International Journal of Serious Games

ISSN: 2384-8766
https://journal.seriousgamessociety.org/

Article

Role-Playing in Teacher Education with InCoLearn and its Qualitative Usability

Charlotte Knorr[1] and Bernd Zinn[1]

[1]*Department of Vocational Education focused on Teaching Technology (BPT), University of Stuttgart, Stuttgart, Germany*
charlotte.knorr@ife.uni-stuttgart.de; bernd.zinn@ife.uni-stuttgart.de

Keywords:

Digital Game-Based Learning
Role-Playing Game
Teacher Education
Vocational Education
Usability
Qualitative Games User Research

Received: September 2023
Accepted: February 2023
Published: March 2024
DOI: 10.17083/ijsg.v11i1.681

Abstract

Role-playing has long been a method in the professionalization of educators to emulate or anticipate everyday school and teaching situations. This paper introduces InCoLearn, the first multiplayer online role-playing game (RPG) developed for student teachers to facilitate professional and action-based knowledge of inclusion and heterogeneity in the classroom with non-linear storytelling through individual quests. A qualitative usability study conducted via focus groups aims to provide insights into usability. It seeks to answer the research question of which aspects are perceived as the most important and problematic to student teachers using InCoLearn. To answer another research question, whether previous quantitative usability results can be confirmed and explained by the qualitative results, the qualitative results are set into the context of a quantitative usability study.

1. Introduction

The rapid advancement of technology across various sectors has profoundly influenced the education landscape. Among the emerging trends is the integration of digital role-playing games (RPGs) within the realm of serious games for teacher education, showing substantial promise in enhancing learning outcomes [1]. Despite their potential benefits, the total adoption of serious games – mainly digital RPGs – in vocational education remains somewhat elusive. At present, they are primarily confined to research projects and isolated initiatives that cater to larger target groups but are only available to a select few, e.g., *IVT-T* [2] for vocational education and training (VET) of teachers or *Game Bridge* [3] in the context of event personnel. This limitation presents an obstacle to realizing the widespread advantages that serious games can offer. One of the critical challenges in successfully implementing digital RPGs in teacher education is ensuring that these games are accessible and enjoyable for all users. The intricate systems, mechanics, and diverse player base make this task particularly crucial. Notably, in the case of RPGs, players invest significant time – sometimes spending hundreds of hours – exploring game worlds, developing characters, and engaging with storylines [4]. Hence, if the

game's usability falls short, it can lead to frustration as well as disengagement, and ultimately cause users to abandon the game or result in ineffective learning outcomes.

Moreover, the context of teacher education presents unique challenges, such as catering to diverse learners and addressing complex teaching concepts [1], which further complicates the usability issue. Consequently, conducting dedicated usability research becomes imperative to ensure that digital RPGs are thoughtfully designed and implemented to meet the specific needs of teacher education. Integrating digital RPGs into serious games can significantly empower teacher education and enhance learning outcomes. However, their limited availability in vocational education calls for broader adoption and accessibility. Addressing usability concerns through dedicated research and tailored design will be instrumental in harnessing the full potential of these immersive and interactive learning tools in the education sector.

1.1 Background

In recent years, there has been a growing interest in incorporating RPGs into education programs [4]–[5]. RPGs involve participants assuming fictional roles and engaging in collaborative storytelling in multiplayer settings [5]. They offer unique opportunities for pre- and in-service teachers to develop and enhance their professional skills. RPGs provide an experiential learning environment, allowing teachers to engage in authentic teaching situations. By assuming roles, making meaningful decisions, and exploring different paths [5] or consequences to decisions within the game, educators can gain first-hand experience in managing classroom dynamics, addressing diverse student needs, and adapting instructional strategies. RPGs evoke emotional engagement among players [6], simulating the complex social and emotional dynamics in real classrooms. RPGs often present participants with challenging scenarios that require problem-solving and critical thinking skills, enabling collaboration and communication among participants. While the integration of RPGs into teacher education is a relatively new field, research has begun to explore its potential benefits. Existing studies in education programs have indicated several positive outcomes, including critical thinking [7], motivation [7]–[8], existing activity [9], learning performance [8][10], increased self-efficacy [8], improved skills [11], more perceived learning opportunities [12] and enhanced empathy [13], as well as creativity [11] and professional collaboration [14]. Although the positive effects of games in vocational education outweigh the negative effects in general [15], decreased knowledge of students with high prior knowledge and no effects on motivation have also been reported with an RPG in this field [16]. Integrating role-playing into teacher education holds strong potential for improving future teachers' skills, knowledge, and attitudes. The experiential and immersive nature of role-playing provides unique opportunities to develop essential competencies required for effective teaching. Although the available evidence points to the positive impact of integrating role-playing games into teacher education, little is known about the design and use of digital game-based learning (DGBL) with RPGs in professional education (e.g., [17]–[18]).

DGBL with RPGs – also known as serious role-playing games – is acknowledged for its potential to deliver immersive, interactive, and engaging learning experiences [19] that can enhance teachers' professional development. These games serve as entertainment and a medium to impart serious content and train practical skills [19]. Online role-play games, in particular, provide a realistic simulation of typical professional scenarios, enabling preservice teachers to train and reflect on their communicative skills through an avatar [20]. DGBL with RPGs can present authentic and complex scenarios in vocational teacher education that reflect real-world teaching situations. This allows prospective teachers to practice and reflect on their teaching strategies in a safe and controlled environment [1]. Virtual RPGs offer mobile, safe, and continuous environments [19], making them a practical tool for teacher training.

Furthermore, DGBL with RPGs can aid in developing teachers' professional vision, including the ability to perceive and interpret significant features of classroom situations and

make appropriate pedagogical decisions [1]. These games can also assist in training specific social skills [19]. However, despite the potential benefits, implementing DGBL with RPGs in teacher education faces several challenges, including the need for suitable game design, integrating games into the curriculum, and providing adequate support for teachers. Further research is needed to explore the potential of DGBL with RPGs in teacher education and address the existing challenges. This could contribute to developing effective and innovative approaches to teacher education that meet the demands of the 21st century. The three essential dimensions to be measured to investigate game-based learning effectiveness are *learning* (outcomes), e.g., domain-specific knowledge tests, *engagement*, with surveys like the Intrinsic Motivation Inventory (IMI) [21] or Game Experience Questionnaire (GEQ) [22], and *usability* [23], e.g., through usability surveys or heuristics like PLAY heuristics [24].

Usability research is important for the fundamental evaluation of the quality aspects of any software product. It is furthermore an important endeavor in DGBL research [25] and all major stages of game production, such as pre-production, production [26] as well as post-production and games user research (GUR) in general [27]–[28] to enhance the gaming [29] and learning experience. Furthermore, it ensures the utility and acceptability of serious education games from pedagogical, technical, and contextual standpoints [30]. On the one hand, usability can be assessed quantitatively, often realized through surveys. Furthermore, physiological or biometric data and game analytics are often used for enrichment purposes or to answer specific research questions [31] regarding game usability or playability and beyond. On the other hand, qualitative usability research can be realized in multiple ways, e.g., behavior observation and/or think-aloud protocols of video-/audio-recorded playtests, heuristics, (group) interviews, and focus groups [27].

A few research projects and studies examining game-based learning in teacher education have used *Second Life* to develop classroom learning scenarios (e.g., [32]). Among those solutions, VirtualPREX [33] gained more attention in the past. It has been useful to create and conduct specific interventions and enhance university courses for student teachers with role-playing in the digital realm with limited effort. Furthermore, some games with a focus on teacher education have been developed that are available to the public and have also been part of studies. Other research projects include the development of serious games in the context of teacher education only in the scope of research projects, namely *Breaking Bad Behavior* (3B) [34] and *Interactive Virtual Training for Teachers* (IVT-T) [35]. There is another group of commercial serious games for teacher education and teaching simulations, which are, therefore, available to a broader player base and were often former products of research projects that have been further developed. Some of those games for teacher education are *simSchool* [36], *TeachLivE™/ Mursion™* [37]–[38], and *SimInClass* [39].

To provide context, the aforementioned simulations or serious games are compared with *InCoLearn* in terms of purpose, platform, genre, and an excerpt of the core (game) design elements of aesthetic design (2D/3D), narrative, game mechanics (only interactions) to present a short overview of the current landscape of interactive solutions for teacher training.

Table 1. Comparison of simulations / serious games for teacher training. [33]–[39]

Games/Simulations (Year)	Purpose	Platform	Genre	Core (Game) Design Elements
simSchool (2005)	Meaningful interactions with AI students in a virtual classroom	Desktop/Web App	(Classroom) Simulation	2D; individual characteristics and emotional reactions of students; interactions with UI elements and NPCs
VirtualPREX (2011)	Assessment of the virtual professional experience of preservice teachers	Desktop (SecondLife)	Online-RPG with Multiplayer	3D (First/Third Person); “naughty” and “Nice” students in a game with one (main) behavior; interactions with UI elements, objects, real players, and NPCs (“bots”)
TeachLivE™ (2014) / Mursion™	Provide a safe space for the professionalization and soft skill development of teachers	Mixed Reality	(Classroom) Simulation	3D (First Person); focus on pedagogical role-play; interactions with semi-programmed NPCs (puppetry, specialist/instructor needed to simulate students’ behavior)
Breaking Bad Behavior (2016)	Training classroom management skills with a focus on disruptive student behavior for teachers	Virtual Reality	(Classroom) Management Simulation	3D (First Person); the narrative has a minor role: behavior of NPCs through animations/SFX and simple dialogs; interactions with semi-programmed NPCs (puppetry)
IVT-T (2017)	Training of behavior management strategies in high poverty schools for early career teachers	Desktop	(Classroom) Management Simulation, RPG	3D (First Person); individual character backgrounds, non-linear storytelling with various dialogs; interactions with UI elements and NPCs
SimInClass (2018)	Training of classroom management skills according to Technological Pedagogical Content Knowledge (TPACK)	Desktop, VR, and Mobile	(Classroom) Management Simulation	3D (First Person); focus on simulation aspects and the whole class rather than individuals; interactions with UI elements and NPCs
InCoLearn (2021)	Professional and actionable knowledge of student teachers about inclusion and heterogeneity	Desktop	Online-RPG, Multiplayer Serious Game	3D (First Person); individual character backgrounds, non-linear storytelling with individual quests and various dialogs; interactions with UI elements, objects, and real players

Table 1 shows that *InCoLearn* is unique regarding the online multiplayer it offers as a standalone application and serious game. As *VirtualPREX* was created in *Second Life*, it is considered an online multiplayer solution but is limited to the customization opportunities *Second Life* offers its end users. *InCoLearn*, therefore, sets itself apart from the non-linear storytelling created with diverse character profiles, the quest design, and its elaborated dialog system in the context of an online multiplayer game. It offers task types and dialogs

characteristic to RPGs (like object interactions, external events, dialogs, etc.) and decision-making that controls the course of the game.

There has also been some user experience and/or usability research for some of the mentioned games: Some aspects of usability – labeled under the construct of user interaction – were addressed in a quantitative study on design quality study conducted with 22 student teachers and *simSchool* [40]. The answers to five questions show mixed opinions with means ranging from 2.36 ("poor") to 3.23 ("good") on a 5-point Likert scale, with most scores being at around three ("good"). Especially the feedback on student responses is not effectively used (M = 2.36; SD = .90). For *Breaking Bad Behavior* [34], only usability data of the user interface for instructors controlling the semi-programmed NPCs is available from a quantitative study with 11 instructors (10 student teachers) [34]. The results show a low task load (M = 6.42; SD = 3.42) and an overall high subjective consequence of intuitive use over five dimensions (M = 3.83; SD = .68) [34]. Usability research with the end user application has been conducted but not reported in detail [41]. After a heuristic evaluation, *IVT-T* was further evaluated with seven education students applying mixed methods: qualitative concurrent think-aloud protocols and semi-structured interviews as well as two surveys covering system usability and user satisfaction [2]. The quantitative data shows acceptable overall system usability (M = 86.79; SD = 8.75) and medium to high user satisfaction [2]. Qualitative data from the think-aloud protocols indicate most problems with "Learning" [2], i.e., learning controls and navigation. Interview data was used to illustrate character/environment design, narrative, and instructional design impressions collected from the play sessions and think-aloud protocols [2]. A qualitative study with 18 participants (16 aspiring teachers) on user experience after long-term usage of 14 weeks of *SimInClass* [39] included focus groups, interviews, and observations. The evaluation of common technical issues and the used interface shows a general appreciation of clear instructions (one code) and the accelerations feature (one code) opposing a freezing problem (three codes) and high difficulty (11 codes) as well as low visibility of app reports (one code) and not fulfilled expectations in terms of environmental change (one code). A mixed methods approach was used to evaluate the "Mystery Powder task" [42], created with *Mursion*, and its usability and viability with 53 participants (49 preservice teachers). The survey results of using the solution conducted with the preservice teachers produced game experience data rather than usability data. Only one usability-related aspect, "technical issues with the discussion performance" [42], that was not further elaborated, was covered in the quantitative results of the preservice teacher survey. Most of these studies indicate that a larger sample size is needed for future research.

Research on games with a strong focus on narrative, like *InCoLearn,* is currently just emerging [43]. Apart from the examples mentioned above, there is still a scarcity of games addressing game-based learning in the context of vocational education, and their application in vocational institutions is not common [15]. Often, trained instructors and extensive instructions are required to operate or use those games [15], which also applies to most games in Table 1 as the prior experience of student teachers. Their perception of strengths and weaknesses in serious games and simulations for teacher training impacts intentions to use such solutions. Therefore, in the field of technology acceptance, further research on experience and quality perception is required [40]. These research desiderata are addressed in the research design of the qualitative usability study.

1.2 Research Objectives and Questions

Apart from providing a useful tool for aspiring teachers to gain hands-on experience and develop professional and actionable knowledge about inclusion and heterogeneity by playing *InCoLearn* with others, the main research objective is to advance domain-specific research on the application of serious games in teacher education at vocational schools. *InCoLearn* also attempts to contribute to the young research field of serious game narrative with a strong focus

on non-linear storytelling during role-playing. It also offers a game design following established conventions from other RPG titles and onboards players via an instructional scenario or instructional quests that seamlessly integrate with the learning content. This usability study aims to gain a thorough understanding of important usability aspects of such serious games following the quality perception specific to the addressed target group of student teachers, which is subordinate to the overall research goal. The following research questions are addressed in this study:

1. Which usability aspects are most important to student teachers? (RQ1)
2. How can the qualitative analysis further divide the usability constructs of the quantitative usability study [30] into specific usability aspects and explain the quantitative results? (RQ2)

The following methods and material support the investigation of these questions.

2. Methods and Material

A qualitative usability study was conducted via group interviews and resulted in data from ten focus groups with 46 students in total and five students per group. This study was carried out alongside a quantitative usability and player experience study and a qualitative study of attitudes and requirements [44] to combine data in a mixed methods approach. In line with the mixed-methods approach, the qualitative data is integrated with the quantitative data, following the *embedded design* [45]. The main goal is complementarity [46], i.e., deepening the understanding of the quantitative outcomes through contextualization in qualitative analysis. The study was approved by the Ethics Committee of the University of Stuttgart. The *InCoLearn* game will be described and illustrated in the following sections, and the sample, data collection, and analysis will be outlined in more detail.

2.1 The InCoLearn Game

InCoLearn (Figure 1) is a multiplayer serious game and online RPG intended as well as newly designed and developed for – but not limited to – student teachers [18]. It was developed on the theoretical basis of the Integrated Design Framework for Playful Learning, constructivist and social constructivist learning theories [47], and the learning forms of collaborative and explorative learning [18]. It offers non-linear storytelling through quests, quest tasks, and dialogs in the form of a "directed network" [48] and offers a safe space to explore, train, as well as reflect on teaching and different teaching styles [18]. Bridging the diegetic frameworks of *educational role-play*, which strongly emphasize narrative elements to support learning objectives, with those of RPGs [4] that prioritize world-building and interaction with game mechanics, *InCoLearn* aims to transform educational role-play for the digital domain. It aims at fostering professional and actionable knowledge of inclusion and heterogeneity through the experience of different scenarios incorporating selected heterogeneity dimensions of six playable characters. These characters include three pupils from different cultural backgrounds with distinct profiles, some with special educational needs and mental issues. Different cultural and educational backgrounds are also involved in the profiles of the other three characters: the experienced teacher, the social pedagogue, and the student teacher. The interactive 3D setting of a fictional vocational school serves as the backdrop. Players are connected via voice chat and navigate freely, embodying one of the six characters. They experience a day/hour in the life of the character in this virtual school, with each role offering unique perspectives and challenges based on their distinct character traits, backgrounds, and attitudes. Particularly notable are the students, whose individual needs and qualities emerge through spontaneous actions, dialogues, and quest-driven activities. These interactions reveal each character's complexities to others and add depth as well as intricacy to the group's overall experience [18].

The audio design [18] features three unique zones – the classroom, hallway, and schoolyard – each with isolated voice chats and ambient sounds, creating a lifelike environment. Sound effects for footsteps, object interactions (foley sounds), and game events further immerse players and provide feedback. The prototype was created with Unity 2020.2.0f1 and the Universal Render Pipeline (URP) for optimal PC performance with heterogeneous hardware specifications [18]. It uses the *Photon Engine* for networking, through the *Photon Unity Networking 2* plugin, and incorporates *Photon Voice* for voice communication [18] of players and the following participants.

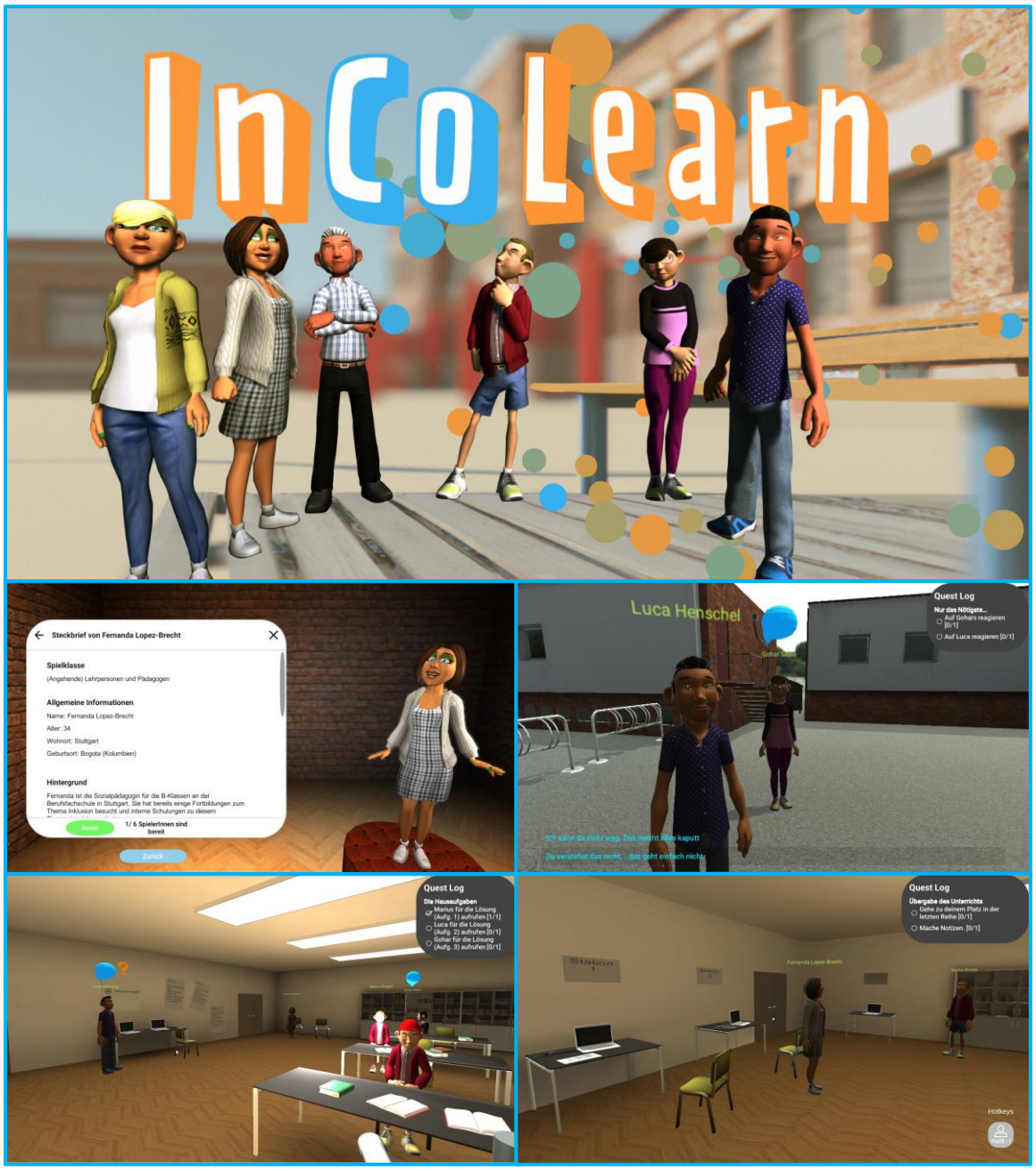

Figure 1. InCoLearn preview with the title (top), character selection and introduction (center left), dialog on schoolyard (center right), classroom layout in the introduction scenario (bottom left), and classroom layout for station learning in the first learning scenario (bottom right).

2.2 Participants

Potential participants were recruited through convenience and purposive sampling from three seminars at the Institute of Educational Science of the University of Stuttgart, focusing on

didactics along with inclusion and heterogeneity. This was mainly due to the reason that it was easier to obtain a larger sample size during the COVID-19 pandemic, as most of the participants would not have appropriate hardware at home (e.g., macOS or a headset) or had problems with installing *InCoLearn* or configuring audio settings in their operating systems. Furthermore, network problems or bugs in the prototype could potentially require a restart. These issues could be addressed more easily on-site.

53 teacher education students at vocational schools were part of the study, with 46 consenting to the analysis of their gaming session recordings. The demographic data collected in the quantitative study [44] included 23 undergraduates in Vocational and Technical Education and 30 graduate students in Technical Education. Females represented 58.49 % (31 students) and males 41.51 % (22 students), with an average age of 24.91 years (SD = 3.71; range: 20–37 years). The investigation into gaming habits revealed that hobby gamers (15.09 %, eight students) averaged 5.71 hours of weekly playtime, while casual gamers (28.30 %, 15 students) played for 2.57 hours on average. Non-gamers, making up 56.60 % of the participants, were split into novices (41.51 %, 22 students) and those with some gaming experience (15.09 %, eight students). This categorization is based on self-assessment involving gaming experience and usage. Considering the genre preferences of the experienced non-gamers, 62.5 % (8 students) primarily played sports and racing games, while only 12.5 % were involved in RPGs. Hobby gamers preferred strategy games (62.5 %, five students), with RPGs also being popular (37.5 %, three students). Casual gamers, representing 28.30 % of the participants, displayed a balanced preference for sports/racing and strategy games (53.33 %, eight students), with a third (five students) favoring RPGs. The computer was the most common gaming platform across all groups, particularly among 75 % of experienced non-gamers and hobby gamers (six students) and 53.33 % of casual gamers (eight students), with the latter equally engaging in console gaming.

2.3 Data Collection

As a basis for the discussions, a preceding playtest of the *InCoLearn* prototype of around 45 minutes – depending on the groups' pace – took place with teams of six people immediately beforehand. For the game sessions, each of the six participants chose one of the available characters to play in an introductory game scenario and played online with other players supported by voice chat.

Apart from statements by participants that were stated without explicitly asking, two main questions aimed at prompting discussion about usability through low-key entry points and via understandable questions without mentioning or explaining the usability aspects per se. These two questions were: "Which aspects of *InCoLearn* have had a promoting (positive) impact on your gaming experience?" and "Which aspects of *InCoLearn* have had an inhibiting (negative) impact on your gaming experience?"

2.4 Category System and Data Analysis

The category system was developed deductively based on the dialog principles of *ISO 9241-10:1996*, which was also the basis of the adapted usability survey [49] with five constructs used to map the results in mixed-methods analysis. The items of the entire individualization and error tolerance constructs did not fit the 3D game scope. Therefore, only *suitability for the task, self-descriptiveness, conformity with user expectations, suitability for learning,* and *controllability* were used.

The transcription of the discussions was assisted by *f4x* for automatic speech recognition and reviewed as well as corrected with *f4transkript*. After this, the transcripts of the discussions were analyzed via *qualitative content analysis* (QCA) [46][50] in *MAXQDA 2020* and later in *MAXQDA 2022*. Focus groups can also be analyzed with grounded theory and discourse

analysis [51]. As the qualitative data is put into the context of the quantitative usability study results [44] with predefined usability constructs, QCA is used and based on the categories derived from the usability constructs [44]. QCA by Udo Kuckartz [50] was chosen as the QCA methodology. With him being an author of the *MAXQDA* software relying on this methodology, its optimization for qualitative and mixed methods research, and offering features for focus group analysis, it was used for the data analysis of the focus groups.

As analyzing and, more specifically, coding certain segments of the discussion can be rather difficult compared to interviews or other parts of the group interviews, the coding guidelines were adjusted accordingly after analyzing the first 10 % of the material. The two coders, one from the research group specializing in games and one from outside the research group with a pedagogical background, coded the material independently without further alignment throughout the analysis. The coding procedure is also further elaborated in the following results.

3. Results

The QCA was conducted with the coding scheme units defined in Table 2. One particular difference compared to individual or group interviews is the consideration of multiple text passages of the transcribed discussions, as there can be interruptions or the opening of subdiscussions with new topics addressed. After those slide-ins in the text for the QCA, the discussion on a topic sometimes continues with the unfinished stream – sometimes initiated by the test administration – or stops entirely.

Table 2. Units applied in the coding scheme.

Coding Scheme Unit	Definition	Example
Analysis Unit	A focus group discussion from the respective data collection time point	RTF file of the Focus Group Discussion
Context Unit	Connected/adjacent sentences from a speaker within the group of test participants that are thematically related to a specific point (even in cases of interruptions by other participants)	W1: So, with the given text, I find it easy to empathize. I had Mr. Rubens, for example. Test Leader: Hm (affirmatively). W1: Um, because I had information that he is, for example, strict. Test Leader: Hm (affirmatively). W1: And that Mrs. Lopez-Kraft or -Brecht, I believe, doesn't really like him that much... Test Leader: Hm (affirmatively). W1: ...um, I could roughly estimate what I should respond. Test Leader: Okay. Hm (affirmatively). (...) Could the others also empathize well? M1: Well, I find the text absolutely helpful for creating a bit of a backstory, which you can get used to or adapt to. Test Leader: Hm (affirmatively). W1: So, in my case, it was mentioned that, uh, my German language skills are not yet, uh, perfect.
Coding Unit	Propositions from a speaker within the group of test participants	Having information like him being strict, for instance, and that Mrs. Lopez-Kraft or -Brecht, I think, doesn't have a very high opinion of him allowed me to roughly gauge what my responses should be.

To assess intercoder reliability, three aspects were taken into account: the presence of codes in documents, the frequency of codes within each document, and the overlap of code assignments in document segments with a minimum percentage of 50 % considering rather complex coding units with multiple interjections by other speakers. The intercoder reliability values of the coded material are 61.76 % for code occurrence and 29.41 % for code frequency. For code overlaps at a minimum of 50 % segment level, intercoder reliability is 41.82 %, and

10.91 % for a perfect code overlap (100 %). The chance-corrected Kappa [46] value of κ_n = .27 for code overlaps (50 %) shows a sufficient agreement with $\kappa > .20$ [52].

The quantitative results of the other usability study [44] extended with the data on the two aspects of *conformity with user expectations* and *suitability for learning* depicted in Figure 2 and Table 3 show another perspective on important areas of usability. They indicate that *conformity with user expectations (M* = 4.77*; SD* = .73) was rated best on the six-point Likert scale, and *controllability (M* = 3.78*; SD* = .97) has the strongest potential for improvement, despite still being rated with "somewhat agree" on average. Furthermore, *conformity with user expectations* and *suitability for learning* differ from the other constructs as their minimum values are higher ("somewhat disagree" (three)). Therefore, their ranges and standard deviations are smaller.

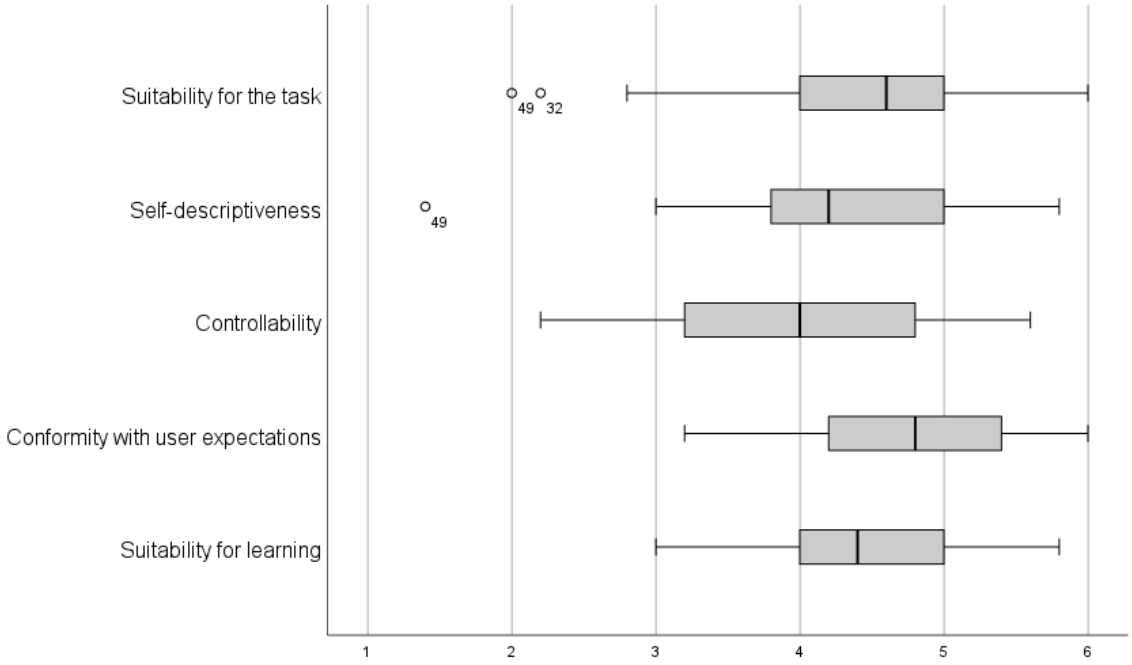

Figure 2. Quantitative results of the usability survey [44] with additional conformity with user expectations and suitability for learning

Table 3. Quantitative results of the usability survey [44] with additional conformity with user expectations and suitability for learning (six-point Likert scale)

Construct	*n*	*M*	*SD*	*IQA*	*Min*	*Max*	*Skewness*	*Kurtosis*
Suitability for the task	49	4.40	.80	.80	2.00	6.00	–.94	1.50
Self-descriptiveness	45	3.96	.94	1.30	1.40	5.80	–.35	.25
Controllability	30	3.78	.97	1.65	1.80	5.60	.11	–.74
Conformity with user expectations	44	4.77	.73	1.20	3.20	6.00	–.49	–.58
Suitability for learning	44	4.50	.77	1.15	3.00	6.00	.19	–.68

Based on the number of codes, *suitability for learning* (21 codes) holds the strongest relevance with the most statements in the focus groups, closely followed by *controllability (19)*. *Suitability for the task* and *self-descriptiveness* are both assigned ten codes. *Conformity with user expectations* (eight codes) has the fewest statements and, therefore, the lowest

relevance in the focus groups. *Controllability* was rated worst in the quantitative data [44] and has a relatively high number of codes. The major discussion topic of *suitability for learning* still reached an average of $M = 4.50$ ($SD = .77$) and ranked second best in the quantitative data. *Suitability for the task* ($M = 4.40$; $SD = .80$) and *self-descriptiveness* ($M = 3.96$; $SD = .94$) both rank in the middle area with a relatively moderate ten codes. The low concentration on *conformity with user expectations* (eight codes) is also represented by the best score in the quantitative data ($M = 4.77$; $SD = .73$). In summary, the quantity of codes in the qualitative data matches well with the quantitative results, except for *suitability for learning* ranking highest (21 codes). The following deep dive into the qualitative results at the code summary and statement levels will draw a detailed picture of the focus groups on the usability of *InCoLearn* and the especially important category of *suitability for learning*. Still, they will enrich the quantitative data and explain it to some extent.

Starting with the most codes (21) in the category of *suitability for learning*, participants provided diverse feedback with only seven codes highlighting limiting factors or problems with learning the game. Seven codes directed at the settling-in period were diverse, with four codes indicating a difficult or long settling-in period (five codes), contrasting with those who found it easy or short (two codes). The statements regarding the fact that the concept of quests was initially unknown (one code) and that the tutorial video was sufficient for understanding the game (one code) only partially explain the settling-in experiences. Four codes revolve around how participants got into role-playing: Three participants mentioned that the game provided easy entrance into role-play, while one still found it difficult. Seven codes about supporting aspects of learning the game – such as settling into the game and role-play – in the form of quest descriptions (one code), dialog options (three codes), and the character profile (short key) (three codes) shed light on how learning was facilitated.

In terms of *controllability* (nineteen codes), there are positive remarks about simple and intuitive controls (six codes), but concerns were more frequent (thirteen codes). Players were satisfied with the keyboard and mouse controls (four codes) as well as how quests (one code) and dialogs (one code) were controlled. Four codes highlight the inconvenience of mouse control, and there were multiple mentions (five codes) about the limitations when the quest window was open, affecting both character and camera movement. Apart from that, two participants did not know how to sit down, and one stated how the game is controlled and that it is not comparable to AAA titles.

The statements coded as *suitability for the task* (ten codes) show a divided image, as six concerns are included. These include the notion that it was unclear what was expected in quests (two codes) and what to answer when speech bubbles were out of sight to provide dialog options (one code). It was also difficult to find out how to fulfill tasks (one code), expressing that more hints are needed (one code). One participant reflected on concentration on task fulfillment as it was difficult for them to maintain it without prior game experience. However, fulfilling the tasks was easier when dialog options were followed (one code). Furthermore, some tasks allowed the freedom to act out (two codes), and fulfilling tasks supported the game flow (one code).

Only three codes in *self-descriptiveness* (ten codes) show that the game is easy to understand (one code) or self-descriptive (one code). Some participants were unsure if it was necessary to click the dialog options (two codes) or follow the character profile (two codes). One participant contradicted the latter. Besides that, the game does not indicate who is speaking (one code). Moreover, two participants could not find out what to do sometimes (one code) or where to find the next quest giver (one code).

In terms of *conformity with user expectations* (eight codes), the participants were satisfied with the separation of the audio stream into different rooms (one code), interactions with game objects (one code), and the collision with other characters (one code). However, information locations (one code) and available dialog continuation were not obvious (two codes). Others

were annoyed by obstacles/invisible walls (one code) and some tasks not being fulfilled as expected (one code).

To illustrate the most important code summaries, example quotes have been put together in Table 4. Furthermore, a focus on the deeper understanding of the worst aspects within *controllability* and the best aspects in *conformity with user expectations* have been added according to the quantitative data [30], which is briefly discussed beforehand and will be discussed in more detail in the following chapter.

Table 4. Example quotes in the categories translated from German to English.

Category	Example Quotes
Suitability for the task	“At first, I found it a bit strange because I wasn't sure what I was supposed to do. Whether I should first look at the quest... or if it's generally a question directed at my avatar, exactly.”
	“And I actually quite liked that in the end, it just said to divert the lesson and do something else. I really appreciated that because there was also a bit of creative freedom there.”
Self-descriptiveness	"At the beginning, I really didn't know if I should, you know, those things there, I didn't know if I should, for example, click on the blue text. Or if I should say it or something. I was a bit confused about that. It just wasn't entirely clear to me what I was supposed to do."
	"I wasn't sure if we were supposed to play a role, like, um, mine said something about a mean teacher. Um, I'm not sure now... um, if we should just do it ourselves. I was never really clear on what was meant. I wasn't sure if I should decide for myself how to react or if I should, um, decide based on the profile of this, um, character."
Controllability	"For me, everything was actually very intuitive and self-explanatory with WASD and all. Yeah, it was somehow NOT a problem at all."
	"Exactly, and those that involve mouse controls. So, my mouse was often not in the center of the screen but somewhere at the edge, and you can't really get it back to the center because then you have to look around to press on that quest or anything at all. You have to first position it right at the edge of the screen and then press there. It was kind of inconvenient."
	"The only issue was, when the camera is fixed, you end up looking at the ceiling or the floor, for instance, and then you have to move around like that."
Conformity with user expectations	“It was quite good that you could do things. For example, opening doors was possible, and you could throw objects.”
	"With the communication, it wasn't entirely clear to me either. Usually, I could click on a blue droplet above the head and then speak, or sometimes there was a speech bubble, and then other times there wasn't."
Suitability for learning	"I think at first I was a bit overwhelmed with 'how do I walk, how do I... until I figured out how to sit down. So, for me, just getting the hang of the controls was a challenge, but I think it's also because I have very little experience with it. As a result, I was focused on other things and not so much on the character I had taken on."
	"The text is provided for you. So, you don't have to think as much as, for instance, in school where you really had to write a role-play about what you'd say. And here, it's much simpler."
	"I thought it was quite good that there was this 'I' (profile short key) to retrieve the information again, so you wouldn't forget things."

4. Discussion

The results of the quantitative study [44] that was used with the qualitative usability results indicated that *controllability* had the strongest room for improvement, followed by *self-descriptiveness* and *suitability for the task*, all rated with "somewhat agree" on the six-point Likert scale. However, the number of codes in the qualitative results indicates that the student teachers talked most about *suitability for learning*, which was the second best in the quantitative results. This category was followed by the lowest-rated constructs from the quantitative results, with *controllability* having almost double the number of codes compared to the other two categories. *Conformity with user expectations* was rated best in the quantitative study and had the fewest codes in the focus groups. Generally, the number of codes in categories from the discussions matches well with the rating of the same constructs in the survey. By contrast, *suitability for learning* holds a strong interest for the target group and could, therefore, be argued to deviate from this matching.

Although 21 codes of this category could be identified, only seven stated limiting factors or problems with learning the game. The second-best rating of 4.50 in the survey can, therefore, be described well with the mixed experiences of student teachers on the length and difficulty of the settling-in period into the game and role-play. Furthermore, there are seven codes on how certain game aspects are supporting these. *Controllability* – the lowest-rated construct in the survey – has the most concerns (thirteen codes) compared to the other categories. Its low rating can be well explained by the inconvenience of mouse control and the limited character as well as camera movement when the quest window was open. With a moderate rating, *suitability for the task* also shows a divided image, with six out of ten codes communicating uncertainty about what to do in quests and how to act in certain situations to fulfill tasks. By contrast, some statements mention tasks supporting freedom of acting and the game flow.

The code distribution, considering both endorsed and problematic aspects of *self-descriptiveness*, aligns with the second-lowest rating. It underscores the numerous uncertainties regarding which dialog options or character profiles should be followed, who is speaking, and where to find the next quest. Being best rated in the survey despite the lowest number of codes, co*nformity with user expectations* includes more critical statements, unobvious dialog continuation, obstacles, and vague statements like not knowing where to obtain information or not fulfilling tasks. For the latter, the context – which specific piece of information and why tasks were not fulfilled – bug or usage problems could not be identified. The codes supporting the good rating disclose satisfaction with the divided audio stream for rooms, object interactions, and collisions.

It needs to be noted that the importance of code frequency can potentially resemble different aspects of why student teachers mention them often. With expectations attached to a usability study, the topics or categories addressed multiple times and by various speakers in a focus group are often due to more problems with certain aspects. Individual interests in specific usability aspects – possibly shaped by prior game experiences – can be another influencing factor. Inherent to focus groups is the possibility that dominant speakers in a focus group could drive code frequency toward a thematic direction.

The results are limited by the short playtime of 45 minutes per group before the focus groups. The intercoder reliability is only in the acceptable range and somewhat lower compared to the intercoder reliability of the attitudes and requirements from the same focus groups, as thoroughly analyzed in [44]. These deviations among coders result in risks to the internal study quality. As with the requirements for game design [44], there is a divergence in the categorization of non-specialist expressions by the speakers to specific game design elements [44], which can also be stated for the usability aspects. One of the reasons might be the different backgrounds and expertise levels considering game usability or development. As QCA prevails in the qualitative analysis of these games (e.g. [2] and [42]), the analysis method selected still

seems adequate despite the improvable intercoder reliability. It could be enhanced by revisiting the analysis results of the initial coding rounds of each coder and starting another iteration after realignment. Although it would have benefited the study quality, it was impossible to obtain coders with comparable expertise in software/game development. Furthermore, properly coding group discussions with complex conversational structures with multiple interjections is challenging [51]. Investigating the long-term effects of InCoLearn on student teachers through a longitudinal study with more play sessions could offer valuable insights into its impact on teaching skills and knowledge application.

C*onformity with user expectations* and *suitability for learning* in the quantitative data were used along with the other usability aspects as a starting point for deeper qualitative analysis. However, their internal consistency is not in the acceptable or excellent range of the other constructs [44][53]. Instead of the standard *ISO 9241-10:1996* [54] of the usability survey being used [49], the standard *ISO 9241-110:2020* [55] with interaction principles instead of dialog principles should serve as a future basis for further usability inquiries. The category system of the qualitative analysis was intentionally developed based on *ISO 9241-10:1996* [54], as this allowed the matching of the quantitative data from the other usability study [44] with this data to gain a deeper understanding of both usability perspectives.

The study results were already placed into the context of the results of a quantitative usability study [44] to gain a thorough understanding of game usability. These results support the game design adjustments for the development of the alpha version of *InCoLearn*. To address the topics covered in the focus groups on *suitability for learning*, the game should streamline the onboarding process, possibly by enhancing quest descriptions and offering more robust tutorial materials that familiarize players with the quest concept and role-play mechanics. This responds to the mixed feedback on the settling-in period and requests for better support in understanding the game's learning aspects. *Controllability* requires attention to improve intuitiveness and ease of use. This includes optimizing mouse control and alternative control schemes that align with AAA game standards. Considering *suitability for the task* and *self-descriptiveness*, quest expectations and requirements should be clarified, possibly with hints to guide task fulfillment. The interface and feedback systems should be re-designed to be more self-explanatory, indicating currently speaking characters and providing clear instructions for the next steps. To address the issues of (*conformity with*) *user expectations,* the visibility of information locations, obstacle placement, and ensuring tasks can be completed as expected without unanticipated barriers. Some solutions to these issues have already been implemented: For example, eleven of thirteen concerns in the qualitative results regarding *controllability,* camera movement via mouse, and the movement and camera freeze system when working with different types of windows in the game – e.g., the quest information window – were entirely reworked and key controls for important actions (sitting-down, profile) are now displayed at all times.

The sample sizes of the study results of serious games or simulations for teacher training (mentioned in 1.1 Background) are – although being mostly quantitative studies – rather low compared to the quantitative study (n = 53) [44] and this qualitative study (n = 46) on usability. Nevertheless, as the sample consists of volunteering student teachers from the University of Stuttgart recruited via convenience sampling, a possible bias and risk for the generalizability and external validity of the study that lies within the sampling strategy [56] is to be expected. The generalizability might, therefore, be limited to the students of University of Stuttgart where *InCoLearn* is planned to be used in the curriculum, e.g., in the seminars the participants were recruited from.

The mixed methods analysis of IVT-T also included applying the constructs from the user satisfaction survey to be used as deductive categories when coding the material of the concurrent think-aloud protocols [2]. The greatest problems users had with usability in *IVT-T* in the concurrent think-aloud protocols are in the category of “Learning” (the application) [2].

Although this category also has some controllability-related codes, this is in accord with *suitability for learning*, having the highest code frequency (21 codes) and *controllability* including the most concerns (thirteen codes). “Other design issues” [2] and “Screen design and layout” [2] have no comparable categories in the usability categories of *InCoLearn*. “Terminology and system information” [2], as well as “System capabilities” [2], however, can be compared with *self-descriptiveness* and *conformity with user expectations* based on the available example observations [2]. In each application, both categories have the lowest code frequencies and problems.

The qualitative study with *SimInClass* also includes some qualitative data on usability. However, the issues addressed in the categories of “technical issues” and “interface” are mostly very specific to that application [39]. Only three aspects can be compared [39]: The freezing problem (three codes) in *SimInClass* was caused by performance problems; the perceived freezing of the screen when the quest or dialog window was opened in *InCoLearn* (five codes) was an intended switch from FPS to window mouse movement. The visibility issue in *SimInClass*, considering the application report (one code) based on overlooking, is similar to the speech bubbles overlooked by players in *InCoLearn*. Clear instructions were provided by SimInClass (one code). In contrast, *InCoLearn* has numerous weaknesses in *suitability for the task, such* as issues with information on how to progress in the game (six codes).

Another investigation was conducted in the following months to measure action-based knowledge of inclusion and heterogeneity in *InCoLearn* based on professional knowledge, and the alpha version was developed accordingly. In a qualitative approach, experts in the field have developed and validated a thorough observation instrument for analyzing the teaching quality and professional behavior in serious games concerning inclusion and heterogeneity. Furthermore, the observation instrument is then used in a qualitative study to consistently evaluate the defined aspects of action-based knowledge through rating the content of recorded play sessions by student teachers playing the trainee character (Johanna Kracht). A fixed pool of actors for the five fellow players in each session and a professional knowledge test of inclusion and heterogeneity complete the basic study design.

5. Conclusions

InCoLearn is the first multiplayer online role-playing game (RPG) newly developed to enhance teacher education students' professional and action-oriented knowledge in dealing with the issues of inclusion and heterogeneity in the classroom with non-linear storytelling through an elaborate quest system. A qualitative usability study resulted in the focus group data of 46 student teachers from the University of Stuttgart. The data was matched with the quantitative data of a usability survey [44]. The results show that *suitability for learning* and *controllability* were by far the most important to student teachers, with around double the number of codes compared to the other categories (RQ1). Additionally, the quantitative research results resemble the qualitative results (RQ2) in such a way that usually, the lowest rated usability constructs from the survey have the highest code count as qualitative categories. Apart from *suitability for learning* in terms of code frequency, the most critical statements in codes were still encountered in *controllability*. This leads to an addition to RQ1 that a higher code frequency represents usability problems rather than importance in specific usability categories. The short playtime, improvable intercoder reliability, and questionable generalizability as well as external validity through convenience sampling limit the study results. This investigation into usability reveals explicit directives for game design enhancements to develop the alpha version. Key focus areas include refining the learning curve, controllability, and task-related feedback mechanisms and ensuring the game's mechanics as well as the interface align with user expectations. Detailed user feedback is a roadmap for iterative design improvements that

could significantly elevate the gaming experience. Apart from improvements to *InCoLearn* itself, the discovered usability aspects can potentially be of value for game designers and developers of serious role-playing games or simulations for, but not limited to, teacher training to streamline their usability requirements during game design. As research in games for vocational teacher education and especially usability research is limited, researchers may use the results to plan qualitative usability research and ultimately draw comparisons. Another planned qualitative study, including behavior analysis of in-game teaching, aims to measure the application of a new observation instrument on teaching quality and professional actionable knowledge in serious games.

Acknowledgments

We thank the German Federal Ministry of Education and Research (Bundesministerium für Bildung und Forschung; BMBF) for sponsoring the project LEBUS² [Teacher education at vocational schools; Lehrerbildung an beruflichen Schulen; funding indicator: 01JA1902] which made the development of and research with this serious game possible.

Conflicts of Interest

The authors have no conflicts of interest to declare.

References

[1] K. Stürmer, B. Marczynski, C. Wecker, M. Siebeck, and S. Ufer, "Praxisnahe Lerngelegenheiten in der Lehrerbildung - Validierung der simulationsbasierten Lernumgebung DiMaL zur Förderung diagnostischer Kompetenzen von angehenden Mathematiklehrpersonen," in *Vielfältig herausgefordert. Forschungs- und Entwicklungsfelder der Lehrerbildung auf dem Prüfstand*, N. Beck, T. Bohl, and S. Meissner, Eds., Tübingen, Germany: Tübingen University Press, 2021, pp. 57–69, doi: 10.15496/publikation-52636.

[2] E. S. Shernoff, K. Von Schalscha, J. L. Gabbard, A. Delmarre, S. Frazier, C. Buche, and C. L. Lisetti, "Evaluating the usability and instructional design quality of Interactive Virtual Training for Teachers (IVT-T)," *Educational Technology Research and Development*, vol. 68, pp. 3235–3262, 2020, doi: 10.1007/s11423-020-09819-9.

[3] R. H. Hämäläinen, M. Niilo-Rämä, T. Lainema, and K. Oksanen, "How to Raise Different Game Collaboration Activities: The Association Between Game Mechanics, Players' Roles and Collaboration Processes," *Simulation and Gaming*, vol. 49, no. 1, pp. 50–71, 2018, doi: 10.1177/1046878117752470.

[4] G. C. B. Winardy and E. Septiana, "Role, play, and games: Comparison between role-playing games and role-play in education," *Social Sciences & Humanities Open*, vol. 8, no. 1, Article No.: 100527, 2023, doi: 10.1016/j.ssaho.2023.100527.

[5] R. H. P. Prager, "Exploring the use of role-playing games in education," *Master of Teaching Research Journal*, no. 2, 2019, pp. 1–8.

[6] L. Zhong, "Investigating Students' Engagement Patterns and Supporting Game Features in a Personalized Computerized Role-Playing Game Environment," *Journal of Educational Computing Research*, vol. 61, no. 3, pp. 578–604, 2023, doi: 10.1177/07356331221125946.

[7] H.-L. Chen and C. T. Wu, "A digital role-playing game for learning: effects on critical thinking and motivation," *Interactive Learning Environments*, vol. 31, no. 5, 2023, pp. 3018–3030, doi: 10.1080/10494820.2021.1916765.

[8] D. Zou, R. Zhang, H. Xie, and F. L. Wang, "Digital game-based learning of information literacy: Effects of gameplay modes on university students' learning performance, motivation, self-efficacy, and flow experiences," *Australasian Journal of Educational Technology*, vol. 37, no. 2, 2021, pp. 152–170, doi: 10.14742/ajet.6682.

[9] L. Hidayati and P. Pardjono, "The implementation of role play in education of preservice vocational teacher," in *Materials Science Engineering*, in IOP Conference Series, vol. 296, 2018, doi: 10.1088/1757-899X/296/1/012016.

[10] J. C. Yang and B. Quadir, "Effects of Prior Knowledge on Learning Performance and Anxiety in an English Learning Online Role-Playing Game," *Educational Technology & Society*, vol. 21, no. 3, 2018, pp. 174–185.

[11] A.-J. Moreno-Guerrero, C. Rodríguez-Jiménez, G. Gómez-García, and M. R. Navas-Parejo, "Educational Innovation in Higher Education: Use of Role Playing and Educational Video in Future Teachers' Training," *Sustainability*, vol. 12, no. 6, 2020, Article No.: 2558, doi:10.3390/su12062558.

[12] K. Gerhard, V. Heuser, J. König, C. Melzer, and K. Kaspar, "Eine quasi-experimentelle Studie zur Förderung des pädagogischen Wissens für inklusiven Unterricht durch ein rollenspielgestütztes Lehr-Lernformat. Empirischer Beitrag zu Grundlagen, Rahmenbedingungen und Herausforderungen," *Herausforderung Lehrer*innenbildung - Zeitschrift zur Konzeption, Gestaltung und Diskussion*, vol. 2, no. 3, 2019, pp. 346–364, doi: 10.4119/hlz-2521.

[13] M. Bearman, C. Palermo, M. Louise, and B. Williams, "Learning Empathy Through Simulation A Systematic Literature Review," *Simulation in Healthcare: The Journal of the Society for Simulation in Healthcare*, vol. 10, no. 5, 2015, pp. 308–319, doi: 10.1097/SIH.0000000000000113.

[14] P. Kilgour, D. Reynaud, M. Northcote, and M. Shields, "Role-Playing as a Tool to Facilitate Learning, Self-Reflection and Social Awareness in Teacher Education," *International Journal of Innovative Interdisciplinary Research*, vol. 2, no. 4, 2015, pp. 9–21.

[15] F. Dahalan, N. Alias, and M. S. N. Shaharom, "Gamification and Game Based Learning for Vocational Education and Training: A Systematic Literature Review," *Education and Information Technologies*, vol. 29, 2023, pp. 1279–1317, doi: 10.1007/s10639-022-11548-w.

[16] P. Wouters and E. S. van der Meulen, "The Role of Learning Styles in Game-Based Learning," *International Journal of Game-Based Learning*, vol.10, no.1, pp. 54–69, 2020, doi: 10.4018/IJGBL.2020010104.

[17] B. Dalgarno, S. Gregory, V. Knox, and T. Reiners, "Practising teaching using virtual classroom role plays," *Australian Journal of Teacher Education (Online)*, vol. 41, no. 1, 2016, pp. 126–154, doi: 10.3316/informit.184882045640447.

[18] C. Knorr and B. Zinn, "Design and Development of a Collaborative Serious Game to Promote Professional Knowledge Acquisition of Prospective Teachers," in *Mobility for Smart Cities and Regional Development - Challenges for Higher Education. ICL 2021. Lecture Notes in Networks and Systems*, vol. 389, M. E. Auer, H. Hortsch, O. Michler, and T. Köhler, Eds., Cham, Germany: Springer, 2022, pp. 890–901, doi: 10.1007/978-3-030-93904-5_87.

[19] J. Othlinghaus-Wulhorst and H. U. Hoppe, "A Technical and Conceptual Framework for Serious Role-Playing Games in the Area of Social Skill Training," *Frontiers of Computer Science.*, vol. 2, Article No. 28, 2020, doi: 10.3389/fcomp.2020.00028.

[20] J. Kunze, S. Mohr, and A. Ittel, "Online-Rollenspiele in der Lehrkräfteausbildung," *Beiträge zur Lehrerinnen- und Lehrerbildung*, vol. 34, no. 2, pp. 188–195, 2016, doi: 10.25656/01:13941.

[21] R. M. Ryan, "Control and information in the intrapersonal sphere: An extension of cognitive evaluation theory," in *Journal of Personality and Social Psychology*, vol. 43, pp. 450-461, 1982, doi: 10.1037/0022-3514.43.3.450.

[22] W. A. IJsselsteijn, Y. A. W. de Kort, and K. Poels, "The Game Experience Questionnaire," Technische Universiteit Eindhoven, 2013.

[23] G. Gris and C. Bengtson, "Assessment Measures in Game-based Learning Research: A Systematic Review," in *International Journal of Serious Games*, vol. 8, no. 1, pp. 3–26, 2021, doi: 10.17083/ijsg.v8i1.383.

[24] H. Desurvire and C. Wiberg, "Game Usability Heuristics (PLAY) for Evaluating and Designing Better Games: The Next Iteration," in *Online Communities and Social Computing*, vol. 5621, A. A. Ozok and P. Zaphiris, Eds., Lecture Notes in Computer Science, Berlin, Heidelberg, Germany: Springer, 2009, pp. 557–566, doi: 10.1007/978-3-642-02774-1_60.

[25] J. L. Plass, B. D. Homer, R. E. Mayer, and C. K. Kinzer, "Theoretical Foundations of Game-Based and Playful Learning," in *Handbook of Game-Based Learning*, J. L. Plass, R. E. Mayer, and B. D. Homer, Eds., Cambridge, MA, USA: MIT Press, 2020, pp. 3–24.

[26] R. Bakhtiari and Z. Habibzadeh, "Designing a framework and validating a tool for evaluating the educational quality of serious games: a meta-synthesis," in *International Journal of Serious Games*, vol. 10, no. 2, pp. 61–83, 2023, doi.org/10.17083/ijsg.v10i2.576.

[27] M. C. Medlock, "An overview of GUR methods," in *Games User Research*, A. Drachen, P. Mirza-Babaei, and L. E. Nacke, Eds., Oxford, UK: Oxford University Press, 2018, pp. 98–114.

[28] V. Zammitto, "Games User Research as part of the development process in the game industry," in *Games User Research*, A. Drachen, P. Mirza-Babaei, and L. E. Nacke, Eds., Oxford, UK: Oxford University Press, 2018, pp. 14–30.

[29] S. Kirginas, "User Experience Evaluation Methods for Games in Serious Contexts," in *Software Engineering for Games in Serious Contexts*, K. M. L. Cooper and A. Bucchiarone, Eds., Cham, Germany: Springer, 2023, pp. 19–42, doi: 10.1007/978-3-031-33338-5_2.

[30] I. Daoudi, "Learning analytics for enhancing the usability of serious games in formal education: A systematic literature review and research agenda," *Education and Information Technologies*, vol. 27, pp. 11237–11266, 2022, doi: 10.1007/s10639-022-11087-4.

[31] P. Chalfoun and J. Dankoff, "Developing actionable biometric insights for production teams," in *Games User Research*, A. Drachen, P. Mirza-Babaei, and L. E. Nacke, Eds., Oxford, UK: Oxford University Press, 2018, pp. 300–321.

[32] K. Moser, K. Chambless, and N. Miller, "Creating a Virtual Space in World Language Teacher Preparation? Considerations for Practice-Based Rehearsals," *Journal of Technology and Teacher Education*, vol. 28, no. 4, pp. 613–638, 2020. [Online]. Available: https://www.learntechlib.org/primary/p/216918/

[33] S. Gregory et al., "Changing directions through VirtualPREX: engaging preservice teachers in virtual professional experience," in *Proceedings ascilite 2011 changing demands, changing directions*, 2011, pp. 491–501.

[34] J.-L. Lugrin, M. E. Latoschik, M. Habel, D. Roth, C. Seufert, and S. Grafe, "Breaking Bad Behaviors: A New Tool for Learning Classroom Management using Virtual Reality," *Frontiers in ICT*, vol. 3, p. 26, 2016, doi: 10.3389/fict.2016.00026.

[35] A. P. Delamarre et al., "An Interactive Virtual Training (IVT) Simulation for Early Career Teachers to Practice in 3D Classrooms with Student Avatars," in *Proceedings of the Thirtieth International Florida Artificial Intelligence Research Society Conference*, 2017, pp. 400–403.

[36] M. Zibit and D. Gibson, "simSchool: The game of teaching," *Innovate: Journal of Online Education*, vol. 1, no. 6, Article No.: 4, 2005. [Online]. Available: http://nsuworks.nova.edu/innovate/vol1/iss6/4.

[37] L. Dieker, J. Rodriguez, B. Lignugaris-Kraft, M. Hynes, and C. Hughes, "The Potential of Simulated Environments in Teacher Education: Current and Future Possibilities," *Teacher Education and Special Education*, vol. 37, no. 1, pp. 21–33, 2014, doi: 10.1177/0888406413512683.

[38] A. T. Hayes et al., "Ludic Learning: Exploration of TLE TeachLivE™ and Effective Teacher Training," *International Journal of Gaming and Computer-Mediated Simulations*, vol. 5, no. 2, pp. 20–33, 2013, doi: 10.4018/jgcms.2013040102.

[39] Ö. Kelleci and N. C. Aksoy, "Using game-based virtual classroom simulation in teacher training: User experience research," *Simulation & Gaming*, vol. 52, no. 2, pp. 204–225, 2021, doi: 10.1177/10468781209621.

[40] F. Badiee and D. Kaufman, "Design evaluation of a simulation for teacher education," *Sage Open*, vol. 5, no. 2, 2015, doi: 10.1177/21582440155924.

[41] M. E. Latoschik et al., "Breaking bad behavior: Immersive training of class room management," in *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology*, 2016, pp. 317–318, doi: 10.1145/2993369.2996308.

[42] J. N. Mikeska, H. Howell, and D. Kinsey, "Examining the usability and viability of using a simulated classroom environment to prepare preservice science teachers during and after the COVID-19 pandemic," *Disciplinary and Interdisciplinary Science Education Research*, vol. 4, no. 1, pp. 1–20, 2022, doi: 10.1186/s43031-022-00054-1.

[43] M. D. Dickey, "11 Narrative in Game-Based Learning," in in *Handbook of Game-Based Learning*, J. L. Plass, R. E. Mayer, and B. D. Homer, Eds., Cambridge, MA, USA: MIT Press, 2020, pp. 283–306.

[44] C. Knorr, "Eine explorative Untersuchung zu einem digitalen Rollenspiel in der beruflichen Lehrpersonenbildung," *Zeitschrift für Berufs-und Wirtschaftspädagogik, vol.* 119, no. 1, pp. 136–171, 2023, doi: 10.25162/zbw-2023-0006.

[45] U. Kuckartz, *Mixed Methods*, Wiesbaden, Germany: Springer, 2014, doi: 10.1007/978-3-531-93267-5_3.

[46] S. Rädiker and U. Kuckartz, *Analyse qualitativer Daten mit MAXQDA*, Wiesbaden, Germany: Springer, 2019, doi: 10.1007/978-3-658-22095-2.

[47] C. Knorr, "Lerntheorien im Game Design digitaler Lernspiele," in *Virtual Reality, Augmented Reality und Serious Games als Educational Technologies in der beruflichen Bildung*, B. Zinn, Ed., Stuttgart, Germany: Franz Steiner Verlag, 2023, pp. 95–116, doi: 10.25162/9783515135481.

[48] M. L. Ryan, *Narrative as Virtual Reality: Immersion and Interactivity*, Baltimore, MD, USA: Johns Hopkins University Press, 2001.

[49] Q. Guo, B. Ditton, and B. Zinn, "Eine Anwendung der Virtuellen Realität in der beruflichen Bildung im Kontext physikalischer Lerninhalte," in *Proceedings of DELFI Workshops 201*9, S. Schulz, Ed., Bonn, Deutschland: Gesellschaft für Informatik e.V., 2019, pp. 173–185, doi: 10.18420/delfi2019-ws-119.

[50] U. Kuckartz, *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*, 4th ed., Basel, Germany: Beltz Juventa, 2018.

[51] N. T. Ochieng, K. Wilson, C. J. Derrick, and N. Mukherjee, "The use of focus group discussion methodology: Insights from two decades of application in conservation," *Methods in Ecology and Evolution*, vol. 9, pp. 20–32, 2018, doi: 10.1111/2041-210X.12860.

[52] J. R. Landis and G. G. Koch, "The Measurement of Observer Agreement for Categorical Data," *Biometrics*, vol. 33, no. 1, pp. 159–174, 1977, doi: 10.2307/2529310.

[53] D. Krebs and N. Menold, "Gütekriterien quantitativer Sozialforschung," in *Handbuch Methoden der empirischen Sozialforschung*, N. Baur and J. Blasius, Eds., Wiesbaden, Germany: Springer VS, 2019, pp. 489–504, doi: 10.1007/978-3-658-37985-8_35.

[54] *Ergonomische Anforderungen für Bürotätigkeiten mit Bildschirmgeräten - Teil 10: Grundsätze der Dialoggestaltung*, DIN EN ISO Standard Nr. 9241-10:1996-07, Deutsches Institut für Normung, July 1996. [Online]. Available: https://www.beuth.de/de/norm/din-en-iso-9241-10/2810291.

[55] *Ergonomie der Mensch-System-Interaktion - Teil 110: Interaktionsprinzipien*, DIN EN ISO Standard Nr. 9241-110:2020, Deutsches Institut für Normung, October 2020. [Online]. Available: https://www.beuth.de/de/norm/din-en-iso-9241-110/320862700.

[56] C. Andrade, "The Inconvenient Truth About Convenience and Purposive Samples," *Indian Journal of Psychological Medicine*, vol. 43, no. 1, pp. 86–88, 2021, doi: 10.1177/0253717620977000.