba Heidari, Nan Jiang, Markus Mentzel, Huangpan Zhang, Hao Zheng, Viet Hung Dinh, and Anne Mainz (CuCoMaG).

## REFERENCES

Abdul-Kader, S. A., and Woods, J. (2015). Survey on chatbot design techniques in speech conversation systems. *Int. J. Adv. Comput. Sci. Appl.* 6, 72–80. doi: 10.14569/IJACSA.2015.060712

Annetta, L. A. (2010). The "I's" have it: a framework for serious educational game design. *Rev. Gen. Psychol.* 14, 105–113. doi: 10.1037/a0018985

Augello, A., Gentile, M., and Dignum, F. (2016). "Social agents for learning in virtual environments," in *International Conference of Games and Learning Alliance* (Cham: Springer), 133–143.

Aylett, R. S., Louchart, S., Dias, J., Paiva, A., and Vala, M. (2005). "FearNOT! - an experiment in emergent narrative," in *International Workshop on Intelligent Virtual Agents* (Berlin; Heidelberg: Springer), 305–316.

Beaumont, R., and Sofronoff, K. (2008). A multi-component social skills intervention for children with asperger syndrome: the junior detective training program. *J. Child Psychol. Psychatry* 49, 743–753. doi: 10.1111/j.1469-7610.2008.01920.x

Behler, P., Börsting, I., Choi, H., Fricke, E., Liszio, S., Klöpfel, C., et al. (2013). "Eigentlich geht es mir gut - Entwicklung eines Serious Games zur patientenzentrierten Gesprächsführung," in *DeLFI 2013: Die 11. E-Learning Fachtagung Informatik*, eds A. Breiter and C. Rensing (Bonn: Gesellschaft für Informatik e.V.), 11–23.

Bellotti, F., Berta, R., De Gloria, A., and Primavera, L. (2010). Supporting authors in the development of task-based learning in serious virtual worlds. *Br. J. Educ. Technol.* 41, 86–107. doi: 10.1111/j.1467-8535.2009.01039.x

Bezuijen, A. (2012). *Teamplay: the further development of TeamUP, a teamwork focused serious game* [Master's thesis]. Delf University of Technology, Delft, Netherlands.

Bosse, T., and Gerritsen, C. (2016). "Towards serious gaming for communication training - a pilot study with police academy students," in *International Conference on Intelligent Technologies for Interactive Entertainment* (Cham: Springer), 13–22.

Boud, D., Keogh, R., and Walker, D. (1985). "Promoting reflection in learning: a model," in *Reflection: Turning Experience Into Learning*, eds D. Boud, R. Keogh, and D. Walker (London: Kogan Page), 18–40.

Brusk, J., and Björk, S. (2009). "Gameplay design patterns for game dialogues," in *Proceedings of the 2009 Digital Games Research Association Conference* (London).

Cho, Y., Im, I., Hiltz, R., and Fjermestad, J. (2002). "An analysis of online customer complaints: implications for web complaint management," in *Proceedings of the 35th Annual Hawaii International Conference on System Sciences* (Big Island, HI: IEEE), 2308–2317. doi: 10.1109/HICSS.2002.994162

Combs, M. L., and Slaby, D. A. (1977). Social-skills training with children. *Adv. Clin. Child Psychol.* 1, 161–201. doi: 10.1007/978-1-4613-9799-1_5

Corti, K. (2006). Games-based Learning; a serious business application. *Informe PixelLearn*. 34, 1–20. Available online at: http://www.pixelearning.com/docs/seriousgamesbusinessapplications.pdf

De Freitas, S., and Oliver, M. (2006). How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? *Comput. Educ.* 46, 249–264. doi: 10.1016/j.compedu.2005.11.007

Dell'Aquila, E., Marocco, D., Ponticorvo, M., Di Ferdinando, A., Schembri, M., and Miglino, O. (2017). *Educational Games for Soft-Skills Training in Digital Environments*. Cham: Springer. doi: 10.1007/978-3-319-06311-9

Di Ferdinando, A., Schembri, M., Linchan, C., Linehan, C., and Miglino, O. (2011). *Learn to Lead - A Web Based Game to Teach Leadership Theories in Vocational Courses. Games and Creativity in Education and Training*. Naples: Fridericiana Editrice Universitaria.

Dillenbourg, P. (1999). What Do You Mean by Collaborative Learning? *Collaborative Learning: Cognitive and Computational Approaches.* Oxford: Elsevier, 1–19.

Dillenbourg, P., Baker, M., Blaye, A., and O'Malley, C. (1996). The Evolution of Research on Collaborative Learning. Learning in Humans and Machine: Towards an Interdisciplinary Learning Science. Oxford: Elsevier, 189–211.

Doberstein, D., Agreiter, N., Bäumer, M., Cui, M., Abdollahzadegan, S., Heidari, D., et al. (2016). "CuCoMaG - group reflection support in role-playing environments," *in DeLFI 2016 - Die 14. E-Learning Fachtagung Informatik*, ed U. Lucke et al. (Bonn: Gesellschaft für Informatik e.V.), 327–329.

Emmerich, K., Neuwald, K., Othlinghaus, J., Ziebarth, S., and Hoppe, H. U. (2012). "Training conflict management in a virtual environment," *International Conference on Collaboration and Technology* (Berlin; Heidelberg: Springer), 17–32.

Even, C., Bosser, A. G., Ferreira, J., Buche, C., Stéphan, F., Cavazza, M., et al. (2016). "Supporting social skill rehabilitation with virtual storytelling," in *Proceedings of the 29th International Florida Artificial Intelligence Research Society Conference (FLAIRS 2016)* (Key Largo, FL), 329–334.

Festinger, L. (1954). A theory of social comparison processes. *Hum. Relat.* 7, 117–140. doi: 10.1177/001872675400700202

Flynn, R., McKinnon, L., Bacon, E., and Webb, J. (2011). "Maritime city: using games technology to train social workers - some initial results," in *International Conference on Entertainment Computing*. (Berlin; Heidelberg: Springer), 415–418.

Gebhard, P., Schneeberger, T., André, E., Baur, T., Damian, I., Mehlmann, G., et al. (2018). Serious games for training social skills in job interviews. *IEEE Trans. Games* 11, 340–351.

Gee, J. P. (2007). *Good Video Games and Good Learning: Collected Essays on Video Games, Learning, and Literacy*. New York, NY: Lang.

Gelernter, D. (1985). Generative communication in Linda. *ACM Transac. Prog. Lang. Syst.* 7, 80–112. doi: 10.1145/2363.2433

Glock, F., Junker, A., Kraus, M., Lehrian, C., Schäfer, A., Hoffmann, S., et al. (2011). ""Office Brawl" - A Coversational Storytelling Game and its Creation Process," in *Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology (ACE 2011)* (New York: ACM).

Gresham, F., and Elliott, S. N. (2008). *Social Skills Improvement System (SSIS) Rating Scales*. Bloomington, IN: Pearson Assessment.

Gunter, G. A., Kenny, R. F., and Vick, E. H. (2006). A case for a formal design paradigm for serious games. *J Int. Dig. Media Arts Assoc*, 3, 93–105. Available online at: https://www.academia.edu/download/49480493/Gunter_20Kenny_20Vick_20paper.pdf

Haferkamp, N., Kraemer, N., Linehan, C., and Schembri, M. (2011). Training desaster communication by means of serious games in virtual environments. *Entertain. Comput.* 2, 81–88. doi: 10.1016/j.entcom.2010.12.009

Harteveld, C. (2011). *Triadic Game Design*. London: Springer.

Holtgraves, T. M., Ross, S. J., Weywadt, C. R., and Lin, H. T. (2007). Perceiving artificial social agents. *Comput. Hum. Behav.* 23, 2163–2174. doi: 10.1016/j.chb.2006.02.017

Hubal, R. C., Frank, G. A., and Guinn, C. I. (2003). "Lessons learned in modeling schizophrenic and depressed responsive virtual humans for training," in *Proceedings of IUI* (Miami, FL), 85–92. doi: 10.1145/604045.604062

Hunicke, R., Leblanc, M., and Zubek, R. (2004). "MDA: a formal approach to game design and game research," in *Proceedings of the Challenges in Games AI Workshop, 19th National Conference of Artificial Intelligence* (San Jose, CA: AAAI Press), 1-5.

Ijsselsteijn, W. A., De Kort, Y. A., and Poels, K. (2013). *The Game Experience Questionnaire*. Eindhoven: Technische Universiteit Eindhoven.

Jennett, C., Cox, A. L., Cairns, P., Dhoparee, S., Epps, A., Tijs, T., et al. (2008). Measuring and defining the experience of immersion in games. *Int. J. Hum. Comput. Stud.* 66, 641–661. doi: 10.1016/j.ijhcs.2008.04.004

Jeuring, J., Grosfeld, F., Heeren, B., Hulsbergen, M., Ijntema, R., Jonker, V., et al. (2015). *Communicate! - A Serious Game for Communication Skills. EC-TEL 2015*. Cham: Springer, 513–517.

Johnson, C. I., Bailey, S., and Van Buskirk, W. L. (2017). "Designing effective feedback messages in serious games and simulations: a research review," in *Instructional Techniques to Facilitate Learning and Motivation of Serious Games*, eds P. Wouters and H. van Oostendor (Cham: Springer), 119–140.

Jonassen, D., Mayes, T., and McAleese, R. (1993). "A manifesto for a constructivist approach to uses of technology in higher education," in *Designing Environments for Constructive Learning*, eds T. M. Duffy, J. Lowyck, and D. H. Jonassen (Berlin; Heidelberg: Springer), 231–247.

Kenny, R., and Gunter, G. (2011). Factors affecting adoption of video games in the classroom. *J. Interact. Learn. Res.* 22, 259–276.

Kerly, A., Hall, P., and Bull, S. (2006). "Bringing chatbots into education: Towards natural language negotiation of open learner models," in *International Conference on Innovative Techniques and Applications of Artificial Intelligence* (London: Springer), 179–192.

Kesti,M. O., Leinonen, J., and Kesti, T. (2017). "The productive leadership game: fromtheory to games-based learning," in *Public Sector Entrepreneurship and the Integration of Innovative Business Models*, eds M. Lewandowski and B. Kozuch (Hershey: IGI Global), 238-260. doi: 10.4018/978-1-5225-2215-7.ch010

Kickmeier-Rust, M. D., and Albert, D. (2010). Micro-adaptivity: protecting immersion in didactically adaptive digital educational games. *J. Comput. Assist. Learn.* 26, 95–105. doi: 10.1111/j.1365-2729.2009.00332.x

Knode, S., and Knode, J.-D. (2011). "Using a simulation program to teach leadership," in *Proceedings of the 2011 ASCUE Summer Conference* (North Myrtle Beach, SC), 86–92.

Kortmann, R., and Harteveld, C. (2009). "Agile game development: lessons learned from software engineering," in *Proceedings of the 40th Annual Conference of the International Simulation and Gaming Association* (Singapore).

Kruse, J. (2000). *Diagnostische Falleinschätzung bei Patienten mit Psychischen und psychosomatischen Beschwerden und Störungen in hausärztlichen Praxen* [Professorial dissertation]. Heinrich-Heine-Universität Düsseldorf, Düsseldorf, Germany.

Lane, H. C., and Hays, M. J. (2008). "Getting down to business: teaching cross-cultural social interaction skills in a serious game," in *Workshop on Culturally Aware Tutoring Systems (CATS)* (Montreal, QC), 35–46.

Leary, T. (1957). *Interpersonal Diagnosis of Personality: Functional Theory and Methodology for Personality Evaluation*. Oxford: Ronald Press.

Lim, M. Y., Aylett, R., Enz, S., Kriegel, M., Vannini, N., Hall, L., et al. (2009). Towards intelligent computer assisted educational role-play. *Edutainment* 2009, 208–219. doi: 10.1007/978-3-642-03364-3_27

Linssen, J., de Groot, T., Theune, M., and Bruijnes, M. (2014). "Beyond simulations: serious games for training interpersonal skills in law enforcement," in *Proceedings of European Social Simulation Association*, 604–607. Available online at: http://ddd.uab.cat/record/125597

Malzahn, N., Buhmes, H., Ziebarth, S., and Hoppe, H. U. (2010). "Supporting reflection in an immersive 3D learning environment based on role-play," in *European Conference on Technology Enhanced Learning* (Berlin; Heidelberg: Springer), 542–547.

Marocco, D., Pacella, D., Dell'Aquila, E., and Di Ferdinando, A. (2015). "Grounding serious game design on scientific findings: the case of ENACT on soft skills training and assessment," in *Design for Teaching and Learning in A Networked World (9307)*, eds G. Conole, C. Klobučar, J. Rensing, J. Konert, and É. Lavoué (Switzerland: Springer), 441–446.

Marr, A. C. (2010). *Serious Games für die Informations- und Wissensvermittlung. Bibliotheken auf Neuen Wegen*. Wiesbaden: Dinges & Frick.

Martens, A., Diener, H., Malo, S., Pan, Z., Cheok, A., Müller, W., et al. (2008). Games-based learning with computers - learning, simulations, and games. *Transac. Edutain* 1, 172–190. doi: 10.1007/978-3-540-69744-2_15

Mateas, M., and Stern, A. (2003). "Façade: an experiment in building a fully-realized interactive drama," in *Proceedings of the Game Developers Conference (GDC '03)* (San Jose,CA).

Mayer, I., Bekebrede, G., Harteveld, C., Warmelink, H., Zhou, Q., van Ruijven, T., et al. (2014). The research and evaluation of serious games: Toward a comprehensive methodology. *Br. J. Educ. Technol.* 45, 502–527. doi: 10.1111/bjet.12067

McMahon, M., and Ojeda, C. (2008). "A model of immersion to guide the design of serious games," in *Proceedings of the World Conference on E-Learning in Corporate, Government, Healthcare and Higher Education* (Las Vegas), 1833–1842.

Memarzia, M., and Star, K. (2011). "Choices and voices - a serious game for prevent violent extremism," in *Intelligence Management: Knowledge Driven Frameworks for Combating Terrorism and Orgamized Crime*, eds S. Yates and B. Akhgar (Heidelberg: Springer), 130–142.

Michael, D., and Chen, S. (2006). *Serious Games: Games that Educate, Train, and Inform*. Boston, MA: Thomson Course Technology.

Mitchell, A., and Savill-Smith, C. (2005). *The Use of Computer and Video Games for Learning. A Review of the Literature*. London: LSDA.

Murray, J. H. (2017). *Hamlet on the Holodeck: The Future of Narrative in Cyberspace*. Cambridge; London: MIT press.

Naido, S., Ip, A., and Linser, R. (2000). Dynamic goal-based role-play simulation on the web: a case study. *J. Educ. Technol. Soc.* 3, 190–202. Available online at: https://www.jstor.org/stable/10.2307/jeductechsoci.3.3.190

Othlinghaus, J., and Hoppe, H. U. (2016). "Supporting group reflection in a virtual role-playing environment," in *International Conference on Intelligent Technologies for Interactive Entertainment* (Cham: Springer), 292–298.

Othlinghaus-Wulhorst, J., Jedich, A., Hoppe, H. U., and Harrer, A. (2018). "Using eye-tracking to analyze collaboration in a virtual role play environment," in *International Conference of Collaboration and Technology* (Cham: Springer), 185–197.

Othlinghaus-Wulhorst, J., Mainz, A., and Hoppe, H. U. (2019). "Training customer complaint management in a virtual role-playing game: a user study," in *European Conference on Technology Enhanced Learning* (Cham: Springer), 436–449.

Pivec, P. (2009). *Game-Based Learning or Game-Based Teaching*? British Educational Communications and Technology Agency (BECTA), Report No. 1509.

Prensky, M. (2000). *Digital Game-Based Learning*. New York, NY: McGraw-Hill.

Proksch, S. (2010). *Konfliktmanagement in Unternehmen. Mediation als Instrument für Konflikt- und Kooperationsmanagement am Arbeitsplatz.* Heidelberg: Springer.

Rahim, A., and Bonoma, T. V. (1979). Managing organizational conflict: A model for diagnosis and intervention. *Psychol. Rep.* 44, 1323–1244. doi: 10.2466/pr0.1979.44.3c.1323

Ribbens, W., and Malliet, S. (2010). Perceived digital game realism: a quantitative exploration of its structure. *Pres. Teleoperat. Virtual Environ.* 19, 585–600. doi: 10.1162/pres_a_00024

Ritterfeld, U., Cody, M., and Vorderer, P. (2009). *Serious Games: Mechanics and Effects.* New York, NY: Routledge.

Romero, M., Usart, M., Ott, M., Earp, J., and de Freitas, S. (2012). "Learning through playing for or against each other? Promoting collaborative learning in digital game based learning," in *Proceedings of the 20th European Conference on Information Systems (ECIS) 2012* (Barcelona). Available online at: http://aisel.aisnet.org/ecis2012/93

Salen, L., and Zimmerman, E. (2004). *Rules of Play: Game Design Fundamentals*. Cambridge; London: The MIT Press.

Schatzki, T. R. (1996). *Social Practices: A Wittgensteinian Approach to Human Activity and the Social*. Cambridge: Cambridge University Press.

Schmidt, D. C. (2006). Model-driven engineering. *IEEE Comput.* 39:25. doi: 10.1109/MC.2006.58

Schuster, R. J. (2010). Gruppenreflexion als Kommunikationsinstrument. Management und Wirtschaft. *Schriftenreihe zur Wirtschaftlichen Forschung*, 13, 7–23. Available online at: http://www.fhvie.ac.at/Forschung/Publikationen/Schriftenreihe

Schwantes, U., and Kampmann, M. (2007). Gesundheitsorientierte Gesprächsführung. *Der Mensch* 38. Available online at: https://www.dachverband-salutogenese.de/cms/der-mensch/heft-38-12007/

Shawar, B. A., and Atwell, E. (2007). Chatbots: are they really useful? *Ldv Forum* 22, 29–49.

Shute, V. J. (2008). Focus on formative feedback. *Rev. Educ. Res.* 78, 153–189. doi: 10.3102/0034654307313795

Squire, K., and Jenkins, H. (2003). Harnessing the power of games in education. *Insight* 3, 5–33.

Stauss, B., and Seidel, W. (2010). "Complaint management," in *Introduction to Service Engineering*, eds G. Salvendy and W. Karwowski (Hoboken, NJ: John Wiley & Sons), 414–432.

Susi, T., Johannesson, M., and Backlund, P. (2007). *Serious Games: An Overview. Technical Report HS-IKI-TR-07-001*. Skövde: School of Humanities and Informatics, University of Skövde.

Sutcliffe, M. (2002). "Simulations, games and role-play," in *Handbook for Economics Lecturers*, ed P. Davies (Bristol: The Higher Education Academy, The Economics Network), 1–26.

Totty, M. (2005). Better training through gaming. *Wall Street J. Eastern Ed.* 245:R6.

Vaassen, F., and Wauters, J. (2012). "deLearyous: training interpersonal communication skills using unconstrained text input," in *Proceedings of ECGBL* (Cork), 505–513.

Van Dijk, D., Dick, R., Hunneman, R., and Wildlevuur, S. (2008). Self City: "Training social skills in a game," in *Proceedings of Second European Conferences on Game-based Learning* (Academic Publishing Limited) (Barcelona), 481–488.

Walk, W., Görlich, D., and Barrett, M. (2017). "Design, dynamics, experience (DDE): an advancement of the MDA framework for game design," in *Game Dynamics* eds O. Korn and N. Lee (Cham: Springer), 27–45.

Wallace, R. (2004). *The Elements if AIML Style*. San Francisco, CA: ALICE AI Foundation.

Weinbrenner, S. (2012). *SQLSpaces - A Platform for Flexible Language-Heterogeneous Multi-Agent Systems* [Doctoral dissertation]. Verlag Dr. Hut, Munich, Germany.

Wendel, V., Gutjahr, M., Göbel, S., and Steinmetz, R. (2013). Designing collaborative multiplayer serious games. *Educ. Inform. Technol.* 18, 287–308. doi: 10.1007/s10639-012-9244-6

Whitton, N., and Hollins, P. (2008). Collaborative virtual gaming worlds in higher education. *ALT-J* 16, 221–229. doi: 10.3402/rlt.v16i3.10900

Winkler, R., and Söllner, W. (2018). *Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis.* Chicago, IL: Academy of Management Annual Meeting (AOM).

Yee, N. (2006). The demographics, motivations, and derived experiences of users of massively multi-user online graphical environments. *Pres. Teleoperat. Virtual Environ.* 15, 309–329. doi: 10.1162/pres.15.3.309

Yusoff, A., Crowder, R., Gilbert, L., and Wills, G. (2009). "A conceptual framework for serious games," in *2009 Ninth IEEE International Conference on Advanced Learning Technologies* (Riga: IEEE), 21–23. doi: 10.1109/ICALT.2009.19

Ziebarth, S., Kizina, A., Hoppe, H. U., and Dini, L. (2014). "A serious game for training patient-centered medical interviews," in *14th International Conference on Advanced Learning Technologies* (Athens: IEEE), 213–217. doi: 10.1109/ICALT.2014.69

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.