Doctoral Dissertation
Doctoral Program in Computer and Control Enginering (33.th cycle)

Augmented reality and serious games for learning: exploring potentialities, assessing effectiveness, and investigating user experience.

Francesco Strada

* * * * * *

Supervisor
Prof. Andrea Bottino, Supervisor

Doctoral Examination Committee:
Prof. Laura Anna Ripamonti, Referee, Università degli studi di Milano
Prof. Zhijie Xu, Referee, University of Huddersfield
Prof. Ivan Martinez-Ortiz, Complutense University of Madrid
Prof. Ronald Poppe, University of Utrecht
Prof. Maurizio Morisio, Politecnico di Torino

Politecnico di Torino
November 8, 2021

I hereby declare that, the contents and organisation of this dissertation constitute my own original work and does not compromise in any way the rights of third parties, including those relating to the security of personal data.

. . . .

Francesco Strada
Turin, November 8, 2021

Summary

The evolution of digital systems has shaped the way we engage with information, stimulating the proposal of new learning methodologies. Among them, thanks to the growth of the multi-media PC market, the idea of learning through entertaining digital applications (i.e., edutainment) rose to popularity through the 1990s. Then, at the turn of the century, powered by the ever-growing video-game industry, the concept of edutainment evolved into what they are now called serious games (SG), which are games designed with a purpose other than pure entertainment. Although these "serious" purposes can be varied (e.g., physical exercise and rehabilitation, and marketing), they undoubtedly include learning.

Through the years, SGs have leveraged a vast set of technological platforms such as computers, consoles, mobile devices, and, more recently, augmented reality (AR). The interest in this technology has several explanations. First, AR has shown positive evidence as an ideal tool for learning purposes. Second, the development of AR-based applications has become easier and more accessible. Third, consumer-level mobile devices can now support AR applications. Finally, both SG and AR educational applications can create motivating and engaging educational activities in which users can explore the proposed contents through an interactive and immersive digital experience that supports experiential approaches. Although the number of developed learning experiences combing AR and SGs (AR-SGs) is growing, this research field is still at an early stage and rapidly evolving as AR technology continues to improve.

Based on these premises, this Ph.D. thesis aims to advance the current state of the art in AR-SGs research exploring new potentialities, addressing possible issues, and evaluating the outcomes of combining these two tools/approaches in an educational context. More formally, this work wishes to provide empirical evidence to answer the following research questions:

- RQ1: Can we establish synergies between AR and SGs in order to develop effective AR-SGs in a variety of learning contexts?
- RQ2: Are AR-SGs effective?
- RQ3: Usability can be an issue in AR-SGs, can it be softened?

This work adopted a case study approach through the design, development, and assessment of four AR-SGs. Each of these research "instruments" provided evidence in support of all the RQs, allowing their analysis under slightly different lenses.

To answer RQ1, I have selected a subset of learning domains in which either AR or SGs had already been primarily exploited as "individual" but not widely in combination with one another. They are procedure learning, complex systems learning, and soft-skills learning. For each context, I developed one or more applications leveraging synergies between AR and SGs, where for synergies, I mean the creation of an effective "relationship" between these two approaches/tools, where the affordances of either one would be exploited to enhance the other. Thanks to the established synergies, the proposed solutions immersed and engaged the users in the learning experience, motivated them to play (and learn), and fostered several group dynamics beneficial to the overarching learning objective. Although these outcomes have been achieved in specific topics, many of the design choices and mechanics proposed can be transferred, with similar positive results, to other topics within the same learning domain.

Through RQ2, I wished to explore the effectiveness of the proposed AR-SGs. In other words, could they achieve the purpose for which they were designed? In the context of AR-SGs, these purposes can be divided into an entertaining and educational one. The latter can then target different outcomes according to the specific learning domain addressed by the AR-SGs. These outcomes can be broad, ranging from the acquisition of concepts to the development of 21st-century skills. Finally, an essential aspect of AR-SGs (and SGs in general) is the relation that must be established (and assessed) between entertaining and educational purposes as these two elements must promote one another, i.e., the entertaining experience elicits the educational one and vice versa. To explore the effectiveness of the proposed AR-SGs I adopted an evaluation procedure that featured the collection of a comprehensive set of data, both subjective (i.e., questionnaires) and objective (i.e., application logs and annotated gameplay recordings). By analyzing these data separately and by exploring possible relations, I could conclude that the proposed AR-SGs were fun and engaging, they achieved their pedagogical goal, and these two outcomes were positively interconnected as they both influenced each other. Finally, similar conclusions achieved in the design of our applications can be made for the evaluation methodologies. I believe that several experimental design choices and employed instruments could be exploited and adapted by future practitioners who aim to assess outcomes similar to those I propose.

The last research question originated from evidence in the literature reporting compromised learning outcomes caused by inadequate usability levels or poorly designed UX in AR learning applications. I approached this problem by developing the four AR-SGs through a careful process that placed users' needs at the cornerstone of my design choices. The outcomes of this process were tested through

usability studies conducted for each AR-SGs. Users' evaluations were highly positive, even when targeting users with no or limited prior knowledge of AR. As an output, we propose several guidelines that future practitioners could follow when embarking on the delicate task of developing AR-SGs where usability issues do not hinder the overarching entertaining and learning experience.

Concluding, this thesis provides empirical evidence that from the combination of AR and SGs, new and unique learning experiences can be established.

Acknowledgements

Now that I am (almost) at the end of my PhD, I can finally take a moment to relieve these years and reflect on all the people with whom I have shared this journey. I would like to take this opportunity to thank all those who have pushed me in many ways to achieve this result.

First and foremost, my thanks go to my supervisor, Prof. Andrea Bottino, who always trusted me and paved a solid path for me to follow, pushing me to do things I could hardly have imagined on my own. Most of all, I thank him for his kindness and his ability to always find the right words in any situation. You have been such a great inspiration.

A special thanks to those who have been directly involved in the many projects we have worked with, and whose contributions can also be found in this work. To begin (in strict chronological order). I would like to thank the GRAINS research group, particularly Prof. Fabrizio Lamberti and Gabriele Pratticò, the SIMNOVA Centre, in particular Prof. Pier Luigi Ingrassia, and finally the team of the Huddersfield Centre for Visual and Immersive Computing, Prof. Carlo Fabricatore, Maria Ximena Lopez and Dimitar Gyaurov.

A special thanks go to Alysson, whose meeting I would describe as an epiphany, as he gave me the (unexpected) opportunity to delve into the incredible and fascinating world of serious games. I thank him not only for the numerous projects we have worked on together but more importantly for being that “big brother” I hope every PhD student crosses paths with. I will never forget your incredible caipirinhas.

Also, I would like to thank all the past and present members of Lab1, but also apologize for the immense annoyance I caused them with all the user testing I did in their workspace. In particular, I would like to thank Edoardo for being a precious addition to the CGVG research group. I hope I have also been a good “big brother” to you and Riccardo for the profound and pointless conversations we had during our countless cigarette breaks.

Last but not least, I would like to thank my family for their constant support regardless of the distance; all my friends here in Turin for making this city a second home for me and filling my days with joy and lightness; and all my friends in Bologna who have always enriched my visits back home and for their unforgettable

surprises.

Finally, I thank Martina for her patience in dealing with me on a daily basis, despite my emotional swings. But most of all, I thank her for showing me large fields of red poppies, “a les belles choses, aux beaux jours”. Lastly, I thank my dogs, Kanga and Paolo, for always lying under my desk while writing this thesis.

To nonna Aurora, the
first person to teach me
how to play.

Contents

List of Tables

List of Figures

Acronyms

AED automated external defibrillator.

AR Augmented Reality.

AR-SG Augmented Reality Serious Game.

BIM Building Information Model.

BLS basic life support.

BLSD Basic Life Support and Defibrillation.

CCC continuous chest compression.

CPR cardiopulmonary resuscitation.

CSCL computer-supported collaborative learning.

CSCW Computer Supported Cooperative Work.

EMS emergency medical service.

ESD education for sustainable development.

FOV field of view.

GA Game Analytics.

GF game feature.

HARUS Handheld Augmented Reality Usability Scale.

HHD Hand-Held Device.

HMD Head-Mounted Display.

IMMS Instructional Materials Motivation Survey.

IPO *input-process-output*.

K-12 kindergarten through twelfth grade.

LRS local reference system.

N2P negative to positive.

NPC non playing character.

P2N positive to negative.

PPQ Pre and Post-test questionnaire.

PPT pre-post test.

RQ Research Question.

SCA sudden cardiac arrest.

SG Serious Game.

SPOI Shared Point of Interest.

SRS shared reference system.

SSG sustainability serious game.

STEM Science, Technology, Engineering, and Mathematics.

SUS System Usability Scale.

TEL Technology-Enhanced Learning.

UI User Interface.

UX User eXperience.

VA Video Analysis.

VC visual cue.

Chapter 1

Introduction

Digital technologies have completely changed our lives. They facilitate our communications and provide easy access to an impressive amount of information, primarily through computers and mobile devices. However, this information is tied to the bi-dimensional nature of these tools. They are windows (i.e., screens) through which we can observe and interact with digital contents based on well-established metaphors (e.g., using a mouse cursor or a touch interface to access and activate the elements displayed on screen). Although we have become familiar with these forms of visualization and interaction, they have two major limitations. First, three-dimensional data (e.g., consumer products, architectural projects, or characters in a video game) are displayed on a flat surface, losing entirely the real sense of depth and the ability to truly understand their physical properties (i.e., how big they are, their spatial relationship with other objects, or how they would fit into the real world). Second, these devices force us to interact with the objects displayed on the screen through specific equipment (e.g., mouses, keyboards, touchscreens) and the interaction metaphors they implement. This constraint requires us to learn how to use the devices and understand the metaphors, rather than interacting naturally with the digital objects (e.g., by reaching out to grasp them as we would a real object).

In 1992 science fiction author Neal Stephenson coined, in his novel Snow Crash [289], the term and concept *Metaverse*, which represents a universe that goes beyond our physical world where digital contents are seamlessly blended on top of the real environment. Here users can interact with data using the same metaphors they use with tangible objects. The *Metaverse* does not exist yet, but one of the key technological enablers is headed in this direction. We[1] are referring to Augmented Reality (AR), which according to researcher Ron Azuma [17] is a technology that (i)

---

[1] I wish to acknowledge that from this moment on, the term “we” will be used instead of “I”, since all the contents which will be addressed have benefit from a collaborative effort of both the research group I am affiliated with and external collaborators.

combines real and virtual content, (ii) is interactive in real-time, and (iii) accurately registers real and virtual objects in 3D.

In 1968, long before Stephenson's novel or Azuma's definition, in a research lab at Harvard University Ivan Sutherland and Bob Sproull developed the first see-through Head-Mounted Display (HMD) capable of creating 3D graphics overlaid on the real world [293]. For the following 30 years, researchers invested their efforts to solve three crucial technological problems which are vital to enable effective AR, namely: tracking [15], display technologies [262], and input devices [109]. However, these early solutions were costly to develop, complex to deploy, and cumbersome to use. The real breakthrough came only at the turn of the twentieth century when the development of computer vision-based tracking solutions first [250], and its deployment on Hand-Held Devices (HHDs) later [315], initiated a strand of mobile-based AR research and applications. Furthermore, the smartphone era that started with the iPhone release in 2007 transformed how people use portable technology. This was the first touch screen device to offer a complete, non-watered down version of the Internet, giving consumers the ability to browse the web just as they would do on a desktop computer. And the technology rapidly evolved from that point, paving the way for a new ecosystem of powerful devices that could efficiently perform complex computational tasks by leveraging a large amount of memory and communication bandwidth, which allowed them to process an amount of multimedia data that was completely unthinkable just a few years before. Recent devices offer, beyond many other features, real-time vision-based tracking, image recognition and the possibility to render complex 3D graphics. These technological enablers have made AR (virtually) available in the pockets of millions of users and have set the stage for further developments. The first Microsoft Hololens, released in 2016, and the mobile markerless tracking SDKs (i.e., ARKit and ARCore) available after 2017, were the last giant leap forward that pushed AR in the business and consumer sector. Thanks to these developments, in 2019, AR was removed from the Gartner Hype cycle, thus moving this technology from the emerging to the mature status [119]. Then, as more VR headsets are equipped with embedded cameras for inside-out tracking, we are seeing an increasing presence of VR HMDs with AR pass-through capabilities. Unlike see-through technology (e.g., the Hololens), where virtual images are projected onto transparent lenses and users perceive the real world with their naked eyes. In contrast, pass-through technology displays both virtual and real-world content through the HMDs' internal screens. The built-in cameras capture the outside world, and the digital content is overlaid on top of the camera's video image. However, these devices have some limitations as the quality of the real world seen on the displays is directly dependent on the quality of the cameras and the pixel density of the displays. Therefore, when this experience is embedded in consumer-level technology such as the Oculus Quest 2, the overall quality of AR is not comparable to that of HHDs and other see-through devices (e.g., the Hololens). However, the recently unveiled Varjo XR-3 device, aimed at the enterprise market (priced at $5495), features breakthrough cameras and display technologies that provide users with a truly immersive visual experience. The real world can be displayed in great detail, and furthermore, complex 3D content can be rendered via a workstation. All of this can be experienced with a field of view (FOV) of 110°, which is typical of VR HMDs and was a major limitation of AR see-through HMDs (e.g., the Hololens 2 has a FOV of 43°). Despite this high cost, we can expect prices for pass-through devices to drop when AR becomes more widely available, similar to what happened with VR HMDs.

Despite these technological developments and recognition as a consumer-ready technology, AR can still be considered somewhat in its infancy. In fact, it is only after the first big technological leap (i.e., the availability of AR on mobile devices) that the development and research in the area of AR applications in the most diverse fields actually took off. Since the main technological challenges had been partially (but effectively) solved, practitioners could start approaching AR-related research at a higher level, focusing their attention on application design and development rather than on the development of the enabling technologies.

Among the many areas that have expressed strong interest in new AR technologies is Technology-Enhanced Learning (TEL) environments. Put simply, TEL refers to the use of technology to maximize the student learning experience. Teachers and students within a TEL context leverage digital tools, multimedia, interactive content, and engaging applications (which may also integrate computing and artificial intelligence modules) to achieve their teaching and learning goals. An effective TEL environment takes advantage of the technology to connect people to information and other participants in the educational process. In addition, digital technologies allow access to content anytime and anywhere, and provide the ability to share and edit information easily. The physical environment in which the training phase takes place can have different structures that, in addition to integrating different digital tools, allow for flexible grouping of students and the development of face-to-face, hybrid, or fully online educational approaches.

Based on these characteristics, researchers immediately recognized AR as an ideal candidate to further improve TEL environments and approaches. Therefore, starting with the early 2010's we have witnessed an exponential growth in the literature of works experimenting with the adoption of AR as a learning tool [18, 4]. Through the years, practitioners have applied AR to a variety of educational contexts, encompassing eHealth [50], science [199], engineering [145] and arts [80], targeting learners across all educational levels, in both formal and informal contexts. Compared to traditional learning methods or other TEL approaches, AR has a unique set of benefits and affordances which make it particularly suited for learning purposes [246]. Firstly, its unique visualization properties, which seamlessly blend virtual contents on top of the physical world, have shown numerous benefits in improving users' understanding of complex phenomena [179, 185, 58].

Also, thanks to the ability to freely explore digital contents from different perspectives (since users control the point of view over digital contents), AR has shown to help develop spatial abilities [246, 134, 72]. Furthermore, by connecting the digital information to the physical world (and the objects it includes), AR can provide an intuitive way of delivering contextual and location-specific information to the user [5]. Therefore, information can be delivered at the right location and the right time (i.e., when and where the user needs it, based on the current task or location), helping learners navigate through a large amount of learning content more easily. Then, in AR, digital contents can be updated in real-time (also) according to the user's interactions, who by seeing direct consequences of their actions are instilled with a sense of self-efficacy [282] which has proven to enhance users' learning performances [20]. These potentialities in improving learning have also been attributed to the overall motivation and engagement fostered by AR technology itself, as users feel immersed in the digital experience making the learning process more transparent reaching more easily a state of "flow" (i.e., users' state of optimal experience, [55]). Finally, a widely recognized affordance of AR learning tools in comparison to traditional approaches is the ability to foster collaborative social interactions [246]. Users can easily engage in face-to-face interactions, and as a consequence, they are more stimulated towards dialogue and collaborative problem-solving. Whereby sharing different perspectives on the same topic strengthens the learning process and stimulates the development of 21st-century skills (e.g., problem-solving, communication, and negotiation abilities) [111].

Among AR-based learning benefits, there is a degree of overlap with another promising learning approach, i.e., that of *Serious Games* (SGs), which are games that do not have entertainment as their primary purpose [86, 194, 88, 206]. When SGs have an educational purpose, they seek to achieve their educational goals through a playful experience that motivates players to actively engage in the learning process. This approach to learning stems from the idea that play and learning experiences are closely related [124, 122]. Indeed, it is assumed that the enjoyment of a game lies in the process of learning how to play the game itself, i.e., "the fundamental motivation for all game-playing is to learn" [70]. Veteran game designer Raph Koster argues that the enjoyment of games, which is also associated with the chemical release of dopamine in the brain, results from the broader learning process associated with solving game puzzles and unraveling the intricate patterns featured in games [178]. Koster states that games are fun because they appeal to our innate instinct to acquire new skills (both physical and mental) and to perceive our own progress. Because play has a powerful influence on learning [255] and can be seen as a learning mechanism common to all human cultures [304], SGs aim to harness this use of play as a teaching strategy [176] by, among other things, taking advantage of the motivational drive fostered by play [203, 6]. Furthermore, when SGs are experienced in digital form, players can take advantage of the high degree of interactivity, as SGs provide players with a sense of self-efficacy because they experience that their actions are effective in the virtual game world (i.e., by interacting with the game, they have the ability to change the virtual environment). This sense of control thus triggers a pleasurable feeling that motivates further interactions and reinforces engagement with the learning content [42]. Then, thanks to traditional game features/mechanics such as narratives, quests, and characters, players (learners) are also engaged on a deep emotional level. This emotional engagement can stimulate learning and allow users to participate in cognitive and affective processes that can influence behavior change and promote positive attitudes [33]. Furthermore, evidence [138, 313, 232] highlights that SGs are ideal for creating situated and socially mediated learning contexts by enabling shared experiences (e.g., by providing multiplayer settings or allowing learners to share information and results through the social networks). SGs can be designed to foster collaborative discussions and reflections, which in turn support critical thinking, deepen understanding, and expand learners' shared knowledge [292, 286].

From these observations, it can be seen that AR-based and SG-based learning approaches have several elements of similarity. First, they both strive to establish an educational context in which the learner is encouraged to interact with the learning contents directly, promoting a process where they construct knowledge through their own experience (and actions) in the digital environment. Second, both approaches aim to motivate the user in the learning process by engaging them in a visually and emotionally captivating experience. Thanks to AR, users are enticed to explore and understand the learning contents by observing and analyzing them in a unique way. Similarly, SGs, establish complex and attractive fictional contexts where the learning process happens invisibly as the user playfully interacts and explores the virtual environment. Thanks to all these elements both AR and SGs, immerse the user in the digital experience, soliciting a state of deep concentration and optimal experience (i.e., the "flow" state). Then, due to these characteristics, AR and SGs can both sustain situated learning activities (i.e., learning that takes place in the same context in which it is applied) of complex real-world problems, difficult to experience otherwise. Finally, both AR and SGs rely on collaborative social interactions to make learning more meaningful and pleasant. They both show affordances particularly suited to establish this behaviour. AR facilitates face-to-face interaction, whereas SGs exploit collaborative game mechanics (e.g., roles, shared resources and objectives) to promote interaction and collaborative problem-solving. Therefore, the possibility to combine AR features into SG experiences into *Augmented Reality Serious Games* (AR-SGs) is rapidly gaining momentum as researchers are increasingly proposing this new approach to a variety of learning contexts [188]. In an attempt to map the current state of the art of AR-SGs, different reviews have been proposed [188, 300, 180, 235, 184]. From these reviews, it emerges that, despite the growing attention addressed to AR-SGs, this field is still in its infancy, as suggested by the numerous formative evaluations, which mainly describe early prototypes rather than more detailed experimental assessments [184].

Despite this, works that gathered experimental evidence have shown that (as hypothesized) AR-SGs are indeed able to inherit most of the benefits that separately characterize AR and SGs [272]. At the same time, however, the greater complexity involved arising from their combination also raises questions that have not yet been fully addressed.

First, AR-SGs have been shown to promote motivation and engagement. However, it is unclear if this stems from single affordances of AR and SG elements or from the combination of affordances from both fields [234]. Then, in AR educational research, many authors warn practitioners that the reported engagement might merely result from the (stunning) novel technology that attracts users' attention [18]. Nevertheless, this novelty factor is inevitably going to diminish as users become more acquainted with AR. Thus, one possible question is: will AR still be able to stimulate user motivation in AR-experts? Is it possible that the SG component of AR-SGs will be able to make up for the disappearance of the "wow effect" of AR, and if so, how? These questions still have no clear answer. Moreover, although SGs are inherently games designed to solicit fun and engagement (which are drivers for motivation), the "fun" dimension of AR-SGs seems to be underexplored in the literature, as this element is almost never reported in reviews that address research on AR-SGs.

Second, despite AR-SGs have been reported to promote collaborative social interactions, the number of works addressing collaboration is still limited [300]. As a consequence, the effectiveness of AR-SGs in this regards remain unclear in some aspects. Most importantly, researchers urge to understand better (or validate) if and how collaboration, in AR-SGs, contributes to the improved learning performances and which elements of AR-SGs enhance (or inhibit) collaborative interactions [300, 188]. Also, of the solutions which have addressed collaboration, the majority seem to focus on experiences where collaboration is happening more frequently outside the digital environment rather than inside (i.e., users sharing the same augmented digital environment). This limitation is mainly due to the complex technical features required to deploy shared AR experiences. Virtual contents need to be aligned between different devices (which requires the problem of registering the local reference system of each device within a common global reference system). The state of the shared application (which evolves according to both internal rules and users' interactions with digital contents) must be synchronized between participants. Then, shared AR requires a solid networked infrastructure to allow exchanging information between devices (e.g., who has interacted with which object or in which position of the digital simulation is each user located). The development of an AR-SGs with these characteristics can be challenging, costly, and time-consuming [94]. However, the latest advances in AR technology and development tools have established a favourable context for developing applications with such advanced characteristics, paving the way to novel research aimed at better understanding the benefits of collaborative AR-SGs.

Moreover, assessing these elements, given their inherent complexity, also calls for sound evaluation methodologies. In particular, current mixed-methods studies can benefit from the sensors embedded in recent AR devices, which are capable of capturing what is happening in the physical world, what the users are doing in it and how they are interacting with digital content. Thus, it is possible to integrate for such an analysis in-game data (i.e., gameplay interactions and viewpoint over the virtual environment), users' physical movements, verbal and non-verbal communication and many other data, providing more objective evidence for the verification of novel hypotheses under test.

Finally, AR-SGs have also inherited limitations that currently affect AR learning systems. Among them, there is usability [300]. Although the features of AR certainly have positive effects on learners, the fact that this technology is relatively new creates barriers for the need of interacting with a system that users are not experts in. The difficulties of learning and managing these new modalities and metaphors of interaction implies an additional cognitive load for the learner that may even come to hinder the overall learning activity [246]. Indeed, designers of AR-SGs need to maximize usability so that learners are not forced to put more effort into figuring out how to do things rather than focusing on the learning content. By correctly approaching design and placing the user at the center of the process, these difficulties can be overcome, eventually achieving the overall goals of AR-SGs without setbacks. However, designing highly usable applications can be time-consuming, as it requires numerous iteration cycles to fine-tune the application experience to meet user needs as best as possible. We believe that the early stages of research on AR-SGs partly overlooked this fundamental aspect, and further attention and research focused on this goal is needed.

Another AR-SGs limitation stems from a similar problem affecting the AR learning field, i.e., the clear imbalance of educational topics addressed through these applications. The literature reports an evident prevalence of approaches in Science, Technology, Engineering, and Mathematics (i.e., STEM disciplines) [300, 188, 18]. This issue derives primarily from the many possibilities that AR has in describing scientific concepts. AR makes it possible to visualize unobservable objects or phenomena (e.g., molecular structures, organs, and the orbit of planets) [3]. Thus, through complex and interactive 3D representations, students can more easily grasp these concepts that would otherwise require their imagination or be poorly represented in traditional media such as books and videos. In contrast, SGs have a long history of successful applications in many domains beyond STEM disciplines. More importantly, SGs have also been applied in domains that are not strictly focusing on hard-skill acquisition. For example, SGs have been successfully applied in the process of extracting and structuring knowledge from a source [301, 19], to foster behavioral and attitudinal changes [141, 156, 301] and to learn more intangible skills such as communication skills, negotiation, and critical thinking [191, 133]. Given the potential offered by SGs in addressing these areas, we believe that the synergies between AR and SG can offer designers, educators, and practitioners novel and exciting possibilities for developing effective, engaging, and (why not?) fun educational paths in a large variety of disciplines.

1.1 Methodology

The main goal of this thesis is to gain better insights into the world of AR-SGs as we firmly believe that they can become one of the most effective learning platforms/approaches in many educational contexts, both formal and informal. Therefore, we aim to address some of the main problems currently faced by AR-SGs by analyzing experimental data collected from a selection of applications designed and developed from scratch for this research. The following section will introduce the Research Questions (RQs) that guided the choices made during this Ph.D. program. To better inform the reader on how we addressed these RQs, we will briefly introduce the AR-SGs we designed and developed, as they are the foundational "tools" employed in this research. These tools served as demonstrators of different approaches to solving the problems identified. From this point of view, it should be emphasized that the challenges and complexities to be addressed are closely connected and correlated (e.g., usability clearly affects learning effectiveness) and, as such, are difficult to address as individual problems. For this reason, the different AR-SGs we developed were envisioned as case studies that can provide evidences on a variety of open problems, each of which can be addressed differently in different demonstrators. These developed tools, which will be extensively detailed and discussed in subsequent chapters, are the following.

Holo-BLSD is an AR-SG that provides a self-instruction and self-assessment environment for Basic Life Support and Defibrillation (BLSD). Holo-BLSD approaches the general problem of procedure learning, i.e., learning those educational, working and organizational activities that are defined as sequences of operations or actions performed according to a specific sequence and certain rules. Therefore, its results are easily transferable to any domain where procedure learning is concerned.

Sustain is a collaborative multiplayer AR-SG aimed at promoting sustainability awareness and commitment. Sustain addresses the general problem of learning about complex systems, i.e., those requiring the ability to approach them from multiple and sometimes competing perspectives and may have multiple possible solutions. Complex systems are characterized by dynamics that cannot be predicted simply by examining the isolated behaviors of their individual parts, thus challenging educational methods based on direct instruction and requiring constructivist perspectives in addressing them. In particular, dealing with complex systems requires involving students in collaborative and cooperative activities, which encourage discussion and reflection, helping to generate deeper insight and understanding. To this end, one of the main objectives in the design of Sustain was to develop a game environment capable of fostering and supporting collaborative learning and problem-solving among co-located players.

Asteroid Escape and **ARScape** are two AR-SGs that focus on team activities to improve learners' soft skills (in this particular case, communication and collaboration skills). The two works analyze different facets of the problem. In particular, ARScape focuses on analyzing means to improve awareness in co-located collaborative environments, an element that is conducive to supporting effective collaboration among participants.

1.2 Research Questions

Now that we have introduced the "instruments" used to address open problems in AR-SGs, let us more formally present the top-level RQs we wish to answer throughout this thesis. For greater clarity, we will also underline which (parts) of the above AR-SGs have been used to answer the following RQs.

RQ1: *Can we establish synergies between AR and SGs in order to develop effective AR-SGs in a variety of learning contexts?*

As we have introduced in this section, the breadth of domains AR-SGs are currently employed is mainly limited to STEM knowledge acquisition. Nevertheless, thanks to their relatively long line of research, SGs have been successfully applied to a far more comprehensive range of domains with different topics and different "learning" requirements. SGs have been used to promote behavioral and attitudinal changes in socially delicate topics (e.g., sustainability, gender inequality, and race discrimination) and have been used to promote 21-century skills. On the other hand, specific learning contexts have largely exploited AR technologies' affordances with little attention to SGs. Among these, we can identify all those learning activities where individuals (or groups of individuals) are required to master a set of actions that must be performed in sequence (e.g., industrial maintenance, assembly, and medical operations). In most cases, these activities involve actions that must be performed in the physical world. Hence the AR affordances make this technology particularly suited to support learning in such contexts. For instance, the possibility to contextualize digital (learning) content according to the physical context where the learning happens helps strengthen the knowledge acquisition process as learners can connect the theory to the practice more efficiently.

In conclusion, there are learning domains in which AR or SGs have been successfully adopted but only individually, and examples combining their strengths are rare or nil. Our hypothesis is that the unique characteristics of AR and SGs can be leveraged in many of these domains to establish a more synergistic relationship in which possible problems or needs of either approach can be met by the other.

RQ2: *Are AR-SGs effective?*

Before analyzing this RQ, let us first briefly describe what we mean by effectiveness in the context of this thesis. Effectiveness can be defined as a product, task, or process's ability to generate its desired result. In other words, one of the elements mentioned above (product, task, process) can be considered effective when it has achieved the purposes for which it was designed. At a higher level, AR-SGs have two core purposes to fulfill, an educational and an entertaining one. Entertainment relates to the fact that an AR-SG is a game at its heart and should be fun and engaging. At the same time, AR-SGs should also strive to achieve their educational ("serious") purpose, which indeed can be very broad. Nevertheless, these two elements should intertwine as the educational outcome can be strengthened by the entertaining one and vice-versa.

RQ3: *Usability can be an issue in AR-SGs, can it be softened?*

Usability is currently one of the most significant problems faced by AR-based learning applications. While engaging with a novel technology like AR, users might place too much cognitive effort in understanding what to do and how to do it rather than focusing on the learning content. Thus, we ask ourselves if, by addressing users' needs in the design process, AR-SGs usability issues can be softened. To answer this question, we will explore the decisions taken throughout the design of all the proposed AR-SGs and discuss the results of our usability studies.

1.3 Contributions

The overarching objective of this thesis has been the exploration of possibilities offered by AR-SGs as novel learning tools. We now present the main contributions organized according to the top-level RQs discussed in the previous section.

Addressing RQ1: Can we establish synergies between AR and SGs in order to design and develop more effective AR-SGs?

To evaluate possible synergies between AR and SGs, we designed and developed four AR-SGs addressing learning domains that were given little attention by previous research. Therefore, one first contribution to AR-SGs' state of the art has been collecting practical evidence that positive synergies can be established between AR and SGs within learning approaches that extend the "mere" knowledge acquisition process traditionally applied in science learning. We identified three learning domains (i.e., procedure learning, complex systems learning, and soft-skills learning) for their unique learning requirements, which made them ideal candidates for leveraging an AR-SG approach. Despite our AR-SGs specifically address a specific topic, the design features and mechanics we propose could be generalized to other topics within the same learning domain. For example, the core mechanics of Sustain could be transferred to other topics of complex systems learning. Alternatively, the learning path and procedural model of Holo-BLSD could be easily migrated to other domains in procedure learning.

Addressing RQ2: Are AR-SGs effective?
For all the proposed AR-SGs, we conducted extensive experimental assessments by exploiting a rich set of data collected from numerous sources encompassing: (i) subjective data from questionnaires (ii) users' behaviour observational data (collected and coded from the video recording of play sessions) and (iii) objective measurements collected from the in-game interactions. These data sources were first analyzed individually, and then relations among them were searched. This approach provided us with more robust results, pointing to the positive relationship between the game and its outcomes which were proven successful for all the objectives each AR-SGs aimed to accomplish. We could establish that AR-SGs can be, first and foremost, fun and at the same time effective in numerous ways (e.g., promote collaboration, communication capabilities, behavioural changes, and train on complex procedures).

Addressing RQ3: Usability can be an issue in AR-SGs, can it be softened?
Other research in the literature report compromised learning outcomes caused by inadequate usability levels [246]. However, this thesis shows that this problem can be approached through a careful design process focused on satisfying users' needs. Then, by collecting users' feedbacks throughout the development process, these design choices can be iteratively optimized. As a result, we obtained AR-SGs whose usability levels were positively evaluated and did not affect users' learning experience. The details of the design process provided in this thesis can serve as guidelines for future researchers for developing better AR-SGs that are effective in their intended purpose without hindering the user experience. We also explored possibilities of enhancing usability levels in collaborative AR-SGs by designing and testing novel visual cues to promote users' mutual awareness and improve the overall user experience.

As a final note, this research cannot continue if new questions do not arise. Thanks to the theoretical study and the practical developments throughout the years leading to this Ph.D. thesis, we have found new areas or issues requiring further investigation. Unfortunately, they extended beyond this thesis's scope or would have required too much time to be addressed appropriately. As a consequence, they remain unanswered. We proposed them in Section 7.2, hoping that they can provide directions for the AR-SG research community.

In conclusion, by answering our RQs and reporting the solutions found to different AR-SGs' issues, we have added critical elements that future practitioners embarking on AR-SGs research can exploit. Therefore, although the *Metaverse* might still seem far, we believe that learning will be one of its crucial elements and, through this dissertation, we have (hopefully) helped get a little closer.

1.4 Document Structure

The rest of the thesis is organized as follows. Initially, Chapter 2 presents the necessary background of AR and SGs individually, followed by a description of AR-SGs' state of the art. The chapter concludes by emphasizing the open problems this thesis aims to address. Then Chapter 3 describes the possible synergies which can be established between AR and SGs, contextualizing them to three distinct learning domains, namely: procedure learning, complex systems learning and soft-skills learning. For each of these domains, we will (i) describe the learning requirements and the possible affordances offered by AR and SGs individually and how a synergic relationship can be established, (ii) present the state of the art of either AR or SGs to that particular learning domain (iii) outline the key characteristics and design choices of the AR-SGs used as a case study to explore the feasibility of our envisioned synergies and (iv) describe the experimental methodology and discuss the collected results. Then, Chapter 4 describes the problem of users' awareness in a shared digital collaborative environment, highlighting which areas of research in this specific topic have been given little attention. We then describe our proposed VCs for users' awareness. We present the evaluation environment and methodology adopted to assess the feasibility of these newly proposed VCs. Finally, we describe their use in an AR-SG, specifically designed as a testbed to verify the effectiveness in the context of this thesis. In Chapter 5 we summarize and discuss the most relevant result of the performed experimental validations focusing on the (positive) outcomes of both the ludic and educational outcomes and how these relate to one another. This summary is intended to provide the reader with a broader view of (i) the different assessment methodologies applied, (ii) how different data sources can be effectively combined for achieving a deeper understanding of the effectiveness of AR-SGs and (iii) highlight similarities and differences of the outcomes we were able to achieve. Chapter 6 details the design process and decisions made to optimize the User eXperience (UX), followed by a discussion of the (positive) implications of our choices also supported by data collected from usability assessments performed on each AR-SG. Finally, Chapter 7 summarizes the contributions of this thesis and discusses future possibilities.

Chapter 2

Background

Over the past two decades, researchers and practitioners have focused much attention on AR and SGs as effective tools for supporting users improve their knowledge, skills, attitudes, and behaviours. Research has focused primarily on understanding which key features and affordances of these technologies are most effective in enhancing learners' learning experiences. The main reason for this interest is that AR and SGs can create motivating and engaging educational activities where users can explore the proposed content through an interactive and immersive digital experience that supports experiential approaches. These features enable the creation of compelling educational pathways and favourable learning environments, as demonstrated by several findings in the literature that emphasize how these approaches can lead to better learning outcomes associated with high levels of user enjoyment and increased engagement in the proposed activities.

However, despite the positive results, it is essential to emphasize that research into the combined use of AR and SGs is essentially still at an early stage. One of the main reasons that have limited the development of this integration in the past has been the lack of practical tools to manage the AR component. Devices often have poor accuracy in tracking and registration of augmented components, interaction tools are poorly suited for intuitive and straightforward enjoyment of digital content, ultimately resulting in (often) frustrating UX. For this reason, AR and SG have often been considered separate entities in the past. The situation has changed radically in recent years thanks to technological developments that have finally allowed AR to become an (almost) established technology, with effective tools that enable accurate tracking and precise registration, equipped with high-resolution and high-quality displays, and with new interaction capabilities that exploit metaphors that are familiar and natural to users. These developments have also increased the level of acceptability of AR technologies by the end-users and, in terms of the specific scope of this thesis, have sparked a strong interest in combining AR and SGs to investigate augmented reality serious games (AR-SGs). As stated in Chapter 1, the main goal of this thesis is to advance, with empirical evidence, the current state of the art of AR-SGs.

Therefore, to provide the reader with the necessary epistemological context, this chapter will introduce the fundamental findings and theories related to AR and SGs one by one. Then, through an overview of the current state of the art of AR-SGs, we will highlight what the literature suggests are the current shortcomings of these technologies/approaches, shortcomings that will be addressed in the remainder of this thesis.

We would also like to make one final comment. In the following chapter, we will not present an exhaustive list of the applications proposed in the literature for AR learning, SG, or AR-SGs. However, we will show several targeted examples to emphasize the statements and concepts discussed later. More comprehensive background overviews, however, will be introduced in Chapter 3. Here we will focus on specific AR-SGs' educational domains; therefore, for each of these domains, we will provide the reader with a more comprehensive overview of how AR, SGs, or AR-SGs solutions have been used.

2.1 Augmented Reality for Learning

One of the most widely accepted definitions of AR is that of Azuma [16], which defines AR as a system with the following properties: "combines real and virtual objects in a real environment, runs interactively, and in real-time, and registers (that is, aligns) real and virtual objects with each other". The development of a system capable of performing all of these operations is the result of years of AR research that has focused primarily on improving the underlying technologies (i.e., tracking, displays, and input devices) and making them affordable and easy to use.

Once a solution to these problems was found, and development tools that could effortlessly create AR applications were available to practitioners, a new wave of research began, mainly committed to understanding the strengths and benefits of applying AR systems to a broad spectrum of domains. Similar to what happened with computers in the past, enhancing learning environments is a practice that immediately attracts interest as soon as a novel technology becomes available (and sufficiently mature). With AR, research reports a dramatic increase in AR education-related work starting in 2011, which increased exponentially over the years [4]. This increase in interest may have two explanations. First, the availability of AR features on modern mobile devices (e.g., smartphones and tablets) has made AR even more affordable and easier to use [198], facilitating their adoption in both formal and informal learning contexts. Second, researchers believe that AR has numerous benefits and opportunities for learning [328], as well as unique affordances prompted by the specificity of its visual and interactive capabilities [58].

The corpus of works addressing AR for learning is vast, and recent efforts have focused on (systematically) surveying these numerous researches and formulating comprehensive literature reviews [18, 246, 4, 82, 187]. In order to provide the reader with an exhaustive overview of the state of the art, we summarize the main findings from these works. First, we review the learning contexts and topics in which AR has been most widely used. Then, we review the main benefits reported in the literature, followed by an overview of AR affordances for learning purposes and how they (positively) affect learning outcomes. Finally, we outline current limitations that require further investigation.

2.1.1 Target Audience and Educational Fields

Within formal education (i.e., education provided in institutions such as schools or universities), AR-enhanced learning activities are directed primarily at students between kindergarten (K) and twelfth grade (12) [4, 149, 18]. This age group is also formally known as K-12. The greater prevalence (compared to other levels of formal education) of AR approaches in the K-12 age group may have several explanations. First, according to Piaget's stages of cognitive development, students in this age group are in a stage of developing knowledge through the use of their senses [197]. Therefore, the powerful and stimulating visual and auditory properties of AR play an essential role in the learning process at this stage. Another potential explanation is the widespread use of digital games at this age. Therefore, students may find it engaging to learn through a digital (and interactive) artifact that is familiar to the experienced digital activities they do at home [187]. College students are the second most targeted age group, which provides evidence that AR learning can also effectively impart knowledge about topics with a level of complexity and depth appropriate for students at this educational level [187].

Finally, in informal learning contexts (i.e., outside the classroom), there is an increasing number of approaches aimed at both non-professionals and professionals. In the latter, professionals in a wide range of fields, from medicine to industry, need to practice complex activities that primarily involve the correct (and timely) performance of a series of tasks. For example, in a maintenance task, a technician must follow a predefined sequence of steps to correctly repair a malfunctioning machine. With the advent of AR, new forms of training are being explored that provide greater contextualization of the instructional material that seamlessly integrates with the physical objects and environment in which the task takes place. In non-professional contexts, we see (again, thanks to the proliferation of AR-enabled mobile devices) an increasing proliferation of AR educational applications that individuals (of any age) can experience at home or in museums. In this particular case, however, AR is more often embedded in digital games to further engage users in the learning activity.

As for the educational domains involved, AR learning solutions have been mainly explored in STEM related topics [18, 300]. One explanation for this comes from a large body of accumulated evidence showing that AR can be particularly effective when the topic being taught requires visualization of complex concepts that are not (easily) seen in the real world (e.g., electron motion or magnetic fields)[325] or are poorly represented in traditional media supports (e.g., textbooks or videos). Rather than using their imaginations, students can use AR to obtain a rich, three-dimensional representation of the concept they are learning that enhances their scientific spatial skills, and conceptual understanding [58]. For example, the authors of [225] have used AR to teach students the 3D arrangement of crystalline structures. Then, the interactivity features of AR applications play another essential role in science education [219]. By allowing students to change the state of the observed simulation in real-time, they are placed at the center of the meaning-making process as they witness the consequences of their actions. No other field of study has been reported to have the same (or comparable) level of interest as STEM disciplines, as research in educational AR for other disciplines from arts and humanities to health and welfare is relatively sparse [4].

2.1.2 Benefits and Affordances

The great (and rising) interest in AR for learning across a variety of age groups and learning domains stems from the numerous benefits reported in the literature that have led researchers to further explore this new approach to learning. First of all, AR approaches have the potential to increase students' performances as many studies report that AR technology leads to improved learning outcomes [4]. These results are thought to come from improved understanding of the content, which in turn leads to long-term retention of the topics learned [246]. Then, AR solutions can stimulate students' motivation to learn, improve their positive attitudes, and increase their satisfaction with the learning process [18]. According to [59], compared to traditional learning tools, where students are forced to search for relevant information on their own, AR can guide students more effectively by providing them with immediate and relevant information, which in turn increases their motivation to learn. Interestingly, user motivation remains significantly higher even when the AR system is rated as more challenging to use compared to non AR alternatives [246].

Overall, these positive attitudes fostered by the AR learning experience contribute to students feeling more engaged and experiencing a sense of enjoyment as even the less enticing parts of learning can become entertaining [148]. Therefore, due to the increased learning interest and the ability to evoke positive psychological states, students are better able to focus on the topic at hand [330]. Moreover, AR is reported to promote student-centered learning by making students self-responsible for their progress and enabling them to learn independently and individually [82].

All of these benefits are the result of the numerous affordances prompted by

AR technologies. First, due to its unique way of displaying information superimposed on the physical environment compared to other media supports, AR can contextualize virtual information with the environment or a physical object [271]. According to the philosophy of contextual learning, linking instructional content to a familiar context can lead to more effective learning experiences. Also, as this digital information is presented in 3D, students can analyze and explore the content from different perspectives, which facilitates their understanding [325]. Then, as mentioned in the previous section, 3D representations also foster spatial skills and the ability to visualize unseen phenomena.

Thanks to these unique visual properties and the ability to provide real-time feedback on users' actions, AR systems can encourage students' sense of immediacy [179]. This feature is crucial because previous research suggests that immediacy is essential to foster the affective side of learning [21]. Moreover, thanks to the rich sensory stimuli elicited by AR experiences, combined with the ability to manipulate the digital contents, AR promotes a sense of immersion and presence for students. Not only does this engage students more fully in the learning process, but it can also promote situated learning activities (i.e., learning that takes place in the same context in which it is applied) about complex real-world problems that are otherwise difficult to experience. For example, in [56] the authors aim to support student learning on social science issues related to radiation pollution and nuclear energy use in the context of the nuclear accidents at the Fukushima nuclear power plant following the 2011 earthquake in Japan.

Another key affordance that almost all researchers agree on is the ability of AR to foster and enhance collaboration among learners [246]. According to [92], this arises from the ability to talk face-to-face when using mobile devices and through the sharing of different roles that have different and incomplete information. In this way, individuals are motivated to solve problems through dialog and collaboratively hypothesize, draw conclusions, and share knowledge. Similarly, [325] concludes that mobile AR systems could enable ubiquitous, collaborative, and situated learning through the use of handheld devices, wireless connectivity, and location-based technology. Although the possibility of enhanced collaboration seems to be widely acknowledged, research indicates a need to understand better how to make these collaborative activities even more effective. A key issue to address is how to promote user awareness and, in particular, how to effectively use visual and auditory information to support collaboration [100].

2.1.3 Limitations

One of the most acknowledged limitations described in the literature is usability. Authors report that it is challenging for learners to interact with the proposed learning tools and that they have to exert excessive cognitive effort to adequately engage with the learning content, which undermines the overall learning experience and thus the positive outcomes described earlier [246, 325]. Most authors believe that the critical factors for this limitation are technological barriers and AR educational applications that have been developed without properly considering UX [4, 18].

Therefore, on the one hand, it is necessary for developers and practitioners to carefully consider usability issues when developing such solutions and try to adopt a more user-centred design approach. Thus, to achieve an appropriate level of usability, numerous iterations are required to adapt the software product according to users' feedback. This, of course, requires a significant amount of time and is sometimes overlooked by practitioners. However, if the UX is appropriately designed and tested, it is possible to create AR experiences that do not hinder the user's learning activity.

In addition to making applications understandable and accessible, researchers believe that future technological advances will also mitigate specific usability issues reported by users over the past decade. In recent years, we have seen two key technological leaps for AR. One is the ability to support markerless AR on smartphone devices. Thanks to state-of-the-art computer vision algorithms (which can now run on mobile CPUs), smartphones and tablets are capable of capturing their surroundings in great detail, simply using the embedded cameras. This improvement has two significant consequences. First, it makes mobile experiences even more convenient to use, as users do not need to frame particular images (markers) to activate and track enhanced content. This allows for a more natural experience whose physical boundaries can more easily extend across an entire room. In addition, mobile devices can more accurately identify what is in front of them. This allows overlap between digital and physical elements to be resolved (i.e. digital content can be partially or fully obscured by physical content if it is behind it). This feature ultimately provides a higher level of immersion by making the simulation in front of the user's eyes even more believable. As mobile AR is currently the technology of choice for educational purposes, these advances will enhance the user's experience of learning with AR.

The second significant technological advance has been the release of various consumer-level HMDs (e.g., the Hololens or the Magic Leap). These devices, first, allow for a more immersive experience without having to view the world through a window (i.e., the tablet screen). Second, they provide the same markerless experience (and benefits) as described above. Finally, and more importantly, users have their hands free to interact with both digital and real-world elements. Recently, HMDs have also started to come with full hand-tracking capabilities that allow for natural interaction with the digital world, further enhancing the sense of immersion and removing some interaction limitations faced on HHDs. However, due to their high cost, AR HMDs are still far from being widely deployed in schools or other contexts that require many devices and/or have strict economic budgets. In a school, for example, we can assume that in a classroom of 20 students, [228] all should have one AR HMD or at least one every other student (i.e., the fewer devices per student, the longer students have to wait to experience the AR content, increasing the complexity of managing the class). By using the latest HHD (i.e., smartphones), we can provide an AR experience with rich graphics, markerless tracking, and occlusion management with an investment of between \$2500 and \$5000 [1]. If these devices were to be replaced by HMDs that still offer a similar (and improved) AR experience, it would require an investment of \$45900 to \$70000, as the two most accredited AR HMDs are the Magic Leap One and the Hololens 2, which cost \$2295 and \$3500 respectively. However, it should be noted that there are few more affordable AR HMDs. However, they have some drawbacks which make them not worth choosing over an AR experience delivered via a smartphone. For example, a relatively affordable solution could be the EPSON Moverio Smart Glasses, which are sold at different prices depending on their feature set. Although the lowest price is €646.6 (and the most expensive is €1220), these devices are mainly intended as drone accessories and for industrial solutions that can only be developed with proprietary tools that prevent the use of common game engines. The cheapest solution available is the Merge AR /VR headset (\$49.99). This headset is simply a container into which a smartphone (purchased separately) is inserted to provide a pass-through AR experience. Although this device can be considered an AR-HMD, it has several (and crucial) drawbacks. The first is the overall pass-through experience, which relies on the smartphone's monoscopic front cameras, resulting in an uncomfortable viewing experience and potentially triggering motion sickness. Also, this HMD has a limited input system as the smartphone's touchscreen is not accessible to the user, and tracking relies solely on marker-based algorithms. Overall, these considerations lead us to believe that large-scale AR classroom activities with AR HHD will (and should) continue until AR HMDs achieve sufficient market penetration that results in lower production costs and thus lower prices without compromising the quality of the AR experience. For this reason, we are seeing increased experimentation with HMD-based solutions in informal learning environments (e.g., hospitals, industry) where economic resources are typically higher than in educational institutions and where effective training practices ultimately lead to a higher return on investment as staff skills improve.

In addition to usability, the effective use of AR learning activities in the classroom is another widely acknowledged limit. This limit is multi-faceted because it includes the teacher, the students, and the interactions between them. First, managing classroom activities based on AR applications can become complex, requiring inordinate amounts of additional instructional time to become familiar with these

---

[1] This estimate was calculated considering 20 devices at an average price of \$250, taking into account the cost of three smartphones: Huawei Honor 10, Huawei Matte 20, and the Motorola moto g9 plus. These devices were selected among the most affordable Android AR-enabled devices listed by Google: https://developers.google.com/ar/devices

novel tools [214]. This complexity is further compounded by findings from the literature [18, 149], which suggest that students can become distracted by AR content as they become too cognitively and emotionally engaged in the digital experience and lose interest or pay less attention to the learning content. Another major limitation of current AR learning tools is the inability of teachers to adapt (or create new) content to fit their curricula or methods [71]. In addition, most of the current tools have limited tracking and assessment features that can inform the teacher of student progress [210]. Therefore, in this scenario, teachers are on the periphery of the learning process rather than at its center, acting mainly as coordinators while the activity unfolds. As these learning activities are mainly carried out in class, there are few examples of how they can be continued outside class as homework.

As with the discussion of usability, the research community believes that technological advances (which we are already experiencing) will help overcome these problems. In these more systemic areas (i.e., the AR learning process as a whole), the problems will be overcome as new tools become available to both teachers and students. Most importantly, the development of the tools/applications used in the classroom must be simplified. This would allow teachers to (easily) extend and adapt the learning content as they see fit, without requiring the help of technical experts or developers [210]. If these tools could then be made easily accessible to students, we would be able to establish a more bidirectional interaction between teachers and students, where both can contribute to the improvement and refinement of learning content/activities, as this has been shown to lead to better learning outcomes [180]. Then, since AR should not be considered as a substitute for traditional learning methods, it should be analyzed how to properly and most effectively integrate the two activities (i.e., AR and non-AR) [325, 246].

In conclusion, usability is now one of the most critical limitations to consider today in improving AR learning activities. This can (and should) be addressed through more user-centric designed applications and by leveraging the latest technological advances, considering both mobile and HMD devices depending on the target user, context, and learning activity.

2.2 Serious Games

In the 1990s, the growth of the multimedia market PC ensured the rising popularity of edutainment, a TEL approach that delivers education through entertainment. The first edutainment solutions were aimed primarily at young students and focused on teaching reading, math, and science. However, edutainment applications were not as successful as hoped, as they were described as “boring games and drill-and-kill learning” [304] and thus failed to achieve their entertainment and educational goals. Nevertheless, thanks to the ever-growing video game industry and the collected feedback from failures in edutainment applications, the concept of SGs was re-examined at the turn of the century. According to Michael and Chen [206], SGs involve the same goals as edutainment but go far beyond teaching facts and concepts. Based on these premises, they define SGs in 2005 as "games that do not have entertainment, enjoyment, or fun as their primary purpose". From this definition, which is one of the most widely used, it is clear that SGs are games designed and developed with a specific purpose in mind, which can vary widely, as SGs can be used for education, information, physical training, and rehabilitation, or as marketing platforms.

In the context of this thesis, to better align with AR learning background, we will mainly focus on SGs for educational purposes (in the broadest sense). Therefore, the terms SG and game-based learning will be used interchangeably, as it has been reported in the literature that they can now be considered synonyms when SGs have an educational "serious" purpose [36].

2.2.1 Means of Engagement

Regardless of their purpose, SGs should be fun and evoke a sense of enjoyment. It is believed that these positive feelings make users more engaged with SGs, which in turn promotes motivation and ultimately leads to successful "serious" outcomes. The motivational virtues solicited by video games were the main features that have attracted educators to game-based learning approaches, but there is much more to SG than just fun as a means to engage learners [67], and many authors believe that specific (and unique) SG elements and affordances are responsible for the motivation users feel when playing SGs [243, 207]. The features that make SGs engaging and effective for their "serious" purposes can be summarized as follows.

First, the core of games is *gameplay*, i.e., "the experience of a game set into motion through the participation of players" [266]. Unlike books, videos, or other unidirectional learning experiences (i.e., source to the learner, stop), the player (learner) can interact with the subject matter presented in the game. This interaction allows him/her to experiment and play with different outcomes of his/hers actions, mimicking how humans learn. In addition, games allow players to experiment with real life situations without the consequences and dangers of the real world. In the real world, children learn by interacting with objects around them, sometimes resulting in an injury from a sharp knife or bruises from falling down stairs. In digital games, on the other hand, the phenomena that occur in the virtual environment can be (safely) analyzed in detail, from different perspectives, and with the ability to highlight their properties.

Second, games create a fictional environment where the virtual experience gives meaning to the game and provides a context for the actions and decisions a player must make. This fictional environment can also be seen as a metaphor that helps visualize and understand complex issues (e.g., learning). Furthermore, the story and setting of a game are effective ways to create an immersive experience. Players are more (emotionally) connected to the characters in a game and what happens to them when they care about them. And then there is also a vast and beautiful world to navigate, which appeals to our primary motivation to explore and have new experiences. All of these elements increase player engagement, immersion, and retention. The virtual environment and game provide a meaningful context in which learning takes place, and learning within a relevant context is more effective than learning outside of it, a point long argued in theories of situated cognition [123].

Games can then be equipped with progression systems that appeal to and satisfy players of all skill levels. A game should bring the player into a perfect balance of skill and challenge. The player should neither lose interest because they are too bored (i.e., the game is easy) nor be frustrated because they are faced with a task that is too complex. This optimal experience state is also called *flow* state [55]. Games are designed to keep the player in this state because the difficulty of the challenges increases harmoniously with the player's progress, keeping the player motivated. Digital games can use data about the state of the game and the user's actions to determine the player's learning progress in real-time which can also allow the game's difficulty to be automatically adjusted, creating an experience tailored to the user. This same data can provide visual and auditory feedback to inform the player of their actions and ultimately reward them for their performance. This feedback triggers a release of endorphins, making the player feel better and creating an endless cycle that encourages and motivates the user to keep playing and thus learn.

Finally, evidence [313, 232] suggests that SGs are ideal for creating situated and socially mediated learning contexts by enabling shared experiences (e.g., with multiplayer games or by allowing learners to share information and outcomes via social networks). SGs can be designed to foster collaborative discussions and reflections, which in turn support critical thinking, deepen understanding, and expand learners' shared knowledge [292, 286]. Furthermore, these social interactions can also be competitive in nature, as players must compete against each other to succeed. This competitive side of games has also been shown to influence analytical skills and motivation to invest more energy and focus in an activity [115].

2.2.2 Outcomes

The above elements characterize both entertainment-oriented games and SGs alike. However, in SGs they also serve a "serious" purpose. Therefore, the challenge in SGs is to embed the educational content (learning path) in these characteristics, to seamlessly integrate fun and learning, and to present the learning content as something that is neither external to the game nor a juxtaposition of entertaining sequences and educational material. Therefore, SG should be designed with the idea of "stealth-learning", where the learning process becomes invisible to the users

[42].

When this delicate goal is achieved, SGs can have numerous positive outcomes [65]. One of the most commonly reported is knowledge acquisition. In line with the features described above, SGs can reduce cognitive workload and help clarify concepts and patterns that would be difficult to identify in other forms. This process has been shown to be effective in short-term knowledge acquisition, and some examples also report benefits in long-term knowledge retention [10, 222]. The effectiveness of SGs in sustaining knowledge acquisition outcomes is often compared to traditional learning approaches. The extensive literature and evidence collected in numerous fields provide us with mixed results in this area, so it is trivial to paint a general picture [60]. What can be concluded is that SGs are not necessarily a better solution, although this may depend on the topic being addressed. However, they are experiences that are widely preferred by learners [60].

Although knowledge acquisition outcomes are comparable to non-SGs learning approaches, SGs are characterized by a range of other outcomes beyond simply knowledge acquisition [65]. These SG-specific outcomes can operate at perceptual, cognitive, behavioral, and affective levels. In addition, SGs can promote the acquisition of social and soft skills [65, 37].

Perceptual and cognitive outcomes refer to the (reported) benefits of SGs in improving users' attention and visual perceptual abilities [77] (e.g., central and peripheral visual acuity and contrast sensitivity) in conjunction with a variety of cognitive processes that can be trained and strengthened thanks to SGs. These can be memory and reasoning skills, problem-solving skills, and task-switching performance [34].

In addition, SGs can influence learners at the behavioral level by affecting players' attitudes, and behaviors [33]. Immersion in the fictional digital experience creates connections between behaviors in the virtual world and the personal motivations that drive actions in the real world. Once this connection is established, it can ultimately lead to a kind of transfer to the real world of positive actions performed in the virtual one [274]. Thanks to these capabilities, SGs have been used on very delicate and sensitive topics, such as substance abuse education (alcohol and drugs) [309], self-esteem enhancement [74], or gender equality awareness [22], to name a few.

Furthermore, SGs can elicit immediate emotional responses to the elements depicted or the underlying story. These affective outcomes can have several benefits. First, they are closely linked to behavioral outcomes. If a person is emotionally moved on a sensitive issue, he or she may be motivated to change or adjust his or her behavior. Second, the educational process benefits from the positive emotions that the game can induce because when one feels good, learning performance is enhanced [212]. Third, affective engagement has been shown to improve user immersion [159], which, as we have already seen, promotes overall engagement and motivation, which are all positive states for the ultimate learning goal.

Finally, there has recently been a growing interest in the social outcomes promoted by SGs. By offering experiences where individuals can interact simultaneously, SGs encourage players to engage in social interactions in either collaborative or competitive ways. This unique feature of SGs is often exploited for soft skills training, an evolving area of interest [36, 114, 263]. For example, evidence shows that SGs can teach how to communicate and express emotions [128], strengthen team leadership skills [11], and strengthen group decision making [191].

Most importantly, it should be emphasized that these outcomes are not mutually exclusive. Instead, they integrate each other, and therefore playing a SG can help learners simultaneously increase their knowledge of a topic, promote behavioral and attitudinal change, and also develop various skills such as strategic thinking, planning, and communication [286].

2.2.3 Target Audience and Educational Fields

Thanks to the multi-facet set of outcomes just described, SGs are considered unique tools for educational purposes [87]. Consequently, the research community's (and industry's) interest in SGs has grown exponentially since the early 2000s [183]. Over the years, practitioners have used SGs in many different educational settings, targeting users of all ages [87]. There are several explanations for this proliferation and the expanded audience, ranging from preschoolers to adults [248]. First, the users who are now (potential) SGs players have grown up and socialized with digital media such as PCs, the Internet, and gaming consoles. They are often referred to as "digital natives" [244] and games have become part of their everyday culture, making them particularly apt, and familiar with the language of games [227]. Secondly, SGs have also penetrated informal learning contexts such as museums [230], businesses [265] and are also played by individuals in their leisure time [259]. Finally, closely related to this large audience are the extensive domains in which SGs are used, satisfying players of different ages with very different needs. Domains can range from physics, mathematics, history, languages, arts, health habits to social issues.

Although examples of SG can be found in numerous learning domains, some authors emphasize that the main focus is on processes leading to knowledge acquisition [37]. Therefore, the SG research community believes that more evidence should be gathered to assess the feasibility and effectiveness of SGs in achieving a broader range of outcomes, such as learning 21st century skills and promoting behavior change [37, 87]. In light of this observation, we aim in this paper to contribute to the current state of the art by providing evidence for the feasibility of SGs in relation to these outcomes.

2.3 AR-SGs

In the previous sections, we have presented and explained the contextual background of the individual use of AR and digital games for learning purposes, in a broad sense that includes knowledge acquisition, skills training, behavior change, and 21st-century skill development. In general, these two technologies/approaches have similar goals and affordances to achieve them. They both aim to motivate and engage the users in meaningful and stimulating interactions in a digital environment where they actively participate in the knowledge-building process through experience. Considering these similarities, we can find more and more solutions in the literature where both are synergically intertwined. However, this particular line of research (i.e., AR-SGs) is still in its infancy and is a frequently mentioned future direction addressed in various reviews that focus on AR or SG alike [30, 246]. Henceforth, quantitative and qualitative evidence should be collected to define how AR-SGs and their affordances or groups of related affordances support or hinder teaching and learning [272]. This section aims to provide an overview of current efforts in AR-SGs research, summarizing the most important findings, and highlighting open issues collected in various literature reviews [188, 184, 234, 180, 235, 300]. Given this comprehensive picture, we will emphasize how this thesis attempts to fill (some of) the gaps.

2.3.1 Target Audience and Educational Fields

A comparison of the state of the art in AR-SGs and in the separate domains of AR and SGs leads to similar conclusions. One of these is the apparent abundance of AR-SGs applied to science-related topics [300]. A direct explanation for this lies in the numerous affordances AR systems yield when applied to science learning (Section 2.1.1). Therefore, it can be hypothesized that when practitioners began to explore AR-SGs applications, they immediately turned their attention to areas that already had a long history of successful evidence in educational AR research. This observation suggests that the SG counterpart to these experiences played a "secondary" role, as it was probably used primarily for engagement purposes to make the AR science learning experience more playful. However, compared to AR for learning purposes, SGs have a relatively long history, having been designed and developed on technological supports released many years before the advent of AR (e.g., computers and consoles). Because they were not bound by the technological constraints of an innovative technology like AR, practitioners had more freedom to explore and experiment with (digital) SGs.

Thanks to this (longer) experimentation, SGs were applied to a broader variety of topics than AR-based learning. Evidence accumulated over the years suggests that SGs can be unique and viable approaches to learning. SGs expand the scientific domain and enable higher-level learning that goes beyond mere knowledge acquisition. By leveraging their unique affordances, SGs can provide experiential learning activities that help learners understand and grasp more complex problems and master skills that are not based solely on the accumulation of individual pieces of information. These problems and skills may require understanding complex relationships of cause-effect that can only be thoroughly understood through experience. On this basis, the authors of [180] suggest that AR-SGs should broaden their breadth of application domains to include a more comprehensive range of options. The development and evaluation of AR-SGs in disciplines beyond STEM should begin by identifying domains in which SGs (in non AR forms) have been extensively researched and empower the SG-based approaches with AR features and key affordances. However, this integration can also be done the other way around. There are areas of learning where AR-based learning has the upper hand over SGs alternative. For example, practical skills training, where AR solutions are preferred. Even in these contexts, SGs elements could be combined to enhance what has more often been achieved only through AR. The possibility of using this synergic relationship between AR and SGs in different domains will be extensively discussed in Chapter 3.

Another similarity between AR-SGs and AR-based learning approaches is the predominance of applications targeting the K-12 age groups [188, 184, 300]. However, compared to the AR-only option, AR-SGs find little to no application in higher educational contexts. This finding is more difficult to justify, as SGs have long been well suited for students in this age group. One possible explanation may arise from the tendency to present playful AR activities to younger students through an initial exploration of possibilities. Indeed, creating AR-SGs for more sophisticated audiences may involve a higher degree of complexity, as striking the right balance between AR and SG while targeting topics that require a level of deepening suited for university students can be complex. This consideration also stems from the fact that most AR-SGs found in the literature are primarily formative evaluations [184] (i.e., evaluations of prototypes that serve as the basis for further development).

Concluding, these results show that we are only at the beginning of the exploration of AR-SGs. The proposed applications/games are relatively simple in gameplay and are mainly targeted at young audiences (i.e., the content and interactions are simplified). Therefore, there is an urgent need to extend these approaches to a wider variety of contexts, possibly using the insights gained through SGs developed for such contexts.

2.3.2 Motivation and Engagement

As detailed in the previous two sections, key benefits of AR and SGs include the ability to strongly motivate students in the learning process by creating visually and emotionally engaging experiences that foster a sense of immersion, presence, and self-efficacy, which can ultimately impact achievement and learning success.

Recent literature reviews [234, 300, 188] show that the same conclusions can be drawn for AR-SGs. Although a contrary result would have been odd (i.e., AR-SGs not eliciting these positive feelings), this finding requires further attention, as the elements responsible for this engagement are not yet fully understood. Many believe that in AR-only applications, the "novelty" of this technology (i.e., the "wow" factor) may be a major factor in user engagement and motivation [18]. Although this is a positive element, researchers are concerned that this novelty effect will inevitably diminish in the long run as users become more familiar with the technology. As a result, various authors warn practitioners that this phenomenon may interfere with the positive results reported so far [18].

However, as we have seen, the same elements that are responsible for the sense of wonder (e.g., complex and interactive 3D visuals contextualized with the physical environment) and consequently for engagement are also the affordances that make AR learning applications effective for explaining complex topics. Thanks to rich (3D) visualization, users can easily observe and analyze phenomena that are impossible to experience in other ways. On the other hand, SGs also aim to engage users, and they achieve this through a number of unique features that make the experience fun and engaging. Compared to AR, these features (i.e., gameplay, rewards, fictional settings, and characters) stimulate user motivation at a deeper level than "simply" through engaging visual stimuli. Since positive psychological feelings are fundamental to the overall learning activity, it is hypothesized that embedding the unique (and appetible) AR features in SGs, i.e., creating AR-SGs, could maintain users' engagement and motivation even as they become more experienced AR users.

Finally, while motivation is a crucial aspect for AR-SGs, this positive outcome may be hindered by usability issues. When users encounter difficulties interacting with a digital tool, the sense of immersion and presence is interrupted, which in turn can detract from the learning experience. Usability issues can affect both SGs and AR applications. However, as we described in section 2.1.3, this is currently a widely recognized limitation in AR, mainly due to the novel technology. Therefore, usability is an open problem that should also be addressed in AR-SGs. Particular attention should be paid to mitigating the usability implications of the AR technology itself and other factors associated with the game experience.

2.3.3 Collaboration

We have seen that a unique affordance of AR is to encourage and sustain collaborative activities, as it provides natural instances of face-to-face interactivity and the freedom to explore the digital simulation through physical movements (e.g., orbiting a tablet around a 3D object or moving around a table of augmented digital contents). On the other hand, one of the different approaches to designing effective SGs is to design multiplayer activities where joint discussions and problem-solving lead to a deeper understanding of the subject addressed by the SG. Based on these observations, it seems evident that fruitful synergies can be created between AR and SGs to create even more compelling collaborative experiences.

The various reviews presented in this section report mixed and contrasting findings on collaboration in AR-SGs. Li *et al* [188] reports that almost half of the 26 identified research articles show some degree of collaboration, while only 3 works (out of 21) reviewed by Pellas *et al* [234] address collaboration, and Tobar [300] reported 7 games with these features (out of 30). One possible explanation for these mixed results is the authors' different inclusion criteria for selecting works from the literature.

Another significant difference between the mentioned reviews is how collaboration is considered. Tobar [300] classified works as collaborative when the AR-SGs exhibited collaboration as a game feature, in contrast to [234, 188], which viewed collaboration (and more generally "collaborative social interactions") as a detected positive outcome solicited by AR-SGs. To better understand this difference, we should clarify what it means for collaboration to be a game feature. If the design of SG includes collaborative mechanisms, the game requires multiplayer interactions that each player must perform to successfully progress in the game. In other words, the experience cannot be completed by a single user. In [41], for example, players take on different roles and must complete both individual and collaborative quests. The individual quests are specific to each role, and by solving them, players unlock various information that must be shared with others to solve the collaborative puzzles.

Nonetheless, an AR-SGs activity can be designed in such a way that it does not strictly enforce collaboration but still encourages "collaborative social interactions". For example, in the game described in [12], three players must solve a puzzle, but in this case, only one player controls a mobile AR device through which information is provided. The other two players are designated as collaborators and must help the other player by taking notes and solving the tasks together. This example shows that it is possible to play AR-SG using the power of AR to support and encourage communication, leading to positive "collaborative social interactions".

Another factor that could explain the differences in the collaborative experiences reported in the selected reviews is the technological filters used to select the works from the literature. From a visualization perspective, [300] focused strictly on mobile AR-SGs experiences, while [234, 188] also included projected AR and computer-based solutions. Both included, for example, the work described in [101]. In these AR-SGs, the augmented content is overlaid on the classroom floor and displayed via a projected screen. Using marker-based technology, students can collaboratively place tracked objects on the floor and apply "virtual" forces to them. By comparing their predictions with the simulated results, students improved their understanding of forces, friction, and two-dimensional motion.

Despite the differences in the selected papers, the results of these reviews are similar. First, they all agree that supporting, promoting, and improving collaboration and related skills (e.g., communication, negotiation, and mutual understanding) has remarkably positive effects. However, they believe that the contribution of collaborative social interactions in AR-SGs has not yet been fully explored.

A crucial element that has been given little importance is whether and how collaboration contributes to the overall goal of AR-SG [188, 300]. In other words, can collaboration lead to better learning outcomes? Can collaborative activities lead to more engagement and fun? Can collaboration be enhanced by playing AR-SGs? Answers to these questions require careful analysis and robust evaluation methods, as they involve both what happens inside the digital simulation (i.e., game activities and achievements) and what happens outside (i.e., social interactions).

Then, although not strictly declared by the authors, by analyzing the collaborative games selected, we find that a limited number of the selected works include collaborative environments where users act and interact together in the same digital augmented environment (i.e., implement scenarios where the digital simulation is shared among users). Creating a shared augmented digital gaming experience can be a challenging, costly, and time-consuming activity [94] that requires solving various technological problems. Digital content must be synchronized between different AR instances, application state must be synchronized across all devices, and interactions must be coordinated. However, it is not enough to solve these technical problems because other challenges target the UX. One of the most recognized problems in AR-based collaborative experiences is mutual awareness (Chapter 4), which also requires attention and investigation in an AR-based gaming context.

2.3.4 Assessment

The evaluation of AR-SGs is still in its infancy, as numerous papers found in the literature describe mainly experimental evaluations that focus on formative evaluations of early prototypes [184]. This finding highlights that research on AR-SG is still in its early stages, and most authors are still experimenting with how to effectively combine AR and SGs.

The few detailed assessments reported focus on the hoped-for positive effects of AR-SGs, namely improved learning outcomes and higher levels of motivation/engagement fostered by AR-SGs [234, 188]. We believe that this is another indication of the current novelty level of AR-SGs. Researchers and practitioners are mainly focused on evaluating the primary assumption behind AR-SGs, i.e., that by combining AR and SGs, their individual achievements and benefits can continue to be recognized and even strengthened by this integration. Although initial results support this hypothesis, better evidence is needed (e.g., to determine whether learner motivation is enhanced simply by a novelty effect, as described in Section 2.3.2). Nevertheless, there are also areas of assessment that have attracted little to no attention. To name a few: how do usability issues affect learning and motivation? How do positive social interactions affect learning outcomes? Can AR-SGs be considered reliable instruments for assessing learning outcomes?

The complexity involved in assessing some of these items also requires appropriate evaluation methodologies. Questionnaires in pre-post test (PPT) studies have been a significantly exploited tool to assess learning achievements [188] (i.e., students' knowledge was measured before and after the AR-SGs, and the differences between the two measurements explain the learning outcomes). While pre-post questionnaires fit perfectly with the needs of the STEM disciplines (which involve a process of knowledge acquisition), some fields deal with more "intangible" skills (e.g., problem-solving, communication, or negotiation skills) that require more complex and robust methods/techniques. In particular, these approaches need to integrate questionnaires with various data collected both in-game (e.g., by recording all player activities and events) and out-of-game (e.g., by analyzing the recording of player movements, verbal and nonverbal interactions).

Chapter 3

AR-SGs Synergies

The primary objective of this Chapter is to present evidence that can help us answer *RQ1: Can we establish synergies between AR and SGs in order to develop effective AR-SGs in a variety of learning contexts?*. By synergy, we mean the creation of an effective "relationship" between these two approaches/tools for learning, where the affordances of either one would be exploited to enhance the other. To select the most appropriate learning contexts for answering RQ1, we identified learning areas in which either AR or SG had already been primarily exploited as "individual" but not widely in combination with one another. In the end, we identified three possible domains which met our needs. They are procedure learning, complex systems learning, and soft-skills learning.

In procedure learning, students and professionals are trained to perform a sequence of operations or actions correctly according to a specific sequence and certain rules. In this field, AR can offer unique learning possibilities. For example, users can receive just-in-time information (i.e., when a specific operation of the sequence should be performed) and contextualize it to the physical world and objects where the procedure will eventually occur. Thanks to these features (and others which will be extensively discussed in Section 3.1), AR is rapidly becoming one of the most acknowledged systems for procedure learning. Another benefit of AR for procedure learning is the increased motivation it fosters in learners, which, as described in Section 2.3, can be further enhanced by SGs. Indeed, their features solicit more intrinsically users' engagement at both a rational and emotional level. However, despite that, the adoption of SG-based solutions in procedure learning is still sparse, and therefore, we believe this can be a relevant field where synergies between AR and SG can (and should) be explored.

Complex systems learning aims to teach about the nonlinear relationships which govern complex dynamic systems. These systems cannot be simplified through a linear cause and effect relationship. In fact, complex systems effects can result from a multi-facet (and unobservable) set of phenomena. Traditional methods which rely on direct instructions have shown several limits in addressing learning in this domain as they can be ineffective in making such complexity understandable for students. Recent research recognizes that the educational requirements for learning about complex systems can be more effectively satisfied by constructivist approaches, which are based on the assumption that learners construct knowledge and skills as they try to make sense of their experiences (i.e., experiential learning). Among these approaches, SGs play an essential role as they feature a set of characteristics that meet the requirements of constructivism and are therefore recognized as ideal candidates for complex system learning. SGs allow players to build knowledge by experiencing and interacting with virtual environments that mimic complex real-world scenarios. Similar to procedure learning, but the other way around, complex systems learning has mainly focused its attention on SGs without paying attention to the novel possibilities offered by AR. Among these, one of the most important is the proven benefit of AR systems to promote users' collaboration.

Learning and training soft-skills is a field gaining momentum as organizations have recognized the importance and benefits of individuals working in teams. Being effective in teams requires a specific set of skills (soft-skills) like communication, negotiation abilities, collaborative problem-solving. Unfortunately, these skills are complex to teach through traditional instructional approaches. Similarly to complex systems learning, they require an experiential approach. Therefore, among the specific interventions aimed at teaching and practicing soft-skills, organizations are increasingly turning to game-based methods to make these activities more fun and engaging. In particular digital SGs offer endless possibilities of establishing engaging and motivating experiences. On the other side, the use of AR for soft-skills learning has been largely unexplored, even though AR offers specific affordances that seem tailored for this learning domain, like engagement, face-to-face collaboration, and so on.

In the following sections, we will address each learning domain separately as follows: procedure learning (Section 3.1), complex systems learning (Section 3.2), and soft-skills learning (Section 3.3). For each domain, we will first introduce the learning context in greater detail. Then we will describe the background usage of AR and SGs for that specific domain, emphasizing how a synergic relationship can be established between them, and discussing the design and implementation of the AR-SG crafted as a case study for that particular learning context. Finally, we will present and discuss the results of their evaluations.

Concerning the developed AR-SGs we underline, as mentioned in Section 1.1, that they serve as demonstrators and their evaluation results help us answer all the RQ proposed in this thesis, which (as discussed in Chapter 1) are tightly interconnected. Therefore, while drafting this thesis, we were faced with two options. First, divide the experimental results among chapters addressing specific RQs; second, introduce all the necessary information (methods and results) to answer RQ1 first and then recall the most important findings to answer RQ2 and RQ3 in dedicated chapters.

In the end, we opted for this second option since we believe that exhaustively presenting all the works and their experimental assessment in a more compact way will help the readers easily recall information associated with a specific AR-SGs.

3.1 Procedure Learning

In many areas, learning, training, and the management of work and organizational activities can be defined in terms of procedures. A procedure is a sequence of operations or actions performed according to a specific sequence and certain rules. An example in the industrial field is maintenance operations, defined by procedures that identify a sequence of clear, well-defined instructions detailing the work to be performed. These instructions serve to ensure that operations are performed following the maintenance strategy, policies, and programs.A possible difficulty is that these procedures are often quite complex and are influenced by various situations and boundary conditions, resulting in execution flows that are difficult to manage and memorize. The consequence of this observation is that new staff needs to be thoroughly trained, while resident staff skills often needs to be continuously refreshed (and possibly re-trained in case of changes to the procedures' definition). These mechanisms apply to other areas, both in industry and elsewhere. We have activities such as repair, assembly, and quality assurance in the industrial field. Other scenarios include learning safety and medical procedures, training personnel on the operations to be performed, and improving their decision-making processes.

Procedure learning has more general implications because, ultimately, in almost everything we do, we do it at least in part using a procedure (i.e., following a series of steps). Think, for example, of activities such as reading, writing, doing arithmetic operations, driving a car, cooking our meals, and getting dressed. We also use procedures to classify concepts. In essence, procedures are the form of almost everything we do. Given the importance of procedural skills in our professional and personal lives, it is essential to find ways to teach them well.

To support procedure learning, teachers (including computer-supported teaching approaches) need to provide a set of activity-specific instructions and describe how these activities are logically linked according to the form of the procedure. These learning objective can benefit from a set desirable features. Among them, the possibility to deliver *multimedia* and possibly *interactive information*, to leverage *context awareness* (i.e., recognizing the state of the learner's operations and activities, managing and verifying the correct sequence of operations and activities performed), and to benefit from *adaptive information communication* (i.e., modulating the communication of information based on the state of execution of the activity and the surrounding environment).

Based on these requirements, AR can play a relevant role in strengthening procedure learning. Through AR systems, information can be spatially anchored to objects in the environment, complementing the physical world with virtual constructs that assist trainees in following and understanding the learned procedure. Thus, AR maintains the physical context, realism, and multi-sensory interaction of a task, while adding virtual elements such as overlaid instructions and feedbacks. Thanks to this, users can familiarize themselves with spatial relationships between the actions to complete, the objects to perform them on, and the tools to use.

The ability to overlay all this (spatially coherent) information is supported by the numerous embedded sensors that enable the continuous tracking of AR devices, as well as capturing the state of the actions that users (learners) are performing in both the physical (e.g., where is the user located) and virtual environment (e.g., when the user interacts with a given virtual object). Thanks to this information, the AR system can be aware of the users' actions and verify their correctness, and, thus, it can (i) provide information contextualized to the current step of the procedure, (ii) support the learner in case of mistakes or difficulties in proceeding with the activities, and (iii) automatically assess the user's performance while completing the procedure.

Furthermore, growing evidence shows that people learning procedures through AR can eventually perform them faster and with fewer errors once faced with the actual task [161]. Literature suggest that integrating AR approaches into existing curricula can improve retention, attention and satisfaction [270]

Finally, another factor that accounts for the positive outcomes reported in AR procedure learning is the increased motivation that learners experience while training through this technology. Compared to other digital or non-digital practices, in AR, trainees feel more engaged by the highly interactive and immersive experience tailored to their individual needs. Therefore, trainees embrace an active role throughout the course feeling more motivated to learn procedures that can be long, complicated, and (at times) also tedious.

As described in Section 2.3, the motivational experience delivered by AR learning systems can be enhanced by means of SG features that more intrinsically solicit users' motivation and engagement at both a rational and emotional level.

However, creating "traditional" SGs, which establish a fictional setting composed of characters, narratives, and quests, which take place in an imaginary virtual world, can result as counter-effective when training in highly skilled domains (e.g., medical and industrial training). In these domains, trainees need to learn a complex set of activities that will eventually be performed in real life. Therefore, a digital simulation that tries to mimic the real world is preferable because it simplifies the transfer of skills from the virtual simulation to the real world.

Despite the potential drawbacks of SGs, other game-based approaches can be used effectively in this learning context. One of these is Gamification, i.e., the practice of using game elements in non-gaming systems to enhance the UX and increase user engagement, loyalty, and enjoyment [76]. Gamification aims to harness the motivational power of games to drive participation, persistence, and performance. A key element that games and gamification practices have in common is the purposeful structuring of activities (i.e., goal-oriented) [127]. When players (learners) have a clear goal defined by a "victory" condition, they are guided toward mastery of a topic or skill, which in turn leads them to spend more time learning (topics) or practicing (skills). An example of how this process is supported in gamification interventions are reward systems that award points, trophies, and badges for players' actions. For example, points (i.e., scoring systems associated with correct/wrong actions, such as the steps of a procedure) can be included in a leaderboard of all participants in a training activity. Comparing one's score with that of other participants arouses a (positive) sense of competition in the learner, which is triggered by the human need for social recognition and social status and leads to more engaged activities [254]. Similarly, trophies can be awarded in conjunction with goals achieved (i.e., goal-oriented activity) that unlock customizations or additional activities tailored to the characteristics and needs of different players. Overall, players are encouraged to acquire knowledge through these mechanisms. Gamification systems not only show how the player is performing in comparison to others, but also use the idea of progress tracking, which shows (constantly or at specific points in time) how the player is performing in relation to the end goal. Thanks to this "comprehensive" picture, learners are given a context in relation to the long-term goal while also being able to develop their own as they progress through the experience. Finally, a key feature of gamification interventions adopted from games is the acceptance of failure and building a positive relationship with it [48]. In other words, games can also involve trial and error, leading to failure or success through practice, experience, reflection, and learning. In this context, failure is not seen as an end but as a drive for game mastery. By building rapid feedback cycles in games and gamified learning interventions, failures can therefore be highlighted (positively) to reinforce understanding of why they occurred and prevent their reoccurrence [186].

Although the benefits of gamified learning have been widely exploited both in education [51], and in the specific case of procedural learning [217], the combination of game elements with AR procedural learning has not yet been fully explored [217]. Nevertheless, there are examples in the literature that suggest that using game elements in AR procedure learning can yield better outcomes. For example in [39] authors describe an AR-based (i.e., Hololens) training application for a warehouse picking task. The same experience was designed with and without different game elements, namely leaderboards and badges. The performed evaluation highlights that the gamified solution significantly improved user performance and motivation over the non-gamified one.

3.1.1 Holo-BLSD, Learning and Evaluating Emergency Response Skills

The first example of procedure learning we present is a system for self-learning and self-assessment of emergency procedures in the field of resuscitation and specifically for first aid in the event of sudden cardiac arrest (SCA). In such a situation, the heart abruptly stops beating, halting the blood flow to the brain and other vital organs, eventually causing the victim's death if not treated within minutes. In 2017, the estimated annual incidence of SCA victims in Europe and the United States was, respectively, 300,000 and 347,000 [299, 26], confirming SCA as one of the major causes of death in adults in developed countries [327].

In case of out-of-hospital cardiac arrests, early recognition and intervention are critical for patient survival. In these situations, the combination of cardiopulmonary resuscitation (CPR) and the use of an automated external defibrillator (AED) can more than double the victim's chance of survival [257]. Unfortunately, only a minority of arrest victims receive bystander CPR, mainly due to the lay rescuers' fear or inability to perform this procedure [126]. Thus, increasing the percentage of the population able to deliver proficient CPR in an emergency can help save lives [132].

To this end, a primary educational goal in resuscitation is to train laypeople in BLSD, i.e., the sequence of operations aimed at recognizing a patient in cardiac arrest and managing the first response emergency procedures. Currently, the gold standard for BLSD training is instructor-led courses, which include theory lectures and demonstrations. These courses integrate different simulations involving script-based role play [61] and the use of high-fidelity manikins, thus providing an experiential learning experience and allowing the hands-on practice of several critical care procedures [213]. Regrettably, instructor-led courses are time-consuming, and costly [314, 283].

Well-designed self-instruction contents represent a practical alternative for both laypeople and healthcare providers. Multiple studies have shown that their learning outcomes (in terms of cognitive effort, skill performance at course conclusion, and skill decay) are comparable with that of traditional courses [327]. Among the BLSD self-instruction methods, the computer-based ones offer several interesting features. Besides enabling the development of practical learning approaches that integrates multimedia and interactive contents, they can be deployed on mobile devices [291, 90] and can also encompass the use of manikins for hands-on learning (like in traditional courses) [182]. Students can complete the program at their own pace, repeat the training at will, and monitor their progress through the application's self-assessment features. Furthermore, a computer-based method is cost-effective in many ways, as it does not require the physical presence of an instructor, it can reduce travel time, and it allows developing the application once and using it for many learners.

The recent advances in VR and AR technologies fostered their introduction in BLSD learning. VR allows the development of experiential learning environments where different scenarios can be recreated, including the possibility to simulate (real) situations involving various risks for both the patient and the rescuers in a (virtual) safe environment. AR can be even more powerful for the scenario at hand since it allows users to enjoy the learning experience within the current real environment, which is "augmented" with virtual elements re-creating the context of the simulated emergency scenario.

These premises led us to analyze the possibility of developing HoloBLSD, an AR-SG for self-instructional training and self-assessment of BLSD skills. Holo-BLSD is an application for the Microsoft HoloLens AR HMD that combines a standard low-cost CPR manikin with holographic interactive contents reproducing a realistic emergency scenario (Fig. 3.1). The choice of using this device in particular was dictated by several needs. First, the BLSD procedure requires several activities to be performed with the hands, on top of all the CPR steps. Moreover, these activities can (and should) benefit from haptic feedback (i.e., the manikin). Therefore, we thought that an HMD-based solution would be the most appropriate, as it guarantees a hands-free experience. When we developed HoloBLSD, the number of AR HMDs was still quite limited, and the Hololens had just entered the market. This device offered a number of features that other HMDs did not have at that time. One of these is the need for an untethered experience (i.e., computations are managed entirely on the device without relying on an external workstation) to ensure maximum freedom for users as they move around and interact with the physical manikin. Since BLSD training can be targeted at non-medical personnel, these learning activities are also conducted in non-hospital settings such as schools or other public spaces. Thus, to maximize the portability of the learning experience, we needed something simple to set up (simply placing the device on the learners' head) and easy to transport (the Hololens fits in a regular backpack). For these reasons, the Hololens was chosen over the Meta 2 (which was popular at the time but is now discontinued), which is (was) a tethered AR HMD. Compared to other untethered HMDs, we preferred the Hololens over devices like the EPSON Moverio Smart Glasses because they have limited development tools, rely on proprietary solutions, and do not provide SDKs for popular game engines (e.g., Unity or Unreal). In contrast, the Hololens has a stable and feature-rich SDK for the Unity game engine, an environment in which we had already prepared content for this learning activity (i.e., 3D graphics and application logic). Therefore, we felt it preferable to proceed with this development platform without converting all the work we had previously done to the Moverio environment. Finally, since the Hololens had just been launched, we used the experience of developing HoloBLSD as a stress test to assess the limits of the device and gain insights for future developments (e.g., Asteroid Escape, described in Section 3.3).

Within the application, learners can use natural gestures, body movements,

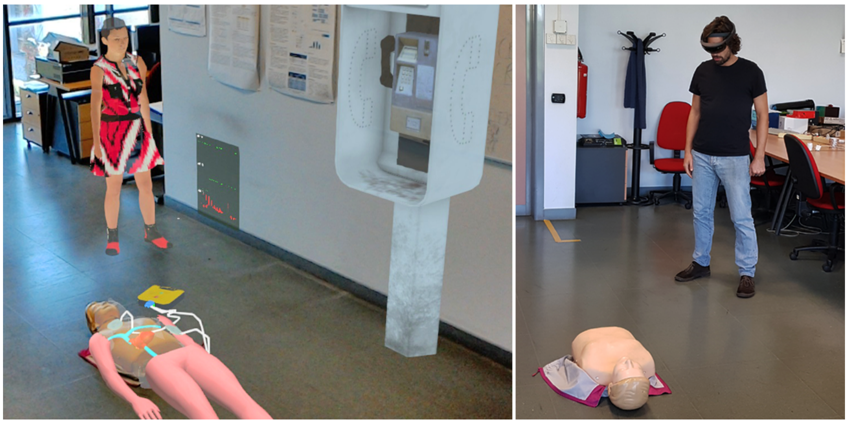


Figure 3.1: Holo-BLSD: a view of the AR environment (left) perceived by the learner (right).

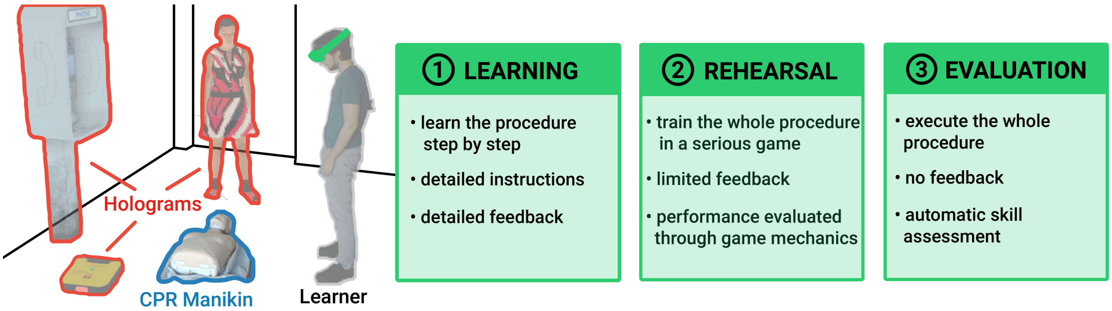


Figure 3.2: Holo-BLSD: AR environment (left) and outline of the proposed learning path, including three modes (right).

and spoken commands to perform their tasks. The rationale for choosing AR as technology is twofold. First, as far as training involves movements in a real environment, AR guarantees users higher confidence than VR in performing their tasks. Second, the possibility to interact with digital elements in the real, physical environment can improve the knowledge transfer from the virtual to the real world.

The learning path of Holo-BLSD[1] offers three different modes (Fig. 3.2):

- *learning*, where users are guided step-by-step through the correct sequence of actions they have to perform to complete the procedures of interest;
- *rehearsal*, envisioned as a gamified experience where users can train the intervention procedures they have learned in the previous phase;
- *evaluation*, where the system automatically assesses trainees' BLSD skills.

---

[1] For a general video description of Holo-BLSD learning path, see `http://tiny.cc/HoloBLSD_Description`

Users are free to repeat learning and rehearsal until they feel confident with the procedures and are ready to be evaluated.

The main novelty of our work is that, to the best of our knowledge, Holo-BLSD is the first system in the domain of BLSD learning that exploits AR to provide an all-in-one and complete self-learning, self-training, and self-evaluation tool. Its main features can be summarized as follows:

- the actions of the emergency procedure can be tailored to different audiences (e.g., laypersons and healthcare providers), thus allowing the adaptation of the educational contents to the actual learners' need [132];
- its AR environment is easily reconfigurable to simulate a variety of emergency scenarios (e.g., indoor, outdoor, with different hazards and victims' consciousness states);
- it guarantees hands-on learning through the use of a standard, low-cost CPR manikin;
- it features a set of game elements which help learners practice and improve the required skills;
- it helps learners to monitor their training progress by providing feedback on performed actions, and features an automatic self-assessment tool of the learned skills that reproduces traditional evaluation, thus limiting the impact of subjective tutors' measurements;
- it includes a debriefing companion application that allows learners to review and critically analyze what they did.

We thoroughly assessed Holo-BLSD through a user study that involved a panel of volunteers with no previous knowledge in BLSD and no previous experience with the HoloLens device. Our experimental results demonstrate the usability of the proposed tool and highlight its capability to stimulate learners' attention to levels similar to those achieved with traditional training. Furthermore, the comparison between traditional and automatic measurements indicates that Holo-BLSD is both an effective learning method and a reliable self-evaluation tool.

3.1.2 Background

Several VR and AR based approaches to CPR and basic life support (BLS) learning have been proposed in the literature. Mini-VREM [278] is a VR-enhanced manikin capable of rendering the main patient's vital signs and reactions to treatments. At the same time, the software provides real-time feedback on the quality of the chest compression during CPR by means of low-cost motion capture technologies based on the Microsoft Kinect sensor.

Method	**Problem addr.**	**Virtual envir.**	**Self learn.**	**Self train.**	**Self assess.**	**Exper. valid.**	**Manikin**	**Voice int.**	**Gesture int.**	**Mouse & keyb.**	**Ext. dev.**	**Game mech.**
VR												
MINI-VREM	CPR	✗	✗	✓	CPR rate/depth	✗	✓	✗	✗	✗	✓	✗
RELIVE	CPR	✓(PC)	✗	✓	CPR rate/depth	✗	✓	✗	✗	✓	✓	✓
EMERGENZA	BLSD	✓(PC)	✗	✓	Limited feedback	✗	✗	✗	✓	✓	✓	✓
LISSA	BLSD	✓(PC)	✓	✗	✓	✗	✓	✗	✗	✓	✓	✓
AR												
ARLIST	CPR	✗	✗	✓	✗	✗	✓	✗	✗	✗	✗	✗
Park et al. [233]	CPR	✗	✗	✓	CPR rate/depth	✗	✓	✗	✗	✗	✓	✗
Higashi et al. [146]	CPR	✗	✗	✓	CPR quality	✗	✓	✗	✗	✗	✓	✗
HoloCPR	CPR	✗	✓	✗	✗	✗	✓	✗	✓	✗	✗	✗
MyCardiacCoach	CPR	✓(mobile)	✓	✗	CPR rate	✗	✗	✗	✓	✗	✗	✓
HoloBLSD	BLSD	✓(HMD)	✓	✓	✓	✓	✓	✓	✓	✗	✗	✓

Table 3.1: Comparison of CPR/BLSD applications.

A similar approach is pursued by RELIVE [277], a VR learning tool that leverages the Kinect to assess the correct arm pose and chest compression rate. Compared to the work reported in this paper, these approaches support only CPR training. Moreover, they merely provide a feedback on the performed actions, not a full assessment of learners' skills.

EMERGENZA [110] is a VR system that allows to simulate different first-aid scenarios and procedures, including BLS. Learners' performance can be assessed both on-line, by analyzing the actions performed in the virtual environment, and off-line, in a debriefing session based on the logs recorded during the training session. Compared to the devised Holo-BLSD application, this system focuses only on the training phase and does not provide a built-in support for learning. A more complete approach is proposed by LISSA (LIfe Support Simulation Application) [319], a desktop VR application that manages self-learning and self-assessment of BLSD procedures. The main difference with Holo-BLSD is the lack of a self-training part. Moreover, the self-assessment includes feedback aimed to help learners in their activities (opposite to the solution discussed in the current work, where no feedback is provided) and, most importantly, its reliability has not been validated through comparison with expert ratings.

As for AR, one of the first solutions to be proposed was ARLIST (Augmented Reality for Life Support Training, [245]), which exploits projector-based AR to augment a standard manikin with vital signs and facial expressions. More complex implementations of the approach adopted in ARLIST, which include also feedback for CPR rate and depth (based on sensors embedded in the manikin) as well as CPR trainee's posture (by means of a RGB-D camera), were discussed in [233] and [146]. The main limitations of these solutions are that they provide only real-time feedback on the CPR quality, and lack any self-instruction module. HoloCPR [260] is an AR application that exploits the Microsoft HoloLens to provide contextual information and real-time instructions for delivering CPR; however, AR is only used to overlay hints on the actions to perform and no evaluation of the learners' skills is actually considered. Given the recent capabilities (and performance) of mobile devices, developers also started to create AR apps for smartphones and tablets. An example is My Cardiac Coach [8], an app that combines interactive lessons with an AR-based CPR training system allowing users to practice CPR on a virtual victim and obtain feedback on their performance.

Table 3.1 summarizes the main features of all the above methods, by indicating the problem(s) addressed, the possible use of a virtual environment to boost the sense of presence (and the device used to present it to the user), the support for self-learning, -training and -assessment phases (or the lack of it), the availability of an experimental comparison between automatic and expert ratings, i.e., of a validation, the use of manikins, the interactions considered (voice, gestures or mouse & keyboard), the need for external devices and the use of game mechanics within the learning process.

Based on the review performed, Holo-BLSD appears to be the only tool to provide complete coverage of the various steps involved in the users' learning path, i.e., receiving instruction on the main actions to perform, train the whole procedure(s), assess the achieved skills and critically review performance in a debriefing session. It is also the only AR tool to provide context for the learning experience through a virtually-recreated emergency scenario and further boost users' immersion through various natural interaction techniques.

Furthermore, to the best of our knowledge, it is the only tool for which a confirmation of the reliability of automatic assessment has been obtained through a comparison, on a common rubric, of computer-based and traditional (that is, performed by instructors via visual inspection) evaluations. Then, it does not require the use of any external sensor or device, which makes it significantly easier to deploy and use. Finally, of the selected AR works, we stress that only one features game elements. This finding supports our statement that the combination of AR and SGs in procedure learning contexts is still a research area requiring further investigation.

3.1.3 Application and Environment Design

Holo-BLSD has been envisaged as an application that should allow users experiment the BLSD procedures in their entirety and in different scenarios (e.g., hospital, street, mall, and so on). As a further requirement, it should be able to target different potential learner categories, such as laypeople and healthcare providers, each with its own peculiar learning requirements. Furthermore, various clinical cases should be simulated, including, e.g., a conscious victim requiring no intervention or an unconscious victim showing vital signs, which only requires the learner to call the emergency medical service (EMS).

Thus, Holo-BLSD has been designed to be flexible enough to support the different configurations, scenarios and action sequences mentioned above, as well as to facilitate the introduction of new elements, thus enabling future extensions (e.g., to support emergency team training).

The overall logical architecture of the system is shown in Fig. 3.3. The simulator module is responsible of managing all the activities involved in the specific learning scenario and learning mode selected. The session parameters can be selected through a configuration menu within the application. The user performs his or her activities in an augmented world that can generate different scenarios at varying complexity and that can be adapted to the actual physical location where the training is delivered (through an integrated editor). The interaction module is responsible for managing all the interactions between the learner and the augmented environment. Rehearsal and evaluation sessions can be recorded to be later analyzed in a debriefing step. The application has been created in Unity.

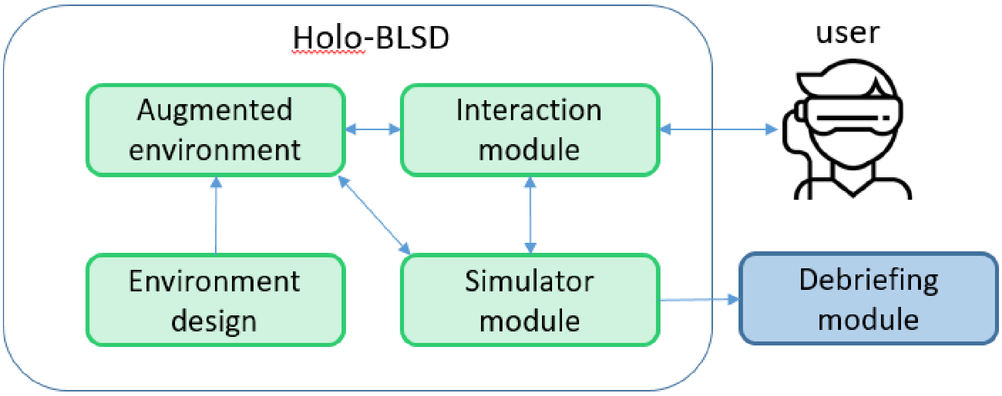


Figure 3.3: The architecture of Holo-BLSD.

The rest of the section presents some of the technical aspects of the above modules, focusing in particular on the design of the simulator, of the AR environment and of the User Interface (UI). Presentation makes reference to the simulated scenario exploited for the experimental observations, which involved volunteers with no previous knowledge of BLSD; for this reason, the considered procedures include a modified version of CPR named continuous chest compression (CCC), which does not require interrupting the chest compression for rescue breathing. CCC is easier to perform than standard CPR, and was likely to be more acceptable for a layperson since it does not require mouth-to-mouth breathing [32].

In this emergency scenario, an adult is lying on the floor. The learner should first check the scene safety (eventually removing potential hazards such as wet, sharp objects or broken glasses), then evaluate the responsiveness of the victim or his or her vital signs (e.g., moving, coughing or breathing). Afterwards, the learner should call the local emergency number for getting professional support and ask a bystander, if present, to get an AED; otherwise, he or she should get one by himself or herself only if within reach. If an AED is not readily available, the learner should start chest compression immediately. When the AED becomes available, it should be operated by first switching it on, then plugging and placing the pads, and finally checking the victim's heart rhythm. Upon machine instruction, the learner must control that no one is touching the victim, deliver one electric shock and then restart chest compression until the AED suggests delivering another shock. This cycle must be repeated until the victim starts to breathe or the EMS arrives.

Simulator Module

The BLSD procedures, irrespective of the actual clinical case and simulation scenario, require to carry out certain actions in a specific order. The completion of an action involves the interaction with the environment (and objects in it), the possible bystanders and/or the victim.

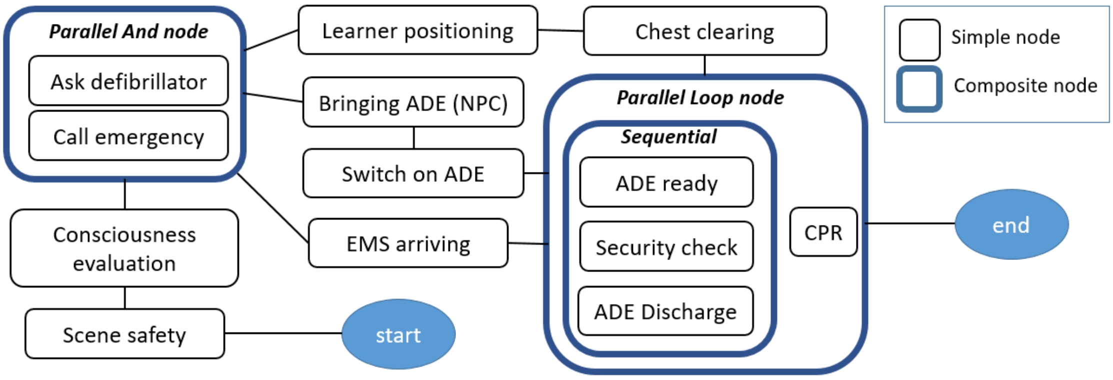


Figure 3.4: Action graph for the clinical case corresponding to a patient unconscious and not showing vital signs. Boxes with a thin outline represent simple nodes, boxes with a thick outline are composite nodes. Start and end nodes represent, respectively, the beginning and end of the procedure.

In order to keep into account the adaptability and extensibility constraints previously mentioned, in designing the simulator we modeled dependencies between actions as a directed graph, where nodes represent individual actions and edges correspond to dependency requirements. The simulator is then responsible for the action scheduling. Each action is implemented as a single software module, and its execution flow is a function of the user's interactions with the virtual environment as well as of internal and external events. To further gain in flexibility when defining the action graphs of the procedures, we included the possibility to use composite nodes, which gather various sub-nodes implementing different algorithms for managing them (e.g., sequential or parallel execution and loop management). An example of action graph is shown in Fig. 3.4.

With this approach, the procedures related to different clinical cases can be simply modeled as different action graphs, and targeting a novel audience (e.g., security officers in a chemical plant, etc.) can be addressed by implementing additional and target-specific actions (e.g., managing the spill of toxic substances).

Augmented Environment

The augmented environment manages all the virtual objects that are necessary to create the context of the simulated scenario. The environment can be also populated with non playing characters (NPCs) that can take different roles in the simulation. In order to improve the user's sense of presence, high-quality and realistic holographic contents were used. For the same purpose, the avatars included in the environment were animated using motion-capture data, and were provided with the capability to have voice-based and realistic (though limited) social interactions with the user.

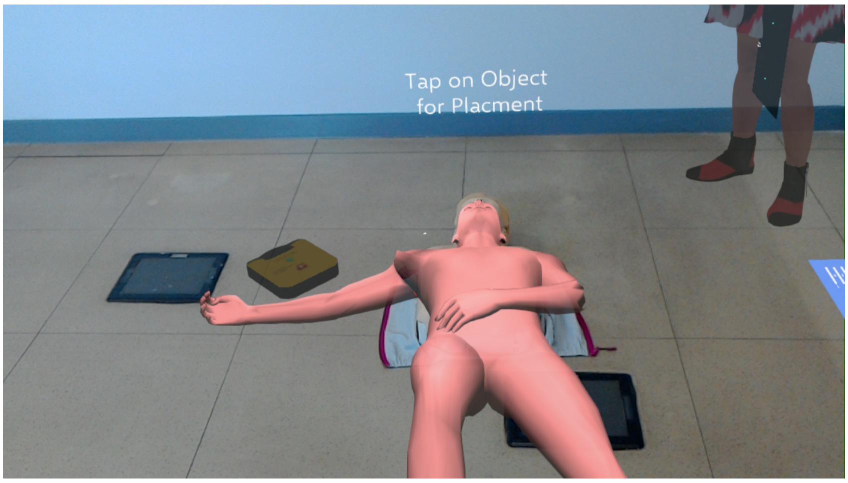


Figure 3.5: An example of the interactive placement of virtual objects.

The possibility to deploy the system in different real environments requires to adapt the placement of virtual elements in the emergency scenario to the actual setting. To this purpose, Holo-BLSD includes an environment design feature that exploits the HoloLens's *spatial anchors*, i.e., geometric descriptors that allow to register one or more persistent holograms with the real-world surfaces reconstructed by the device.

Holo-BLSD provides two environment design modes. The first is an interactive placement mode (Fig. 3.5), which leverages basic HoloLens features and offers the designer an easy and quick way to reconfigure the virtual environment at the cost of a coarse object placement. A second design mode has been implemented ad-hoc for situations requiring precise and accurate anchor placement. This mode consists in elaborating offline (in a 3D editor) the geometric reconstruction of the real environment obtained from the HoloLens, then storing the result in a file that is processed by the application at startup.

In the emergency scenario, the CPR manikin is augmented by superimposing to it the full body of the victim in AR. Currently, rather than registering the victim's body with the manikin (which would require the availability of an extra tracking device, reducing portability and flexibility), the manikin is manually aligned with the virtual body in the real world. Any possible misalignment introduced during the execution of the procedures can then be easily adjusted at the beginning of a new session.

Interaction Module

In Holo-BLSD, users are engaged with an (augmented) environment which is experienced through the HoloLens device. The user controls both the camera position and the navigation inside the virtual environment with his or her own movements.

Optical sensors are used to compute spatial mapping and positional information (via inside-out tracking) as well as to perform gesture recognition. A microphone array is used to capture the user's voice.

It should be noted that working with a holographic headset for the first time may represent a barrier for the users, and newbies often need significant practice and training before they get comfortable with this type of devices. This is a relevant issue that needed to be addressed in a careful way in our case. Users should be enabled to operate the system efficiently in the shortest time possible. In this way, they can spend their mental effort in learning and performing the actions required during the simulated emergency procedure, rather than in trying to understand how the AR system works.

Thus, the design of the UI was subject to the following interwoven constraints, all aimed at boosting intuitiveness and maximizing the UX. First, users should be provided with a limited set of interaction modalities in order to reduce their cognitive load[2]. Second, the system and the UI design should foster the learnability and memorability of the interface. Third, every type of interaction should include some sort of prompt and clear feedback from the UI; feedback is necessary to (i) inform the users when an interaction is available, (ii) allow them to predict the result of this interaction, and (iii) notify the success/failure in performing a task.

According to the aforementioned constraints, the set of required interactions was limited to object selection & dragging and to voice interaction. These interactions build upon the ability to select the object to interact with. Selection is controlled by the user's gaze. An highlight was added to any interactable object to signal users when it is "active", i.e., selected (Fig. 3.6).

Actions can be taken on the selected object either by means of a tapping gesture (an "air tap") or by pressing the button of the Clicker, a peripheral device of the HoloLens. In both cases, a selection event is triggered and proper per-object audio and/or visual feedback is provided. As an example, a specific sound underlines when a broken glass is selected to be removed from the scene (as requested by the procedure).

Users can also interactively move objects around the environment by dragging them. Dragging is performed as follows: once the object has been selected, its position is "hooked" along the current gaze direction at a suitable distance from the user; the drag is then completed by "clicking" with the air tap or with the Clicker. As an option, it is also possible to add a target in the environment to indicate where the object should be dropped. This is done, for instance, to indicate in learning mode where to place AED pads.

The last interaction considered is based on voice and it is used in two different

---

[2] In UX design, the term cognitive load refers to the amount of mental resources that is required to operate a system.

Figure 3.6: Highlighting of a gaze-selected, or -targeted, object (the white dot is the gaze cursor).

ways. The first one is to recognize phrases in simulated dialogues, which have to be managed, for instance, when the user calls the EMS or asks a bystander (NPC) to get an AED. This modality is first activated by selecting a voice-responsive object (e.g., a telephone box or a bystander). When the volume of the user's voice is higher than a suitable threshold, the beginning of a phrase is detected; when the volume returns below the threshold, the end of the phrase is marked. Although extremely simple, this method is indeed effective for the task at hand. The second use of the voice is to recognize two simple keywords, "yes" and "no", which are used to reply to direct questions asked by the system (i.e., to check if the victim is conscious or shows vital signs). Using only these two simple keywords makes sure that the user's commands are interpreted unambiguously. In both cases, a small volume meter in the UI indicates when voice interaction is enabled; proper per-object feedback is provided when voice interaction is completed.

Interaction Training

Users can get acquainted with the interaction system through a training session. All interaction types are introduced individually (and incrementally) in the following order: targeting, air tap, Clicker, drag, voice interaction. Users first receive detailed instructions on the actual interaction method they are going to experience through voice and visual clues (animated GIFs). Then, they are requested to try each interaction at least two times. Every interaction element is characterized by an icon, which is introduced in this step and then used in the learning phase as a reminder of the types of interaction required to complete an action (Figure 3.7).

Users can repeat the interaction training until they become confident with the various methods. Since starting any session requires users to push a button in the

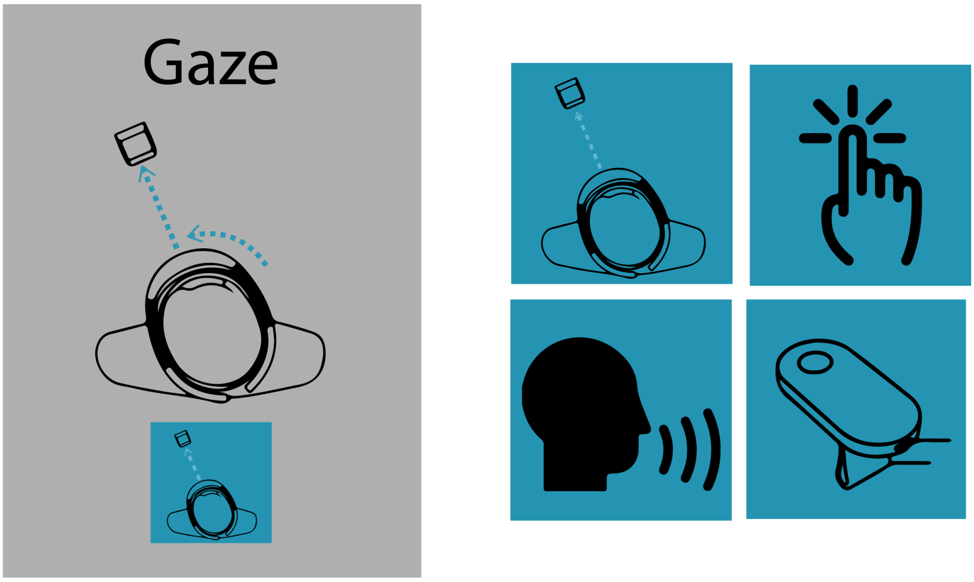


Figure 3.7: An example of the panel introducing an interaction element and of the icons characterizing various interaction types.

main menu, which is an ability they might not have learned yet, if the system does not detect any user interaction within a predefined time interval, it automatically starts a short voice tutorial and then asks the learner if it should activate the interaction training session.

CPR management

The quality of a cardiac massage can be evaluated by measuring a set of variables like compression frequency (and variance), depth and chest recoil. These data are relevant for all the Holo–BLSD sessions since (i) they can provide users with immediate feedback of the correctness of their actions (in learning and rehearsal), and (ii) they can be used to measure and evaluate the learner performance (in rehearsal and evaluation). Since the Holo-BLSD has been envisioned as a low-cost self-learning tool, it should not rely for such measurements on the availability of professional training manikins (or external devices) and their software. However, this is a severe drawback for the devised system, which asks for the implementation of alternative methods to obtain the same data in a reliable way using merely the sensing capabilities of the Hololens (if possible). In the following, we briefly describe the design and implementation of a method for directly measuring frequency and variance of the cardiac massage with the Hololens. These variables are used to provide a real-time on-screen feedback of the actual chest compression quality.

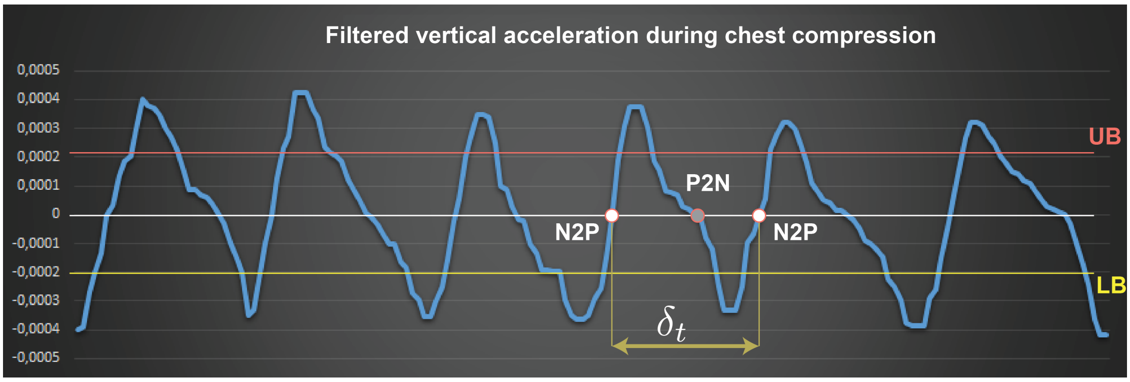


Figure 3.8: Plot of the filtered values of the horizontal head acceleration during chest compression: The plot highlights also the lower and upper bound and the length δ_t corresponding to the interval between two consecutive N2P crossings.

The measurement of the above variables is based on a heuristic method, which proved to be very reliable. First, the vertical acceleration of the head as the second derivative of the position provided by the Hololens is computed. Noise is removed by applying a median filter over the last five samples recorded. During chest compression, these values are characterized by a sequence of zero-crossings (Figure 3.8), where the lowermost head elongation corresponds to a negative to positive (N2P) cross and the uppermost elongation to a cross in the opposite direction, positive to negative (P2N). The positive and negative acceleration peaks are, on the contrary, somewhat related to the compression depth.

On the basis of these observations, an N2P crossing is identified as a valid compression iff (i) the time distance (δ_t) from the previous N2P crossing recorded is lower than a threshold ϵ_I, and (ii) the two acceleration peaks in the interval are, respectively, lower and higher than two suitable bounds LB and UB. Since the application is running, on average, at 50 frames/sec, the maximum delay with which a valid compression is identified is around 40 ms.

In order to guarantee robustness and independence from the user's characteristics,suitable values of the method hyperparameters (ϵ_I, LB and UB) were estimated through preliminary tests which involved both expert and non-expert users of different sex, age, weight and height. Afterwards, system was validated through further tests, by involving only expert users who were required to perform 50 compressions. This process was repeated until obtained 100% correct compression detection was obtained.

Work was started also for developing similar methods to analyze compression depth and chest recoil, but their assessment asked for the use of professional hardware capable of extracting validated measures to be used as a ground truth. Unfortunately, activities could not be completed before the start of the experimental phase; thus, it has been left as future work. Alternative approaches could have been

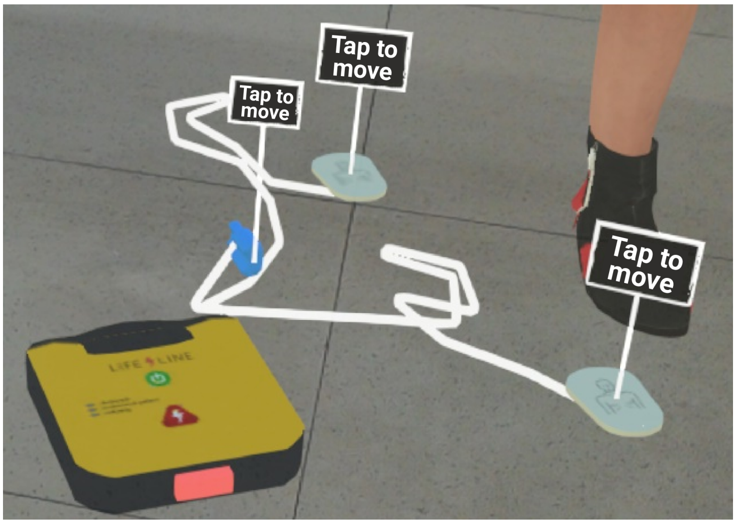


Figure 3.9: Virtual AED: Tooltips attached to the defibrillator paddles.

to implement methods based either on low-cost motion capture devices (like in [278, 319, 277, 233, 146]) or on off-the-shelf devices readily available to learners, such as a smartphone [121]. However, even these solutions are not optimal, since they complicate the setup for non-expert users. Thus, again, the possible integration of such systems will be considered in future works.

3.1.4 Holo-BLSD Learning Path

As said, the learning path of Holo-BLSD includes three modes, which are designed to allow learners to (i) receive step–by–step instructions on the procedure they have to perform (learning mode), (ii) train on that procedure in the context of a serious-game (rehearsal mode), and (iii) get an automatic assessment of their BLSD skills (evaluation mode). In the following, the three modes are described in detail. A debriefing phase is also discussed.

Learning Mode

In a learning session, users are guided through the various actions of the given emergency procedure. Each action is introduced by visual and audio hints, aimed at explaining learners what they have to do, why, which are the objects they have to interact with and through which mechanics.

Objects and hints are introduced one at a time, with the aim to keep users' cognitive load low. Objects are presented as interactable (a glow effect is added when they are gazed) and provide consistent feedback in response to users' operations. Icons and other graphic signs (e.g., tooltips) are exploited as remainders of the interaction required by/allowed on a given object (Fig. 3.9). When dragging

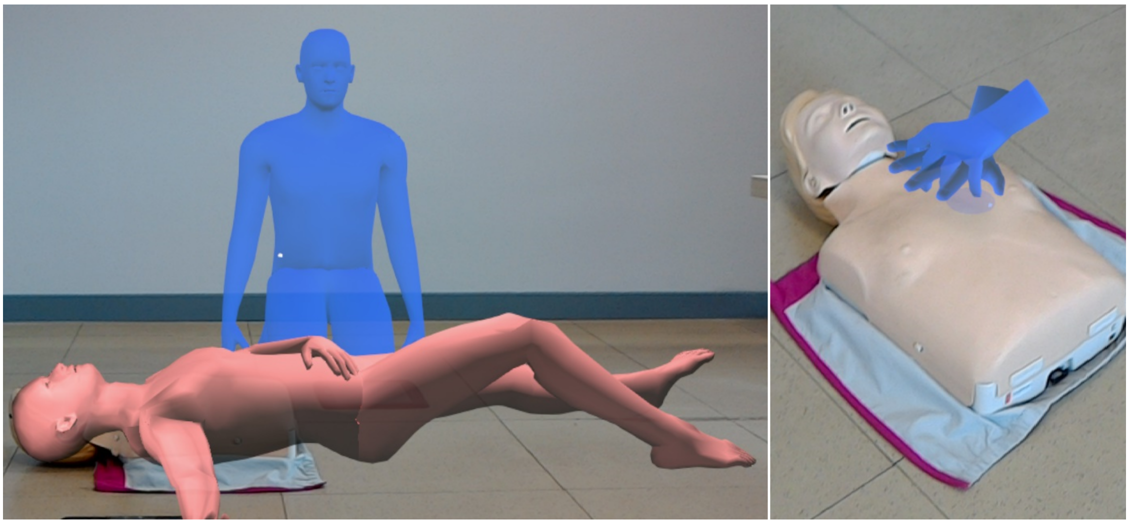


Figure 3.10: Visual hints indicating the correct learner's position (left) and hands placement on the upper chest (right) when executing the CPR.

needs to be used on an object, a virtual target is displayed to indicate where it can be dropped. Since interaction is situated in a full 360° space, voice prompts were integrated to encourage users to look around and explore the whole environment, thus helping them to find possible off-screen objects.

As an example, when an AED has to be requested and no actions are recorded within a certain time interval, the learner is guided to look in the direction where a bystander can be found.

In the BLSD training, learning how to perform a correct chest compression is crucial; thus, this action received a particular attention. Learners are first instructed to position themselves in the correct way. This goal is accomplished with the help of an avatar, which shows the learners how to kneel beside the victim's upper chest and to place their hands in the proper place (Fig. 3.10). Audio and visual instructions explain all the details of an effective chest compression. An audio feedback is provided for every chest compression detected by the system, and a metronome at 110 BPM helps learners to keep the right rhythm. Learners also receive a feedback on the actual quality of the cardiac massage in terms of compression frequency, which is displayed on a virtual panel placed in front of them. The method used to monitor the chest compression is discussed in the Appendix. It is worth noting that compression frequency data are relevant for all the Holo-BLSD modes; besides serving to provide users with an immediate feedback on the correctness of their operations in the learning and in the rehearsal modes, they are used to measure and automatically evaluate the learners' performance in both the rehearsal and evaluation modes.

The simulated clinical case (victim unconscious, conscious or showing vital signs) is generated randomly at every given session with a non-uniform distribution favoring the most complex case (victim unconscious not showing vital signs, which is proposed in 70% of the sessions). This way, learners are forced to experiment how to deal with different emergency situations characterized by varying difficulty.

Rehearsal Mode: A BLSD Gamification Approach

In this mode, users can train the procedures they have learned in the previous mode. Considering the relevance of gamification interventions in educational contexts (discussed in section 3.1), the rehearsal phase has been designed according to (i) the guidelines proposed in the literature [127, 48] and (ii) the results described in [170], which report the effectiveness of two game design patterns used to learn BLS procedures, namely a timer pattern and a score pattern. In this phase, the learners' goal is to maximize their score by performing the sequence of actions required for the procedure under consideration in the right way and in a timely manner.

Therefore, this rehearsal phase leverages a reward system, which is one of the possible game mechanisms that can be included in gamification practices [127]. This system uses a scoring system that relates to the correctness and speed of the steps performed in the procedure, as follows. Players can perform any action, but they cannot benefit from the visual and auditory aids available in the learning mode. The game logic maintains an action timer and verifies the completion of the action using an internal scheduler. Players receive points for completing an action in a reasonable time and in the correct order, as well as for performing actions correctly (e.g., performing chest compressions at the correct frequency). When the time for completing a particular action expires, the player loses a certain number of points per second. To motivate players to improve, at the end of the game they can see their overall ranking based on their final score (i.e., players are motivated by the challenge to beat their peers'). As shown in [170], the visual and auditory presence of a timer (in conjunction with the game score) had positive effects on users' knowledge growth. Therefore, the timer (and current score) are displayed on a panel placed in the virtual environment to provide auditory and visual feedback on specific game events. For example, an alarm sound signals the expiration of the timer, while another sound informs players when new points are obtained.

Finally, as [48] points out, an important aspect of gamification experiences is the acceptance of failure, i.e. failure should not be a punishment but a moment of reflection and understanding of mistakes to drive future improvements. Therefore, in Holo-BLSD's rehearsal phase, players may make mistakes of varying severity, which are treated as either warnings or errors and immediately reported by the system. For example, starting chest compressions before you have secured the scene will be treated as a mistake, while doing so before asking for an AED will result in a warning. Warnings only result in a point deduction, while errors result in the session being terminated (to ensure that actions are only performed in the correct manner).

Evaluation Mode

The aim of this mode is to enable the assessment of learners' BLSD skills. Trainees are presented with the same emergency situation experienced during learning and rehearsal sessions, and are asked to complete all the required actions without any audio or visual help from the system. As in the rehearsal mode, learners are free to execute actions in any order, with the difference that in the evaluation mode warnings and errors contribute (negatively) to the final assessment without compromising the session (which is not terminated).

Evaluation is based on the assessment form reported in Table 3.2 (which will be discussed in Section 3.1.5), whose items are rated based on actual users' activity. In the experiments performed in this work, the clinical case always encompassed an unconscious, non-breathing victim, in order to evaluate the learners on the most complex and challenging case available. Nonetheless, the evaluation mode can, in principle, consider any of the clinical cases and scenarios available.

Session Debriefing

The Holo-BLSD tool includes a debriefing companion application, which relies on the analytics collected during the rehearsal and evaluation sessions. The availability of a debriefing step is extremely relevant for knowledge retention, since it helps learners to reflect on what they did, get insights from their experience and make meaningful connections with the real world, thus enhancing transfer of knowledge and skills. Even when results are not as successful as the learners hoped, debriefing can still promote active learning by helping them to analyze mistakes made and explore alternative solutions [118].

The Holo-BLSD debriefing application is a server component that runs on a PC and allows for the visualization of an interactive Gantt-like chart showing the begin and end of each action, as well as all the events related to it (e.g., the removal of a debris, the placement of an AED pad, the detected chest compressions and so forth). The chart can be synced with a video recording of the session, captured from the Hololens: this feature can be used to visually inspect a particular action in detail and to possibly create a repository of training events.

3.1.5 Experiments

In order to analyze the suitability of the proposed tool, a user study was performed by involving 58 volunteers selected among Health-care Nursing first year students of the University of Eastern Piedmont in Novara, Italy. Volunteers underwent either a traditional instructor-led course or a self-training delivered by the developed AR-SG. The main aim was to investigate:

- to what extent the devised AR-based training course can stimulate learners' attention compared to traditional training;
- the usability of the devised tool from multiple perspectives, including, among others, the learnability of the interaction techniques, the realism of the simulation and the flexibility of the experience;
- how the employed gamified systems have affected users' motivation and the learning process
- how results obtained by learners in the AR training course compare to those obtained with traditional training;
- whether automatic assessment of learners' performance made by the tool is consistent with the evaluation performed by examiners through visual inspection.

In the following, the methodology adopted for the study will be first introduced, by also presenting the objective and subjective (questionnaire-based) metrics devised to analyze the above aspects. All the questionnaires and video material related to the experiments are available for download[3].

Methodology

All the volunteers were first provided with an introduction to the BLSD procedure, which was delivered by a medical instructor with the support of ad hoc video contents. Afterwards, volunteers were randomly split in two groups, later referred to as the "traditional training" and the "AR training" groups.

Traditional training group. This group included 29 volunteers aged between 19 and 47 (M = 21.34, STD = 5.59), 5 males and 24 females. One of the volunteers had a previous knowledge of the BLSD procedure and was not included in the evaluation.

Learners were organized in small groups of 2-3 people. First, an instructor showed them how to carry out the whole procedure, which includes the 21 actions listed in Table 3.2. All possible situations were illustrated (i.e., conscious, unconscious but breathing and unconscious, non-breathing victim). Afterwards, the instructor invited the learners to individually carry out all the actions, by answering possible questions and correcting them as needed. This phase was repeated two times. Then, each learner was asked to perform the procedure alone, without any instructor's feedback. In case of a severe error, the learner was stopped and asked to repeat the session until the procedure was successfully completed.

---

[3] `http://tiny.cc/holo-blsd`

Table 3.2: List of Actions of the BLSD Procedure and Evaluation Criteria.

Action	**Evaluation criteria**
1. Scene safety	In the scene there are always 3 objects to remove. If no object is removed by the learner, 0 points; 1–2 objects removed, 1 point; 3 objects removed (action completed), 2 points.
2. LOC evaluation	Learner is requested to assess level of consciousness by shaking the victim and ask if he or she is all right with loud voice. Action is completed only if both the operations have been accomplished: for instance, calling the victim without shaking him or her (or vice versa) means that action is not completed (1 point).
3. Vital signs evaluation	Learner has to observe the victim at least for five seconds to see whether he or she is breathing or not (2 points), otherwise the action is considered as not started (0 points).
4. Emergency call	Action is assigned 2 points is the call is completed. If the learner makes errors in informing the medical services about victim's LOC and vital signs, score is reduced to 0.
5. Get AED	If the learner asks bystanders for the defibrillator, 2 points, otherwise 0 points (the lack of the defibrillator will make it impossible for the learner to start and complete steps 12–15 and 20–21).
6. Clear chest	If the learner clears victim's chest from the obstructing arm (which blocks CPR), 2 points, otherwise 0 points.
7. 1st CPR – Start	If compression is initiated within 30 second after having asked for the defibrillator, 2 points, otherwise 0 points.
8. 1st CPR – Rate	Score is assigned based on average compression rate (BPM): 95 <BPM <125, 2 points; 80 <BPM <95 or 125 <BPM <140, 1 point, BPM <80 or BPM >140, 0 points.
9. 1st CPR – Depth†	If the compressions are 5 cm deep, 2 points, otherwise 0 points. Not assessed by the AR tool.
10. 1st CPR – Expansion†	If the chest returns in a neutral position after each compression, 2 points, otherwise 0 points. Not assessed by the AR tool.
11. AED turned on	If the defibrillator is turned on, 2 points, otherwise 0 points.
12. Paddles placed	If the defibrillator's paddles are placed correctly on the victim's chest, 2 points, otherwise 0 points.
13. Paddles plugged in	If the paddles' connector is plugged in to the defibrillator, 2 points, otherwise 0 points.
14. 1st security protocol	Defibrillation security protocol requests to move away from the victim, use loud voice to ask bystander to do the same, look around to make sure that nobody approaches the victim. The AR tool only assesses the last two operations: if both of them are performed, 2 points; if just one of them is performed, 1 point; otherwise, 0 points.
15. 1st defibrillation	If the defibrillator is discharged after having been invited to do that by the device (ready signal), 2 points. If the defibrillator is never discharged, 0 points.
16. 2nd CPR – Start	If compression is initiated within 30 second after having discharged the defibrillator, 2 points, otherwise 0 points.
17. 2nd CPR – Rate	Same as for action 8.
18. 2nd CPR – Depth†	Same as for action 9.
19. 2nd CPR – Expansion†	Same as for action 10.
20. 2nd security protocol	Same as for action 14.
21. 2nd defibrillation	Same as for action 15.

Finally, the evaluation was started. As said, the clinical case simulated always encompassed an unconscious, not breathing victim. This way, the whole BLSD procedure could be assessed, including the use of the defibrillator. During the evaluation, the learners were requested to carry out the BLSD procedure autonomously. An examiner observed the operations without interacting with the learners and assigned a score to every action in the list (evaluation criteria are reported in Table 3.2 and will be discussed in detail in Section 3.1.5).

AR training group. This group included 29 volunteers, aged between 19 and 34 (M = 20.76, STD = 2.85), 13 males and 16 females. One of them had a previous knowledge of the BLSD procedure and was not included in the evaluation. Concerning technology awareness, 6 of them have had a previous experience with VR (3 of them using hand controllers); similarly, 6 of them have used already AR applications (4 on mobile devices, 1 on wearable devices and 1 on both the devices). Out of the 28 volunteers involved in the evaluation, only 26 filled in the questionnaire used to collect subjective measurements.

Learners were first given time to familiarize with the HoloLens by running the interaction training (Section 3.1.3). The session was repeated twice to let learners get acquainted with both the air tap gesture and the Clicker device. Then, they were allowed to choose the interaction means they preferred.

After that, volunteers were invited to run the learning session, where the actual BLSD training is delivered (Section 3.1.4). Differently than in the traditional training, there was no interaction with a human agent. Learners were allowed to execute this session at least twice. In fact, it is not uncommon for users who experiment AR for the first time to simply disregard provided instructions while they explore the virtual world. During the session, different clinical cases (conscious victim, unconscious breathing and unconscious non-breathing) were randomly selected. It was guaranteed that each volunteer experimented at least once the clinical case considered in the evaluation, which was the unconscious non-breathing case like in the traditional training.

At the conclusion of the learning session, volunteers were asked to engage in the rehearsal session (Section 3.1.4). Like in the traditional session, high-severity errors terminated the session and learners had to repeat rehearsal until they were able to successfully complete a session that contemplates the evaluation case. Finally, volunteers entered the evaluation session, where they had a six-minute time limit to complete the procedure; afterwards, the session was automatically terminated. Like with the traditional training group, during the evaluation session an examiner observed the learners and assessed their performance against the score sheet in Table 3.2.

Additionally, the same evaluations were collected automatically by the AR-based tool for most of the actions (as it will be shown in Section 3.1.5). This way, it was possible to study to what extent the devised tool is able to replicate the examiner's evaluation and, hence, serves as a reliable self-assessment tool.

Evaluation Metrics

Evaluation encompassed both objective and subjective measurements. Metrics adopted are reported in the following sub-sections, whereas results obtained are discussed in Section 3.1.6.

Objective measurements. As introduced in Section 3.1.5, learners' performance was evaluated in objective terms against the score sheet in Table 3.2. In particular, each action in the BLSD procedure could be assigned a score from 0 to 2, with 0 meaning that the action was not started, 1 that the action was started but not completed (when a partial completion is possible), and 2 that the action was completed.

These scores could be modified to account for errors made by the learner. In general, errors correspond to actions that are started earlier than expected. When this is the case, if the action was completed its score is lowered from 2 to 1 (if action was not started or not completed, score remains unchanged). As an example, suppose that the learner calls the EMS before securing the scene. Since the first action follows the second in the procedure, the score of the two actions is lowered by 1 even though they have been both completed.

Actions concerning the CPR were scored in a different way. In particular, for actions 7 and 16 the timely start of chest compression was evaluated: if the compression was initiated within 30 seconds after the conclusion of the previous action, 2 points were assigned, otherwise 0. For actions 8 and 17, the quality of the chest compression rate was considered, and score was assigned based on the distance between the observed/measured rate and the advised one.

It is worth recalling that BLSD actions were evaluated using the same metrics by both the examiner (through visual inspection) and the AR-based tool (in an automatic way). However, as explained in the Appendix, at present the AR-based tool is not able to measure the CPR compression depth and chest recoil and, thus, it cannot assess actions 9–10 and 18–19 (marked by † in Table 3.2). Hence, when comparing the examiner's and the AR tool's evaluations, these actions were not considered and the maximum score that could be reached was 34 (17×2).

The learner passes the examination if he or she obtains at least 60% of the available points (i.e., 21 or higher) and no major error was made. Major errors are as follows: none of the CPR phases was ever started; security procedure was never started; emergency call was never made; defibrillator was never requested; defibrillator was never switched on; paddles were never placed on the victim's body nor plugged into the defibrillator; defibrillator was never discharged.

Subjective measurements. The suitability of the devised AR-based tool (and its contents) for self-training in the context of BLSD was analyzed through a questionnaire based on the Instructional Materials Motivation Survey (IMMS) [171]. This questionnaire was delivered to learners in both the AR training and traditional training groups. To this aim, statements in the original IMMS were slightly adapted to match the way training was actually delivered and to the material used.

A usability questionnaire was then delivered to the sole users of the AR-based tool to identify possible issues and drive future developments in the field. The questionnaire was organized in four sections. The first two sections analyzed usability in broad terms by considering the System Usability Scale (SUS) [43] and the five attributes defined by Nielsen [220], i.e., learnability, efficiency, memorability, (possibility to recover from) errors and satisfaction. The third section explored ergonomics aspects concerning the interaction with the device through statements derived from the ISO 9241-400 standard [153]. The fourth section explored in detail a number of usability aspects concerning virtual/synthetic environments defined in [166]. In particular, the questionnaire focused on user input (gaze, gestures and voice), sense of immersion/presence, system output (display), user guidance and help, consistency, simulation fidelity, flexibility, error correction/handling, i.e., robustness, and overall usability. Aspects in [166] related to functionality were not considered since they were addressed already in previous sections.

Lastly, a further questionnaire was delivered again only to the learners in the AR training group to assess their perception of the gamified learning approach adopted in the rehearsal session.

3.1.6 Results

In this section, the results of the subjective evaluation concerning learners' motivation levels, usability of the AR-based tool and suitability of the gamified learning approach will be discussed first. Afterwards, objective measurements collected by examiners in the two groups will be compared, and agreement between manual and automatic evaluations will be determined.

Subjective Results

The IMMS includes 36 statements organized in four sub-scales, which are aimed to investigate learners' motivation levels based on several principles of instructional design, i.e., attention (12 statements), relevance (nine statements), confidence (nine statements) and satisfaction (six statements). With the exception of 10 reverted statements, the higher the score the learner gives to a statement, the higher his or her motivational score is. Statements are evaluated on a scale from 1 (strong disagreement) to 5 (strong agreement).

A scale reliability test was first conducted on the overall IMMS scale (36 statements). A standardized Cronbach Alpha equal to 0.876 ($n = 28$) and 0.919 ($n = 24$, 2 excluded) was calculated, respectively, for the traditional and the AR training groups, suggesting a good to excellent reliability of the results.

Overall motivation levels are reported in Table 3.3 on a 5-interval scale from low to high. For all learners except one in both groups, motivation was from medium-high to high (summing up to 96.43% and 96.15% of the respondents for, respectively, the traditional and AR training groups). These numbers indicate that both training methods were largely able to positively stimulate learners' motivation.

Table 3.3: Learner's Motivation Levels for the Two Groups: Overall Results (Number of Occurrences and Percentages).

Motiv. level	**Scores**	**Trad. training**	**AR training**
High	4.21–5.00	18 (64,29%)	13 (50.00%)
Medium-high	3.41–4.20	9 (32,14%)	12 (46.15%)
Medium	2.60–3.40	1 (3,57%)	1 (3.85%)
Medium-low	1.80–2.59	0 (0%)	0 (0%)
Low	1.00–1.79	0 (0%)	0 (0%)

Averaged results for individual IMMS sub-scales (attention, relevance, confidence and satisfaction) are summarized in Table 3.4. Analyzing data using unpaired t-tests, the only differences that can be considered as statistically significant ($p < 0.05$) are those related to the confidence and satisfaction perspectives. In particular, AR learners reported a confidence and satisfaction lower than that of users in the traditional training group. These findings could be explained by the fact that the questionnaire asked respondents to focus only on course material. However, learners in the AR training group had to face the difficulties posed both by the BLSD contents and the use of a new tool and, in some cases, it was difficult for them to isolate content- from technology-related aspects.

Although the differences are not significant, slightly higher results can also be observed in the other two sub-scales for learners in the traditional training group.

For instance, considering the attention perspective and focusing on statistically significant statements, it can be noticed that learners in the AR training group found that "the amount of repetition caused them to get bored sometimes" and said that "the amount of information was so high that it was irritating". These outcomes are not surprising. In fact, the AR-based tool is currently designed to present material always in the same way (in terms of both visual and audio contents), independent of the fact that concepts have been assimilated or not. Moreover, in the experiments, learners were requested to carry out the BLSD procedure several times with only slight modifications (concerning victim's conditions and the use of gamification elements). This fact indeed contributed at making them perceive contents as repeated.

Similar considerations hold for the relevance dimension. For instance, higher scores assigned by learners in the AR training group to statements like "there are explanations or examples of how people use the knowledge of this course" can be easily explained by the lack of such contents in the AR-based procedure (whereas they could be provided by the instructor in the traditional training course, at least upon learners' request).

Despite these differences, it shall be noticed that, for all the scales, average levels in the AR training group are in the medium-high interval (according to the 5-interval scale used in Table 3.3), which confirms the important level of motivation that the proposed tool can guarantee.

Table 3.4: Learner's Motivation Levels for the Two Groups: Results for Individual Sub-scales (Mean Values and Standard Deviations Reported). Rows marked with "*" show a statistically significant difference ($p < 0.05$) among the groups

Sub-scale	***p*-value**	**Trad. training**	**AR training**
Attention	0,061	4.51 (0.52)	4.25 (0.47)
Relevance	0,230	4.39 (0.40)	4.24 (0.82)
Confidence*	0,013	4.15 (0.42)	3.83 (0.48)
Satisfaction*	0,045	4.57 (0.54)	4.23 (0.66)

As previously introduced, usability was assessed through a questionnaire (delivered only to the AR training group) that included four sections based on (i) SUS, (ii) Nielsen's attributes of usability, (iii) the ISO 9241-400 standard about man-machine ergonomics and (iv) the VRUSE tool. All items had to be scored on a scale from 1 to 5 (with the same meaning of the IMMS tool).

Concerning SUS (first section), learners were asked to score the 10 items in Table 3.5. The normalized result (with odd items reverted) equal to 72.03 in the 0–100 range can be regarded as an indication of the usability of the designed tool (according to [45], a score above 68 shall be considered as above average).

Table 3.5: Statements in the SUS Tool (Mean Values and Standard Deviations Reported).

Statement	**Score**
I think that I would like to use this system frequently	3.89 (1.05)
I found the system unnecessarily complex	1.67 (0.92)
I thought the system was easy to use	3.89 (0.93)
I think that I would need the support of a technical person to be able to use this system	2.78 (1.09)
I found the various functions in this system were well integrated	3.63 (0.93)
I thought there was too much inconsistency in this system	1.48 (0.75)
I would imagine that most people would learn to use this system very quickly	3.44 (1.19)
I found the system very cumbersome to use	1.74 (0.90)
I felt very confident using the system	3.67 (1.00)
I needed to learn a lot of things before I could get going with this system	2.04 (1.09)

Similar conclusions can be drawn by considering Nielsen's attributes of usability (second section). Mean scores were as follow: learnability 3.63 (SD = 1.01), efficiency 3.93 (SD = 0.73), memorability 3.89 (SD = 1.01), possibility to recover from errors 3.04 (SD = 0.85) and satisfaction 4.44 (SD = 1.01). The low score assigned to the possibility to recover from errors can be explained by the fact that learners were generally not allowed to recover from mistakes made with the AR tool. In fact, these mistakes could correspond to errors in the BLSD procedure, which had to be recorded and evaluated as explained in Section 3.1.5.

As for ergonomics (third section), learners were asked to evaluate the interaction with the HoloLens based on the four statements in Table 3.6. Scores are in the medium-low to medium range, suggesting that developments are still required in the field of head-mounted AR devices.

Table 3.6: Statements from the ISO 9241-400 Standard (Mean Values and Standard Deviations Reported).

Statement	**Score (SD)**
The wearable device is very cumbersome / heavy	2.15 (1.13)
Mental effort required to operate the device is very high	2.41 (1.05)
Physical effort required to operate the device is very high	2.26 (0.86)
I would feel comfortable using the system for long times	3.07 (1.04)

The interaction with synthetic and virtual environments (fourth section) was analyzed through statements, adapted from VRUSE, belonging to 10 categories. For each category, learners were asked to additionally express their overall level of satisfaction on a scale from 1 (poorly satisfied) to 5 (very satisfied). Categories and overall scores are given in Table 3.7.

According to the scale used in Table 3.3, all the categories received medium-high to high scores. This fact indeed represents a further confirmation of the appreciation and the usability level reached by the devised AR tool. Despite that, interesting insights for driving future developments can be obtained from the statements that obtained the lower scores in each category (although it is worth observing that worst scores were in the 2.60–3.40 medium range).

With respect to user input, learners found that the modality based on finger gestures was not ideal for interacting with virtual elements. Nevertheless, this fact was not particularly critical, since learners were allowed to choose between gesture- and Clicker- based interaction. However, similar concerns were raised also for voice interaction. In fact, learners stated that they did not feel to have always the right control on what they wanted to do while interacting with the system, mainly because of language recognition/understanding issues.

Learners stated that being immersed in the virtual experience was important for completing the assigned task, confirming the importance of using AR to create a simulated scenario. Regrettably, they also indicated that the characteristics of the

Table 3.7: Categories from the VRUSE Tool (Mean Values and Standard Deviations of the Overall Satisfaction Reported).

Category	**Score (SD)**
User input (gaze and taps)	3.58 (0.76)
User input (voice)	3.77 (0.76)
Sense of immersion/presence	3.54 (0.81)
System output (display)	3.96 (0.65)
User guidance and help	4.00 (0.78)
Consistency	3.81 (0.83)
Simulation fidelity	3.89 (0.93)
Flexibility	3.73 (0.78)
Error correction/handling and robustness	3.58 (0.81)
Overall usability	4.36 (0.64)

screen, in particular its FOV, partially reduced their sense of immersion (confirming that further advancements in wearable AR technology are needed).

Concerning system output, learners rated graphics quality as appropriate for the task, though realism was not judged as particularly high. They also stated to be able to read and understand the information displayed by the AR application. None of them experienced motion sickness or eye fatigue.

Although they did not find it difficult to learn how to use the system and to use it, learners said they needed external help while using it (notwithstanding, it is worth recalling that they had to carry out the learning, rehearsal and evaluation sessions without any human intervention).

Learners found that, in general, the system behaved in a consistent way, but in some cases the meaning of visual and audio cues was not as straightforward as they were expecting.

Concerning simulation, they found that the BLSD procedure was simulated with the appropriate fidelity and that simplifications that were possibly implemented did not impact their performance. However, they felt that virtual elements did not always behave and move in a natural way.

Scores regarding flexibility are not very high. In fact, learners felt that interaction modality sometimes interfered with their activity. Thus, they had to adapt their behavior to the system, and this fact sometimes prevented them to achieve exactly the intended result. They also found that the system lacks shortcuts to perform given operations.

With respect to error correction/error handling and robustness, learners lamented the fact that the system is not adopting strategies able to prevent them to make silly mistakes and that it is difficult to recover from errors. Although strategies could be devised to deal with errors deriving from interaction issues, it is worth recalling that means to recover from procedural errors were intentionally avoided (see discussion above).

Lastly, considering overall usability, learners found that the system's responsiveness partially impacted on their performance (time is needed to advance in the simulation, to react to user's interactions, etc.). Learners also stated that the system does not presently do all what they would expect. Indeed, motivations for this result could be identified in the issues discussed for previous categories.

As for the various problems highlighted in the previous paragraphs, it is worth saying that a possible way to tackle them could be to introduce in the AR tool some of the features that have been considered positively by the traditional training group. Examples of such features are the possibility to interact with a question-and-answer mechanism (e.g., mimicking the presence of a human instructor through conversational agents) and to let the learners' tailor training to their actual needs (e.g., by selecting particular actions in the procedure to experiment with), or the adoption of different ways to present the same content (to avoid repetitions), and so on.

Concluding, the subjective assessment of Holo-BLSD confirmed that the proposed AR-based learning path is able to stimulate learners' attention to levels similar to those achieved with traditional training. Results also demonstrated the usability of the devised tool. Nevertheless, the same results also allowed us to identify aspects that shall be considered to enhance the suitability of the proposed AR-based BLSD training in terms of both contents and technology (e.g., avoiding repetitions, improving interaction, etc.).

Objective Results

Objective results collected during the experiments are summarized in Table 3.8. The first column reports the actions of the BLSD procedure. The second and third columns tabulate scores assigned by the examiners in, respectively, the traditional (TE) and AR (AE) training groups (mean values and standard deviations reported). Statistical significance of the differences between TE and AE scores is measured using unpaired t-tests ($p = 0.05$). The fourth column provides calculated p-values. The fifth column (AT†) reports the automatic scores assigned by the tool in the AR course (actions that cannot be assessed so far are marked with † in the first column). The sixth column reports p-values calculated on the comparison between examiner's scores for the traditional training course (TE†) and automatic scores for the AR training group (AT†). Finally, the last column provides inter-rater agreement between the examiner (AE†) and the tool (AT†) when scoring the AR training group. These agreements were calculated using weighted Cohen's k. The last two rows of the table report the overall evaluation results, considering or not the actions that cannot be assessed automatically by the tool.

Table 3.8: Objective Results for Individual Actions and for the Whole BLSD Procedure.

Action	**Trad.–Ex. (TE)**	**AR–Ex. (AE)**	***p* (TE–AE)**	**AR–Tool (AT†)**	***p* (TE†-AT†)**	***k* (AE†-AT†)**
1. Scene safety	1.97 (0.19)	2.00 (0.00)	0.322	1.97 (0.19)	1.000	0.926
2. LOC evaluation	2.00 (0.00)	1.86 (0.35)	0.039*	1.93 (0.26)	0.161	–0.101
3. Vital signs evaluation	1.97 (0.19)	1.72 (0.45)	0.011*	1.93 (0.26)	0.562	–0.124
4. Emergency call	2.00 (0.00)	2.00 (0.00)	–	2.00 (0.00)	–	1.000
5. Get AED	1.86 (0.35)	1.90 (0.41)	0.732	1.90 (0.41)	0.732	1.000
6. Clear chest	1.93 (0.26)	1.97 (0.19)	0.561	1.90 (0.31)	0.647	0.055
1st CPR – Start	2.00 (0.00)	1.90 (0.31)	0.078	1.93 (0.37)	0.326	0.374
1st CPR – Rate	1.90 (0.31)	1.79 (0.41)	0.285	1.69 (0.66)	0.135	0.329
1st CPR – Depth†	1.83 (0.38)	1.76 (0.51)	0.564	–	–	–
1st CPR – Expansion†	2.00 (0.00)	1.90 (0.41)	0.179	–	–	–
AED turned on	1.97 (0.19)	1.86 (0.44)	0.249	1.83 (0.47)	0.149	0.877
Paddles placed	1.97 (0.19)	1.83 (0.47)	0.146	1.79 (0.62)	0.161	0.699
Pladdles plugged in	1.86 (0.44)	1.86 (0.44)	1.000	1.93 (0.37)	0.977	0.651
1st security	1.86 (0.35)	1.79 (0.56)	0.576	1.72 (0.65)	0.320	0.836
1st defibrillation	2.00 (0.00)	1.66 (0.77)	0.019*	1.72 (0.70)	0.043*	0.869
2nd CPR – Start	2.00 (0.00)	1.66 (0.77)	0.013*	1.66 (0.72)	0.023*	0.879
2nd CPR – Rate	1.76 (0.51)	1.48 (0.78)	0.118	1.45 (0.74)	0.021*	0.620
2nd CPR – Depth†	1.62 (0.56)	1.59 (0.73)	0.841	–	–	–
2nd CPR – Expansion†	1.93 (0.37)	1.66 (0.72)	0.072	–	–	–
2nd security	1.45 (0.74)	1.38 (0.90)	0.751	1.48 (0.87)	0.299	0.873
2nd defibrillation	1.62 (0.78)	1.52 (0.87)	0.635	1.52 (0.87)	0.055	1.000
All actions	39.48 (2.50)	37.07 (7.07)	0.088	–	–	–
All actions except †	32.11 (2.28)	30.34 (5.60)	–	30.17 (5.41)	0.109	0.794

As a first comment, it can be seen that the average overall examiner scores of the two groups (second and third columns) are rather close (39.48 for the traditional training group, 37.07 for the AR training group, on a maximum score of 42, and that their difference is not statistically significant. This finding suggests that the learning outcomes achieved by the instructor-led and the AR-based courses are overall comparable, which is also a possible indication of the effectiveness of the proposed approach as a learning tool.

However, it is also worth observing that standard deviation is much higher for AR learners. This difference could be due to difficulties that (some of the) learners may have experienced in interacting with the AR tool, e.g., due to missed recognition of gesture and voice inputs. Another explanation could be the higher complexity for examiners to judge learners' performance. In fact, the execution of many actions required learners to interact with virtual elements and, even though examiners were allowed to see the point of view of the learners, delays due to data transmission made it difficult in some cases to fully appreciate their actual behavior. Lastly, this result may also indicate a lower capability of the AR tool to level learners' abilities, which could be due, among others, to the lack of mechanisms for adapting contents and their presentation to learners' actual needs.

Similar considerations can be made when analyzing the results of individual actions. The cases in which differences are significant are only four, namely LOC and vital signs evaluation, start of first defibrillation and of second CPR (rows marked with * in the fourth column). In some cases, differences can be explained again with the difficulty for examiners to judge learners' operation. For instance, in the LOC evaluation, learners were expected to shake the victim and call him or her loud; in some cases examiners judged the force applied or the voice level used as not appropriate. However, examiners could not rely on any quantitative information in the assessment. Furthermore, differently than in the traditional training group, users of the AR tool did not had a ground truth, since they had not seen the instructor execute those actions. In the evaluation of the victim's vital signs, learners' have to observe the victim for five seconds; even using measurement instruments, examiners tended to approximate actual time. While performing the defibrillation, in some cases examiners were not able to determine whether actions had been performed in the proper way due to the subtle movements involved or fine precision requested (e.g., in plugging paddles' connector, or placing paddles on the victim's chest). With respect to CPR, it was quite difficult for the examiner to determine when to start measuring time. In the above situations, the AR tool was able to assign a truly objective score, based on measurements collected by internal sensors and rules defined in the application (e.g., on interactions performed, threshold levels passed, and so on).

A support for these explanation can come from comparing the examiner's scores in the traditional course (TE†) and the AR tool scores (AT†). It can be seen that only the differences concerning the first defibrillation and the second CPR remain significant (sixth column, rows marked with *). However, also this result can be easily explained, since the AR tool is not able to assign a score for the second CPR if defibrillation was not completed (CPR time is measured from discharge).

It is also worth observing that, for both courses, there are a number of actions that received low scores. This outcome is particularly relevant, since it suggests those parts of the learning path that should be improved, no matter how the course is going to be delivered. Other interesting insights can be obtained about the effectiveness of Holo-BLSD as a self-evaluation tool. Comparing the overall scores assigned in the AR training group by the examiner (AE†) and the tool (AT†), it can be observed that mean values are comparable (30.17 vs. 30.34, with no significant difference), like standard deviations. This finding is confirmed by a Cohen's k value equal to 0.794 (last column), which suggests a quite high inter-rater agreement. When considering individual actions, it can be noticed that there are a number of situations for which inter-rater agreement is lower. Based on discussion above, this result can be largely explained with the difficulty in providing objective results for such actions.

With the aim to summarize this discussion, several insights can be obtained from the analysis of the objective results. First, by comparing scores assigned by human examiners in the traditional and AR-based courses, it was observed that learners achieved comparable results, thus confirming the suitability of the devised self-learning tool for training considered skills. Second, since scores automatically computed by the tool were found to be largely consistent with those assigned by the (human) examiners to the same learners, it can be concluded that the proposed system can also be regarded as a reliable instrument for self-assessment.

Effectiveness of the Gamification Systems

The analysis of the last questionnaire, whose ten items are reported in Table 3.9 with assigned scores, suggests a positive users' perception of the devised gamification approach. Item 1 highlights that learners had fun (4.31) while using the game to learn the intended content. Also, they considered it a beneficial tool for educational purposes (item 4, 4.19) and preferred it mainly to a more traditional approach featuring a human instructor (item 2, 2.08) or through books, notes, and slides (item 3, 1.73). When we analyze the single elements of the gamification approach, we can emphasize that the scoring system was the most appreciated by users. Trying to improve their result was evaluated as a critical factor to help them increase their learning (item 8, 4.04). This statement is further supported by item 6 (4.08), which shows that the possibility to repeat the game experience several times (we hypothesize to improve their result) was also a determinant for better understanding the correct sequence in which actions should be performed. The possibility to compare individual scores with those of other learners (item 5) had a slighter lower result (3.92), suggesting that users perhaps felt more engaged from competing with themself rather than with others. However, this result is still high, indicating that the leaderboard system was an effective design choice. Finally, the timer was the gamification element which reported the lowest score (3.42), and users had mixed feelings (SD = 1.27) whether this feature helped them in completing actions rapidly, which we underline is a critical skill to train for the BLSD procedure. However, we believe this result was not associated with stressful sensations prompted by the presence of the timer (item 10, 1.88). Finally, users emphasized that the education and entertainment aspects were well balanced (item 7, 2.69). We stress that, in the context of a gamification activity (or more generally a SG), this is a promising result since these two must be harmoniously combined without making the user believe that the experience he is carrying out is too "unserious" or too boring as associated to a pure learning activity.

Table 3.9: Statements used to Assess Perception of the Gamified Approch for Training Purposes (Mean Values and Standard Deviations).

#	Category	Score (SD)
1	I found it fun to use a game to learn intended content	4.31 (0.93)
2	I would have preferred a different learning modality, based on the presence of an instructor	2.08 (1.29)
3	I would have preferred a different learning modality, based on books, notes, slides, etc.	1.73 (1.15)
4	The designed game is a valid tool to learn intended content	4.19 (0.80)
5	The possibility to compare my score with other learners' scores made me try to improve my results	3.92 (1.06)
6	The possibility to play the game several times made me better understand the correct sequence in which actions have to be performed	4.08 (0.89)
7	I found the tool more a game than a system suitable for training	2.69 (1.23)
8	Trying to improve my results in the game let me learn intended content better	4.04 (1.04)
9	The presence of a timer stimulated me to quickly carry out required actions	3.42 (1.27)
10	I found the timer a stressful element	1.88 (1.20)

3.1.7 Transferring Holo-BLSD Procedure Learning Approach to other Domains

One of the relevant outcomes of the work done on Holo-BLSD is that many of the research results involving procedure learning can be transferred to many other application domains. The possible uses of this kind of approach to training and teaching are many.

Given these possibilities, the ability to have a software framework that facilitates the management of these approaches to learning and training is desirable since it helps decrease the costs and efforts related to their design and implementation. To define this framework, one solution is to build on the software architecture implemented in Holo-BLSD.

Another result of the Holo-BLSD experimentation that is transferrable to other contexts is the possibility of using game mechanics and gamification approaches to make the training activity more engaging, more immersive, and therefore more effective in terms of learning. In particular, the feedback from our users and the quantitative results (in terms of correctness of the execution of the procedure and timely completion of the activities that make up the procedure) lead us to suppose that the game mechanics used in the Rehearsal session can also be transferred to other domains where procedure learning is of interest.

To demonstrate these hypotheses (i.e., the reduction in development time granted by the availability of a specific framework for procedure learning, and the advantages, in terms of knowledge transfer, of using game mechanics in the training phase), in the following sections, we present and briefly discuss an example of application in a different context than that of Holo-BLSD. This case study (which will be described in the following sections) presents a training system for firefighters to teach intervention procedures in a public building following the outbreak of a fire. Before describing this application, we provide details of the design of our framework for procedure learning management.

Procedure learning management: framework design

We begin the description of the implemented framework by taking up what must be its main characteristics. The framework must be designed to support different configurations, scenarios, and activity sequences in a simple and flexible way. To ensure this goal, our design models activities as nodes that are organized into a graph, whose edges defines dependency requirements. Each activity can then be defined by a single class, which extends the basic `ActivityNode` class and implements its own execution flow. The activity flow management is controlled by the trainee's activity and the state of the learning environment, which is a function of the user's interactions and the internal and external events occurring during the simulation.

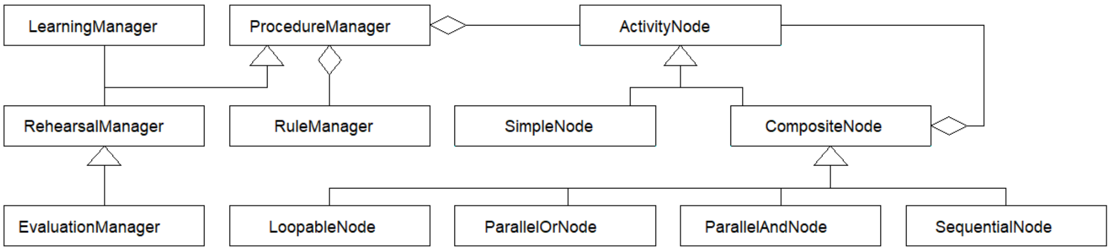


Figure 3.11: UML design of the software framework for procedure learning management.

The activity graph is managed by an instance of `ProcedureManager` class, whose main responsibility is to define the list of activities it has to control and the rules for scheduling them. Each of the different subclasses of `ProcedureManager` defines the different activity graphs (according to the specific application requirements) and their control logic for a specific session (interaction training, learning, rehearsal and evaluation). For instance, with reference to the Holo-BLSD application, in rehearsal mode, checking the consciousness of the victim before having secured the scene causes a game over, while the same event in evaluation mode affects the final evaluation without ending the session. In order to simplify the graph management, the nodes (i.e., activities) hierarchy implements the composite design pattern [117], which aims at treating in the same way both simple (`SimpleNode`) and more complex (and aggregated) nodes (`CompositeNode`). Each `ActivityNode` can have one of four state: inactive, running, completed and aborted. All nodes are initially inactive and when the session begins, the root node is started. When a simple node is started, it remains in the running state until its execution is completed or it has been aborted due to some internal condition or external events. The composite nodes gather various nodes and implement different algorithm for managing them. The `SequentialNode` provides, as the name suggests, a sequential execution of a stream of activities, where each activity can be started only when the previous one has successfully completed its job. The sequential node is in running state until there is a running child and is completed when all of its children are marked as completed. If any child switches to the abort state, the sequential node stops its execution and becomes aborted as well. Both the `ParallelAndNode` and `ParallelOrNode` start all their children in parallel when they are activated. The difference between the two is in their termination conditions. The `ParallelAndNode` is completed when all children have completed and is aborted if any of its children switches to the aborted state. On the contrary, the `ParallelOrNode` is completed as soon as one of its children is completed. When this happens all its running children are marked as completed. Instead, the node is marked as aborted only if all of its children have aborted. The `LoopableNode` allows to repeat the execution of its nodes according to different control logic, thus allowing to implement `for` and `while` loops. Any child aborting causes the interruption of the `LoopableNode`, which otherwise remains in the running state until the flow control terminates the loop.

Finally, the `RuleManager` class is the class devoted to define and control the game rules possibly implemented by the learning activity. To ensure maximal flexibility, these rules can be either hard-coded or defined into an external xml file. This allows, for instance, the definition of different game rules for the same scenario and different players, or different rules for the different users' skill levels.

Firefighter training: a proof of concept of the developed framework for procedure learning

In this Section, we briefly present a use case in which the framework described in the previous Section was used as a demonstrator of its effectiveness in supporting the rapid development of a training application on complex procedures. The work arose in the context of the IDEal reSCUE project (ERC Starting Grant), and required the development of a training activity for teaching firefighters the emergency procedures to be implemented in public buildings following the outbreak of fires. For this purpose, we developed a VR application in which to maximize the user's sense of immersion, different modes of locomotion management in VR were implemented and compared. The development of this project had a very tight timeline (two months to get a functional prototype), which made it a particularly challenging test-bed where rapid prototyping was an essential requirement.

The rationale for using VR-based firefighter training can be summarized as follows. First, firefighters need to learn how to address critical situations where they can risk their life, and VR provides a safe training environment where emergencies can be simulated without putting trainees in any real danger. Second, a VR application allows trainees to experience different scenarios and emergency procedures, and the same application can be used to train different categories of personnel. Third, VR training can also be delivered remotely, thus helping reduce costs and overall training times, and it can even support collaborative activities, which are beneficial for learning [261]. Then, VR offers compelling experiential learning environments that improve the learner's engagement and, ultimately, the learning outcomes. Another advantage is that VR promotes repetition (i.e., users can repeat the training session at their own pace), which, in turn, improves content retention. Finally, training sessions can be logged and reviewed in a debriefing session, where learners can critically analyze what they did, get insights from their experience, analyze mistakes, and explore alternative solutions. Thus, debriefing sessions enhance the transfer of knowledge and skills from the virtual to the real world and promote active learning [118].

Despite the many advantages, there are a number of weaknesses that still affect current VR systems in this specific context [99]. One relevant issue is that, despite the many efforts, there are no devices or interaction metaphors that allow natural movements in large VR spaces, such as those typical in firefighter training scenarios. Moreover, some of these navigation approaches might even induce cybersickness [249], with a negative effect on learning. Therefore, one of the goals of this work is to assess the usability of three different VR systems (using different hardware configurations) in this specific context. In details, these systems are a desktop VR, and two immersive environments, both leveraging an HMD for visualization and handheld controllers for interaction but exploiting two different locomotion techniques: a gaze-directed-steering metaphor [35] and an active re-positioning technique [221] that allows trainees to naturally walk through large virtual environments leveraging an omni-directional treadmill. The training scenario used in the experiments reproduces a fire-following-earthquake event in a school.

The educational path of the application comprises a learning mode (where trainees are instructed step-by-step on the sequence of actions they have to perform in the emergency procedure), and an evaluation mode, which leverages a serious game to assess the learned skills. As for Holo-BLSD, the game design leverages two main mechanics. A *timer* enforces a time limit for completing a given task. A *score value* translates the player's success in the game into a numerical representation, which is then used as an automatic assessment of the learned skills.

The three VR systems have been compared through a user-evaluation study that involved 45 volunteers divided into three separate groups (each using a different system). Experimental results showed differences in the usability of the three systems under analysis, highlighting that the way locomotion is managed is the most critical parameter that affects usability and users' achievements in terms of learning outcomes.

VR environments

For the evaluation, the application was deployed in the three types of VR environments depicted in Figure 3.12, whose main differences consist in the interaction techniques and the level of immersion offered.

The first is a desktop VR system (referred to as DVR in the following), where the virtual environment is displayed on a large monitor. The mouse movements control camera orientation, while the player's spatial movements are handled with the keyboard arrow keys. The immersive VR system (IC) uses an HMD for visualization and hand-held controllers for managing interaction and locomotion. Since the active area of the HMD trackers is not large enough to let user explore the VR by physical walking, locomotion is managed by using the joystick included within the controllers to translate the user in the current gaze yaw direction.

The last system (KAT) is similar to IC since it leverages HMD as a display system and hand-held controllers for interaction with objects but uses a KATWalk treadmill to manage locomotion. Treadmills are "body-centric" re-positioning systems [221] that translate physical gestures (e.g. walking or running) into virtual movements. With the KATWalk, the user is strapped into a harness, attached to a supporting structure, and slightly lifted over a concave platform (Figure 3.12). The platform has a low friction surface that prevents the forces generated during each step from physically moving forward the user. The walking gestures are captured by two inertial sensors placed on special overshoes, while a third sensor, placed on the back of the harness, tracks the movement direction and triggers the locomotion input when the user tilts forward or backward.

Experiments

The scenario selected for our experiments is a fire-following-earthquake event in a school. In order to make it as realistic as possible, the emergency scenario is set into an actual building (the Mascagni middle school of Melzo, Italy), which was recreated in the virtual environment importing its Building Information Model (BIM). In the emergency scenario under analysis, a short circuit caused by an earthquake generates a fire inside a classroom located on the school's first floor. Some students in the classroom are wounded or trapped by fallen debris and, thus, they need to be assisted and rescued, while the remaining students escape from the classroom and evacuate through the external emergency stairs. The emergency procedure activities (and their completion times) have been defined with the help of expert firefighter trainers according to the characteristics of the building where the procedure takes place. Some snapshots of the procedure actions can be seen in Figure 3.13. We underline that the scenario taken into consideration allows us to stress the interaction in general and the locomotion interfaces in particular since users had to (virtually) walk for long distances and move inside narrow spaces (thus requiring a fine-grained and detailed control of their movements) in order to

Figure 3.12: VR environments considered in this study: desktop VR (DVR, left), immersive VR (IC, center) and treadmill based immersive VR (KAT, right).

Figure 3.13: Snapshot of the VR training scenario (from left to right): the main entrance of the school, the computer room where the fire started, the evacuation meeting point seen from the emergency stairs and the fire-extinguishing step.

complete the assigned tasks.

In order to compare the different setups described in Section 3.1.7, we performed a user study that involved 45 volunteers (35 men and 10 women, aged between 29 and 30) selected among students and personnel of our University. Six of them are frequent VR users, 24 have had a previous experience with VR, and the remaining

15 had never experienced VR before. Users were divided into three groups of 15 people, each experiencing the training application through a different system. The experimental protocol applied was the following. First, since users are likely to have low familiarity with the VR devices, the learning experience started with an interaction training session in a test environment where users could get acquainted with the interaction and locomotion modalities available. When users felt confident with the system, they were invited to repeat twice the learning session and, finally, perform twice the evaluation session.

The assessment of the different systems encompassed the analysis of both the learning outcomes and the system's usability. As for the expected learning outcomes, since trainees should learn how to execute the various steps of the procedure in the correct order and within a predefined time interval, their assessment was based on the analysis of users' behavior in terms of completion time, the correctness of the performed procedures and learning progresses. These metrics rely on quantitative and objective measures obtained from in-game analytics. Usability was assessed through standardized questionnaires (i.e., leveraging subjective measurements) that users were required to fill after they completed the experience. The questionnaire structure is the same used in Holo-BLSD (i.e., a combination of SUS [45], Nielsen [220], statements derived from the ISO 9241-400 standard [153]) and VRUSE [166], followed by a short custom questionnaire to analyze the appreciation of the game mechanics used). All questionnaire items had to be scored by users expressing their agreement on a five-point Lickert scale (1, totally disagree; 5, totally agree).

Results

In the following, for the sake of brevity, we will not report all the experimental results obtained and their analysis (the interested reader is referred to [66] for the details).

In brief, we found different level of usability for the different VR interfaces used. In particular, the the KAT SUS score is significantly lower than that of the other two systems (DVR 86.0, IC 81.7, KAT 68.3, $p = 0.00002$) and, thus, its usability can be considered as barely acceptable. Similar conclusions can be drawn by considering Nielsen's attributes, where the KAT had significantly lower values than DVR and IC regarding efficiency and was also perceived as having lower possibilities to recover from errors than IC, and the ergonomy questionnaire, where the KAT scored significantly lower than its competitors. The analysis of the quantitative data showed similar difference between KAT and other systems for the percentage of actions completed in time, with a negative consequences on the learning outcomes. Trainees should not only learn to perform all the actions required (and in the correct order), but they should also learn to complete them in a suitable time, since (for some actions) time is a critical parameter (e.g., in our case, the room where fire started should be reached as soon as possible in order to provide immediate first-aid to the people in that place). An in depth-analysis of the path travelled highlights that KAT users were less capable than others to exert a fine-grained control of their movements, which in turn led to increase the completion times of individual actions and, thus, of the full procedure.

The VR application was effective in producing the expected learning outcomes. The percentage of actions completed in the correct order in the second evaluation run was 100% for all users except one volunteer of the DVR group who made a single mistake. The percentage of actions completed in the correct order and in a timely fashion was close to 100% for all groups, except for KAT where this value was only 83% (and significantly different from that of the other groups). This result, as we discussed before, was mainly due to the issues with the locomotion device.

Statement	**Score (SD)**
I had fun using a game to learn intended content	4.56 (0.58)
I would have preferred an instructor-based learning modality	2.22 (0.94)
I would have preferred a learning modality, based on books, notes, slides, etc.	1.38 (0.74)
The designed game is a valid learning tool	4.42 (0.61)
The possibility to compare my score with other learners' scores made me try to improve my results	4.40 (1.04)
I found the tool more a game than a system suitable for training	2.78 (1.09)
Trying to improve my results in the game let me learn intended content better	4.02 (1.02)
The presence of a timer stimulated me to quickly carry out required actions	4.07 (1.06)
I found the timer a stressful element	1.78 (0.84)

Table 3.10: Statements used to assess users' appreciation of the gamified approach (mean values and standard deviations).

As for the game mechanics, we can analyze their contribution to the learning outcomes through the results of the concluding section of the questionnaire, in which we asked volunteers to express their level of agreement with the 9 statements reported in Table 3.10. We found that the results have no significant differences among the three groups and, thus, we reported them as the cumulative averages among all users of the three groups. These results can be summarized as follows. The game features made the learning process more fun and the approach proposed to deliver the educational contents was appreciated by users. The game elements introduced to solicit the competitiveness (the score and the leaderboard) were able to improve the learners' engagement and foster repetition, and the timer was considered as an effective element in encouraging users to complete the procedure actions in time. Finally, volunteers found that there was a good balance between educational and entertainment elements.

Finally, although we did not collect specific measurements within this work, the design process was extremely simplified by the availability of the framework. Its underlying structure allowed designers and developers to focus on the definition of the rules and assets involved in the various activities rather that on re-thinking from scratch how to manage their delivery to learners and control their advancement. As a result, the "time to market" was extremely shortened, as demonstrated by the fact that we could meet the initial (tight) deadline assigned. We think that these positive results enables further developments of the framework. In particular, our plan is to extend it with the capabilities to manage collaborative activities that involve multiple users (either co-located or remote).

3.1.8 Conclusions

In this section, we have described Holo-BLSD, an AR-SGs for procedure learning. This specific learning context was selected as we deemed it a relevant testbed to emphasize the positive synergies that could be established between AR and SG. AR has numerous affordances (e.g., context awareness, adaptive information, and spatialized digital contents) that can be exploited in procedure learning activities. For this reason, AR solutions are rapidly gaining interest, especially in specific professional sectors like health and industry. Growing evidence [39, 161, 270] is showing that these approaches can lead to better learning performances, improved retention, and satisfaction. In return, companies can have a more skilled professional faster and at a lower cost. Another factor driving all these positive outcomes is the more significant users' motivation to engage with the augmented learning content. Furthermore, also SGs aim to solicit user engagement through elements that intrinsically stimulate learners' motivation (e.g., points, leaderboards, and narratives). However, as we have pointed out, procedure learning approaches are not extensively adopting gaming elements to engage learners, either through SGs or through gamification practices. Therefore, through the experience and evaluation of Holo-BLSD, we wished to investigate if different game elements can be synergically combined in more "traditional" procedure learning activity (i.e., BLSD) as means to furtherly stimulate users' motivation and engagement and possibly learning outcomes.

HoloBLSD provided evidence for answering RQ1. Experimental results show the positive synergies between AR and gaming elements in addressing the needs of procedure learning. Users emphasized that the education and entertainment aspects were well balanced. Furthermore, users had fun and felt engaged by playing the game, stating that the SG elements contributed to making learning more effective and less tiresome compared to other traditional approaches. Thanks to the gaming experience, users were more encouraged to repeat the rehearsal phase several times, driven mainly by a sense of competition with themself or others. Repeatability was mostly promoted by the scoring system and the leaderboard feature. These perceived positive outcomes, fostered by the gaming experience had a tangible reflection in the objective evaluation users received at the end of the learning path (more detail on these results will be discussed in Chapter 5).

Although we are aware that one piece of evidence is not enough to generally state that combining AR and SG is always a practical solution, we believe that these are promising results. We hope future practitioners will benefit from our results to apply similar approaches to other applications addressing procedure learning and AR. Indeed, we stress that we could transfer the learning framework of Holo-BLSD to a completely different field (firefighter training) using a different technology (VR). The positive results show that the developed framework for supporting the rapid prototyping of procedure learning applications is agnostic to the targeted domain and underlying technology.

3.2 Learning about Complex Systems

Complex systems are typically defined as those requiring the ability to approach them from multiple and sometimes competing perspectives and they also may have multiple possible solutions [290]. These characteristics mean that many fundamental concepts related to complex systems can be difficult for students to learn [155], and they can even be counter-intuitive or conflict with commonly accepted beliefs [112]. For example, many people tend to establish linear relationships between actions and effects (i.e., a “small” action has a “small” effect), whereas conversely, in complex and dynamic systems, a small action can generate a series of chain reactions that have large-scale, relevant effects (i.e., the so-called “butterfly effect”). Another obstacle in approaching complex systems is the tendency of students to favor simple, reductive explanations that assume central control and single, deterministic causality. In contrast, approaching complex systems (e.g., designing a city so that goods and services are maximized and the sustainability of the city is maintained at sufficient levels) requires solutions that consider the overall system, driven by a decentralized control and often governed by the interconnectedness of multiple (and often random) factors.

Thus, since complex systems are characterized by dynamics that cannot be predicted simply by examining the isolated behaviors of their individual parts [290], the challenges that education about complex systems has to deal with rule out educational methods based on direct instruction, which analyze wholes in parts and structure learning in terms of the gradual accumulation of pieces of information [1]. In contrast, recent research highlights the potentialities of constructivist perspectives to help learning about complex systems [154]. Constructivist teaching is based on the assumption that learners construct knowledge and skills as they try to make sense of their experiences. That is, learners are the makers of meaning and knowledge [130]. Developing an effective constructivist educational approach to complex systems requires to take into account several elements. These include the need (i) to create *experiential learning environments* (where students can directly experience and analyze phenomena related to complex systems), (ii) to make the *core concepts explicit* to the students (thus, unveiling the connections between the phenomena observed and their underlying framework) and (iii) to involve students in *collaborative and cooperative activities*, which encourage discussion and reflection, helpful to generate deeper insight and understanding.

Based on these premises, recent research [192, 29, 107, 62] recognizes that SGs offer unique possibilities for learning about complex systems. There are a number of reasons to support this argument.

First, SGs can be seen as a perfect example of experiential learning environments. They allow the creation of virtual representations of complex scenarios that can be explored and analyzed by learners to highlight the dynamics and interactions between the elements and the actors involved. The effects of players' actions on the system can become readily understandable and learners can analyze them on both a global scale and a large time span. As opposed to the real world, players can analyze things repeatedly from different observation points and explore different solutions in a safe scenario (i.e., without actually endangering the real system).

Then, the game scenario, the storytelling and the gameplay can be effective to motivate learners and engage them in interactive and dynamic activities, which in turn provide benefits for the development of cognitive skills (e.g., players will learn to deal with complex facts because they need this knowledge to progress in the game).

In addition, computer games can exploit visual communication, which has three potential effects [310]: *cognitive* (since it increases the information available, reduces the cognitive workload and clarifies patterns of value and relationships), *affective* (being able to trigger instant emotional responses to displayed elements) and *behavioural* (being able to influence players' attitudes and behaviours [33]).

Finally, computer games allow the creation of situated and socially mediated learning contexts by enabling shared experiences (e.g., by providing multiplayer settings or allowing learners to share information and results through the social networks).

These characteristics of SGs make them ideal candidates for complex system learning. Among the different instances of complex problem[4] learning, in the following, we investigate education for sustainable development (ESD), which is characterized by closely intertwined and often conflicting aspects and requirements considering the different interests and points of view of various stakeholders. Thanks to these attributes, there is a growing interest in exploiting SGs for ESD. However, as we will detail in Section 3.2.2, the use of AR-SGs to tackle this problem is still largely unexplored. On the contrary, we believe that the unique affordances of AR are ideal to fulfil complex problem learning requirements. As we have seen,

---

[4] From now on, the term complex problem(s) and complex system(s) are used interchangeably, as they refer to very similar concepts

AR is an excellent tool to promote face-to-face communication and collaboration. Then, it can facilitate users' understanding of the complex system under analysis through its visualization and interaction properties. These elements help enhance those experiential learning activities that are essential for effective learning about complex systems. Therefore, ESD seems an ideal testbed to evaluate if effective synergies can be established between AR and SGs. In the following sections, we will introduce Sustain, a collaborative AR-SG application that aims at fostering collaborative problem-solving among co-located players and promoting awareness and commitment towards sustainability-related issues.

3.2.1 Sustain: an AR-SG to Foster Sustainability Awareness

In 1987, the global community formally recognized the importance and urgency of promoting sustainable development, understood as the "development that meets the needs of the present without compromising the ability of future generations to meet their own needs" [320, p. 43]. Since then, ESD has been progressively theorized, operationalized, implemented, and evaluated [320, 302, 303, 298, 47, 226, 95].

ESD aims to enable individuals and societies to regard our planet as a global community with a shared future, protect the ecosystem, value diversity, identify collective needs, and pursue common interests [303, 47]. Reaching these objectives requires tackling complex and interconnected problem scenarios such as global warming, pandemics, and geopolitical crises. These scenarios present distinctive characteristics that make them extremely difficult to tackle. First, they are defined by interdependent and often conflicting environmental, economic and social factors. Accordingly, sustainability issues should be addressed by adopting systemic perspectives that are appropriate for promoting and balancing three closely interrelated elements:

- *economic sustainability*, through continuously producing adequate volumes of goods and services, while maintaining manageable levels of governance and debt, and avoiding sectorial imbalances that compromise agricultural or industrial production;
- *environmental sustainability*, through maintaining a stable resource base, avoiding over-exploitation of renewable resources, depletion of non-renewable resources, and preserving the ecosystem;
- *social sustainability*, through achieving equitable distribution of resources and prosperity, adequate provision of social services (e.g. healthcare and education), equal opportunities, and political accountability and participation [143].

Second, sustainability scenarios involve multiple stakeholders with diverse and possibly clashing interests and needs [297, 64, 284, 164, 105, 147, 85, 108]. Hence, tackling sustainability issues requires exploring and harmonizing different worldviews through involving all the affected stakeholders in collective dialogue and collaborative decision-making [64, 105, 147, 108]. Finally, sustainability scenarios are ever-changing [164, 108, 104]. The factors that define sustainability issues change over time, and so do the way such factors interact [256, 202, 108, 104]. Furthermore, seemingly simple, small-scale interactions can originate emergent global phenomena, whose large-scale effects can neither be traced back to their origins nor predicted based on the observation of small-scale interactions [202, 108, 104]. Consequently, sustainability problem scenarios are highly uncontrollable and cannot be definitively solved [256, 108, 104]. Therefore, they should be tackled through iterative problem management approaches, suitable to achieve and maintain satisfactory conditions through continuously planning, acting and adapting strategies based on ongoing monitoring of evolving circumstances [256, 164, 108, 104].

In this context, ESD should empower people "to make informed decisions for environmental integrity, economic viability and a just society for present and future generations, while respecting cultural diversity", and act accordingly [47, p. 20]. Despite the promising advances made in the past two decades, ESD remains in many ways an open challenge [47, 226, 95]. In particular, innovation is required to find more effective ways to engage learners in real or simulated sustainability problem scenarios [47, 2, 201, 108]. The aim of these actions is threefold: (i) promote *awareness* of and *commitment* to sustainability issues [302, 303, 2]; (ii) generate opportunities to *explore* complex, multi-stakeholder sustainability problem scenarios that integrate economic, environmental and social aspects, and promote the ability to comprehend their mechanics, relationships and implications [302, 298, 201]; and (iii) foster learners' ability to tackle these issues through collaborative approaches.

Given the relevance in ESD of these elements, defining them in detail is appropriate. *Sustainability awareness* refers to the ability to acknowledge the existence of critical issues that affect sustainable development, recognize the relevance of their implications, and understand key mechanics that define their complexity and the ability to address them [47, 201]. *Commitment* represents a disposition to act to improve sustainability issues, individually or collectively [298, 47]. As for *collaboration*, we refer to iterative and collective process whereby participants collaborate continuously to identify, achieve and maintain desirable states of affairs, coping with uncertain and unpredictable conditions [256, 164]. Such process integrates collective evaluation of environmental conditions, planning and action, and requires continuous adaptation of goals, plans and underpinning assumptions based on the interpretation of changing circumstances [57, 164].

With these observations in mind, sustainability serious games (SSGs) represent an ideal candidate for facilitating ESD. Games can engage players in collaborative sense-making processes, promoting their *holistic understanding* of things, relationships and events in the game space, *sense of caring* towards what happens in the game space, and *commitment* to improve game situations [323, 106]. Hence, by simulating authentic sustainability scenarios that integrate social, economic and environmental aspects, SSGs can promote awareness of complex sustainability issues, and commitment to address them both individually and collectively [107, 156]. Furthermore, games can involve players in open-ended problem scenarios, requiring them to collaborate to manage complex situations, adapting to changing and unpredictable conditions to achieve and maintain desirable states of affairs [287, 104, 137]. Finally, recent technological advances provide multiple potentialities to enhance game-based learning experiences. In particular, the increased multimedia processing and handling capabilities of mobile devices, and the recent developments in AR technologies can be exploited to provide immersive narrative-based experiences suitable to promote a closer coupling between real-world and simulated situations [307, 129]. These possibilities can make game-based learning more meaningful and promote knowledge transfer [106]. Additionally, they enable creating technologically-enhanced spaces suitable to promote face-to-face social exchanges, encourage interactions with and across material and digital worlds, and consequently foster meaningful collaborative activities [68, 78, 129].

Since ESD requires tackling complex systems, the literature indicates that, to turn these potentialities into effective learning approaches, SSGs should require players to: (i) engage in situations emerging from the interplay of authentic social, economic and environmental issues that involve multiple stakeholders and reflect conflicting interests, needs and constraints; (ii) plan and act in changing and unpredictable conditions, relying on uncertain and incomplete information; and (iii) progress through engaging in collaborative activities [105, 147, 104].

In particular, as described in Section 3.2, we think that AR-SGs represent a valuable solution to meet these requirements. With these observation in mind, we present the design and evaluation of Sustain, a collaborative and co-located multiplayer SSG. Sustain is an AR-based urban development game created to pursue the above-mentioned ESD objectives: (i) promoting players' awareness of and commitment to sustainability issues related to urban development; and (ii) fostering players' ability to carry out collaborative processes and support their cooperative actions.

In the following, we introduce the core features of Sustain, the theoretical framework and rationale underpinning its design, a multi-modal evaluation of its impacts on players' learning, and present some conclusive reflections.

3.2.2 Related Works

In the last decades, SSGs have been increasingly developed with the specific purpose of promoting the integrated development of capacities and sensibilities relevant to engage in sustainability problem scenarios [156, 85, 287]. A growing body of empirical evidence has demonstrated that SSGs can produce attitude change [156], foster apprehension of sustainability knowledge [326, 287], promote awareness and understanding of sustainability issues [168, 195], foment collective activity [168, 98], and stimulate critical capabilities such as problem-solving, communication, negotiation and the monitoring of emotional intelligence [168].

All accounted for, collaborative SSGs have been successfully leveraged to promote three key types of sustainability-relevant learning: (i) *cognitive*, consisting in the acquisition of new knowledge or restructuring of existing knowledge (e.g. understanding the mechanics of climate change); (ii) *normative*, which concerns learners' ability to change their assumptions and worldviews (e.g. development of sense of commitment towards sustainability issues); and (iii) *relational*, which reflects learners' ability to understand others' mind-sets better, build trust and willing to collaboration with others [138]. Despite these achievements, recent literature reviews have shown that novel empirical research is needed to address key open problems that exist in each of these areas, and improve methodological approaches for the design and evaluation of SSGs [138, 85, 287, 141].

Concerning cognitive learning, SSGs should promote the holistic comprehension of sustainability issues, accounting for the complex interactions between social, environmental and economic factors [105, 85, 287]. Albeit past SSG research has address all these three themes [141], developed SSGs have too often compartmentalised them [287]. Hence, there is a need for more empirical research that tackles social, economic and environmental sustainability in an integrative way, through engaging learners in scenarios that are dynamised by the interplay of all these three dimensions and, consequently, require their holistic understanding [287].

Normative learning is crucial to adapt and thrive in changing, unpredictable and non-fully-controllable conditions [108]. Accordingly, SSGs should be leveraged to develop learners' ability to adapt their worldviews and assumptions [104], and promote attitude changes beneficial to value and commit to sustainable development [156]. However, too few studies have focussed on normative learning, and more empirical research is required to address this shortfall [138].

Given the importance of collaboration to manage sustainability problem scenarios [95], relational learning should be a central focus of SSGs [138]. However, after review 42 empirical studies on collaborative SSGs, Haan and Voort [138] found that only 17 of them assessed relational learning outcomes. The authors consequently highlighted the need for further research focussed on both promoting and evaluating the development of sustainability-relevant social skills, including collaboration.

Other reviews echoed this call, stressing that promoting meaningful social interactions and collaboration skills through SSGs remains a pressing open problem [e.g. 85, 287].

Concerning methodological issues, there is a need for more robust empirical SSG research, which should not only assess the nature and magnitude of the effects on learners, but also "how and under what conditions those effects are achieved" [141, p. 13]. Accordingly, SSGs' should be designed and evaluated to uncover which design features can effectively foster sustainability-relevant skills and attitudes, and an holistic understanding of sustainability issues [85, 287, 141]. Furthermore, the efficacy of SSGs should be assessed considering changes in learners' knowledge, attitudes and behaviours related to specific sustainability issues and scenarios, and accounting for both short-term and lasting effects [156, 85, 141]. To this end, the evaluation of SSGs' effects should be multi-modal, integrating multiple measures, and leveraging in-game learning analytics as much as out-of-game assessment approaches [138, 85]. Finally, there is a need for more evidence-based prescriptive SSG research, suitable to guide the effective design, evaluation and use of SSGs [141].

AR-based SSGs: unexploited potentialities.

The recent advances in AR technologies offer significant potentialities to enhance the effects of SSGs, particularly in terms of promoting players' engagement in collaborative problem solving activities, and fostering relational learning accordingly [85]. Recent reviews have highlighted that non-digital role-playing and board games can be effective means to engage learners in sustainability problem scenarios and promote meaningful social interactions, active negotiation, and collaboration [287, 141]. Through the augmented tabletop format, AR allows integrating these forms of gaming with digital gameplay activities in simulated mixed-reality scenarios, thus combining and potentially enhancing the effects of digital and non-digital game-based learning [189]. Furthermore, by overlapping a simulated task space with a physical one, AR allows promoting more effective collaboration [28, 316, 193]. Specifically, the literature suggests that in AR environments, as opposed to traditional screen-based collaborative environments, people tend to relate to their peers in manners similar to those occurring in a face-to-face interaction not mediated by technology [193].

Despite the potentialities offered by AR, there is relatively little research that exploits this technology in SSGs. For example, EcoCampus [13] and GreenDesign [267] are two AR games aimed at teaching, respectively, sustainable construction design and sustainable engineering practices in formal education. Although both games contemplate the possibility for players to interact, none of them supports social interactions through AR. FunergyAR [113] combines an individual quiz-based activity with a multiplayer card game to promote energy-efficient behaviors in school pupils. However, FunergyAR uses extensively AR in the individual quiz games but merely to unlock specific cards and thus not to support group activities.

City of Life [142] is a sustainable urban development game which consists in a main city-building gameplay activity, and several AR-based mini-games through which players must fulfil specific sustainability goals to accumulate resources required to build the city. Thus, AR is not integral to the main gameplay activity. Finally, EcoGotchi [242] is a Tamagotchi-style mobile phone AR game, wherein the player has to take care of a living creature whose well-being is dependent on the player's eco-friendly skills, knowledge, and behavior. The game is single-player, but it allows sharing achievements through social networks.

All these works demonstrate a tendency to use of AR in a limited way, lacking support for collaborative and group activities. Furthermore, learning outcomes were evaluated only in two cases. In EcoCampus students playing the AR game were able to produce more creative designs in a shorter time, and had better learning outcomes than two control groups that used pen and paper [13]. City of Life generated an overall increase in sustainability knowledge, even though the small participants sample size (six volunteers) prevents the generalization of these results [142].

The needs identified in the current state of the art motivated us to design Sustain as a multiplayer AR-based collaborative SSGs, and evaluate its effects through a multi-modal approach. Specifically, in order to enhance cognitive learning, Sustain aimed at promoting holistic awareness of sustainability by engaging players in urban development scenarios that integrate social, economic and environmental factors. To foster declarative learning, Sustain aimed at promoting attitude change focusing on commitment to sustainability. To this end, the game was designed to engage players in activities requiring them to progressively explore alternative perspectives, conflicting interests, and the negative effects that individualistic choices can have on the common good. To enhance relational learning, Sustain was designed to promote players engagement in collaborative problem management processes, requiring them to impersonate different roles, and continuously negotiate and harmonize conflicting interests to pursue shared goals. AR was used to develop Sustain as a virtual co-located board game, leveraging augmented contents to simulate the complex scenarios to be tackled, define the mechanics required to explore and act upon them, and to elicit and support meaningful face-to-face interactions and collaboration among players. Finally, we adopted a multi-modal approach to evaluate Sustain learning outcomes, integrating questionnaires and in-game learning analytics to explore the game effects on players' awareness and commitment towards sustainability, and their ability to engage in collective problem management processes.

3.2.3 Design Guidelines

Designing SSGs. Sustain was designed based on a set of guidelines extrapolated from sound conceptual frameworks. Several conceptualisations have been proposed for the design and analysis of serious games (e.g. [324, 53]). However, to the best of our knowledge, the only framework expressly focussed on SSGs is the one proposed in [105]. This framework presents game design principles beneficial to promote sustainability learning through defining and integrating four key aspects of SSGs scenarios: *sustainability contextualization* (defined by settings, player roles, motives, and action possibilities); *player agency* (defined by nature of gameplay problems presented to players, and approaches allowed/required to tackle them); *player adaptivity* (defined by environmental conditions eliciting players' adaptation); and *sociality* (defined by gameplay situations and mechanics promoting meaningful social interactions, collaborative or competitive). We extrapolated from this framework the design guidelines in table 3.11. This framework also emphasizes the importance of engaging players in collaborative activities, albeit without providing detailed design indications apropos.

Table 3.11: Design guidelines: game features fostering sustainability learning

What to design	**How**
Sustainability contextualization	(DG1) Contextualise gameplay activities in scenarios that: (i) involve settings and game objectives reflecting real-world issues and integrate social, economic and environmental sustainability dimensions; and (ii) offer to players multiple sustainability-relevant roles, underpinned by different interests/needs and associated with specific actions.
Player agency	(DG2) Present to players: (i) opportunities to choose roles and self-define game goals and strategies; (ii) ill-defined problems, requiring players to continuously explore and critically interpret changing events and relationships in the game world, reflecting on present and past game states, and forecasting future states; (iii) actions/interactions possibilities mimicking the real-world, with the opportunity to act across real and virtual environments; and (iv) gameplay actions influenced by conflicting constraints and requiring management of limited resources.
Player adaptivity	(DG3) Plan gameplay activities: (i) affected by changing environmental conditions, due to events that players cannot fully anticipate nor control, and which require adaptation of their goals and plans; and (ii) structured in phases, allowing players to explore and evaluate gameplay situations without strict time-limits.
Sociality	(DG4) Define game objectives, tasks and environmental conditions: (i) requiring the engagement of multiple players in different roles; and (ii) requiring players to collaborate as well as deal with conflicting interests.

Designing for collaboration. Given the relevance of collaboration in ESD and SSGs, multiplayer SGSs should be designed relying on guidelines specifically conceived to promote the development of collaboration skills [138, 85]. The literature has proposed several design frameworks useful to promote collaboration in games [85]. For the purpose of our work, we relied on the proposals by [258], [251],

[329], and [104]. These works present design guidelines (summarized in Table 3.12) that are particularly suitable to promote in SSGs those collaboration capabilities emphasized by the literature on ESD [298, 47, 108].

In particular, [104] expanded the analysis of collaborative activities in games. The authors stressed that complex problems like sustainability scenarios should be continuously managed through iterative and adaptive processes [256, 164], involving collaborative meaning-making, action and learning [57, 147]. Accordingly, the authors indicated that gameplay tasks aimed at fostering collaboration should be articulated as iterative and collective processes, requiring players to continuously explore and make sense of problem situations, plan action, implement plans and adjust plans and underpinning assumptions, based on ongoing monitoring of environmental conditions [104]. These principles can be operationalised by designing gameplay tasks according to the framework proposed by [106]. Based on a human factors perspective, this framework conceptualizes game-based learning as the outcome of meaning-making processes that drive all gameplay activities, and are defined by context, structure, and contents of gameplay tasks. Throughout gameplay tasks, players interpret in-game information flows to continuously create and adjust mental models of gameplay situations, events and entities, developing awareness of and attitude towards their characteristics, functions and interactions. This constitutes a meaning-making process that drives players' tasks, and unfolds through players' interactions with the game environment. Meaning-making integrates players' cognition and affection, since their mental models reflect how they think and feel about the world that they perceive. Based on these principles, the framework in [106] indicates that, in order to promote game-based meaning-making and related learning, gameplay tasks should be articulated in iterative phases, through which players (i) define/revise goals, (ii) evaluate environmental conditions, (iii) define/revise plans to achieve goals, (iv) act based on plans, and (v) evaluate the results of their actions (Figure 3.14). Task contents and settings should then be designed identifying "learning foci" (i.e., the knowledge, attitudes, and skill sets that SSGs should foster), and conveying throughout gameplay tasks information flows relevant to promote desirable mental models (i.e. models aligned with the learning foci), and useful for players to progress the game.

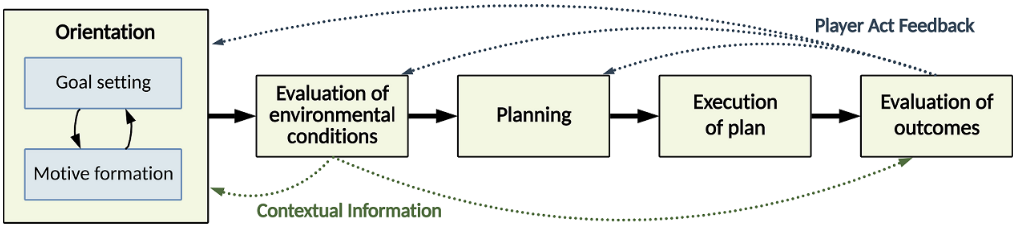


Figure 3.14: Articulation of the gameplay activity - adapted from [106]

Table 3.12: Design guidelines: game characteristics fostering collaboration

What to design	**How**
Structure of gameplay tasks	(DG5) Define gameplay tasks as iterative processes: (i) driven by and promoting collective meaning-making, i.e. formation and update of shared knowledge representations and attitudes; (ii) requiring progressive construction and adaptation of solutions, involving iterative definition and revisions of plans and objectives, and driven by continuous evaluation of activity outcomes and relevant environmental information (Figure 3.14); and (iii) eliciting activity-based learning as the development of knowledge representations, skills and attitudes relevant to pursue shared goals.
Information flows	(DG6) Define contents and provision of information flows to: (i) support meaning-making involved in the definition and revisions of plans and objectives, and in the evaluation of environmental information; and (ii) promote mental representation relevant to to orient the development of desirable knowledge, attitudes and skills.
Diversity ad complementarity of player roles	(DG7) Offer to players different roles that provide specific and complementary skill sets and actions [258], in order to: (i) Help players to be more involved in their role; and (ii) promote communication and co-operation among the different roles, making them necessary for the effective coordination of players' complementary actions.
Shared and constrained goals	(DG8) Define shared goals for all players and provide them with *limited resources* to accomplish collective tasks [258], in order to: (i) emphasize players' need to team up and find creative solutions to gameplay challenges; and (ii) prevent a single player from taking control of the group by performing all actions or commanding the others.
Traceability of gameplay payoffs to players' decisions	(DG9) Offer to players opportunities and means to reflect on the consequences of their actions [329], in order to enable players to: (i) understand how their decisions affected themselves and others; (ii) identify wrong decisions; and (iii) experience "expectation failures", when they realise that the outcomes of their actions are not as good as expected, or even detrimental.
Collective progression	(DG10) Articulate collective gameplay activities in phases, enabling players to engage in individual tasks and decision-making processes, but requiring players to converge, share information and coordinate their activities before moving on to the next phase [251]. To this end, control transition across phases through *Gathering gates*, i.e. gameplay situations requiring players to wait for others because they can only progress together [251].

3.2.4 The Sustain AR Game

Sustain is an AR-based city management game for three players collaborating to expand a fictional urban area by assuming different planning and executive roles. Players act in a real-world scenario where their actions and decision must guarantee a sustainable balance between various elements (i.e., housing, production, resource exploitation, environmental pollution, quality of life, land development, costs and population growth). The game is organised in turns divided into sequential phases (action planning, execution, and evaluation) where individual and collective activities interweave. At the end of each turn, players can monitor their progress through three main variables (environment, happiness, and population number) describing the city's global state, whose value must be maximised to succeed in the game.

Since a non-sustainable city causes unsatisfied citizens to leave, sustainability practices are essential to succeed in the game. Furthermore, fulfilling Sustain's game goal requires active collaboration among players, underpinned by an understanding of the game mechanics that govern different interwoven elements and the complex environmental impacts generated by players' individual and collective decisions. Hence, Sustain targets an audience composed of adults and adolescents, which already reached the formal operational stage of Piaget's cognitive development theory [157].

In the following, we detail how the gameplay and the game context implement the key design concepts and address the design issues described in the theoretical frameworks introduced in Section 3.2.3.

Game contextualisation and game goals

Sustain implements a *sustainability contextualisation* by engaging learners in problem scenarios that mirror real-world issues and involve the interplay of *economic*, *social* and *environmental* dimensions of sustainability [47, 105, 287]. The gameplay and task contextualisation are completed by requiring players to adopt different sustainability-relevant roles. This feature contributes to promoting the meaningfulness of gameplay tasks and the involved learning [106]. Each role is characterised by a set of responsibilities and offers various deployable actions (each of which affects in different ways the game variables) that are progressively made available to the player. The main characteristics of the available roles are the following. The *mayor* supervises houses, factories, public transport, and leisure areas. The *ministry of energy* develops the energy matrix (based on both renewable, wind and solar, and depletable resources, coal, gas, or nuclear power) and manages waste disposal. The *ministry of agriculture* defines the government policy on agriculture, forests, and food production and leads educational campaigns to raise public awareness about healthy behaviours.

The gameplay is set in an urban area and its countryside, each subdivided in a grid of 4 × 4 blocks.Players can deploy actions on these blocks according to their

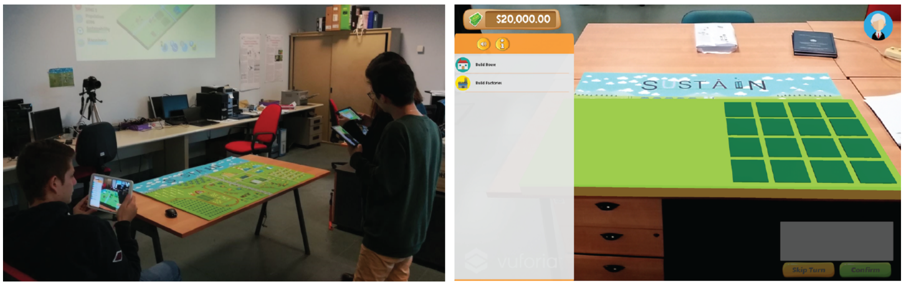


Figure 3.15: Left: a view of the layout of the physical game space, which offers players the possibility to freely move in the environment and observe the AR game board from different perspectives. **Right:** tablet UI view identifying areas where role-specific actions can be applied (in the case displayed, the mayor can identify a city location where to build a house).

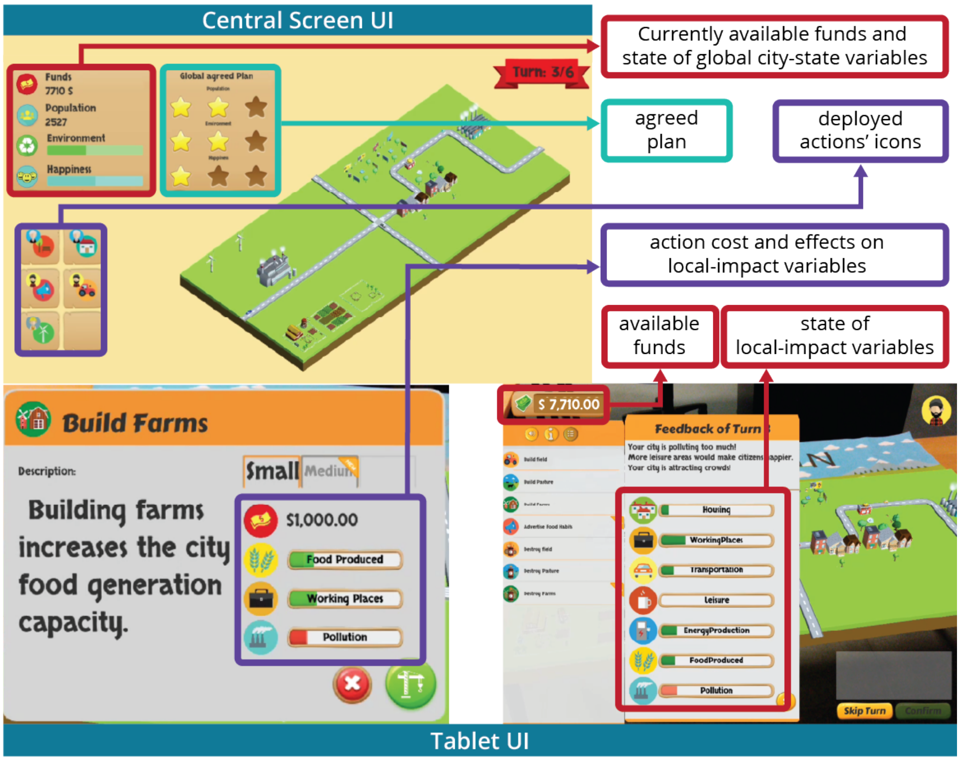
role and the current game turn (e.g., at turn one, the mayor can build a house or a factory on a block, while turn two unlocks the possibility to construct as well public transport). Each action has a cost, which is deducted from the city (shared) funds when the action is executed, and can directly affect seven game variables having a local impact (*transport*, *food*, *energy*, *pollution*, *leisure*, *housing*, and *working places*). Players can visualise (and analyze) the direct local consequences of an action before executing it. In turn, local impact variables determine the value of three variables that describe the global state of the city: *environment*, *happiness*, and *population number*.

These global state variables define the overarching game goal, as the game requires players to maximise the population while at the same time maximising their happiness and the environmental sustainability of the city. Thus, Sustain engages players with a *shared and constrained goal* [258], where planning individual actions involves managing a limited pool of shared resources (city funds) and involves *negotiating priorities* with other players. Indeed, while reflecting a unique perspective on sustainable development issues underpinned by different and potentially *conflicting interests* [105, 108], each role has specific responsibilities and actions that are all equally important to improve the sustainable development of a city. These conditions require players to engage in collective decision-making to devise strategies that seek to integrate individual actions favouring the common good[5].

Altogether, these features promote collaborative *agency* [105]. They also address the ESD requirement to (i) engage players in collective problem-solving processes [105, 108], and (ii) promote meaningful social interactions [105, 108, 85].

---

[5] For a general video description of Sustain gameplay, see http://tiny.cc/Sustain_GameDescription



Building farms increases the city food generation capacity.

Your city is polluting too much!
More leisure areas would make citizens happier.
Your city is attracting crowds!

Figure 3.16: Summary of the in-game information provided to players. **Top:** a view of the central shared screen showing the city-state variables, the available funds, the commonly agreed development plan and the list of actions proposed by players in the current turn. **Bottom:** two different views of the tablets' UI, showing the effect of the selected action on the local-impact variables and their values at the beginning of the current turn.

Game interactions

Visualisation and interaction with the game scenario exploit marker-based AR on mobile devices (tablets). The marker used is a large map placed on a table (Figure 3.15, left). This physical game space layout aims to enhance the players' co-located experience and promote collaborative processes since it improves the level of communication and sociability between unacquainted individuals [83].

The tablet UI displays the AR contents according to the players' viewpoint, allows them to interact with the game scenario (Figure 3.15, right) and visualises both general and role-specific game state information (Figure 3.16, bottom). In this way, AR supports players' individual problem exploration and autonomous decision making (*exploration affordances* [106]), helping them to devise meaningful changes in the game world [107].

Sustain features also a large central screen to provide a global overview of the current game state (Figure 3.16, top). The two types of displays featured (the

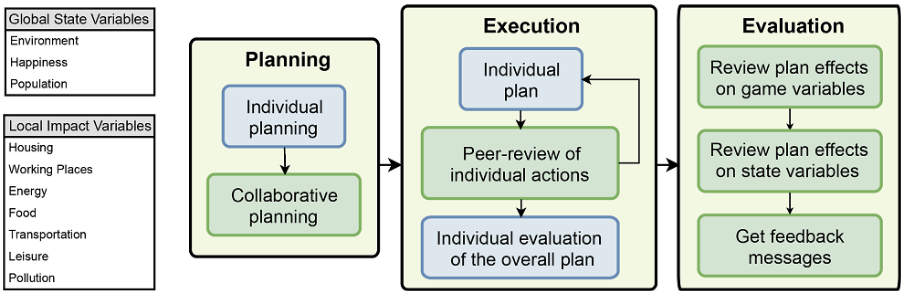


Figure 3.17: Global state variables, local impact variables, and turn flow (where cyan blocks represent individual activities and green blocks the collective ones).

individual one, on the tablet, and the shared one, on the central screen) aim to support different types of collaboration, in both an "action space" (where users perform game actions) and a "reflection space" (where players analyse the results of their efforts).

Gameplay process

A Sustain game is divided into six turns, each corresponding to six months in the fictional urban area's life, and the global state variables value defines the current and final game results. Each game turn is divided into three phases (planning, execution, and evaluation) where individual and collective activities interweave (Figure 3.17).

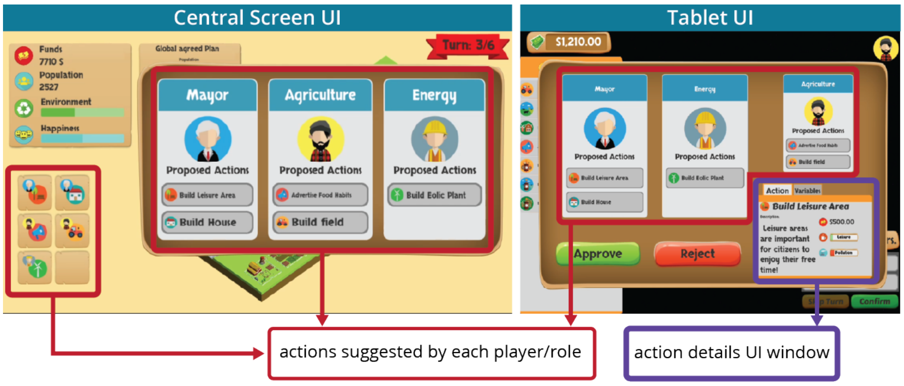
Planning. This phase requires players to collaboratively define an action plan, prioritising the local impact and the global city-state variables they want to improve. A plan defines variable priority in terms of stars assigned to them (Figure 3.18), with a budget of five stars for local impact variables (maximum three per variable) and seven stars for global city-state variables (maximum five per variable). The planning phase is organised by first requiring each player to propose an individual plan, focusing on overall sustainability objectives rather than those related to their roles. Then, players are required to discuss their respective plans and formulate a collectively-agreed one.

Execution. In this phase, each player contributes to implementing the collectively agreed plan by deploying a maximum of two actions. The actions deployed by all players form a procedure, which is then subjected to a peer evaluation process where each player can authorise it or reject it, causing its invalidation and requiring players to discuss and deploy alternative actions (Figure 3.19). This process is repeated until all players validate the procedure, which is finally executed. After that, each user must record an individual evaluation of the collectively-accepted procedure on a "five-star" Lickert scale (1: very unsatisfied, 5: very satisfied).

Evaluation. During this stage (which aims to promote collective and critical discussion about the current development policies, providing information crucial

- Your city is polluting too much!
- The unsustainable city is making citizens sad.
- Your city is attracting crowds!

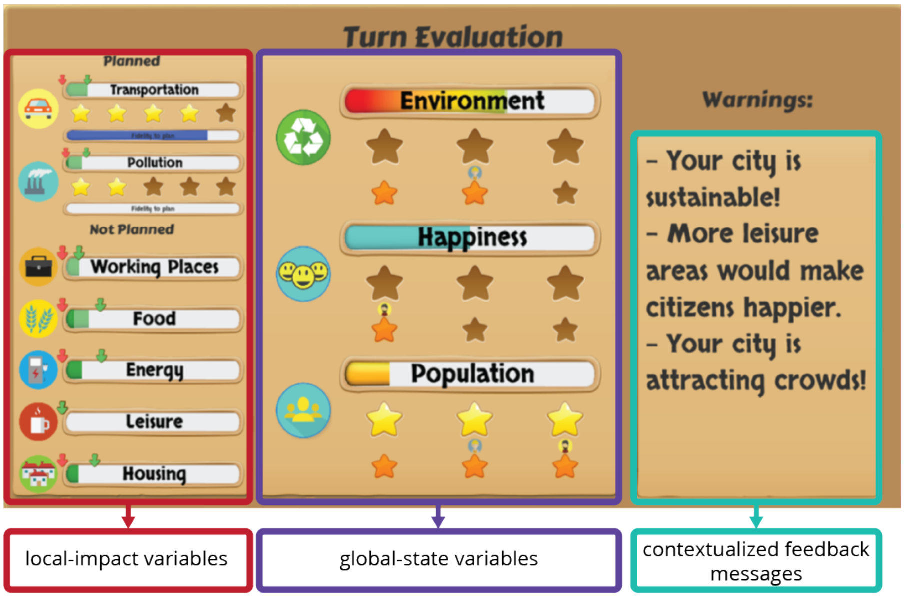
Figure 3.18: The shared screen (**left**) and tablet (**right**) interfaces of the collaborative planning phase. Players can use stars to highlight their priorities in terms of state and game variables. Role icons in the shared screen view represent the priorities that each player assigned through their tablet UI (**left**) during the individual planning phase. To support the planning phase, the central screen also displays the current game state, and recommendations proposed by the game to improve the city.



Leisure areas are important for citizens to enjoy their free time!

Figure 3.19: Peer review phase. The shared screen (**left**) shows the current proposed plan as the list of actions deployed by each role. Each player, through their individual tablet (**right**), can analyze the effects of the actions deployed by other players and eventually approve or reject them.

to redefine or improve them), the central screen provides players with a review of the planning and execution phases. This review is organised into sections, and a new section starts when each player agrees to proceed through their tablet UI. The first section compares the executed procedure's effects on the local impact variables with those agreed upon during the planning phase. The second does the same with the global variables. Finally, the last section presents feedback messages automatically generated by the system to provide clues or warnings about the current city development state. All these sections are depicted in Figure 3.20. At the end of the evaluation, Sustain computes a score based on a heuristic function



- Your city is sustainable!
- More leisure areas would make citizens happier.
- Your city is attracting crowds!

Figure 3.20: The three sections of the executed plan review showed on the central screen.

of the current values of global and local variables. This score is not displayed to players since it is a representation of the current players' achievements used as an internal indicator of a group performance in the current turn.

The turn-sequence mechanic implemented by Sustain addresses the ESD need to engage learners in iterative problem-solving activities, allowing them to learn from the outcomes of past actions in order to define and adapt future strategies [105, 108]. Then, the turn phases mirror the gameplay task model proposed in [106]. The planning and evaluation phases offer a reflection space and elicit "pressure-free" individual and collective exploration, discussion and assessment of the game state (*exploration affordances* [106]), promote collaborative decision-making to plan collective goals [201, 108], and serve to promote, organise, orient and inform meaning-making processes [163, 108].

In the execution phase, players can define individual strategies accounting for personal perspectives and priorities and commonly-agreed objectives while also taking advantage of the opportunity to enjoy collaborative face-to-face discussions in the action space (the table and the AR environment). *Gathering gates* [251] are implemented at each phase to elicit collaborative interactions, ensuring that the game can only progress if players collectively agree on motivations (e.g., priorities of global and local variables), decisions (e.g., action choices), and evaluations (e.g., performance outcomes). Turn phases are not time-constrained to facilitate the individual and collective analysis of game state and gameplay strategies, and the involved *tracing of payoffs* back to actions and decisions [329]. At each stage, players receive individual (on tablets) and collective (on the central screen) just-in-time feedback to ensure that their evaluations and decisions are based on meaningful, sufficient, and timely information [106].

Information flow

The game goal requires players to tackle an open-ended problem, allowing them to select their roles and pursue alternative strategies freely. To achieve their goal, players must iteratively (i.e., turn after turn) plan and execute actions with local effects to indirectly promote desirable global impacts. In doing so, the tablet UI supports them by detailing the local effects of an action as soon as it is selected (Figure 3.16, bottom left). However, as stated before, the local-to-global effects of their actions are not entirely predictable because of the FLS used to determine them. This condition requires that players first understand their actions' local effects and then try to hypothesise their global impact. For example, building a house increases the number of available homes and increases the level of pollution, and both of these variables have different effects on the state of the city (allows the number of inhabitants to increase with a negative effect on the environment). To help players with this task, Sustain provides a variety of context information in the reflection space (the central screen, Figure 3.16, top), and the gathering-gates mechanism aims to avoid the possibility of players losing critical data.

3.2.5 Operationalisation of Design Guidelines

The game features (GFs) implemented in Sustain were designed following sustainability and collaboration guidelines in order to promote ESD, in particular sustainability awareness, sustainability commitment, and collaborative and adaptive problem management. These features cane be summarized as follows.

- **GF1: Gameplay problems mirroring real-world scenarios involving the interplay of economic, social, and environmental dimensions.** Sustain requires players to develop a simulated modern city in pursuit of a shared goal: to maximize the population while at the same time maximizing their happiness and the environmental sustainability of the city. To achieve this goal, players have to manage a system of interconnected local and global variables. These variables have been modelled to mirror authentic sustainability issues and involving all sustainability factors.

- **GF2: Shared and ill-defined goals.** The need to simultaneously increase the environmental sustainability of their city, its population and citizens' happiness represents a shared and ill-defined game goal. Players cannot individually affect the global state of the city, as this is determined at each turn by the individual actions performed by all players. Furthermore, at each turn, there are multiple ways of improving the global state of the game. Thus, there is no way for players to predict what consequences their action plan will have indeed, and there is no indication of optimal strategies that could be pursued. Hence, players are required to individually and collectively self-define their goals and strategies.

- **GF3: Different player roles, with complementary role-specific actions and possibly conflicting interests.** Sustain requires each player to assume a different realistic policy-maker role. Each role has distinctive and complementary capabilities (role-specific actions and resources affecting in-game variables in different ways, making roles complementary and equally important for the city's development. However, each role has specific responsibilities and priorities that could be conflicting.

- **GF4: Shared and limited resources needed for individual actions, requiring negotiation and agreement.** Pursuing the shared goal requires players to manage a limited pool of shared resources for planning and deploying their actions. For example, the game-space locations where players can deploy their actions are development sites that only a single player can use. Then, deploying an action decreases the shared and limited city budget. These constraints elicit collective discussions of game plans to negotiate priorities, and identify and coordinate complementary actions. Thus, to succeed, players have to consider not only the interests and abilities of their role but also those of other roles.

- **GF5: Autonomous exploration and deployment of actions.** In Sustain, even in the presence of collective game goals and shared resources, players are required and encouraged to individually examine problems and make individual decisions to alter the game scenarios in meaningful ways. Players are allowed to explore the game scene individually and select and deploy individual actions at each turn, either according to plans collectively agreed or not. The use of personal tablets, role-specific actions and resources has been planned to provide players with opportunity and information to explore and influence the game autonomously. The mechanism of transferring effects from local-to-global variables allows pursuing alternative courses of action to improve the city, thus ensuring that different individual decisions can indeed lead to alternative success states.

- **GF6: Local-to-global effects determined by a FLS, introducing uncertainty and unpredictability.** The FLS that processes the effects of players' actions on local variables introduces an uncertainty in the game that makes this local-to-global transfer non fully predictable. Thus, a given configuration of local-impact variables may generate multiple and plausible global city-states. At the same time, the system has been designed so that players can intuitively and retrospectively interpret variations in global outcomes and trace them back to other environmental conditions that might have influenced game state changes so that the system can be perceived as non-fully-controllable but not chaotic or random. Altogether, this setting has been planned to require players to explore and become aware of the relationships between local and global variables, trying to understand how their role-specific actions *can directly and certainly* affect local-impact variables, and *may likely and indirectly* affect the global state of the game. By extension, such an exploration promotes discussion and information exchanges across players since each role has different possibilities to influence the global game state.

- **GF7: Turn structure, allowing iterative problem solving and gameplay evolution.** The within-turn structure of Sustain ensures that players can only progress collectively, as it requires them to maintain a shared understanding of individual and collective goals and to iteratively agree on motives (e.g., priorities of global and local variables), decisions (e.g., action choices), and evaluations (e.g., performance outcomes) related to the management of global and local variables.

- **GF9: AR-Based individualized interaction device** Each tablet's UI displays an AR representation of the game scene, general and role-specific information and the list of actions that the player can deploy according to her/his role and the current game turn. The game rules allow players to look at each other's devices and share role-specific perspectives, even though this is not required. Altogether, this feature has been designed to allow players to engage in individual and personalized exploration and interaction with the game environment.

- **GF10: Co-located play space.** The physical game layout and the availability of a shared screen were designed to provide collectively accessible, integrated information concerning individual choices, collective decisions and game state, readily available as a "what did we do, and how did it go?" feedback to support individual and collective decision-making throughout each turn.

Table 3.13 summarizes the mapping between these game features and the design guidelines detailed in Tables 3.11 and 3.12. Although this mapping was already suggested throughout this Section, we think that it is useful to make it explicit here for the sake of clarity.

In brief, *sustainability contextualisation* (DG1) is achieved by integrating GF1 (*gameplay problems mirroring real-world scenarios involving all sustainability dimensions*), GF2 (*shared and ill-defined goals*) and GF3 (*different player roles,* *with complementary role-specific actions and conflicting interests*). *Player agency* (DG2) is fulfilled via the combined impact of GF3, GF4 (*shared and limited resources needed for individual actions, requiring negotiation and agreement*), GF5 (*autonomous exploration and deployment of actions*) and GF9 (i.e., AR visualisation), offering players the ability to freely explore and transform the game environment using their role' abilities, while collaborating with other players. The designed systematic effect of GF2, GF6 (i.e., the uncertainty and unpredictability in determining the global-to-local effects), GF7 (i.e., the turn structure allowing iterative problem solving and gameplay evolution) and GF8 (within-turn phases combining individual and collaborative activities) generates *Player adaptivity* (DG3). Indeed, the integration of these features requires players to iteratively self-define goals, understand complex relationships, and adapt to the game changes. *Sociality* (DG4) is supported by GF3, GF4, GF6 and GF10 (i.e., the physical co-located game layout and the availability of the shared screen).

The *iterative structure* of gameplay tasks (DG5) is fulfilled via the combined impact of GF7 and GF8. *Information flows* (DG6) is generated through the designed systematic effect of all ten core features, allowing players to receive and adapt to individual and collective information related to their roles, resources, decisions and actions, and to the game environment, turns, phases and events.

Diversity and complementarity of player roles (DG7) is achieved by the implementation of GF3, providing players with both conflicting and synergetic unique priorities, resources and abilities. *Shared and constrained goals* (DG8) is fulfilled via the impact of GF4, offering players the challenges of limited group resources but also the support to mediate their deployment towards common goals. The possibilities provided by GF7 and GF8 to iteratively explore and discover new and better strategies for success in the game and to reflect on their plans and actions from previous turns and phases offer players the *traceability of gameplay payoffs* (DG9) Finally, *collective progression* (DG10) is achieved by the implementation of GF8, which requires players to synchronise their activities before moving on to the next phase.

As for the expected outcomes, ESD requires formative scenarios offering systemic, complex and everchanging problems and conditions. Thus, the game features were designed so that their combined effect on players' meaning-making and playing behaviours helped reaching the desired impacts. In particular, we expected that:

- *sustainability awareness* will be promoted by facilitating player engagement with ill-defined problems commonly found in real-world urban development contexts (GF2); involving the interplay of environmental, economic and social aspects (GF1); requiring players to assume different roles with different responsibilities and action capabilities (GF3); and eliciting negotiation between players regarding the use of collective resources (GF4), which will make players create and adapt their mental mental models to incorporate a more holistic understanding of intricate dynamics that govern sustainable development. Players will be encouraged to take into account different perspectives and negotiate solutions according to the available resources, stimulating an awareness of sustainable development as a multidimensional, collective and participatory issue. In addition, implementing an iterative, problem-based turn structure (GF7), and a fuzzy logic approach to determine the relationship between global-local effects (GF6), will demand players learning to deal with uncertainty and unpredictability, and will allow them to grasp the challenges involved in adapting goals and strategies, and predicting global consequences of local actions in conditions of evolution and change.

- *commitment to sustainability* will be fostered by making players take an active role in finding a joint solution to sustainability problems (GF3) that at the same time require autonomous exploration, reflection and decision-making (GF5). By setting the gameplay in a real-world scenario (GF1), it is expected that previous knowledge and values will be challenged and new sensibilities will be developed in the collective exchange of perspectives. These features will support the development a broader comprehension of the impacts of individual and collective actions, providing players with a sense of agency that may affect how their perceive themselves in relation to their own community.

- *collaboration* will be promoted by a combination of features that constitute the core gameplay of Sustain. Due to the within-turn phases that alternate individual and collaborative activities (GF8), players need to shift from autonomously planning and executing game actions (GF5) according to the stakeholder role played in the game (GF3), and discuss, negotiate and agree to define a collective plan of action and the best use of shared resources (GF4). All this is expected to stimulate discussion and collaboration in order to work collectively towards aims defined by the group, incorporating different individual perspectives. Moreover, this will be done in an iterative way due to the iterative turn structure implemented in Sustain (GF7), which will require players to converge, share their views about the outcomes of the previous turn, and adapt their strategies based on the evolution of the game-state. The co-located layout of the game (GF10) and use of AR (GF9) support communication and face-to-face interactions, whereas the central shared screen facilitates players' visualization of the game state, evaluation of impacts and sharing their understanding of sustainability dynamics.

Table 3.13: Mapping between Sustain game features and design guidelines and expected outcomes.

	Design Guidelines										**Outcomes**		
Features	DG1	DG2	DG3	DG4	DG5	DG6	DG7	DG8	DG9	DG10	Awareness	Commitment	Collaboration
GF1: gameplay problems mirroring real-world scenarios and involving multiple sustainability dimensions	✓					✓					✓	✓	
GF2: shared, ill-defined goals	✓		✓			✓		✓			✓		
GF3: different player roles, with complementary role-specificactions, and possibly conflicting interests	✓	✓		✓		✓	✓				✓	✓	✓
GF4: shared and limited resources needed for individual actions, requiring negotiation and agreement		✓		✓		✓		✓			✓		✓
GF5: autonomous exploration and deployment of actions		✓				✓						✓	✓
GF6: local-to-global effects determined by an FLS			✓	✓		✓					✓		
GF7: turn structure, allowing iterative problem solving and gameplay evolution			✓		✓	✓			✓		✓		✓
GF8: within-turn gameplay phases, combining individual and collaborative activities			✓		✓	✓			✓	✓			✓
GF9: AR-Based individualised interaction device		✓				✓							✓
GF10: co-located play space				✓		✓							✓

DG1:sustainability contextualisation, DG2:player agency, DG3:player adaptivity, DG4:sociality, DG5:iterative structure of gameplay tasks, DG6:meaning-making information flows, DG7:diversity and complementarity of player roles, DG8:shared and constrained goals, DG9:traceability of gameplay payoffs, DG10:collective progression

Implementation details

Sustain was implemented in Unity 3D. A client-server architecture supports the multiplayer collaborative interaction, whereby the clients are the players' tablets, and the server handles the simulation state and controls the shared screen. Vuforia[6] was used to manage the marker-based registration of the augmented contents.

Sustain's design and implementation followed the SCRUM agile methodology [169], through an iterative process comprising several cycles of conceptualisation (e.g. brainstorms and discussions around key design aspects), prototyping (whereby design choices were implemented into playable prototypes), playtesting (which involved internal teams, target audience, experienced players and usability experts), and evaluation (to critically review playtesting results and provide feedback to the design and implementation phases). This process allowed the rapid identification of errors and design flaws and the progressive enhancement of the game concept, game usability and user experience. In particular, since a clear and straightforward communication of the necessary pieces of information to players is vital for the success of the game, we carefully designed their introduction and visual representation to guarantee that (i) users immediately perceive the availability of such information and (ii) the communicated data is readily understandable for them.

3.2.6 Experimental Study

We conducted a quasi-experimental study aimed to address the following research questions.

- Does playing Sustain foster a change in players' sustainability awareness? (Section Sustainability Awareness)
- Does playing Sustain promote a change in players' commitment to sustainability issues? (Section Commitment toward sustainability)
- Does Sustain represent an effective tool to support and foster collaboration among users? (Section Impacts on collaboration)

The study involved 99 volunteers recruited among students and PhD candidates of the Politecnico di Torino in Italy. Participants possibly knew one another but had not previously worked together. We randomly assigned participants to 33 groups of three players each. The number of groups (and, thus, participants) was estimated using the G-Power software (based on a priori power analysis for both correlations and pre-post paired comparisons, using a conservative small effect size of 0.3 and small correlation of 0.3, power $(1-\beta) = 0.80$, and $\alpha = 0.05$). As for the

---

[6] `https://developer.vuforia.com/`

participants' characteristics, 76% of them were male, and their age range spanned between 20 and 29 ($M = 23.3$, $SD = 2.02$). Concerning technology awareness, 87 of them had previous experience with AR (four of them frequently, six regularly and the remaining rarely). Similarly, 74 of them are occasional (39) or frequent (35) gamers.

Procedure

Playing sessions were conducted in a laboratory equipped with the devices and setup necessary to play Sustain. Sessions lasted approximately one hour, and they were all video and audio recorded for analysis. Only one group played per session. Before playing the game, we asked volunteers to complete a pre-test questionnaire aimed at evaluating their awareness of sustainability-related themes. Then, participants were given time to familiarize with Sustain through a "training stage". This was based on the first two game turns, and aimed at facilitating the comprehension of the main game mechanics and UI interactions. Players received information regarding the game context, the roles they could assume, and step-by-step instructions on how to use the game interface (provided through a "live" tutorial). To avoid external conditioning, players were not explicitly instructed nor encouraged to discuss the motivation of their actions, or establish shared development policies. After completing the tutorial, each group played Sustain once (six turns). Finally, we asked players to complete a post-test questionnaire to collect information about their gaming experience, and to measure changes in their sustainability awareness.

Data collection

Pre and Post-test questionnaire (PPQ). The questionnaires were created ad hoc for this study, and required participants to express their agreement on several statements using a five-point Likert scale (1 = strongly disagree; 5 = strongly agree). The pre-test questionnaire included eight items organised into two sub-scales: (i) *sustainability awareness* (5 items, Cronbach's $\alpha = 0.60$), exploring players' awareness of sustainability themes such as the relevance of urban planning in achieving environmental regulations, the impact of economic growth on the community's quality of life, and the complexity of the main challenges in the management of a sustainable urban area; and (ii) *commitment to sustainability* (3 items, $\alpha = 0.70$), evaluating participants' motivation to engage with the development of their community. Items for the sustainability awareness and commitment scales were developed consulting validated instruments (i.e. New Environmental Paradigm, NEP, scale [91]; Sustainability Consciousness Questionnaire [125], which were adapted to address principles for sustainable urban design (UN-Habitat report, 2009), and willingness to adopt sustainable behaviors. Finally, participants provided sociodemographic information and familiarity with AR and games.

The post-test questionnaire included 29 items. The first eight were the same statements included in the pre-test questionnaire. The remaining ones were organised into three sub-scales: (i) *elicitation of sustainability thinking skills* (six items, $\alpha = 0.75$), exploring to what extent players perceived that the game required them to use skills associated with sustainability-thinking; (ii) *perceived affordances for collaboration* (five items, $\alpha = 0.67$), aimed at understanding the relevance of the game's design features in supporting collaboration; and (iii) *perceived quality of collaboration* (four items, $\alpha = 0.72$). The questionnaire also included items examining players' perception regarding the use of AR, individual motivations to collaborate, and their overall enjoyment with the game.

Video Analysis (VA). All videos were annotated using ANVIL software [175], based on a coding schema jointly developed by all authors. We formulated an initial set of categorized codes informed by literature on collaboration assessment and computer-supported collaborative learning (CSCL) [151, 200]. This was then refined based on the analysis of three random videos, discussing coding criteria, generating new categories/codes, and excluding ambiguous categories/codes. Finally, three researchers annotated all the session videos, resolving a few disagreements and ambiguities through consensus.

The coding schema aimed at identifying and classifying *verbal activity* and *focus of shared attention* displayed by each group. *Verbal activity* was coded identifying all time intervals where at least one player produced an utterance. The *focus of shared attention* was coded identifying all time intervals in which two or three players focused on the same object, and describing these as a tuple composed of (i) the object of shared attention (e.g. tablets, central screen and other players), (ii) the number of players displaying shared attention, and (iii) the time length of the event.

Game Analytics (GA). Sustain collects and stores a variety of fine-grained data concerning game state, in-game activities and player actions. These include information such as play and debriefing times, number of actions analysed, confirmed and withdrawn actions, evolution of game state variables, and UI interactions.

Indicators and data analysis

Based on theoretical considerations and empirical findings reported in the literature, data collected from different sources were analysed and aggregated to formulate indicators suitable to answer our research questions. These are described below, and summarized in Table 3.14.

- *Changes in individual awareness* and *commitment towards sustainability themes*. These two indicators are considered crucial to create a global sustainable society [73]. In this study, they were defined as the differences of the *Sustainability awareness* and the *Commitment to sustainability* sub-scales between the pre and post questionnaire.

- *Satisfaction with collaboration.* This indicator has been associated with enhanced communication and team dynamics [162, 224]. In this study, we used two measures of satisfaction: (i) players' individual level of *satisfaction with the collaborative plan*, declared by players at the end of each collaborative planning phase; and (ii) the *perceived quality of collaboration* sub-scale of the post-questionnaire.

- *Joint attention* and *coupling style* among team members have been frequently used as key concepts to describe collaboration. *Joint attention* is a disposition of interacting subjects to focus their attention on the same external object, person or event, which allows establishing a joint problem-solving space [264]. Earlier studies found that joint attention is associated with increased comprehension among interacting partners [253], better task performance and perceived quality of collaboration [273]. *Coupling styles* appeared later in the literature to describe the extent to which team members' actions are coordinated to one another [295]. Groups usually shift between loosely and tightly-coupled collaboration according to task characteristics and individuals' need or disposition to collaborate [295]. Research on co-located collaboration associated task success with time spent working tightly-coupled [151, 200]. In our study, *joint attention* was measured as the percentage of gameplay time during which two or three players shared the same focus of attention, whereas *tight-coupling* was defined by the percentage of time in which players concurrently engaged in verbal and *joint attention* activities.

- *Number of interactions with the game environment.* This indicator is obtained from the number of clicks performed in the execution phase, and aims at examining players' exploratory and executive activity. Sustain elicits exploratory and executive activity by requiring players to click to (i) review lists of available actions, (ii) analyse actions through opening and navigating description dialogues, and (iii) select what they consider the best actions to contribute to the collaborative plan. Exploratory and transformative behaviours in complex, meaningful scenarios are needed to develop a deeper comprehension of the intricate dynamics that govern sustainability [47, 107, 78]. Furthermore, increased practice and familiarity with tasks and spaces tend to reduce the amount of exploratory and performatory activity, due to increased skills and accuracy in task performance [139]. Finally, shared exploratory processes and outcomes can stimulate additional individual and/or collective activity [276]. Hence, number of clicks in Sustain can help investigating degree of engagement in learning-conducive activities, increases in attunement with the environment and gameplay mastery, and enhanced opportunities for collaboration.

- *Score quartiles.* Group performances were divided into quartiles and associated with different collaborative related indicators. This analysis was done to verify if better-performing teams were distinguishable by more significant and more effective cooperative behaviors. It is important to recall that the score computed at the end of each turn aimed to keep an internal record of group performance and was not visible to players.

The general approach to data analysis included the use of: (i) descriptive statistics for all variables, (ii) Pearson and Spearman correlations (to explore associations between variables), (iii) one-way ANOVA to analyze the relationship between teams' game performance (categorised according to turn score quartiles) and collaboration behaviours

Specific data analysis techniques were also performed to analyse questionnaire sub-scales and items. For frequency analyses, responses 4 and 5 on the Likert scale were aggregated to indicate agreement or positive viewpoints. The Wilcoxon signed-rank test was used to investigate pre and post differences, as the Shapiro-Wilk test revealed non-normal distribution. Effect sizes between pre and post scores were thus computed using $r = Z/\sqrt{N}$).

Table 3.14: Indicators used to answer our RQs. For each item we report the data source (i.e. Pre and Post-test questionnaires, PPQ, Video analysis, VA, and In-game analytics, GA) and the performed data analysis.

Indicator	Data Source	Data analysis
Sustainability awareness	PPQ	Wilcoxon signed-rank test
Commitment to sustainability	PPQ	Wilcoxon signed-rank test
Satisfaction with collaboration	GA and PPQ	Descriptives, Pearson correlations, One-way ANOVA
Joint attention	VA	Descriptives, Pearson correlations, One-way ANOVA
Tight coupling	VA	Descriptives, Pearson correlations, One-way ANOVA
Interactions with the game environment	GA	Descriptives, Pearson correlations, One-way ANOVA
Game performance	GA	Descriptives, Pearson correlations, One-way ANOVA

3.2.7 Results and Discussion

Engagement in the gameplay experience

Data from questionnaires and in-game analytics suggest that players were highly engaged in exploratory and transformative activities suitable to promote ESD. Almost all players said they had fun playing Sustain (96%), and felt challenged by the game and committed to its objectives (85%). The majority also felt that AR made the game more exciting (69%), allowing them to feel more immersed in the play experience (66%). The overall quality of collaboration was positively judged by more than 85% of players, and 95% perceived they actively participated in group discussion. Altogether, players' perception of enjoyment, challenge, excitement and appreciation for collaboration suggest high affective and practical engagement in collective gameplay activities [103].

Players spent an average of 32'32" (SD = 11'28") playing Sustain. This time was spread evenly across turns (each of which lasted 5 minutes circa). In each turn, during the execution phase, players made on average 20.7 clicks, analysed 6.3 actions, and built 1.6 actions. Since each player would potentially need six to eight clicks to deploy the planned actions, these findings suggest that players actively explored action possibilities to identify how to best transform the game space.

Altogether, these findings suggest that Sustain promoted player engagement in collective gameplay tasks involving both exploratory and transformative activities allowing them to to iteratively define, evaluate and adjust gameplay goals, plans and actions. This is particularly important in ESD, as engagement in these types of tasks is key to promote awareness and attitudes toward sustainability issues [106, 156].

Sustainability Awareness

Developing a holistic comprehension of factors affecting sustainability is crucial to address its complex, multidimensional and evolving nature [108], and SSGs have the potential to make a contribution towards this [141, 138]. As detailed in Section 3.2.5, with this purpose in mind, Sustain was designed to incorporate several design and gameplay features that implicitly integrated the systemic and intricate linkages between environmental, economic and social dimensions underlying sustainability. It was therefore expected that players would show an increase in their cognitive learning, specifically regarding their awareness of the complex relationship between urban development, social well-being, and environmental change.

Consistently with our expectations, we found a significant change in the *sustainability awareness* questionnaire scale after playing the game ($Z = -3.211, p = 0.001$). Although the effect size was small ($r = 0.22$), we think that this is a promising finding, as players only played the game for an average of 32 minutes circa. This suggests that playing Sustain promoted a holistic understanding of the factors underpinning sustainable development, and that it was done in an effective way. We believe that these results are likely to be due to game design features specifically implemented in Sustain to promote the intended cognitive learning outcomes. Firstly, playing to develop a city required players to put into practice the abstract concepts related to sustainable development. This has likely enabled them to grasp the connection between local actions and global consequences [195]. By recurrently planning and implementing similar tasks yet in evolving conditions, players could test their evolving knowledge and progressively acquire a comprehensive understanding of factors affecting sustainable development and their interrelationships, all through concrete situations mimicking real-world dynamics [104]. Sustain gameplay required players to continually shift from individual to collective processes, and from global to local actions. This probably helped players to explore and understand the multiple perspectives and levels of decision-making involved in generating sustainable solutions [287]. Altogether, these features have likely facilitated a swift and holistic understanding of the complex interplay of factors involved sustainability.

The findings from our study are aligned with previous research showing the potential that SSGs have to promote cognitive learning [138], which is particularly relevant in the current sustainability landscape. People with increased awareness of complex sustainable issues tend to have better skills for identifying problems, are more sensitive to their possible consequences, and are more prone to commit to sustainable behaviours [288]. Hence, recent literature has highlighted the importance of raising awareness regarding sustainable urban development - especially in young people - as a way to tackle the challenges that cities are currently facing, and enhance the society's capacity to act [9].

Commitment toward sustainability

Promoting normative learning, intended as changes in people's viewpoints and opinions regarding their role towards sustainability issues, is a crucial but challenging endeavour [73, 138]. Such changes have been emphasized as key drivers towards more sustainable societies, even though they are most difficult to attain [73]. As previously mentioned, SSGs should promote individual understanding of sustainable development, as this is a significant factor for personal behaviour regulation and increased disposition to engage in more sustainable practices [288]. At the same time, SSGs should also incorporate game mechanics that promote reflection, dialogue and negotiation to generate new personal meanings and viewpoints, which are likely to motivate a change in the values system [294]. As detailed in Section 3.2.5, Sustain was designed to alternate between individual and collective tasks, in order to promote dialogue and re-evaluation of personal knowledge and beliefs. Consequently, we expected that players would feel that Sustain stimulated thinking and affective processes conducive to reflection, collective reformulation of ideas and increased commitment with their own community development.

Results from the questionnaire scale commitment to sustainability show a significant increase in players' willingness to commit to the development of their own community ($Z = -2.270, p = 0.023$), although the effect size is small ($r = 0.16$). Previous literature has stressed that changes in viewpoints require time to occur [73, 138]. Thus, greater increases were not expected after a short playing session. However, detecting a small change is an encouraging finding, and suggests playing

SSGs like Sustain may indeed lead to important normative learning. Furthermore, outcomes from the *elicitation of sustainability-thinking skills* questionnaire scale indicate that the vast majority of players (88%) perceived that the game spurred thinking processes and sensibilities described as pivotal for sustainable development [298, 105], for example by stimulating a sense of care for the city they were managing, and fostering understanding of others' goals and perspectives. Overall, this indicates that Sustain's features did promote reflection on personal viewpoints. Since personal transformations are more likely to occur when information about sustainable development is situated in scenarios allowing exploration, dialogue and affective engagement [105], this study shows that SSGs like Sustain can foster personal change by offering meaningful contexts whereby players are required to critically reflect and collaboratively find solutions to sustainability problems.

Impacts on collaboration

Overall, Sustain had a positive impact on collaboration. As detailed in the following, players were highly satisfied with their collaboration. Furthermore, our findings suggest that Sustain promoted verbal interaction, joint attention and tight coupling in collaborative work.

Satisfaction with collaboration. In the post game questionnaire, the overall quality of collaboration was rated as good by the majority of players (85%). Over 80% of players specifically declared that Sustain's game design and visualization features facilitated group discussion and agreement. Furthermore, in each turn, GA show that players were highly satisfied with the devised collaborative plan ($M = 3.96, SD = 0.85$), and this satisfaction level did not change significantly across turns. We also found significant association between players' *perceived affordances for collaboration* and both *perceived quality of collaboration* ($r(97) = 0.43, p < 0.001$) and average satisfaction with collaborative plans ($r(97) = 0.25, p = 0.014$). Altogether, these findings suggest that Sustain's design (i) implemented collaboration affordances that were recognised by players, (ii) promoted consistently positive perceptions on the quality of collaboration processes, and (iii) facilitated discussion and negotiation processes leading to satisfactory collective solutions.

Furthermore, the vast majority of players acknowledged that collaboration was motivated by the need to find the best in-game solution (93%) and the achievement of the game goals (89%), whereas only 11% of players felt obliged to co-operate with other players. This suggests that Sustain's design features made collaboration intrinsically motivated by the game mechanics and hence meaningful to progress in the game, which is an important feature to promote sustainability learning through gameplay [105].

The moderate correlation found between the *satisfaction with the collaborative plan* and *perceived quality of collaboration* measures ($r(97) = 0.38, p < 0.001$)

suggests these two variables are related, but are not the same. This probably indicates that achieving consensus does not require unanimity of satisfaction among participants but a commitment sufficient for the purposes [40], which may result in positive perceptions of team dynamics. This is also consistent with findings from [224], indicating that satisfaction with group performance and perceived quality of collaboration may be interrelated and influenced by higher levels of mutual support and reception to ideas.

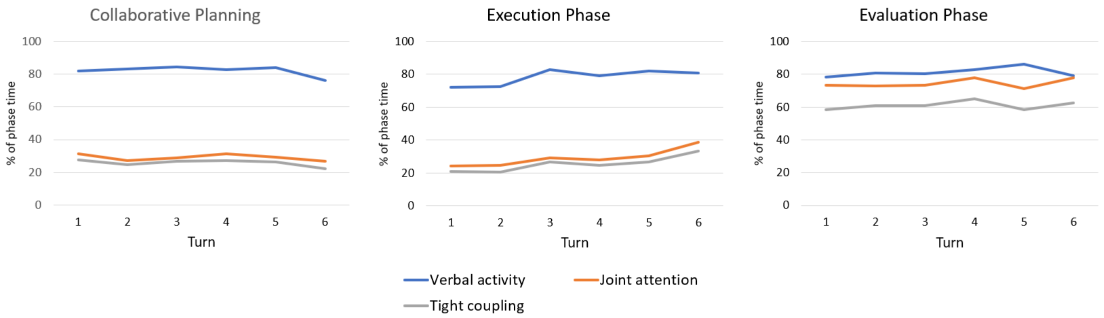


Figure 3.21: Average time spent by all groups in the following activities: *verbal activity*, *joint attention* and *tight coupling*. From left to right the three phases are: collaborative planning (Left), Execution Phase (Center), and Evaluation Phase (Right). Time is reported as the percentage of each activity's recorded length compared to the overall execution phase duration.

Joint attention and coupling. On average, groups spent 75% of the playing time speaking, 36% sharing attention, and 32% working *tightly-coupled*. Based on previous literature [295, 151], we expected significant variations in *coupling styles* across different game phases, with a more loosely-coupled collaboration during individual planning, execution and individual evaluation phases (which did not require players to interact), and a more closely-coupled one during collaborative planning and collective review phases (which were explicitly designed to promote discussion and collaborative decision-making). The results partially confirmed our hypothesis. Groups presented their lowest levels of joint attention and coupling during the individual planning (12% and 9% of the phase, respectively) and individual evaluation phases (10% and 7%), whereas the highest levels of these two indicators were observed during collaborative planning (30% and 27% respectively) and final review phases (76% and 63% respectively). The verbal activity followed the same pattern, with longer speaking times during collaborative planning (82%) and collective evaluation phases (81%), and shorter verbal interactions during individual planning (54%) and individual evaluation phases (51%).

Surprisingly, *verbal activity*, *joint attention* and *tight-coupling* were comparably long during the execution and the collaborative planning phases. This was likely due to players frequently interacting during the execution phase to check each others' devices and share information to identify the best individual actions to pursue the collectively-agreed plan. Monitoring and predicting collaborators' actions has been identified as fundamental mechanisms for coordinated action [311, 276], which in turn affect how each player plans and executes her/his actions. Sustain provides role-specific information through each player's device. For this reason, it is likely that teams used gaze and verbal activity to monitor each other's actions and share information, in order to keep their mental models aligned and coordinate actions accordingly.

If we observe these behaviors across turns and in the three distinct moments of the game (collaborative planning, execution phase, and evaluation phase), we can identify a growing trend in tight coupling and joint attention indicators (Figure 3.21, center). This trend was not detected in the other phases, which required users to either collaborate (Collaborative Planning, Figure 3.21 left) or jointly discuss the outcomes of their actions (Evaluation Phase, Figure 3.21 right). This difference is a promising result, as in the execution phase, users' were not strictly required to collaborate (i.e., they could have completed the phase without interacting with each other). However, the increased time (across turns) users spent working tightly coupled can suggest that maybe the game activity and its features (i.e., shared goals, role complementarity, and shared resources) could be accountable for soliciting users to engage in collaborative interactions.

On average, players made three times more gameplay actions in the execution phase than the minimal number that allows completing it, suggesting that they actively explored different possibilities before agreeing on their final strategy. We also found a positive correlation between *tight-coupling* and the number of clicks performed in the execution phase ($r(183) = 0.41$, $p < 0.001$). This reinforces the idea that players were engaged in collaborative exploratory activities, which are key to promote ESD.

We found small correlations between quality of collaboration perceived by players and the overall time spent working closely-coupled during gameplay ($r(191) = 0.32, p < 0.000$), and in particular with the percentage of time working closely-coupled in the execution phase ($r(184) = 0.23, p = 0.002$). There is also a moderate correlation between the percentage of time working closely coupled and the average number of gameplay actions performed by the teams in the execution phase ($r(184) = 0.41, p < 0.000$).

Collaborative behaviours and game performance. All groups increased their scores throughout the turns, with higher increases seen in the first three turns. Final scores varied greatly between teams (M = 3.7, range = 2.48-4.98). Therefore, we grouped teams by final score quartiles (Q1-Q4) to analyze the relationship between teams' performances and collaborative behaviors. We found interesting trends in the execution phase (Figure 3.22), in which players decided which individual actions would be suitable to achieve shared goals. High-performing teams (Q4) tended to devote more time in this phase and spent significantly more time working closely-coupled ($F = 5.46, p = 0.001$). Moreover, they manifested longer times of gaze coordination ($F = 5.24, p = 0.002$) and verbal activity ($F = 3.31, p = 0.02$) and registered significantly more explorative and executive actions than the lower performing quartiles ($F = 3.004, p = 0.032$). That said, we did not find any significant correlation between perceived quality of collaboration and game performance.

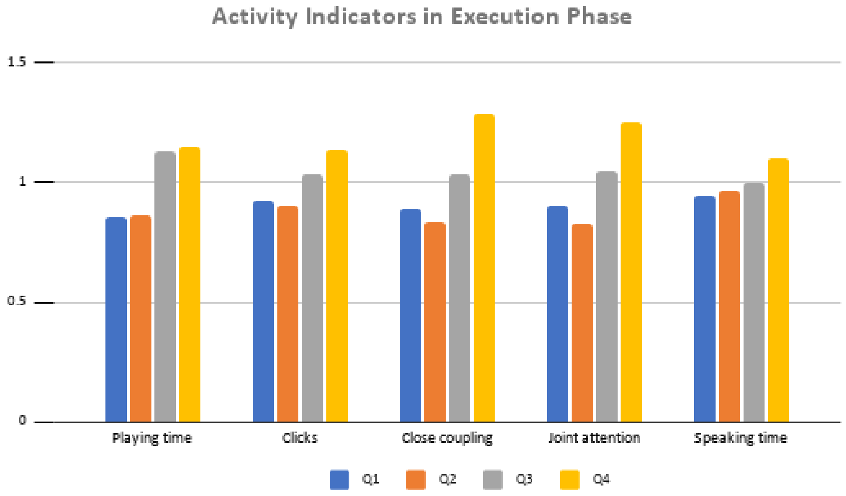


Figure 3.22: Activity indicators in execution phase (values were normalised, with average equal to 1, for comparison)

Overall, the above analysis shows how various facets of complex interaction dynamics can affect (and promote) groups' collaborative behaviors. The results can be summarized as follows.

First, we found that coupling styles can be used to characterize collaboration in our co-located AR-SG. Consistently with previous research [152, 223], our results showed that teams displayed both close and loose coupling styles while performing individual actions to accomplish shared goals. Moreover, these behaviors have to significantly increase over time, suggesting that the game promoted players' communication and joint decision-making. This finding is further supported by the positive association between the time spent working closely coupled and the individual in-game interactions. In fact, this suggests that exploratory activities (i.e., users choosing the optimal action to perform) might have been the result of thorough discussion.

Second, our findings suggest that better-performing teams do work more closely-coupled and invest greater time in analyzing their options to find the optimal solution for a successful outcome. We believe that one possible explanation for this result can be associated with the game providing feedback of the teams' outcome on a per turn basis. Therefore, better-performing teams who could identify positive outcomes at the end of one turn might have been solicited to strengthen their joint effort as they could relate this to tangible evidence of group performance.

Finally, the design features of Sustain, which aimed at promoting collaboration, have been positively recognized and evaluated by players. Also, they encouraged discussion and negotiation processes. These findings indeed support the hypothesis that the AR-SG and its features were accountable for positive collaborative behaviors.

3.2.8 Conclusions

Sustain aimed to be a demonstrator of the effectiveness of AR-SGs in addressing learning about complex systems. In this specific case, we addressed ESD through an AR-based multiplayer SG whose overarching goal is to expand an urban area, maintaining a sustainable balance between environmental, social, and economic aspects. By playing the game, players are expected to increase their awareness of and commitment towards sustainability-related themes. For this, the game was designed based on a sound theoretical framework accurately modeled to effectively and compellingly promote learning about ESD. Sustain requires players to deal with ill-defined problems that are contextualized in an environment that mimics real-world issues and evolves throughout the game due to players' actions. One of the design pillars of the game is to require players to achieve a shared goal adopting three different roles with unique individual responsibilities. By pursuing a common goal through interdependent tasks within a sustainability-related context, players are engaged in collective problem-solving processes, which will eventually promote collaborative agency, enabling the construction of shared knowledge.

Sustain assessment shows that the game effectively changed players' awareness and attitudes towards sustainability themes, eliciting capabilities and sensibilities needed for sustainable development. Our findings show that effective collaborative behaviors were detected through all of our data sources regarding the impacts on collaboration. VA revealed that players spent increasingly more time sharing attention and working tightly-coupled in phases not requiring close collaboration. In-game analytics indicated that collaboration behaviors co-occurred with high levels of interaction with the game environment. Results from the post-questionnaire showed that players perceived that the overall quality of collaboration was good and that the game layout, features, and visualization tools were essential to support collaboration among players. In all, this suggests that Sustain promotes team dynamics which have been described as crucial for both ESD and collaborative learning [298, 264].

We think that, beyond showing their effectiveness in this specific learning scenario, many of the design features of Sustain can be generalized to other contexts involving learning about complex systems. In particular, we highlight the relevance of two elements. The first is the phase-based structure of turns structured around the guidelines for collaboration listed in Table 3.12. Individual and collective activities interweave in this structure, explicitly enforcing the need to collaborate and negotiate to define a final action plan collectively. The second element is the use of AR to manage the gaming environment, which promotes effective collaboration by creating an environment where learners can have face-to-face interactions with their peers while benefiting from a game environment and rich multimedia educational information merged with the physical training environment. Results show that combining these two elements was indeed effective in supporting collaboration and reaching the intended educational objectives. Since both collaborative design guidelines (Table 3.12) and AR approaches are general (and thus transferrable to other learning scenarios), we expect that other contexts can experience the same benefits found in Sustain.

3.3 Learning Soft-Skills : Asteroid Escape

As described in the previous section (3.2), collaborative activities are pillars for establishing an effective learning environment for complex problems (i.e., in our case, sustainability). By sharing and discussing different perspectives, individuals are spurred to coordinate their decision-making, generating more effective learning outcomes. Similarly, over the past decade, organizations increasingly tackle complex problems through team-based (i.e., collaborative) approaches since they enable gathering heterogeneous feedback from stakeholders, promoting workplace synergy, motivating cohesion, and improving overall efficiency and productivity [285].

Nevertheless, an individual requires a series of skills to work effectively and perform well in a team. These skills, usually referred to as soft-skills, comprise a series of personal attributes, personality traits, and communication abilities [144]. However, traditional learning curricula overlook the importance of soft-skills in favor of hard-skill knowledge (i.e., the set of specific and teachable abilities that can be defined and measured). For this reason, in many work environments, there is a rising interest in team-building activities, which aim at improving individuals' soft-skills by involving groups in experiential learning activities where they can challenge their values, and interpersonal dynamics [190]. Furthermore, human resource personal are adopting during the recruitment process various team-building activities in order to assess candidates' team work-related skills.

Research suggests [331, 209] the benefit of improving soft-skills through TEL environments and, in this context, researchers agree over the relevance and importance of SGs in teaching or improving soft-skills in terms of collaboration, communication, and negotiation abilities [87, 114, 263].

Despite some differences, none of the mentioned examples (and the ones reported in Section 3.3.1) exploited AR (or other immersive solutions) for creating a technology-enhanced environment to promote soft-skills learning. However, as we have discussed in Section 2.1.2, AR is an ideal tool for establishing a favorable setting to promote collaborative behaviors and enrich the sense of immersion in the digital experience. Therefore, we believe these characteristics could be exploited in a soft-skills learning context.

In the following sections, we describe Asteroid Escape, a SG aimed at fostering and assessing teamwork competency which leverages a heterogeneous (rich) technological set-up composed of an AR headset (Hololens), smartphones, and an interactive wall projection. In the SG, users are challenged to collaboratively pilot a spaceship, finding a way out of an imploding asteroid. In a group of three people, two in charge of piloting and one of navigation, communication is solicited by both the need to share partial information each member owns and the required coordination in controlling the spaceship piloting commands.

By collecting both subjective and objective measures, we assessed Asteroid Escape in terms of (i) perceived learning effects, (ii) likeability and effectiveness of the game elements for fostering communication, (iii) usability, and (iv) its effectiveness in automatically self assess users' teamwork competencies.

In Section 3.3.1 we present examples found in the literature addressing SGs for soft-skills learning. In Section 3.3.2 we discuss the underpinning rationale Asteroid Escape was designed on. In Section 3.3.3 we describe the game in detail, and in Section 3.3.4 we outline the experimental methodology adopted for the assessment. Finally, in Section 3.3.5 we discuss the obtained results.

3.3.1 Background

To provide the reader with a general picture of how SGs have been exploited in soft-skills learning we describe here some notable works found in the literature. This list is not exhaustive, as it extends the scope of this thesis, however, if the reader wishes to explore further this topic he can refer to the following reviews [236, 14, 79]. Guenaga *et al* [133] developed the SG Let's Team! which focuses on decision making within a team and how it can improve communication among peers. The objective is to collaboratively build a small evolving civilization facing a series of challenges (gather X resources in a given time, build X houses with Y limited resources, and so on). In order to address these cooperative tasks, users can communicate through a chat while they are playing. However, the discussion activity is intended to be also continued in the physical world through group meetings, where they can discuss face-to-face how to approach upcoming challenges. Authors believe that the virtual-physical duality should affect players in their daily activities. However, the authors have not provided any detail of experimentation and assessment of the game and suggest it could be done in future developments.

On the contrary, experimental results are presented in [140] where researchers analyzed the effects of NoviCraft, a team-building serious game where a group of players is asked to collaboratively solve a series of puzzles in order to flee from prison. Each user controls a 3d avatar with whom he can navigate the digital world and interact with the other players. Through a mixed analysis approach, featuring a pre-post questionnaire, in-game log-files, intermediate perceptions, and game session observations, Häkkinen et al. state the game seems promising in both fostering and assessing team competencies. Similarly, [191] explores the use of serious games to support activities such as communication, effective cooperation, decision-making, interpretation, and information processing. Researchers conclude that virtual environments can improve teamwork abilities, especially communication skills, although a minimum level of fidelity in the virtual world is required. Other research [181, 97] focuses on designing virtual activities for distributed work teams. In both examples, the authors state that the game experiences improved individuals' skills to work together as a team.

Finally, we also underline that digital escape rooms are often used for training and practicing soft-skills. In these SGs, players are challenged to collaboratively escape from a "virtual" confined space by solving a sequence of riddles that require players to share information, jointly solve problems, and coordinate their actions. In doing so, they strengthen team cohesion and teamwork-related abilities. A comprehensive overview of these types of activities can be accessed here [308].

As we will shortly see, Asteroid Escape can be considered an escape room activity in a broad sense. Although players are not challenged to escape a room but an imploding asteroid, the same underlying mechanics and benefits apply.

3.3.2 Design Framework

The design of Asteroid Escape largely relied on the *input-process-output* (IPO) framework [204], a sound theoretical framework concerning teamwork and its effectiveness. *Inputs* include individual's, team's and task's characteristics and work structure, whereas a *process* is the set of activities through which a team interacts in order to achieve a final *output* (subjectively and objectively measurable). According to IPO, the following crucial *processes* are involved in an effective team activity: (i) communication, (ii) coordination, (iii) problem solving and (iv) decision making.

Improving these *processes* by organizing *process interventions* (i.e., implementing team building activities) can lead to better team performance. Compared to SGs recalled above, Asteroid Escape's game mechanics addresses directly or indirectly all these processes.

3.3.3 The Game

The game virtually takes place on a mining asteroid on the verge of collapse. The only spaceship left is stuck in the intricate set of mining tunnels and needs to be saved before the asteroid implodes. Asteroid Escape is a three-player game in which players can take over two roles: ship *Crew* member (up to two) and *Navigator*. Each

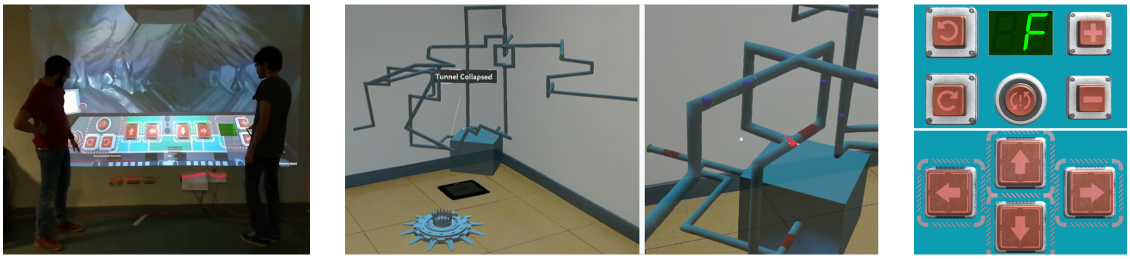


Figure 3.23: Gameplay of Asteroid Escape: *Crew*'s view (left), *Navigator*'s view (center), and pilots' controllers (right).

role has its specific tasks and can access limited information. The *Crew* members are in charge of piloting the ship and have a direct view of the asteroid's tunnels and the obstacles they contain (Figure 3.23, left), but they cannot benefit from a tunnel map during navigation. The *Navigator*, who is virtually located in a remote space control station, must provide the *Crew* with navigation information and early warnings about forthcoming obstacles. To do so, he/she can exploit a 3D tactical map showing (i) the complete tunnel network, (ii) the current position and orientation of the ship, and (iii) a color-coded abstract representation of various types of obstacles that the ship can face during navigation (Figure 3.23, center). A limited time interval to exit the asteroid (13 minutes) was set to solicit fast thinking abilities while preserving control effectiveness and accuracy.

Game design was aimed to foster three out the four critical teamwork abilities in the IPO framework since a successful in-game outcome requires the players to (i) share partial information in a clear way (*communication*), (ii) make participated decisions based on this partial information (*decision making*), and (iii) coordinate their activities in overcoming specific challenges (*coordination*).

Concerning the establishment of effective communication, *Crew* and *Navigator* are required to find a shared vocabulary for two main reasons. First, in order to warn the *Crew* about possible issues during navigation, the *Navigator* must associate the abstract representation of each obstacle to its actual meaning (e.g., gates, tunnel collapses, out-of-control androids wandering in the tunnels, etc.). This meaning can be provided only by the *Crew* when the obstacle is reached. Second, the *Navigator* derives driving directions from the observation of a 3D map (thus, globally) and should be able to translate them for the *Crew* in terms of their local information (i.e., "turn left/right", where left and right are relative to both the forward motion direction of the ship and its orientation along this axis). Since communication is critical for the positive outcome of the game, we included an initial tutorial session where the *Navigator* is instructed to illustrate the *Crew* the mission's objectives and the ship's controls.

Coordination is enforced throughout the game experience in two ways. First, each *Crew* member is provided only with partial and complementary controls (Figure 3.23, right). One pilot controls the orientation (pitch and yaw), whereas the other one manages throttle, roll, and stabilization; missiles can be launched by simultaneously pushing a fire button located on the interactive wall. In this way, pilots are challenged to coordinate their inputs in order to control the spaceship effectively. The coordination between pilots is vital since they must avoid collisions (which may damage the ship and, eventually, destroy it), manage a limited amount of fuel, and promptly react to active dangers (e.g., enemy androids) by shooting at them. Additionally, coordination between *Crew* and *Navigator* is solicited by including tunnel portions where the ship's headlights do not work, thus requiring pilots to rely only on the *Navigator*'s indications.

Finally, in order to encourage *decision making*, all the players are required, during navigation, to discuss and reach a joint decision on which is the best route based on (i) their current position and available information, (ii) the ship's status (fuel and damage) and (iii) the presence of obstacles (enemies or collapsed tunnels). To solicit fast thinking abilities while preserving control effectiveness and accuracy, we decided to set a limited time interval to exit the asteroid (13 minutes). However, this choice made counter-productive the use of puzzles placed along the route, which were initially considered in the design. These puzzles were meant to enforce collaborative problem solving (the fourth relevant teamwork process defined in the IPO framework). However, preliminary tests showed that they were time-consuming and could significantly slow down the game's pace. Despite this change, it is worth observing that problem-solving abilities are addressed implicitly throughout the game, whose goal is to solve an escape problem ultimately.

In a complete game session, all the players are supposed to assume the role of each *Crew* member and of the *Navigator*. By experiencing the gameplay three times, they are expected to understand the difficulties faced by their teammates in previous turns, which should improve team cooperation, cohesiveness and, overall, remove disagreements by favoring critical thoughts on other players' behavior.

Architecture and Technology

We exploited a specific input-output technique for each game role to maximize the sense of immersion while providing an interface as natural and user-friendly as possible.

The *Crew* members are physically located in front of a large touch-sensitive interactive wall, which displays an immersive view from the spaceship's cabin and includes 3D interactive buttons that players can use to pilot the spaceship (Figure 3.23, right). Pilots are provided with a smartphone to operate the main ship's controls (i.e., directions and throttle), whereas the least frequently used commands (i.e., ship ignition and missiles launch) could be accessed on the touch wall. This choice allowed us to minimize the complexity of the hand-held interface while providing a "direct" form of interaction (through the large touch display) with the digital contents.

The *Navigator* is located in a different room than the *Crew* and relies on an AR device (namely, a Microsoft's Hololens) to visualize a holographic representation of the tunnel map. Although indeed, a desktop interface could be sufficient for this task, the AR-based display allowed us to "register" 3D information to the physical space surrounding the player. Thus, he/she can physically move in the environment to inspect the map from different perspectives and distances. This choice enables a natural interaction and fosters spatial presence, which are both key elements for improving immersion [275].

Overall, the system consists of three separate applications connected by a client-server architecture. The server application is hosted on a workstation and manages the current simulation state, controlling the projected interactive wall and collecting in-game analytics. The two remote spaceship's controllers and the AR viewer are connected as clients. *Crew* and *Navigator* communicate using a VoIP chat.

The final setup, here described, is the result of numerous iterations aimed at efficiently combining all the different technological supports employed (i.e., AR headset, interactive wall, mobile application) in Asteroid Escape. The objective was to ensure an optimal level of usability while maximizing the sense of immersion. Details and rationale of the different choices taken are discussed in Section 6.1.3.

3.3.4 Evaluation Methodology

We experimentally assessed Asteroid Escape to evaluate: (i) the perceived learning effects, (ii) the likeability and effectiveness of the game elements for fostering communication, (iii) the effectiveness of Asteroid Escape to automatically self assess users' teamwork competencies, and ultimately (iv) assess if the design choices granted high levels of usability.

To carry out this investigation, we performed a user study which involved 12 volunteers selected among Computer Engineering students of Politecnico di Torino, Italy. Volunteers underwent a game session grouped in four teams of 3 people (with each team member not knowing the others). Each team was requested to go through the tutorial and then play a complete game session by rotating roles three times. To analyze user's verbal interactions each game session was video recorded.

In the following, we describe the recorded metrics used for the study, both subjective and objective.

Metrics

We collected subjective measures employing three separate questionnaires. A pre-experience questionnaire (10 custom questions) aimed to collect information on personality traits [38] and to investigate the degree of relationship relative to other teammates (if any). A second intra-questionnaire (14 questions extracted and adapted from [38, 281]) was delivered after each game run to record personal subjective observations about the team's performance and cohesion. Finally, after the last run, volunteers were asked to answer a post-questionnaire (105 statements) made up of four sections. The first investigated tool's usability from different perspectives by combining the SUS [44], the five attributes defined by Nielsen [220] and questions aimed to assess the perceived sense of immersion/presence and the appropriateness for the tasks at hand, of the given display system [166]. The second section aimed to assess the game's likeability and effectiveness as a tool for developing communication abilities. The third section analyzed the impact of learning effects of the three game runs on the various game elements. The last section evaluated again team skills and performance by means of questions complementary to those asked in the intra-questionnaire.

We gathered quantitative data by visually inspecting the video footage recorded during gameplay and by logging game statistics such as ship collisions, average speed, decision time, tunnels navigated more than once, androids destroyed, and so forth. A game performance score was also computed based on these variables by combining them with positive and negative weights. Visual analysis was performed by an external observer, who labeled players' verbal actions (i.e., events) according to a methodology inspired by [252]. The set of labels considered all possible communication channels between *Navigator* and *Crew*: *Crew* to *Crew* member (C2C), *Navigator* to *Crew* and vice-versa (N2C and C2N respectively). The labeled events were: information sharing (IS) (provide or ask), request action (RA), and complaint (C).

Regarding subjective measures, statements were rated on a Likert scale from 1 (strong disagreement) to 5 (strong agreement). Reverted statements were flipped to normalize scores. A scale reliability test was first conducted on the inter- and post- questionnaires, obtaining a standardized Cronbach's Alpha equal to 0.85 and 0.75, respectively, which suggested an acceptable to good internal consistency of the results (overall). In the end correlation factor analysis where conducted among objective and subjective measurements to gather information on the validity of *Asteroid Escape* as unsupervised assessment tool.

3.3.5 Results

Usability and Immersion

Volunteers had low experience with AR devices (either HMD or screen-based) as 75% of them had never or seldomly used one. On the contrary, 42% of testers used an immersive VR, daily or weekly, and with the same frequency, more than half (66%) played digital games. Concerning usability (first section of the post-questionnaire), the *Navigator* (i.e., the AR-HMD) and *Crew* (i.e., wall projection and touch controls) interfaces were analyzed separately. In the following, they will respectively be addressed as NAV and CRW. A normalized SUS score of 76.67 (NAV) and 74.37 (CRW) in the 0–100 range suggest a good to excellent usability provided by the different interfaces. Similar conclusions can be drawn by considering Nielsen's attributes of usability. NAV mean scores (and standard deviations) were as follow: learnability 3.33 (SD = 1.10), efficiency 4.25 (0.72), memorability 4.33 (0.74), possibility to recover from errors 3.16 (0.69), and satisfaction 4.21 (1.01). For CRW: learnability 3.80 (0.90), efficiency 4.17 (0.69), memorability 4.33 (0.85), possibility to recover from errors 3.08 (0.86) and satisfaction 4.08 (1.32).

These attributes are very similar between the two interfaces. The worst aspect of both has been the possibility to recover from errors. From observing the game recordings, we assume this is mainly due to situations where the spaceship would enter a chain of crashes causing frustration to both the *Crew* members and the *Navigator*. The former had issues taking back control of the ship, and the latter had difficulties telling the other team members how to orient the ship properly, resulting in more crashes. The attribute in which the two interfaces differed the most is learnability (CRW had a higher than NAV). We believe this is mainly due to the lower experience users had with AR applications. Despite this, both interfaces had high levels of satisfaction, efficiency, and memorability.

Finally, sense of immersion/presence was rated 3.75 (0.82) for NAV and 3.83 (0.90) for CRW, whereas appropriateness scored 3.92 (0.64) and 4.08 (0.65), respectively, for the two systems. In both cases, scores were relatively high, confirming that the experience effectively reached a high level of immersion which was one of the main goals throughout the design phase (and its iterations).

Gaming Experience and Communication

Table 3.15 reports users' answers in the post questionnaires section addressing the game experience. From items 1 and 2, it can be seen that communication was evaluated as a key element to succeed in the game (4.67 and 4.25 respectively), and overall, users felt that by playing the game, they improved their ability to work in a team (item 6).

Items 3 and 4 support the assumption we made in designing the game: by switching roles across game runs, users would improve team cohesion due to players better understanding each other perspectives and difficulties faced (i.e., each role had a unique set of challenges to overcome). Moreover, item 8 highlights how team members were committed to the overall game objective. However, users had mixed feelings (i.e., item's 5 high standard deviation, 1.50) whether by repetitively playing in the same role they would have yielded better game performances. Although this might be partially true, as keeping the same role would have improved

Table 3.15: Game Experience

#	Statement	Score
1	To complete the game successfully it was essential to communicate	4.67 (0.47)
2	The game motivated me to find the best way to communicate with the other team members	4.25 (0.82)
3	Having taken on the different roles has been important to understand the other players' perspective on the game	4.58 (0.64)
4	Taking on the other roles helped me to improve my abilities to communicate with the team	4.25 (0.72)
5	Team would have obtained better results by having each player take on the same role all the time	3.08 (1.50)
6	Overall, I think that this game could improve my ability to work in team and to communicate	3.92 (0.64)
7	Reaching the assigned goal was very important for the team	4.33 (0.85)
8	The team worked hard to reach the goal	4.25 (1.09)
9	The game was fun	4.50 (0.50)
10	I liked the game	4.75 (0.43)
11	I perceived the game as mainly educational or ludic	3.92 (1.04)

Table 3.16: Learning effects.

#	Statement	Score
1	Recall of the path/the map	2.00 (1.08)
2	Knowledge of controls/interactions	4.08 (0.95)
3	Coordination within the team	3.92 (0.76)
4	Familiarity in communicating with the team	4.33 (0.62)
5	Effectiveness of communication	3.92 (0.86)
6	Development of a lexicon/of rules in the team	4.08 (1.08)

individual skills in the game, we stress that the core objective of this SG was to solicit and strengthen collaborative behaviors, beneficial for team cooperation, rather than providing a sense of mastery in the game mechanics and scenario. To verify the viability of this assumption, the post-questionnaire's third section investigated the learning effect (self-perceived) solicited by performing multiple runs. From the obtained results (Table 3.16) it can be seen that users did not recall the map or the path leading to the asteroid's exit (item 1). On the contrary, multiple game runs prominently influenced the development of a shared lexicon (item 6) and improved communication and coordination among team members (items 3-5). These observations are furtherly supported by the intra-questionnaire's results (Figure 3.24) where indeed users' evaluation of communication abilities and team cohesion was positively evaluated and improved across multiple runs (mainly between 1 and 2). This outcome roughly suggests that experiencing the game from the different roles contributed to developing collaborative abilities and strengthening uses' sensation

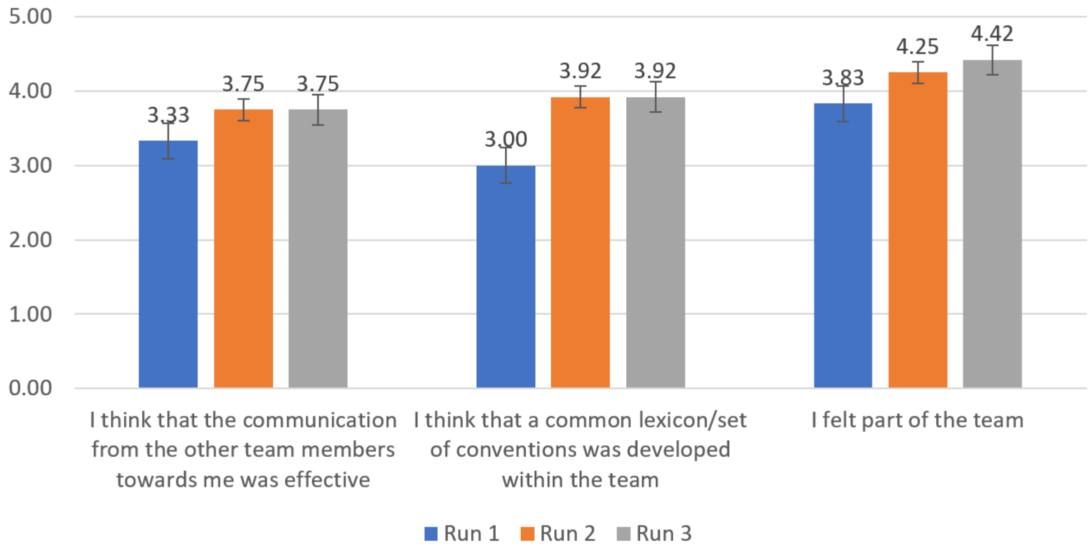


Figure 3.24: Intra-questionnaire, most relevant results (three runs) show the positive progression of the considered dimensions.

of being part of a team.

Finally, items 9 and 10 (Table 3.15) highlight that users' liked the game (4.75) and had fun playing it (4.50). This is an important finding because it suggests that we could combine the serious and playful aspect of the experience, delivering an effective learning activity that users perceived as more ludic rather than educational (item 11). Combining the above results with the high immersion perceived by users helps us suggest the possibilities (and benefits) of exploiting a heterogeneous technological enhanced environment (also featuring AR) to deliver more engaging and motivating SG experiences.

Quantitative Results

Quantitative results (video analysis and game statistics) were first cross-compared and then compared with subjective results. Pearson's correlation coefficients for quantitative results are reported in Table 3.17. As it can be seen, the ship's average speed is a good factor to infer the Request of Action (RA) among *Crew* members (C2C), whereas the number of collisions reflects the overall RA on all the communication channels. This finding suggests that the more users communicated more effective (i.e., faster and precise) their navigation within the asteroid. Similarly, the more users engaged in Information Sharing (IE) events lower the number of tunnel segments navigated more than once. It should be reminded that only one path would lead to the asteroid's exit, and ending up in a tunnel segment

Table 3.17: Correlation analysis on quantitative results.

Factor A	Factor B	r	p
Average speed	RA (C2C)	0.75	0.008
# Collisions	RA	-0.77	0.006
# Segments navigated more than once	E	-0.54	0.086
# Segments navigated more than once	IS	-0.52	0.098
Game Win	E	0.99	0.001

previously explored may suggest users being lost. Likely due to ineffective communication between the *Navigator* and the *Crew* members. Moreover, we detected a strong correlation between encouragement events (E) among the players and a successful game outcome, hinting at a higher performance of teams that were more prone to support each other through positive communication. Nevertheless, the limited amount of data available (i.e., not all teams always could reach the tunnel's exit) could bias this finding. Finally, results indicate that game statistics are quite a good proxy of video analysis data since there was a medium to high correlation between the communication events and the performance score (N2C = 0.73, C2N = 0.48, C2C = 0.79, overall = 0.76, all with $p < 0.05$).

3.3.6 Conclusions

In this section, we have presented the design and implementation of Asteroid Escape, an AR-SG envisioned as a team-building activity for promoting a set of different soft-skills. Two reasons mainly drove the choice to address this particular learning domain: (i) the limited evidence of SG targeting this learning domain [65, 36] and (ii) the belief that a highly immersive environment (also featuring an AR interface) could be beneficial to users' learning and collaborative outcomes.

Our assessment results show that the AR-SG has a promising potential in developing communication interactions and trustiness among members of a forming team. The different game design choices which mainly solicited these results are (i) role complementarity, (ii) partial information sharing, and (iii) collaborative problem-solving. Also, we believe another element that accounted for these positive results was the highly immersive experience delivered to players. The setup involved a large interactive wall, mobile devices which replicated the spaceship controls, and an AR HMD through which players could spatially explore the asteroid's tunnel map. Altogether, these elements might have contributed to making the players feel part of the virtual fictional experience leading to more significant commitment towards the game outcome (i.e., by losing, they would have been destroyed with the asteroid).

3.4 Conclusions

This Chapter explored the application of AR-SGs approaches in three specific learning domains that were given little attention in previous AR-SG research. These are domains where (i) AR-based and SG-based learning were (primarily) used individually, and (ii) is it possible to identify affordances of either AR or SG that could be exploited to enhance the other. We designed and developed an AR-SGs to target each domain, with the overarching objective to explore if powerful synergies between AR and SGs could be established. More formally, we aimed to answer RQ1, which for clarity, we repeat here.

RQ1: Can we establish synergies between AR and SGs in order to develop effective AR-SGs in a variety of learning contexts?

The results reported in this Chapter seem to suggest an affirmative answer to this question. However, let us address this RQ in greater detail.

On the one hand, from the experience of Holo-BLSD, we have emphasized how gaming elements can be effectively exploited in, procedure learning, an educational context typically explored through AR-only solutions. In particular, our approach "enriched" AR-based procedure learning with the following benefits: (i) increased engagement with the learning path, (ii) willingness to repeat the training activities (with positive effects on knowledge acquisition and retention), and (iii) an overall positive sensation of fun, which is beneficial in terms of learning outcomes.

On the other hand, Sustain and Asteroid Escape provide a tangible example of solutions addressing learning domains most frequently explored through SG-based approaches. In both domains, AR adds to the current SG-based solutions two main benefits: (i) facilitate and improve uses' collaboration and (ii) improve the learners' immersion and the contextualization of the learning experience.

Sustain exemplifies how learning about a complex system can be approached with AR-SGs. In particular, Sustain deals with sustainability learning, which, according to previous literature, can significantly benefit from constructivist learning approaches, especially the ones supported by SGs. However, surprisingly, a limited number of solutions using AR have been proposed. Our work showed how the design of Sustain and the synergies between AR and SGs could support the development of an adequate socially mediated learning context that supports face-to-face interactions between peers. The shared game environment allows players to experience rich multimedia educational information merged with the physical training environment, showing how AR affordances can be integrated with SG ones to reach the intended educational objectives.

A similar conclusion can be drawn for Asteroid Escape. Soft-skills learning is a field more and more turning to SG approaches thanks to their unique features.

However, little attention has been given to AR-enhanced SGs, although AR can promote collaborative behaviors and increase the sense of immersion through its unique visualization properties. In soft-skills learning, users are often set in a fictional context where they are challenged to solve a series of challenges, and by doing so, they learn to cooperate, communicate and negotiate. The results collected in the Asteroid Escape evaluation emphasize how enriching an SG with AR features can lead to delivering a compelling and more immersive soft-skill learning activity, rated by players as fun and engaging, where users' felt committed to the overall outcome of the game. We believe that the main factor accounting for these results was the possibility to easily communicate with each other while sharing the same digital experience, which could be experienced from distinct perspectives, roles, and controls thanks to the affordances offered by AR.

Overall we think these results show promising evidence that AR and SG can and should be combined to create more effective learning tools which exploit each other's unique features. We also understand that these assumptions are only supported by a limited number of positive examples targeting particular learning domains. However, since many of the design choices, frameworks, and methodologies employed can be generalized and transferred to other topics within the same (and similar) learning domains, we think our work provides valuable references for researchers and practitioners to expand/adapt our proposed solutions to other AR-SGs in an attempt to advance the current research in this field.

Chapter 4

Awareness in Collaborative AR-SGs

The experimental results we illustrated in the Section 3.2 suggest that the approach followed in Sustain effectively promotes collaboration between co-located users. However, it must be emphasized that the Sustain setting is not adequate to cover all possible modes of interaction and collaboration in AR. Indeed, Sustain uses game mechanics that stimulate discussion and collaboration among players, but it does not consider as available tools for this purpose the interactions between the users and the AR world and the actions that the users perform in the AR world. In Sustain, the AR environment represents the shared game scenario in which individual players can perform the available actions for their role. The moments of collaboration and discussion serve to find an agreement on the actions that individuals can (and must) implement to maximize the results of the collective plan. Nevertheless, in acting and interacting with the AR environment, users do not need support from their peers because they can (and shall) perform these actions independently.

On the contrary, there are several collaborative scenarios where users need to communicate through the AR world or interact together in the AR world. For example, users may need to focus the attention of their peers on a specific point in the environment or a virtual object they all need to operate on, know what virtual object another user is manipulating, and encourage other users to move to precise locations physically.

All of these actions and interactions may be necessary to complete collaborative tasks assigned to users. These considerations introduce a new problem: *developing methods that can encourage and facilitate human collaboration within the AR world* (and not just at its boundaries, as was the case in Sustain). Among these methods, the key element currently being studied is how to promote user awareness, and in particular, how to effectively leverage visual and audio information to support collaboration [100]. These two aspects are closely related to each other. To communicate (and collaborate) efficiently, users must be able to understand how others are interacting with the shared AR environment, and one effective way to support such awareness is to associate users' interactions with visual cues (VCs) that give immediate cognitive feedback on the actions performed (e.g., pointing, annotating or manipulating objects) [173]. Examples of such VCs are displaying digital replicas of each user, identifying their point of attention as they move within the shared virtual environment, and highlighting the objects the user is focused on or is interacting with.

One relevant thing to point out is that these elements related to collaboration and communication in AR are not limited to the field of SGs but extend to more areas, of which undoubtedly the most relevant is that of Computer Supported Co-operative Work (CSCW, [131]), i.e., the research field committed to understanding "how collaborative activities and their coordination can be supported by means of computer systems" [52]. The link between CSCW and AR is that the recent technological advances in AR allow the introduction of disruptive innovation in CSCW thanks to their ability to overlap the task space with the communication space [193], resulting in more effective collaboration [28, 316]. Specifically, the literature suggests that in AR environments, as opposed to traditional screen-based collaborative environments, people tend to relate to their peers in manners similar to those occurring in a face-to-face interaction not mediated by technology [193].

Given the relevance of VCs for increasing awareness in collaborative AR, which specific VCs to use, and how to implement them are design choices related also to the configuration through which these collaborative systems should be deployed. To encompass and describe such configurations, researchers have adopted different taxonomies. One of the most commonly adopted is the *time-space* matrix [160]. This taxonomy categorizes the approaches in terms of two dimensions. The first is the *time* (i.e., *synchronous*, when user interaction occurs at the same time, and *asynchronous*, otherwise), and the second is *space* (i.e., *co-located*, when users share the same space, and *remote*, otherwise). Another interesting dimension for categorization is *symmetry* (first proposed in [27]), which is further distinguished in two subclasses [279]. The first is *technological symmetry*, which can be either *symmetric*, when all users adopt the same hardware devices, or *asymmetric*, when they use different devices. For instance, HMD and HHD imply different interaction and visualization instruments available to each collaborator. The second subclass is the *role symmetry* and indicates when users experience a shared workspace with different roles, each with different assigned tasks, responsibilities, and available actions that can be executed to modify the shared environment.

Based on this premise, one of the needs that emerged from our research work following the analysis and discussion of the results obtained in Sustain was precisely to understand which tools would be most beneficial to support collaboration in those AR-SGs that require shared interactions within the AR environment. Concerning the taxonomies previously presented, the study described in this chapter will tackle scenarios with the following characteristics: *co-located*, *synchronous*, adopting *symmetric technologies* (e.g., using AR HHD for all users), and involving users in both *symmetric-role and asymmetric-role tasks*.

Since most of the research on awareness in AR has been done within CSCW, we first reviewed the literature in this field to gather helpful information. What we found is that recent reviews [100, 279] have analyzed the current state of the art in AR-CSCW, suggesting areas that need further research. In particular, two issues emerge. First, the scenario where users adopt *symmetric technologies* and *asymmetric roles* is mostly unexplored compared to other settings [279]. Second, despite evidence that a team working in a co-located space and synchronous time is more productive [196], recent research is primarily focused on remote experiences (with a certain unbalance towards symmetric configurations), and studies that approach the *co-located context* are significantly older (having been published primarily between 1995 and 2004 [100]). However, in the last fifteen years, AR technologies have made giant leaps forward, moving from emerging to mature status [172, 23], thus requiring novel analyses of the AR-CSCW in this specific context.

In light of these considerations, the current literature lacks a thorough analysis aimed at evaluating the effectiveness of different VCs in improving users' awareness and supporting their communication in collaborative AR scenarios with the specific characteristics required in our SGs (i.e., *co-located*, *synchronous*, adopting *symmetric technologies* and offering both *symmetric-role* and *asymmetric-role* tasks. Therefore, we thought it would be interesting to conduct such an investigation evaluating different VCs inspired by solutions proposed in the literature. This includes the Shared Point of Interest (SPOI), a novel exocentric pointing technique based on the creation of VCs registered in the common (and real) 3D space shared among all users. To comprehensively evaluate the effectiveness of the proposed VCs, we analyzed them in two different contexts, first in a non-SG one, through several evaluation scenarios, and then in an AR-SG one, through ARScape, a multiplayer AR-based Escape Room.

In the evaluation scenarios, users are asked to complete typical and high-level tasks (i.e., they can be applied to any application domain) using the different proposed VCs. For example, users are asked to reach a specific location in the real environment or to focus their attention on a specific element. These scenarios involve users in both symmetric and asymmetric tasks. Thanks to these scenarios, we were able to evaluate the individual contribution of each VC in completing such tasks, which can occur in any AR collaborative application, and among them AR-SGs.

The second step in evaluating the VCs was to assess their effectiveness in ARScape, the AR-SGs developed specifically for this analysis. In ARScape, which exploits HHD to interact with the AR environment, players are involved in various puzzles that they can only solve through collaboration.

The interest in this type of activity in learning soft-skills is evidenced by several papers that have developed Escape Rooms in AR and VR [317, 102, 96]. An Escape Room can indeed be considered a perfect team-building activity. It helps develop the communication skills that are essential for teamwork. Teams must work together to find clues, figure out how they fit together, and learn what strategies are best to get to the puzzle's solution. The tight time constraints make it possible to find out how well (or poorly) the team communicates under pressure, allowing its members to develop effective collaboration skills in these situations. Escape rooms teach players that the group wins (or loses) together. This is a valuable lesson to internalize so that everyone realizes that they are working for the common good and that their work is important to the overall outcome. Finally, one of the most notable features of Escape Rooms is the ability of the group to build trust quickly. This is achieved by analyzing the work that each member does to contribute to the end goal.

The rationale to initially evaluate these VCs using "generic" scenarios and their associated tasks was dictated by two decisions. First, we wanted to gather insights into how and to what extent each VC was helpful in the tasks separately (i.e., evaluated individually). The insights from these analyzes can then be applied to more complex scenarios, such as an AR-SG (i.e., ARScape) that may contain all tasks in the same scenario (e.g., users must simultaneously manipulate objects and position themselves at specific locations in the environment). Second, we wanted to align our experimental methodology with those previously conducted in collaborative AR research, mainly concerned with evaluating VCs in simple and "one task at a time" scenarios.

The proposed VCs were assessed in both contexts by evaluating qualitative and quantitative data from 40 volunteers who first completed the various assessment scenarios using different combinations of VCs. Subsequently, users were able to play ARScape, benefiting from all available VCs. The experimental results show that the proposed VCs help to support user attention in a shared environment by providing contextual and spatial information for all participants. Users rated most of the proposed VCs positively. In particular, SPOI was the most recognized, and once mastered, it also contributed to faster completion of the most challenging asymmetric tasks.

Finally, we note that some of the results of our work, although obtained in a AR HHD environment, can be extended to other scenarios (e.g., remote collaboration) involving other devices (e.g., AR HMDs or a combination of heterogeneous hardware).

The rest of the chapter is organized as follows. In Section 4.1 we provide the relevant background in AR-CSCW research focusing on users' awareness. Then in Section 4.2 we describe the SPOI and present the selected VCs inspired from the literature. Section 4.3 illustrates the evaluation scenarios, the experimental protocol and the collected measures, concluding with the results from this first VC assessment. Finally, Section 4.4 presents ARScape and discusses the results from this second VC assessment.

4.1 Related Works

AR collaboration and awareness. The topic of collaboration in AR, both remote and co-located, has been explored in several studies discussed in the Introduction [100, 279, 238]. In addition, the survey [193] presents interesting considerations and highlights open issues. Some of the key findings are that co-located AR-CSCWs allow users to interact with AR content as naturally as with real-world objects and that AR improves communication by reducing the separation between task space and communication space. Nevertheless, more research is needed to analyze the interaction modalities between co-located (and remote) users and how VCs can improve awareness, which is one of the goals of our work. For the sake of clarity, we start with a definition of awareness, which in CSCWs can be defined as the awareness of people and their interaction with the workspace, rather than just the awareness of the workspace itself (i.e., the perception and understanding of the relevant elements in the environment) [136]. This element is particularly important in shared AR environments where the actions of other users are not fully visible, as it is not always easy to understand where users are looking or what virtual objects they are interacting with.

Some results on awareness in AR-CSCWs are presented in [75], which evaluates the usability of a smartphone-based AR system for crime scene investigation. By placing augmented content in the space, two investigators (one on-site and one remote) collaborate in identifying evidence and potential hazards in a (reconstructed) crime scene. The experimental results highlight the importance of remote assistance in supporting the investigative process and the superior usability of smartphone-based interfaces compared to those using AR HMDs. More recent research [241] explores how different VCs (e.g., avatar, gaze raycasting, FOV volumes) can improve awareness in remote collaborative tasks involving both AR and VR users (such as asking two users to look at the same object to “reveal its secrets”, as done in one of our experimental tasks). Although the authors conclude that the proposed VCs were appreciated by the subjects and were instrumental in improving their performance, this study focuses only on remote collaboration. Moreover, it only highlights the usefulness of VCs without examining which ones are most effective. Finally, the recent survey presented in [247] describes a list of design features for AR HMD collaborative experiences gathered from 92 papers and concludes that further evidence is needed to evaluate the effectiveness of different features in promoting awareness.

AR VCs for spatial awareness. While researchers have thoroughly explored the use of VCs to guide users through a virtual world, the visual language of VCs for spatial awareness in AR is in many ways still in its infancy [100]. Among the reviews of AR VCs, one of the most interesting is [84], which analyzes various VCs in 49 commercial video games and explores their potential use in AR systems. The authors cathegorize VCs along three dimensions: *purpose* (what the cue is for), *trigger* (how and when it is activated), and *markedness* (its visual representation). While VCs' appearance (*markedness*) can be characterized by different nuances, there are three classes for their *purpose*: *Look* (tell the user where to focus their attention in a timely manner), *Discover* (highlight points of interest), and *Go* (to aid navigation). VCs can also have different *triggers*: *user-triggered* (the user actively decides when and where to visualize the hint), *context-triggered* (the VC is generated by the system in response to a specific event), *agent-triggered* (the VC is triggered by another agent or user), or *persistent* (the VC is always visible and usually not spatially registered, e.g. a navigation map or compass overlaid in 2D on the displayed image).

Despite their importance for improving awareness in AR-based CSCWs, most studies in the literature focus on the effectiveness of VCs in single-user scenarios (often involving specific tasks). For example, [89] compares the effectiveness of different marker shapes for spot-welding in automotive manufacturing, emphasizing the dramatic improvement in weld precision provided by a visual reference point. The authors of [312] evaluate the effectiveness of different VCs for a procedural task (identifying the correct sequence of buttons to press), testing both target-based and direction-based cues, and show that the latter are more effective for this type of procedure. The findings demonstrate how clear cues (those with the least ambiguity and interference) lead to better results and fewer procedural errors. The paper [150] analyzes the effectiveness of different VCs in promoting spatial awareness. The study uses projection-based Spatial Augmented Reality, in which a local user (with a projector-based setup) interacts with different remote users using VR or AR devices. The experiments show that a combination of many VCs improves awareness with only two users, but as the number of collaborators increases, excessive visual clutter can become confusing and detrimental to awareness. Another study that combines remote collaboration between AR and VR views is [241], which analyzes different combinations of VCs to promote awareness. According to the results, displaying the FOV and pointing with a Gaze Ray are the most effective.

AR VCs for pointing. Pointing is one of the most important elements for awareness since it can draw attention to objects and locations where an action or task is being performed or planned to be performed [238]. In AR, pointing is managed through egocentric and exocentric metaphors [31]. Egocentric pointing VCs are synchronized with the user's point of view. One solution is to center the VCs on the user's field of view and continuously update their position according to the orientation of the display device [215, 241, 296]. Another solution is to overlay 2D pointing annotations (e.g., arrows, pointers, cursors) on the video stream presented to the user. For example, this solution is used when a remote expert helps a user with an AR device (i.e., the *remote-expert* scenario). The remote expert can attract the attention of the local user by overlaying graphical content on top of a real-time video stream captured by the device's camera, which is then displayed on the assisted user's screen [174]. However, since the overlays are not registered with the real environment, they no longer match the objects they refer to once the user device changes its position or orientation. A similar approach was discussed in [296] for a 360 panorama-based Mixed Reality approach. It was concluded that overlays improve the social aspects of remote collaboration and reduce the required mental effort for the task.

On the contrary, exocentric metaphors detach VCs from the user's point of view. One solution is to anchor the VCs in the environment, ensuring spatial coherence from every point of view. The authors of [120] have developed a system where a *remote-expert*, from a computer workstation, can create spatial pointing annotations that are visible to a local user in AR. In the study, the proposed technique was compared with a 2D egocentric VCs solution. Although the evaluation resulted in a higher preference (80%) for the spatial VCs, the authors emphasized some limitations. The most notable is that annotations could only be instantiated by the remote expert from her computer workstation, highlighting the need to extend this exocentric pointing approach to a multi-user setup. The lack of evaluations of exocentric spatial pointing solutions validated in colocated multi-user settings, has been raised in the review proposed by [279]. Similarly, the authors of [321] also point out that further evidence and more data are needed on the impact of awareness VCs in AR-HHD co-located collaborative environments.

The authors of [306] discuss different approaches to positioning label placeholders, such as free positioning in 3D space, registering labels with 3D objects, and using paper reference markers placed by users in the environment and associated with specific labels. User evaluation showed that volunteers appreciated the proposed VCs, although the small number of participants (six volunteers) prevented generalization of these results.

A different approach is proposed in [211], where the authors use shared virtual landmarks (SVLs) to improve user communication in remote collaborative environments. SVLs are virtual "natural" objects (e.g., lamps, plants, chairs) that augment the visual context shared by participants and allow them to develop a better mutual understanding of their shared space. In addition, they use them as spatial anchors to convey spatial information. Users appreciated SVLs, but also pointed out that SVLs increase the visual clutter of the environment and are static (i.e., they cannot be placed or disabled by users).

Given these observations, the main goal of this paper is to determine whether and to what extent VCs improve user collaboration and mutual awareness of co-located users in experimental settings using AR-HHD devices. In particular, we aim to assess the effects in a co-located setting of the Shared Point of Interest (SPOI), an exocentric spatial pointing technique where each user can mark or highlight an element by creating a 3D VC registered in the (shared) real environment.

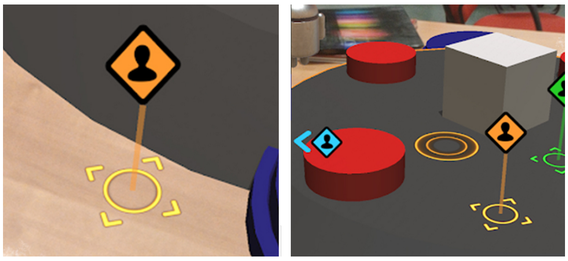


Figure 4.1: Shared Point of Interest (*active* visual cue). *Left:* visual representation of the SPOI when it falls within the FOV. *Right:* Color-coded multi-user visualization showing different SPOIs placed by different users; note the blue cursor on the left side of the image, signaling an SPOI outside the FOV.

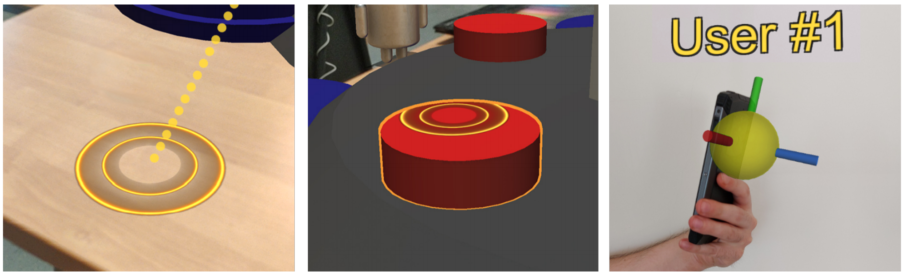


Figure 4.2: *Passive* visual cues. *Left:* Line of Sight. *Center:* Highlight. *Right:* Avatar

4.2 Implemented VCs

Concerning the VCs offered to users in the different assessments (i.e., the evaluation scenarios and the AR-SG), the first is a *user-triggered* VC called **Shared Point of Interest** (SPOI). The adjective "shared" indicates that the POI is visible to all collaborators in the shared augmented environment, and not only in the private space of the user who owns it. Since it is triggered by the user, in the following, we refer to this VC as *active*. The SPOI is an exocentric pointing VC, registered in the 3D virtual space. The user can choose where to place the SPOI by selecting a location on the screen of the device that displays the augmented environment. Double-tapping the screen instantiates the SPOI and determines its 3D position using a raycasting technique that computes the intersection of the line of sight with the virtual objects or environmental planes. Tapping on a placed SPOI (owned by the user) deletes it (Figure 4.1, left). The SPOI is intended to allow users to determine the location where they want to attract the attention of their peers. To avoid visual clutter, each user has only one SPOI available. Since the presence of a placed SPOI should be obvious to all users, a cursor displayed on the HHD screen side (Figure 4.1, right) draws the attention of other users to any SPOI outside their (often limited) FOV.

SPOI is then combined with other VCs from the *persistent* and *context-triggered* classes (hereafter referred to as *passive*). Inspired by the previous literature [84, 100, 312, 241, 150] and taking into account the specificity of the scenarios (which is further explained in the following Section), we selected the following passive VCs as the most suitable ones. A common feature of these VCs is the use of a color code to improve comprehension and information transfer by assigning a unique color to each user and their respective VCs.

- **Line of Sight**: (*persistent*) a segmented ray that highlights the user's line of sight, originates from and is perpendicular to the HHD, ends at the intersection with a virtual object or plane in the environment, and has a limited extent if it does not intersect anything. Finally, if the Line of Sight intersects with a virtual plane or object, a cursor consisting of two concentric circles is displayed at the intersection point (Figure 4.2, left).
- **Highlight**: (*context-triggered*) this VC highlights for a user when an interactable object is in her line of sight by adding a colored outline to the object (with the same color as the user). The same VC is used to signal peers (with the appropriate color) when a user picks up and manipulates a virtual object (Figure 4.2, center).
- **Avatar**: (*persistent*) this VC provides various information, including the position of the user, the user name, and the color used to represent her VCs. The *Avatar* is represented as a colored sphere, rendered at the HHD position and topped by the user's name (Figure 4.2, right). While the *Avatar* is a necessary VC for a remote setting, it can also be useful in a co-located one to improve position and motion understanding when a user and the environment (or other users) are occluded, or to verify that the correct user is moving to the expected location.

4.3 VC Assessment: Evaluation Scenarios

For the first assessment of the proposed VCs, we defined three basic scenarios, each focusing on one of the main purposes for which VCs are used, identified by [84] as *Go* (i.e., encouraging users to reach a specific location in the real-world environment), *Look* (i.e., direct users' visual attention to a specific object or part of the environment), and *Discover* (i.e., direct users' attention to elements that are interactive and can be manipulated). These scenarios involve users in simpler symmetric tasks and more complex asymmetric tasks, where a combination of VCs supports the users' activities. We emphasize that, in addition to serving as toy scenarios, they also reflect a variety of real-world use-cases of collaborative AR applications, some of which are exemplified in Table 4.1.

Figure 4.3: Virtual Envronments for each scenario. *Left:* Navigation Scenario (*Go*). *Center:* Gaze Scenario (*Look*). *Right:* Manipulation Scenario (*Discover*).

Each scenario is divided into different tasks. The first ones are less complex and assign the same role to all users (symmetric role tasks). The last task involves role asymmetry. It is the most complex and therefore the most suitable for evaluating the proposed VCs, as it presents the users with the most challenges in communication and collaboration.

Since we are also interested in exploring the relationships between passive and active (i.e., SPOI) VCs, we defined two distinct visual configurations, V0, which offers users only the passive VCs, and V1, which combines active and passive VCs. As described later, we asked users to repeat each scenario twice using the two different visual configurations.

Navigation Scenario (NS) - *Go*

In this scenario, the VCs are used to encourage users to physically move to a specific spatial location in the real environment. Real-world use cases include any application where navigational guidance is needed. Examples are museum guides in AR, AR-aided sight-seeing for cities or tourist destinations, industrial maintenance tasks (where workers must correctly position themselves relative to machinery or gather in a safe spot during dangerous tasks), and collaborative design (where collaborators should occasionally observe an object from a specific point of view). Similar collaborative navigation tasks can be found in [177] and in the literature related to Collaborative Virtual Environments.

The Navigation Scenario consists of eight platforms arranged in a circle on the ground (Figure 4.3, left) that users must physically reach in order to complete the assigned task.

The Navigation Scenario is divided into three tasks:

- *NS-Task 1*: one platform is chosen at random and marked with a red beacon. All users must simultaneously position themselves on (or around) it for a few

Scenario	Task	Use Cases Examples
Navigation (*Go* purpose)	*NS-Task 1*	Maintenance: all workers must reach a safe spot during a dangerous task
	NS-Task 2	Collaborative Design: collaborators must look at an item from different viewpoints
	*NS-Task 3**	AR-based Museum Guide: a guide tells visitors where they should go next
Gaze (*Look* purpose)	*GS-Task 1*	Maintenance: workers focus their attention on a malfunctioning part of machinery
	GS-Task 2	Collaborative Design: each user focuses on a different part of an item
	*GS-Task 3**	Tour Guide: all visitors should look at a specific exhibit highlighted by the guide
Manipulation (*Discover* purpose)	*MS-Task 1*	Maintenance: a worker must use the correct tool on a specific part of machinery
	*MS-Task 2**	Assembly: a remote expert guides a worker through an assembly process

Table 4.1: Examples of practical use cases for each evaluation scenario and specific task (asymmetric-role tasks are indicated with “*”).

seconds. The task is repeated three times, each time selecting a different platform.

- *NS-Task 2*: the application highlights n platforms (where n is the number of users collaborating), each with the color assigned to a user. Users must position themselves for a few seconds on the platform of their color. The task is repeated three times.

- *NS-Task 3*: platforms’ activation and coloring are the same as in NS-Task 2. However, only one user (*guide*) can see the colored information, whereas others (*followers*) only see the inactive platforms. The guide’s goal is to direct each follower to the correct platform (i.e., the one with the same color as the followers). This task is repeated n times (where n = number of users collaborating), assigning the role of the guide to a different user in each iteration.

Gaze Scenario (GS) - *Look*

In this scenario, we evaluate the effectiveness of VCs in promptly directing the user’s visual attention to a specific object or part of the environment that requires timely action. This feature is fundamental to a wide range of use cases where remote workers or experts need to direct the attention of others to specific objects or their components. Examples of these use cases include maintenance tasks, crime scene investigation similar to [75], location-based collaborative AR games, collaborative design sessions, and tour guides. Other examples of similar tasks in the literature can be found in [177, 211, 150, 296, 120, 321].

In the Gaze Scenario, users must synchronously or asynchronously direct their gaze to one or more cubes arranged on a 4x4 grid (Figure 4.3, center). This scenario is divided into the following tasks:

- *GS-Task 1*: all users must collectively (and for a few seconds) direct their gaze to the cube highlighted by the application. This task is repeated three times, with each repetition selecting a different random cube from the grid to look at.
- *GS-Task 2*: the application selects n random cubes (where n is the number of users), each marked with a different color. To complete the task, all users must simultaneously fix their gaze on the cube of their corresponding color for a few seconds. The task is repeated three times with a different cube selection.
- *GS-Task 3*: the cubes' activation and coloring are as in GS -Task 2, but we design the task as an asymmetric role activity. Only one user (the guide) can see the coloring information on her screen and must guide the other users to gaze the correct cube. The task is repeated n times, where n is the number of users, and at each repetition a different user takes the role of the guide.

Manipulation Scenario (MS) - *Discover*

In several scenarios, users' attention needs to be drawn to objects that are interactive and can be manipulated. The most significant use cases are collaborative maintenance and assembly tasks, where specific objects or parts of an object need to be interacted with or correctly aligned and placed in a specific location in order to complete the procedure. More generally, the ability to clearly show users which objects they can interact with to complete their tasks (or which others interact with for the same purpose) is of utmost importance in any (collaborative) AR environment. For more examples of similar tasks, see [177, 211, 150, 296, 120, 135].

The Manipulation Scenario provides several graspable objects and virtual holes (Figure 4.3, right) with three different shapes (i.e., square, rectangle, or triangle). Holes and objects are also associated with a number. This number is constant for a hole and can be changed for an object by interacting with a virtual red dial on the side of the object. The user's task is to correctly move and rotate an object to insert it into the corresponding hole so that the numerical values of the object and hole match. For our experiments, we defined the two following tasks:

- *MS-Task 1*: user has to place the object with her/his color in the correct hole and with the correct number. The task is repeated three times with different combinations of objects and numbers per user.
- *MS-Task 2*: one user (guide) sees a 4x4 grid with holes, while the other users (followers) only see the outline of the grid, but not the holes. The guide must direct the followers to place their object in the correct location on the grid, with the appropriate orientation and number. The task is repeated n times and at each iteration a new user takes the role of guide.

4.3.1 Implementation Details

The shared AR environment is supported by a peer-to-peer network architecture that manages communication and synchronization of environment state between connected devices. Positional tracking leverages markerless solutions available for smartphones and tablets through recent APIs for mobile AR (i.e. ARCore[1] and ARKit[2] libraries for Android and IOS devices, respectively). However, in markerless tracking, the position of each device is computed in a different local reference system (LRS). Therefore, the alignment of virtual content between different devices is not guaranteed.

Since current APIs do not provide functions to create a common reference system, we developed our own solution to align virtual contents between different AR devices. This solution uses a fiducial marker that defines a shared reference system (SRS). The size of the marker (A4) is large enough to ensure accurate and robust identification of the marker in the physical workspace. The facilitator places the marker in the middle of the workspace before handing out the devices to the participants. Then, the marker needs to be observed by each device i only once at the beginning of the experience to compute the transformation matrix M_i between the i-th local reference system and the SRS. To share a spatial information with other devices, device i can transform a point P from its LRS_i to the SRS according to Eq. 4.1. Applying the inverse operation (i.e., using M_i^{-1}), the coordinates of a point are transformed from the SRS to its LRS (Eq. 4.2). The latter transformation is used to represent spatial information received from peers in the LRS of the device.

$$P^{SRS} = M_i * P^{LRS_i} \tag{4.1}$$

$$P^{LRS_i} = M_i^{-1} * P^{SRS} \tag{4.2}$$

---

[1] https://developers.google.com/ar

[2] https://developer.apple.com/augmented-reality/

The shared AR experience has been implemented[3] with the Unity[4] game engine, which provides a "write once - deploy everywhere" model through its AR Foundation framework that virtualizes a variety of functions from the major AR APIs on the market. Network management was implemented via Photon[5], a peer-to-peer networking library designed and optimized specifically for the Unity development environment.

4.3.2 Experimental Protocol

To assess the effectiveness of the VCs under analysis (and of their combinations), we performed a user study involving 40 volunteers (11 females and 29 males, ranging from 18 to 34 years old). Volunteers were divided into 20 groups, each of which included two users and one researcher. The researcher served as the facilitator, introducing the scenarios, explaining the operations required for each task, and addressing any technical issues that arose during the session. The facilitator was instructed not to interfere with the users' choices and follow promptly their indications. Facilitators did not play the role of guide in the asymmetric-role task of each scenario. To minimize potential differences in interacting with volunteers, communicating information to them, and presenting sessions and tasks, the same researcher assumed the role of facilitator in all experimental sessions.

Methodology

After giving volunteers an overview of the experiment's goals, the scenarios were introduced in increasing order of complexity (i.e., Navigation, Gaze and Manipulation). Since users were likely unfamiliar with the VCs and scenario interactions, we included a simple training task at the start of each experiment to help them become acquainted with these elements.

As previously stated, each scenario was repeated twice, using either the V0 or V1 configuration. To reduce the possibility of learning effects biasing the results (especially the performance metrics), we pseudo-randomly shuffled the order of the executed configurations (i.e., V0-V1 vs V1-V0), ensuring a final balance (50%-50%) of volunteers' groups between the two options.

Evaluation Metrics

Figure 4.4 summarizes the evaluation protocol. The evaluation included qualitative data, objective measurements (completion times for each task), and subjective

---

[3] Source code and data available at: https://gitlab.com/turello.simone/collaborative-ar/

[4] https://unity.com/

[5] https://www.photonengine.com/pun

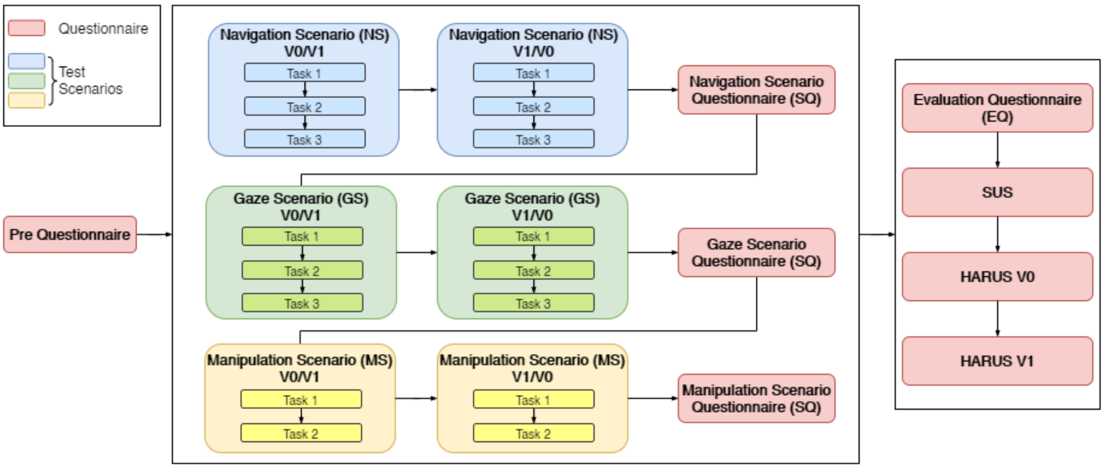


Figure 4.4: Summary of the experimental protocol. After filling a pre-questionnaire, the groups complete two iterations of a scenario (using a different visual configuration at each round) and then fill the SQ. At the end of the experimental tests, users complete the EQ and three usability questionnaires (SUS, and HARUS for both V0 and V1).

measurements. Concerning the latter, we asked volunteers to complete a preliminary questionnaire that included exploratory questions about their personal data (username, gender, and age) and prior experiences with AR technologies. We delivered a questionnaire (scenario questionnaire, SQ) immediately after completing the two iterations of each scenario to collect volunteers' opinions about using the proposed VCs in that specific context (i.e., we collected one such questionnaire per scenario). This questionnaire begins by asking users to express explicitly whether they prefer the V0 (without SPOI) or V1 (with SPOI) configuration. Then, we included three questions about the usefulness of the available visual information in raising awareness and promoting collaboration (*Q1 - For collaboration purposes, I have found the presence of visual information to be significant for the successful completion of tasks; Q2 - The visual information allowed me to better understand the actions of other users; Q3 - The availability of visual information improved the effectiveness of collaboration*). For each scenario, these three questions were combined into a multi-item *collaboration score.*

We also collected qualitative data from visual observations and user comments during the experimental scenarios.

We administered an evaluation questionnaire (EQ) at the end of the entire experimental session, which included two questions for each VC (*Q1 - The use of XX helped me better understand the actions of my teammates; Q2 - I found the XX useful during my collaboration with my teammates*, where XX is the VC under analysis). We then gathered these questions into a multi-item *effectiveness score* for each VC for subsequent analysis.

Finally, we administered two standard usability questionnaires to the subjects.

The first is the SUS [46], a standard de-facto in usability evaluations, which was used to obtain an overall usability assessment of the application. The second questionnaire is the Handheld Augmented Reality Usability Scale (HARUS) questionnaire proposed in [269], which assesses the comprehensibility and manipulability of AR-HHD applications. The HARUS aimed to assist us in gaining a thorough understanding of the comprehensibility of the two proposed visual configurations (V0 and V1), and it was thus administered to users once for each configuration. We emphasize the relevance of usability since, as highlighted in [238], we can only assume a certain amount of technical literacy on the average user, leading to lower adoption due to frustration if usability is not maximized.

Except for HARUS, which is scored on a scale of 1 to 7, and the question about the preferred visual configuration (which can assume only the value V0, V1 or none of the two), all questionnaire items were scored on a Likert Scale from 1 (strong disagreement) to 5 (strong agreement). SUS and HARUS final scores are computed on a $0 - 100$ scale.

4.3.3 Results and Discussion

This section will first discuss the results concerning the overall application usability and the comprehensibility of V0 and V1 visual configurations (Section 4.3.3). Then, we will assess the users' evaluation of the proposed VCs (Section 4.3.3). Finally, we will discuss the completion times recorded for each collaborative task using V0 and V1 configurations (Section 4.3.3).

In the following, unless otherwise stated, the results analyze the differences between V0 and V1 conditions and are reported as the average values per users since, performing a two-sample t-test, we found no statistically significant difference ($p < 0.05$) between groups V0-V1 (i.e., groups executing first the experimental scenario with the V0 configuration and then with V1) and the V1-V0 (i.e., those executing the two iterations beginning with V1 and concluding with V0). Nonetheless, there are cases where we discovered a significant difference between the two configuration orders. These cases are highlighted in the following section, which also discusses the results.

Usability and comprehensibility

The SUS scores were significantly different between the groups V1-V0 and V0-V1 ($p = 0.007$), but both were high, respectively, 83.38 ($SD = 8.52$) and 90.13 ($SD = 6.41$). According to [46], these values highlight the excellent usability of the application (and, consequently, of the proposed VCs). One possible explanation for the difference in SUS values between the two groups is that in the initial iterations, V1-V0 users had to simultaneously learn how to complete the tasks and use the SPOI in the specific scenario context. In contrast, V0-V1 users were first introduced

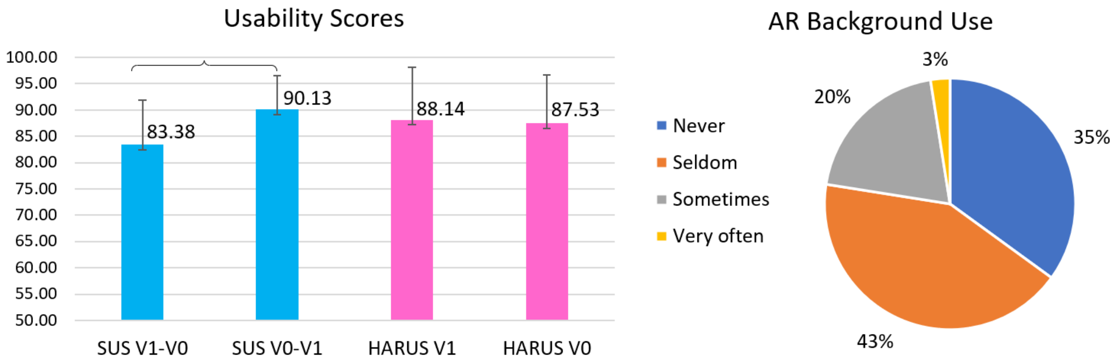


Figure 4.5: *Left:* SUS and HARUS average scores on a scale of 1-100. SUS is reported for the two configuration orders (V1-V0 and V0-V1), whereas HARUS for the two configurations (V0 and V1). Standard deviations are expressed through error bars and significantly different scores through curly brackets. *Right:* Chart showing the frequency of responses to the question: "how often do you use an AR application?".

to the scenario context and then to the SPOI, giving them enough time to become acquainted with the environment, the application interface, and the other VCs.

In terms of comprehensibility, the V1 visualization received a slightly higher HARUS score (88.14, $SD = 9.97$) than the V0 (87.53, $SD = 9.14$), and both can be considered highly positive. This result is intriguing because, on the one hand, it suggests the effectiveness of SPOI. On the other hand, it confirms previous findings [150], which state that combining a reasonable number of different VCs contributes to increasing awareness only if designers can maintain a clear and straightforward visualization (one of the main goals we tried to pursue in the design of our VCs). A summary of the usability scores is shown in Figure 4.5, left.

We conclude by emphasizing that the positive usability and comprehensibility evaluations are even more impressive in light of our testers' lack of AR experience, as 43% and 35% of them had respectively seldom or never used an AR application (Figure 4.5, right). Thus, we were able to design a collaborative AR system and provide a set of VCs that were easy to learn, easy to use, which made users feel in control and ultimately sparked their interest in using a similar system again in the future.

VCs evaluation

The *collaboration score*, i.e., the multi-item scale gathering the last three items of the SQ questionnaire delivered after each test scenario, summarizes the overall users' evaluation of the extent to which, for each experimental scenario, the VCs provided, contributed to foster collaboration. The final average collaboration scores, Figure 4.6 upper left, are high (Navigation: 4.56, $SD = 0.67$, Gaze: 4.68, $SD = 0.45$, Manipulation: 4.72, $SD = 0.35$), have relatively high internal

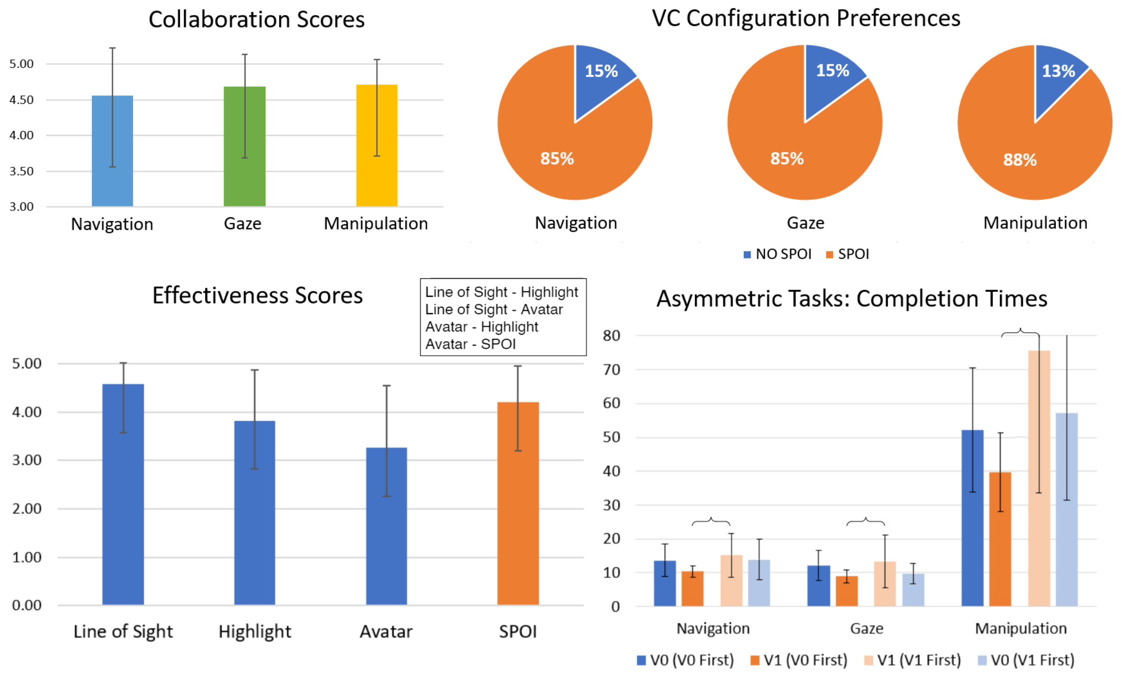


Figure 4.6: Collaboration scores (*Upper Left*) and VC configuration preferences (*Upper Right*) organized per evaluation scenario. *Lower Left:* VC Effectiveness as rated by users. On the top right box of the graph we report couples of VCs which show significant differences. Significance ($p < 0.05$) was evaluated through a Tukey's post-hoc test performed after a One-Way ANOVA test. *Lower Right:* Completion times (in seconds) for asymmetric tasks organized by visualization configuration and configuration order. Significantly different configurations are reported through curly brackets. In all bar charts standard deviations are expressed through error bars.

consistency (Cronbach's Alpha is 0.91 for Navigation, 0.65 for Gaze and 0.60 for Manipulation), and are not significantly different between themselves (differences were tested with One-Way ANOVA, significance at $p = 0.05$). The slightly higher values recorded in the Gaze and Manipulation scenarios can be attributed to the higher complexity of their collaborative tasks, which the users found to be positively supported by the available VCs. Our observations of user behavior provide another indication of how the VCs facilitated collaboration. During the experimental sessions, participants discussed a lot, often looking at each other's faces and turning their gaze away from the AR environment. While communication does not always lead to collaboration, the team cannot work together to achieve a goal without exchanging knowledge and ideas. Then we noticed that communication frequently relied on VCs for assistance (e.g., to indicate objects or places involved in the discussion). However, since the sessions were not videotaped, they could not be annotated in order to collect detailed metrics on the time involved in (and the quality of) this type of communication.

The user assessment of the VCs effectiveness (i.e., the *effectiveness score* gathering the two EQ questions for each VC) is summarized in Figure 4.6 lower left. It can be seen that the *Line of Sight* and the SPOI had high appreciation levels (respectively, 4.58, $SD = 0.45$, and 4.20, $SD = 0.77$). Since both VCs are used for the same function (i.e., pointing in an egocentric or exocentric fashion, respectively), we deemed it interesting to further analyze their different appreciation levels. We believe this difference can be attributed to *Line of Sight* ease and speed in pointing large objects (such as the cubes in the Gaze scenario) or indicating locations (as in the Navigation scenario) by simply orienting the device. Performing the same operation with SPOI, on the other hand, necessitates an additional operation (i.e., explicitly instantiate the SPOI). Nonetheless, we can get further insights from the analysis of the correlation between each VC's effectiveness and the collaboration score (averaged across the three scenarios), which is moderate to high for SPOI ($r = 0.65$, $p < 0.001$) and low for *Line of Sight* ($r = 0.35$, $p = 0.02$). One way to interpret these results is that although *Line of Sight* has higher perceived effectiveness, positive ratings of SPOI effectiveness make the most significant contribution to collaboration outcomes. A second result that points to the greater "real" effectiveness of SPOI (although its perceived effectiveness is lower than that of *Line of Sight*) is the higher percentage of users who prefer the V1 visual configuration to which SPOI belongs (85% in gaze and navigation tasks and 88% in manipulation tasks, as shown in Figure 4.6, top right).

Highlight was less appreciated than other VCs (3.83, $SD = 1.04$), which was an unexpected result given its potential utility in some scenarios (especially for Manipulation). A detailed analysis revealed a possible communication problem with this VC in this specific case. *Highlight* is used both in our implementation and in similar approaches [167, 237, 239, 84] to indicate to a user the objects that can be interacted with, and at the same time it is used to signal peers when the user manipulates a virtual object. However, this mixture of messages (the personal ones and the ones for peers) can confuse communication, despite the colour coding used to distinguish the two cases. This observation points to a design element that can be improved in future implementations. We think this is a helpful hint that other researchers can benefit from.

The *Avatar* received the lowest appreciation (3.26, $SD = 1.29$). This result confirms the lower relevance of this VC in a co-located environment, which is definitely better suited for scenarios that require remote peers to be digitally present in the shared working environment.

Completion times

Although not indicative of the quality of the collaboration and the extent of the increase in awareness mediated by the VCs, completion times are an objective measure of user performance that can shed light on experimental results and facilitate their interpretation.

In all test scenarios, the results show different trends between the tasks where users are provided with symmetric and asymmetric roles. For the tasks with symmetric roles, there was no difference in the completion times for the first and second iterations between the V0-V1 and V1-V0 groups, and there was also minimal difference in the average completion times for the two V1 and V0 configurations for these tasks. These results suggest that SPOIs, although rated positively by users, were not responsible for objective improvements in the completion of tasks with synchronous roles. On the contrary, for tasks with asymmetric roles, significant differences between the completion times of the first and second iteration between the V0-V1 or V1-V0 groups are evident.

We believe that one possible explanation for these observations is that in symmetric tasks, users need tools that support awareness but do not rely on VCs that support directional communication. On the contrary, in asymmetric tasks, the difference between guides and followers lies precisely in the information that the two roles possess, and in the numerous spatial and control information that they must necessarily exchange in order to accomplish the task. In this context, it is essential to provide the guide with tools that enable him or her to provide timely cues (and commands) to help the followers accomplish the task.

The average completion times of tasks with asymmetric roles per configuration order (V0-V1 and V1-V0) are summarized in Figure 4.6 lower right. These results deserve a detailed analysis. In the Gaze and Navigation scenarios, the difference between the completion times for the first iterations of the two configuration orders is not significant. On the contrary, the completion times of the second iterations are significantly different, and the iterations using V1 as the visual configuration consistently take shorter than those using V0. Since both visual configurations include pointing tools (the egocentric *Line of Sight* for V0, combined with the exocentric SPOI in V1), and pointing tools are probably the most relevant type of VCs in these specific contexts, these results seem to point to the contribution of SPOI in supporting communication, spatial information exchange, and awareness, ultimately leading to shorter completion times.

The results of the Manipulation scenario are quite different and more difficult to interpret. Figure 4.6 lower right shows that in the V0-V1 group, the first iteration (V0) has a significantly lower completion time than that of the V1-V0 group, and the second iteration (with V1 and the SPOI) has the lowest average completion time. On the contrary, starting with V1 seems to have a negative effect on completion times. In our experiments, we observed that users were much more challenged by trying to explain (guide) and adjust (the followers) the correct object orientation than by other activities. In the V0-V1 group, users learned effective ways to suggest the correct orientation in the first iteration (with V0 as the visual configuration). Most guides performed this process by pointing to a specific hole corner and telling the follower which corner of the object it corresponded to. In the second iteration, with V1, this solution makes it immediate to use the SPOI for pinpointing the hole landmark. In the V1-V0 group, the highest completion times derived from the trial-and-error process implemented. This process typically began with the guide placing the SPOI in the center of the hole to indicate followers the correct one, and then directing their activities using voice commands (e.g. "rotate a bit left" or "slightly move the object to the right"). However, these voice commands often increased the cognitive load on participants, who had to correctly interpret the relative position cues (with the possibility of misunderstanding them). We also noted a kind of "mental laziness" on the part of the guides, who relied too long on the placed SPOI without immediately understanding that canceling and replacing the SPOI to select a better reference for object positioning could improve communication and speed up activities. Moving to the second iteration (using V0 as the visual configuration), one is forced to rely on the *Line of Sight* as a pointer VC to select an appropriate landmark, which is inherently less accurate than the SPOI due to the instability of the visual reference (controlled by the user's movements), resulting in a slight loss of clarity in the information conveyed. In conclusion, we believe that these empirical observations not only illustrate the initial mental workload required to master the SPOI, but also support its effectiveness in this scenario.

4.4 VC Assessment: ARScape

We conclude this Chapter introducing *ARScape*, a collaborative AR-SG escape room for three players implemented to test our VCs' effectiveness in a real-scenario contextualized to the main topic of this thesis, AR-SGs. Escape Rooms, both digital and non-digital, are considered excellent activities for improving players' collaborative skills [218]. Thanks to this characteristic, Escape Rooms are often used as a medium to strengthen communication and teamwork skills [177], to support team-building activities [318] and to create engaging learning environments [308, 96, 102].

ARScape involves different puzzles that players can only solve by collaboration. Users have 60 minutes to escape from the fictional location and can benefit, in the task accomplishment, from all the passive and active VCs introduced before.

In the first puzzle, players must open a decorated box (Figure 4.7, leftmost). The opening mechanism requires the players to first all look together at the top of the box. This action reveals a series of cylinders that must be properly aligned and inserted into the corresponding holes in the box. The second puzzle revolves around a pedestal with three metal platforms around it (Figure 4.7, center left). Players must first reconstruct a copper pipe whose parts are scattered around the environment. The connected pipe activates the platforms that players must stand on to unlock the next puzzle. The third challenge requires players to take three flashlights and simultaneously illuminate an engraving (Figure 4.7, center right),

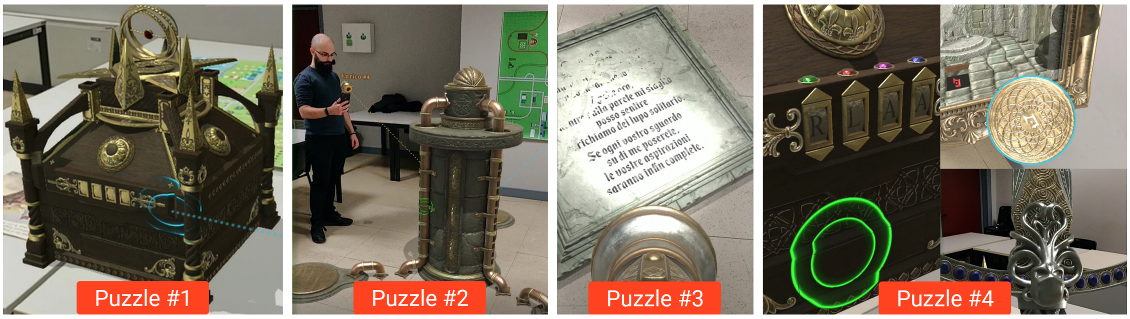


Figure 4.7: Screen captures from the four different *ARScape's* puzzles. From left to right: (*Puzzle #1*) the decorated box users must open, (*Puzzle #2*) the pedestal and the platforms users must position themselves on, (*Puzzle #3*) the flashlight used by players to reveal the hidden inscription and (*Puzzle #4*) different objects used to disclose the final code to be inserted in the box's mechanism.

which reveals inscriptions that help players solve the remaining part of the puzzle. The fourth and final puzzle contains a cipher that reveals a secret code. Players must enter this code into a mechanism in the box (Figure 4.7, rightmost). This puzzle is the most difficult, as the game mechanics (which are not fully disclosed for brevity) require players to simultaneously act on different elements (the cipher, the painting, and the unlocking mechanism) and actively communicate to synchronize their tasks.

It is easy to see in the gameplay of ARScape the main elements of our Navigation, Gaze, and Manipulation scenarios. Unlike the experimental scenarios, in ARScape different tasks are interwoven, often requiring the help of a combination of VCs, and players have no specific guidance on which VC to use to support a particular activity. However, we emphasize that our escape room tasks involve the same symmetric and asymmetric roles that players experienced in the experimental scenarios. Thus, the increased complexity of ARScape is balanced by the mastery of VCs and their use gained in the first phase. Furthermore, the ARScape puzzles are presented in ascending order of complexity (as in the previous phase).

ARScape was tested by all our volunteers who played the game after completing the experimental part and answering our questionnaires. In order not to try the volunteers' patience too much, given the length of the previous testing phase, we gave them a simple questionnaire at the end of the game that allowed us to collect some general information about the game, the gameplay and the usability, and that allowed the players to add free comments.

The results we received were definitely encouraging. The presence of VCs helped players complete tasks that required collaboration (4.47, $SD = 0.64$, on a 5-point Likert scale); the virtual elements were considered very realistic (4.48, $SD = 0.64$), and the game's difficulty was well balanced (3.74, $SD = 0.99$). Overall, players rated their gaming experience as very positive (9.12, $SD = 0.82$, on a scale of

1 to 10), which was reflected in their comments (*"Fancy and utterly brilliant!", "Perfect integration of augmented reality with the application context. Fun and easy to use", "It was the first time I tried a game in AR, and I feel very satisfied", "Congratulations, add some scenes! :)"*).

We underline that the primary rationale that led to the development of ARScape was to test the feasibility of VCs (also) in an AR-SGs context. Hence, we did not assess the effectiveness of ARScape as a team-building activity. Nevertheless, the choice of designing this AR-SG as an escape room stemmed from the growing literature in digital escape rooms, which present them as effective tools exploited in team-building interventions [308]. However, although we have no collected data (VA or GA) from the "live" game session, from our observations, we can highlight that players (effectively) communicated and collaborated as nearly all groups managed to reach the end of the escape room, which, we stress, could not be accomplished by players playing individually.

In conclusion, although preliminary, these results provide initial evidence of the proposed VCs' effectiveness in real and challenging use cases, helping improve users' awareness and support their communication in collaborative HHD-based AR scenarios, where co-located users are involved in both symmetric and asymmetric-role tasks. We underline again that such evidence was missing in the current literature.

Chapter 5

Assessing the Effectiveness of AR-SGs

In general terms, effectiveness can be defined as a product, task, or process's ability to generate its desired result [81]. In other words, any of these elements can be considered effective when it has achieved the purposes for which it was designed. If we focus this discussion on SGs, such a definition of effectiveness can take different forms. First, we need to consider that a SG combines two components, the educational and the entertainment ones. Therefore, in this specific case, a definition of effectiveness must necessarily consider these two aspects, both of which must be fulfilled to the best of expectations and must be harmoniously intertwined.

To better analyze this problem, let us start by underlining that a definition of effectiveness requires specifying the "what" to be measured and the "how" to measure it. Concerning the previous discussion, in a SG, we have to evaluate two different "whats", linked to its *entertainment* and *educational* dimensions (where the term educational is understood in a broad sense since we use it here to indicate the "serious" component of the game and is declined in specific ways according to the specific target the SG was designed for).

In terms of the entertainment-related component, a SG should provide a fun, engaging, and playful experience where players feel immersed, ultimately contributing to the envisaged educational goals. It is clear that, in the context of SGs, this is a fundamental goal that all SGs should aim for.

However, when it comes to the "serious" (*educational*) component of the game, the situation becomes more varied due to the breadth of domains in which SGs can be applied. To name a few, these domains range from Health and Fitness to Personal Social Learning and Ethics to Human Sciences and Cultural Heritage. In all these domains, assessments have perceptual, cognitive, behavioral, affective, and motivational impacts prompted by SG experiences. This observation implies the need to assess more or less tangible outcomes. Among the results that we can consider as "tangible", there is undoubtedly the acquisition of knowledge and understanding of educational content conveyed by SG. These results are those that are most frequently evaluated [36], both because of the greater diffusion of this type of approach and the (relatively) more straightforward evaluation of these outcomes in objective terms, e.g., by being able to compare them with reference values (or reference levels) defined by the designers of the educational path. Conversely, the application areas whose results can be considered more "intangible" (such as those related to changes in behavior and awareness of specific topics and issues and the acquisition of higher-order soft skills) are more challenging to approach.

Regardless of the characteristics of the outcome type (tangible vs. intangible), extensive research [60, 65] has provided experimental evidence on the effectiveness of SGs as a medium to convey *educational* contents. Given these initial promising results and the novelty of applying games for serious purposes, early SG research has attempted to prove the greater efficacy of SGs over traditional learning approaches. However, this research question has remained primarily unanswered or is currently being disputed. Despite that, many authors agree that, when comparable outcomes can be achieved, they way they are achieved becomes relevant. Therefore, opposed to conventional methods, game-based learning approaches provide an experiential and fun learning experience that can ultimately lead to improved knowledge and promote changes at an attitudinal, behavioral, and emotional level. In light of this observation, we wish to underline that we did not aim to prove SG (and AR-SGs in particular) as an alternative to traditional learning approaches. Instead, we wish to provide further evidence of SG's effectiveness, which indeed should not replace yet integrate established educational methods.

Given these general premises and remarking the variety of domains and outcome types, the evaluation of SGs can be a complex process, directly related to the characteristics and feasibility of the approaches (and tools) employed. This evaluation can generally be of two types: (i) *summative*, when outcomes are measured at the end of a given process, and (ii) *formative*, when assessed throughout the process, by continuous monitoring of learners' progress and failures.

In the case of formative assessment, the most widely used technique is the use of PPTs [24], in which the status of a given dimension or variable of interest (e.g., perception, knowledge, awareness, and attitude) is assessed at two different times (i.e., before and after the user's exposure to the SG experience). The two results obtained are then analyzed to look for significant differences in measurable items that can represent a tangible metric of the user's accomplishments, from which, in the second instance, indications of intervention effectiveness can be derived. However, pre-post experimental designs can have several disadvantages. For example, in pre-post questionnaires, it is impossible to determine whether the pre-test can influence the final results, and it can be challenging to understand which elements of the experience contributed most to the results obtained [49].

Unlike summative approaches, the formative ones aim to continuously evaluate users' performance, their activities, and how they affect the system's state and the metrics of interest. This type of analysis allows for collecting "just-in-time" data that can help better understand which elements of the game contribute to the analyzed variables and how this contribution is articulated. Again, however, several difficulties must be taken into account. For example, while being able to capture "live" data from users, the use of explicit forms of in-game evaluation (e.g., using questionnaires) is considered an impractical solution since it interrupts the game flow and the users' engagement with the game experience, destroying the sense of immersion in the scenario and narrative of the game and, thus, definitely compromising the delicate balance between fun and seriousness.

Therefore, to ensure a non-fragmented gaming session, the evaluation must be non-intrusive and, if possible, transparent for the user. A simple solution to achieve this goal is to refer to the GA practices used in the development of Digital Games. GA consists of continuously recording the player's in-game events and activities. These data can then be analyzed with different objectives. One of the primary uses is to allow the early identification of possible flaws in the design, mechanics, and aesthetics of the game [216]. For example, observing that players often die at a specific point in the game may indicate that a level or a quest is too difficult. Thus, developers and game designers can make informed decisions about refining the game design to ensure less frustrating gaming experiences that keep users in the flow state.

However, GAs can also be applied to SGs, since they can provide quantifiable player performance and engagement measures concerning the intended educational goal. For example, the number of interactions with different pieces of information (or repeated analyses of the same piece of information) may provide valuable indications to highlight active participation in the learning process, resulting in positive correlations with the learning outcomes achieved. However, how these objective indications can be derived is closely tied to the type of activity, the type of variables that can be measured, and their correlation to the learning progress. Thus, to correctly interpret GA and gain relevant insights is not a trivial task, as the link between game events (e.g., scores, interactions, time spent) and the object of the evaluation (e.g., knowledge acquisition, behavioral changes) may be blurred. Therefore, the analysis requires combining GA with other data sources (e.g., questionnaires, interviews, or game observations) to reveal hidden connections across data sources. This analysis often involves an iterative assessment process, in which the recorded analytics are fine-tuned to be as accurate as possible predictors of "serious" outcomes [280].

Once the GA have been validated as an accurate and objective means of measuring these outcomes, they can serve two key purposes. One aims at the learner (player), the other at the educator (or evaluator) overseeing the learning intervention. Firstly, from the user perspective, GA can be used to report learning achievements at the end of the game, during a debriefing session, or in real-time, throughout the game experience. In both cases, this stimulates users' motivations as they are confronted with tangible and interpretable indicators of improvement. Moreover, when used in real-time GA can add adaptivity to the educational scenario. Thus, content can be delivered in a way that is appropriate and compatible with the learner's needs, and even the difficulty levels of activities and tasks can be modulated according to the user's difficulty or ease of progress [25].

Finally, learning performances extracted from GA can also inform educators on users' learning accomplishments and improvements, ultimately serving as direct evaluation tools to assess the acquisition and mastering of the knowledge addressed in the SG. The availability of this information can result particularly relevant when more intangible skills must be assessed (e.g., communication abilities, teamwork competency, and soft skills in general). In these domains, the traditional evaluation processes can be highly cumbersome and time-consuming as it relies on repetitive observations and analysis of user's interactions. Therefore, assessing knowledge or skills acquisition simply through in-game outcomes would enormously simplify the evaluation process (i.e., evaluation material is readily available at the end of the learning experience). For example, the evidence that a team is able to overcome different challenges faced in a SG can become a direct method of assessing improvements in collaborative skills.

In addition, the literature emphasizes that special attention should be paid to the transferability of GA-based evaluation metrics since they suffer from low generalization properties. Thus, GA metrics developed in one context and applied to a different one may yield misleading results, emphasizing that it is necessary (in this potential transfer) to keep player characteristics, game mechanics, and gender as similar as possible. This, of course, severely limits the use of validated GA in contexts different from those they were designed and analyzed [7].

Now that the objectives and the methods for assessing the effectiveness of SGs have been introduced, in the following sections, we will briefly recall "how" we evaluated the AR–SGs described in Chapter 3 and discuss the results according to both the *entertainment* and *educational* components as well as their underlying interconnections. The general objective of this Chapter is to collect and jointly present the results collected from all the proposed AR-SGs to answer to RQ2, *Are AR-SGs effective?*.

5.1 Discussion

5.1.1 Ludic Effectiveness of the Developed AR-SGs

We begin our discussion of the ludic effectiveness by briefly taking up what was introduced in the previous chapters about evaluating the play experiences proposed in our papers.

The results useful to carry out this kind of analysis have been collected through questionnaires, proposed to users at the end of each experience. As a general comment, the results obtained were positive. In the context of procedural learning, most users stated that both Holo-BLSD and the application developed for firefighter training were fun experiences for learning the intended content, scoring on average, on a scale of 5, 4.31 for Holo-BLSD and 4.56 for firefighter training. As for soft skill learning, users liked Asteroid Escape (4.75), and had fun playing it (4.50). Similar results were obtained with ARScape, where users rated their game experience as highly positive (9.12 on a decimal scale) and reported very positive appreciations in their free comments at the end of the game session.

Regarding approaches related to complex system analysis, almost all Sustain players (96% of 99 testers) reported that they enjoyed playing the game and felt challenged by the game and engaged in its goals (85%). In addition, the majority of volunteers also felt that AR made the game more exciting (69%), allowing them to feel more immersed in the gaming experience (66%). Furthermore, in Asteroid Escape, users rated the experience as primarily playful rather than educational (on average, 3.92 on a scale of 5). This finding is important because it suggests that the underlying learning experience was transparent to users, who perceived the AR-SG primarily as a fun and engaging game.

This result should be compared and discussed with that obtained in Holo BLSD and firefighter training, where we acquired different indications. In fact, when users were asked to evaluate how they perceived both AR-SGs (on a scale of 5) as either a system suitable for training (a score of 1) or as a playful experience (a score of 5), they mostly favored the former option, with average scores of 2.69 and 2.29 for Holo-BLSD and firefighter training respectively. The motivation for these results is related to two main factors. The first is the design of the training path proposed by these two applications. In both cases, the learners faced a path composed of different phases (a learning phase, a rehearsal phase, and an evaluation phase), in which only the rehearsal phase included a ludic component. The predominance of the more "standard" training component in this sense influenced volunteers' evaluations. The second factor that may explain these results is the design of the ludic component, in which (unlike Asteroid Escape or ARScape) many of the elements typically found in games, such as a storyline, a narrative context, the presence of fictional characters, are missing. In fact, due to the peculiarity and requirements of procedure learning (see Section 3.1), the game design has been focused on the inclusion of game mechanics within the learning phase, thus introducing gamification elements rather than designing a complete game from scratch. However, it should be noted that even though users perceived the non-gaming component as predominant, they still stated that they preferred the proposed SG-based learning mode over the one implemented through a more standard approach (i.e., instructor and textbooks).

Concluding, we think that the results described above suggest that we were able to achieve the planned objective for the ludic components of the developed

AR-SGs.

5.1.2 Educational Effectiveness of the Developed AR-SGs

In our works, the "serious" component of each AR-SG was evaluated through both *formative* and *summative* approaches, encompassing subjective (questionnaires) and objective measures (GA and data extrapolated from video analysis) that were combined to gain better insights into the viability and meaning of the collected results. In the following sections, we outline the most significant results obtained from the experimental assessments of the AR-SGs described in Chapter 3.

Tangible Outcomes

In Holo-BLSD, we aimed at evaluating if the proposed solution could be a feasible instrument for training BLSD procedures. The main measure of educational effectiveness was obtained by comparing the final evaluation results of a control and an experimental group. The control group performed a traditional training with a human instructor and assessed by the same instructor, who filled an evaluation sheet while the trainee performed the BLSD procedure. On the contrary, the experimental group was entirely trained with the proposed AR-SG. As for the assessment, each group member completed the procedure through the AR simulation, with an instructor overseeing all the users' actions (available through a live stream of the users' point of view on the augmented environment) and compiling the same evaluation score used to evaluate the control group. The same form was also automatically compiled by the application, to evaluate the effectiveness of Holo-BLSD as a self-evaluation tool (i.e., by statistically comparing automatic and instructor evaluations of the trainees). In all cases, the final evaluation score was calculated on the basis on the correct completion of each step of the procedure. The results shows that the averages scores assigned by instructors to the two groups are very similar (39.48 for traditional training and 37.07 for AR-SG training, on a maximum score of 42) and that these two results are not statistically different. These findings suggest that the two learning achievements are comparable, indicating the effectiveness of the learning outcomes promoted by Holo-BLSD. In addition, users stated that the game (rehearsal phase) helped them better understand (and learn) the procedure. This indication suggests that the positive learning outcomes noted may actually have been fostered by the integration of gaming and training experiences, again emphasizing the synergies of the serious and playful components of an AR-SG.

With respect to these results we highlight what may be a possible limitation of this evaluation. In fact, as we mentioned, the instructors evaluated the procedures partly through direct observation of the learners and partly through observation of the video stream received from the HMD. However, in an AR context, this form of evaluation would certainly have benefited from the possibility of fully immersing the instructor within the AR world to allow him to observe the learner's actions in their full context, which includes both real and virtual elements. In this way, the instructor would have been able to move freely in the augmented environment and observe it from her/his perspective for more accurate evaluation. Unfortunately, we could not implement this option, mainly due to time constraints and lack of sufficient number of devices. Since we consider this a relevant feature for the assessment phase, future work will definitely include the development of multi-user solutions, possibly based on the integration of different devices (e.g., HMDs and mobile AR devices). We also emphasize that, as described in Section 2.3.3, the industry is also heading in the direction of providing interoperability and content sharing between different VR or AR devices by providing new software frameworks for developing shared AR experiences (e.g., the Microsoft Meshes platform), an option that will be considered in future developments.

Summative Evaluations

In the area of learning soft skills, the results we obtained are just as positive. In Asteroid Escape, users stated that playing the game improved their ability to communicate and work in groups, which were, indeed, the main goals of the educational component. In addition, the session repetition mechanic was probably the main driver of this outcome. We recall that a full game experience includes three runs (i.e., three attempts to escape from the imploding asteroid) in which players switch roles among the three available, and that at the end of each session, users were given a questionnaire to assess their personal observations about team cohesion and performance. These values allowed us to assess how repeated exposure to the game improved users' achievements and show, for all volunteer groups, an increasing trend (Figure 3.24). In addition, players reported feeling committed to the goals of the game that prompted team cooperation toward a positive outcome (i.e., victory). By playing multiple times, players displayed a sense of mastery over the controls and game mechanics, which instilled engagement and promoted motivation, both of which were beneficial (and responsible) for the effective learning outcomes noted. To this end, it should be noted that keeping the same role would have improved their game performance. However, players also acknowledged that switching roles in the repetition of game sessions helped them understand the difficulties that other teammates faced in previous sessions, thus promoting mutual understanding and empatic engagement.

Finally, another learning domain where we gathered positive outcomes was complex problem learning, where, through Sustain, we addressed the specific topic of sustainability. The performed assessment aimed at evaluating: awareness and commitment towards sustainability-related issues. From the pre-post questionnaire, we detected a significant change in the *sustainability awareness* scale. Despite the small effect size ($r = 0.22$), we think this is a promising finding, as players only played the game for an average of 32 minutes. We believe that these results are likely to be due to game design features specifically implemented in Sustain. Firstly, playing to develop a city required players to apply the abstract concepts related to sustainable development. As a result, they deepened their understanding of the connections between local actions and global consequences (i.e., a characteristic of complex systems). Second, by recurrently planning and observing the consequences of their decisions, players could test their evolving knowledge and progressively acquire a comprehensive understanding of factors affecting sustainable development. Finally, Sustain gameplay required players to continually shift from individual to collective processes and global to local actions. This probably helped players to explore and understand the multiple perspectives and levels of decision-making involved in generating sustainable solutions. Altogether, these features have likely facilitated a swift and holistic understanding of the complex interplay of factors involved in sustainability. Concerning commitment, results from the questionnaire *commitment to sustainability* scale show a significant increase in players' willingness to commit to the development of their community. We believe this change might have been stimulated by cognitive and affective processes occurring while playing Sustain. Unfortunately, the pre-post effect sizes is small ($r = 0.16$). However, previous literature has stressed that changes in viewpoints require time to occur [73, 138]. Thus, we suppose that by playing repetitive game sessions (in the experiment, players played only once), this outcome could be different. Furthermore, outcomes from the *elicitation of sustainability-thinking skills* questionnaire scale indicate that the vast majority of players (88%) perceived that the game spurred thinking processes and sensibilities described as pivotal for sustainable development. Overall, the evaluation performed for Sustain shows that SSGs like Sustain can foster personal change and understanding of complex systems by offering meaningful contexts whereby players are required to reflect and collaboratively find solutions to sustainability problems.

GA and Formative Evaluations

In terms of formative evaluation (which aims to continuously assess user activities and their effects on metrics of interest), we proposed using GA collection and analysis, which have been a relevant and valuable procedure in the evaluation of our AR-SGs. As described in the previous section, when combined (and examined) with data collected from other sources, GAs can provide rich insights into user performance and learning outcomes. However, this general statement needs to be analyzed in more detail in the context of the proposed educational activities. Since each of these has its peculiarities and prerogatives, the use of GA and its integration with other types of data can take different forms and explanatory capabilities.

Formative Evaluations in Procedure Learning

In Holo-BLSD (and in procedure learning in general, such as the firefighter training case), GAs were used to record all user actions and interactions within the augmented environment, thus providing an immediate piece of information vital to evaluate the users' accomplishments. Therefore, GAs allowed the application to check if all procedure steps were completed without mistakes, in the correct order, and on time. In this sense, these data represent an objective form of measurement of the educational outcomes and learners' performance once they have been validated. In Holo-BLSD such validation has been performed by comparing GAs-based evaluation with that performed by the human instructors, showing no statistical difference between them. Therefore, since GAs collected are accurate and correctly reflect users' performances (i.e., they are validated measures), they can also be used in future iterations of Holo-BLSD or similar applications as real-time indicators of user's performance and learning rate. This information can ultimately support adaptive and user-tailored learning paths, a feature that would greatly benefit learners, and at the same time, the overarching educational objective of the SG as students' learning outcomes, we believe, could be uniformly leveled.

Formative Evaluations for Intangible Outcomes

The interpretation of GA as measures of learning success becomes more challenging when the SG addresses more intangible skills and knowledge, like collaboration and sustainability awareness (i.e., Sustain) or soft-skills (i.e., Asteroid Escape). In these cases, our approach in understanding how the game actions reflect on the educational effectiveness was to couple GA with measures of players' behavior in the real world extracted from video recording analysis. For example, in Asteroid Escape, we compared GA with measurements of length and occurrences of different communication events between the three players, extracted from the video recordings. We found that the more users would communicate and coordinate their actions, the more effectively they would pilot the ship within the asteroid (i.e., with faster navigation times, a reduced number of collisions, and lower probabilities of getting lost in the intricate tunnel map). Also, team cohesion (i.e., the occurrence of events where players were encouraging each other) more frequently reflected cases of game success (i.e., players successfully escaping the imploding asteroid).

Although the coding scheme used to label users' behaviors while playing Sustain was slightly different, it led to similar conclusions. In this case, we labeled the length of verbal activity (without analyzing its contents) and occurrences of joint attention (i.e., when two or three players focused their view on the same tablet, shared screen, or a point in the physical environment). A posteriori, we considered the simultaneous presence of verbal and joint attention as tight-coupling, suggesting that players would be both visually and verbally engaged while sharing information. A first result was that these collaborative indicators would increase throughout the turns suggesting that the desire to perform well in the game enticed players' collaborative interactions. This finding is further supported by the positive association between the time spent working closely coupled and the individual in-game interactions (i.e., GA). We believe this might suggest that users were more likely to interact in the game due to joint discussion and decision-making. Then, by comparing the different collaborative indicators extracted from the video analysis process with the group scores quartiles, we identified that better-performing teams were characterized by significantly greater collaborative interactions carried out during the game. We believe this can be attributed to the feedback provided by the game at the end of each turn. As players understood that positive cooperative strategies could result in better outcomes, they were more likely to pursue or even improve their collaborative effort. Finally, these findings are also supported by subjective data collected during and at the end of the game. We found positive associations between the occurrences of collaborative behaviors measured in the videos with evaluations players performed assessing the quality of the collaboration. Altogether we believe these are promising results that indicate the generalizability of GAs as an assessment metric to (indirectly) estimate the players' acquisition of the envisaged intangible skills and knowledge.

5.2 Conclusions

In this chapter, we have summarized and jointly discussed the most relevant results from the assessments of the AR-SGs described in Chapter 3. We described the objectives and the tools adopted in the evaluation of the pursued outcomes to given evidence in support of RQ2, *Are AR-SGs effective?*

The evaluation of effectiveness in SGs can be generalized into two core objectives, the "serious" and the "ludic" outcomes and how they correlate to one another. Evaluating the "ludic" dimension means assessing if the final experience is fun and engaging, whereas assessing the "serious" one means establishing if the non-entertaining purposes for which the SG was designed has been achieved.

Concerning the ludic component. All the evaluations we performed led to similar conclusions as all the proposed AR-SGs were, despite the different levels of integrating gaming elements, fun, enjoyable, and overall appreciated by players. Our results suggest that the gaming experience featured in our AR-SGs made the overall learning experience highly engaging and motivating, a result more hardly achievable when delivering the same learning experience through an AR-only application. Finally, we believe that if adequately contextualized, all these are promising results suggesting that AR-SGs can deliver playful activities with a "serious" impact.

Through Holo-BLSD, we were able to show positive "serious" outcomes in the procedure learning domain since learners who underwent training with the AR-SGs had (i) positive results and (ii) these results were comparable to students who performed traditional training (e.g., performed by a human instructor). These results stress the already proven effectiveness of AR-learning approaches highlighting the contribution of gaming elements, which resulted in perceived better outcomes.

In both Asteroid Escape and Sustain, we strive to evaluate their capability to foster users' collaborative interactions by (i) analyzing the contents of verbal information users shared, the length and frequency of their communications, and other non-verbal collaborative cues, and (ii) the relationship between these interactions (happening outside of the game) and other objective data collected from GAs. The collected evidence confirms that both AR-SGs were effective in their envisioned purpose.

Through Sustain, we also explored if the AR-SGs could promote awareness and commitment towards sustainability. Between the pre and the post questionnaires, we detected significant differences on both variables, which we believe were most likely elicited by the game and its design features. Since collaboration is a recognized facilitator for complex problem learning[154], we believe the positive effects on collaboration described above might also have contributed to the effectiveness of Sustain in addressing the awareness and commitment outcomes.

Based on these premises, our results seem to suggest the following answers to *RQ2 (Are AR-SGs effective?)*. First, AR-SGs can be effective in both their "ludic" and "serious" objectives. Second, these two aspects are tightly connected and overall lead to more effective results. Third, AR-SGs are able to produce both "tangible" and "intangible" outcomes.

We wish to conclude this Chapter by highlighting that the assessment process leading to identify the effectiveness of the proposed approaches, results from a mixed-method approach exploiting the collection and combination of data from numerous sources. Due to the complexity of such assessment, the research community in AR-based learning, SGs, and the more recent AR-SGs domains are demanding more robust and sound assessment methodologies [65, 25, 49]. We believe the evaluation methodologies applied in the experimental assessments described in this thesis could contribute to address these needs. Furthermore, we believe the adopted methodologies can be (at least in part) generalized to other domains, providing a useful reference for researchers and practitioners in designing the assessment protocol of novel AR-SGs applications in a variety of contexts.

Chapter 6

Usability

In the previous chapters, we have discussed several applications of AR-based SGs in various domains and with different purposes, ranging from training, education, and support for collaborative activities. In all these areas, AR has disruptive potentials. Nevertheless, at the same time, it requires taking into account some critical factors that influence the success rate of the use of these technologies (such as, for example, the age of the users, their technical experience, and their level of comfort with advanced technologies).

One of the main elements that need to be considered is that AR technologies are pretty new, especially in their daily use by non-tech-savvy users. The consequence of this observation is that the technology risks becoming a barrier that strongly limits users' ability to enjoy the content available in the application and the benefits of using AR-based gaming approaches in different contexts. In this field, the main problem is to ensure that the developed products (and the hardware tools they leverage) are of immediate use for end-users, up to the point that their use becomes absolutely natural (and therefore almost transparent). If this is not the case, or if the learning curve of tools is too steep, there is a substantial risk that users will use most of their cognitive efforts trying to use the tools instead of focusing (as they should) on the contents they convey.

Therefore, in the development of SG in both AR and VR, it becomes necessary to devote adequate time to the study, design, and evaluation of usability and user experience with the application. This consideration is even more critical for all those systems and applications that must adapt to a wide range of uses and an even more comprehensive range of potential users. In addition, in most of our studies, one must also consider the fact that users are first exposed to the application to be used directly in the assessment phase. It is therefore necessary that the design of the application aims at maximizing the most relevant quality elements of the interface, i.e., learnability (how easy users can accomplish tasks the first time they use the application), efficiency (how quickly they perform tasks through the interface), and memorability (how easy is for users to remember how to operate the interface proficiently).

Therefore, it is required that the technology is user-friendly and maximizes the overall user experience with the system since these effects have significant repercussions in terms of the system's effectiveness. Moreover, several studies [205, 305, 208, 158] show that usability and UX play a relevant role in transferring knowledge from the virtual to the real world. This process is complex and involves several elements, such as instructional methods, pedagogical approaches, affective and cognitive factors (e.g., interest, intrinsic motivations, cognitive load). However, even if improving knowledge transfer and learning outcomes cannot be reduced to a mere improvement of usability and UX, on the other hand, benefits can indeed be obtained by maximizing these two elements. A low level of usability can affect how users operate with the application and, consequently, absorb the educational contents the application is supposed to convey. If the application is not usable, learners are forced to spend their time and mental effort trying to (continuously) understand how to operate the application. As a result, users are likely to become frustrated from using the application and certainly not enticed to repeat the experience in the future, which affects their ability to learn the proposed contents negatively.

Besides identifying (and preventing) possible difficulties in the use of technology, usability (and UX) studies allow to reduce misunderstandings in the exchange of information between the application and the user and even risks and dangers related to the use of technology. Consider, for example, the case of VR, where the user is completely immersed within a virtual world. In this case, there are potential health risks since the user, when moving within the physical environment where the simulation is performed, can collide with objects in the environment, stumble and get hurt.

With these observations in mind, in the projects we have developed and presented in this thesis, attention to the ultimate usability of the application (or system) and the UX has always been one of the cornerstones of our design and development process. This chapter aims at discussing in detail how usability and UX issues have been handled in our projects, highlighting (in different cases) the correlations between usability and the final benefits (reported by users, or calculated through quantitative measures) of using the system. More formally, this Chapter presents the evidence required to answer *RQ3: Usability can be an issue in AR-SGs, can it be softened?*

To make the discussion of Section 6.2 self-contained, we start briefly summarizing, for the works presented, how usability and UX have been treated during design, prototyping, and assessment.

6.1 Designing for Usability

6.1.1 Holo-BLSD

In Holo-BLSD (Section 3.1.1), one of the main hurdles to address in order to ensure the final usability of the application and an optimal UX was to allow users an immediate and as smooth as possible use of a hardware tool (an AR headset) that was probably new to them and requires a certain learning curve to be able to operate proficiently. We followed three main guidelines when designing the UI to acheive these goals.

First, we aimed to reduce the users' cognitive load by providing them with a limited set of interaction modalities (i.e., object selection, dragging and voice interaction). The choice of these modalities is again the result of a series of preliminary usability tests on prototype versions of the application, in which we explored with volunteers various alternative solutions and a more significant number of interactions. This analysis allowed us to discard some metaphors that were not immediately understandable or easy to use, given the device's limitations. For example, one of the interactions analyzed for moving objects in the scene was the pinch gesture (to select the object) associated with hand movement tracking (to control the movement). The experimental results showed that this metaphor, while being intuitive, has an extremely low level of usability due to the low accuracy of the tracking data offered by Hololens. In contrast, controlling the movement of objects with the movement of the head is much easier (and more accurate).

The second guideline followed was, besides reducing the cognitive load required to use it, fostering the learnability of the interface by making it as intuitive as possible. As third and last guideline, we carefully designed and included prompt and clear feedback aimed at (i) informing users that an interaction with what is on screen can (and should) be started, (ii) allowing them to predict the result of this interaction, and (iii) notifying them that they succeeded/failed in performing a task. For example, when securing the scene, we noticed that sometimes users did not notice some dangerous object to be removed that, remaining in the scene, prevents the completion of the activity. Therefore, in the learning phase, we included audible alerts informing users of the presence of these objects, which are triggered if no significant activities are recorded for a while. If the user is still unable to identify the objects to be removed, subsequent alerts provide (audio) spatial cues about its location and eventually help the user find it with additional visual cues.

Following the aforementioned design guidelines is necessary but not yet sufficient to allow users to operate effectively and smoothly with the tool right away. For this reason, we have included a training section in the tool specifically dedicated to learning how to use the interactions offered by the application and become familiar with the feedback offered by the system. In this training session, for each interaction type, users first receive detailed instructions on the actual interaction method they are going to experience, and then they can get acquainted with the proposed interaction metaphors and tools by practicing them into a scenario mimicking the real ones. Then, we also associated an individual icon to each interaction used to remind users during the learning phase of the types of interaction required to complete an action.

6.1.2 Sustain

In Sustain (Section 3.2), the collaborative process involving all users bases its foundation on information about the game's current state, the value of the variables involved, the characteristics of the actions available for each role, and how they affect the system variables. In fact, in order to discuss a collective plan, players need to have a clear picture of the general state of the city to understand which elements to prioritize during development. Furthermore, in the execution phase, they need to analyze all the actions made available by their role, and be able to show them to other players and discuss them.

Given these needs, Sustain's UI design phase focused on best supporting the clear and direct sharing of this information and ensuring at the same time that the communicated data is readily understandable for players. This phase has been embodied in an iterative development and prototype analysis process to identify any errors and design flaws as soon as possible. The result is an interface where we tried to make immediately perceivable the presence of the necessary information, using visual representations that were as clear and simple as possible, including indications that would make immediate accessing the information of interest in a consistent way.

6.1.3 Asteroid Escape

The design phase of Asteroid Escape (Section 3.3.3) was probably the one that required the most effort to achieve acceptable levels of usability. The main problem was related to the utterly unconventional setup of the application, in which the three players use different display and interaction systems.

To ensure a high level of immersion, the two crew members are positioned in front of a large interactive screen that depicts the simulated environment in a realistic scale (1:1). Since the screen is made interactive (using an infrared camera and computer vision algorithms for touch detection), the first solution adopted to manage the interaction was to add buttons and interactive elements that allowed users to control the spaceship directly from the touch interface. Unfortunately, the evaluation of this prototype immediately revealed that this type of interface was essentially unusable, especially in the management of operations that required fine control(e.g., navigating the spaceship in the asteroid tunnels). The problem was that the proximity required to operate the touch interface, coupled with the large size of the screen (100 inches), meant that users had a distorted view of the environment outside the spaceship that, in many cases, caused motion sickness. Thus, in the second iteration, the spaceship controls were moved to hand-held mobile devices. This solution has two advantages. First, users can stand at about 2 meters from the screen, ensuring a correct (and not annoying) view of the environment. Second, the hand-held controls guarantee a simpler and more robust interaction than the touch screen system, which often suffered from low accuracy in gesture recognition.

As for the Navigator (the third player, which in the fictional world of the application is in a remote space control station and is actually in a physically separated room), the first hypothesis was to provide her/him with an AR-enabled tablet to see the asteroid map. However, this solution was discarded in the initial tests for two reasons. First, holding a tablet for the entire experience resulted in excessive physical exertion that is not sustainable for the average user. Second, a tablet provides a low level of immersion in the virtual environment. To enhance the sense of immersion, the use an immersive VR HMD was evaluated, but this would have resulted in safety concerns and limitations on the user's movement in physical space to view the map from any possible direction. The user typically observes a central point (the map) and often moves by walking backward to get away from the center. This movement, however, makes the guardian systems used to alert the user that they are moving away from the safety zone highly inefficient. In conclusion, an AR HMD (the Hololens) was deemed the optimal choice in our context (as confirmed by early prototype evaluations). This device provides a natural interaction with virtual content, promotes spatial presence, and provides a stereoscopic view of content, and all of these elements contribute to an increased sense of user immersion and presence.

Similar to what was done in Holo-BLSD, Asteroid Escape includes an interactive tutorial phase to exhaustively describe all the available interactions and audio-visual feedback offered to players. This tutorial is repeated at each run since players swap roles between runs, and each role has its own and unique set of controls. Since Asteroid Escape has been envisioned as a SG to strengthen different communication abilities, the tutorial itself is (implicitly) part of the team-building activity. Indeed, textual instructions are presented only to the Navigator (i.e., the player wearing the Hololens), who has to describe other players the displayed information and test out the described functionalities. At the same time, crew members (physically located in a different room) must describe what appears on the projected display. By engaging in this information sharing, players can learn how to use the application and practice their communication abilities in a less-frenetic environment (where there is no time limit, as opposed to the actual game).

6.1.4 ARScape

In ARScape, the VCs used are intended to improve user awareness in the shared space. To achieve this goal, they must be simple to use, must be easily understandable, and at the same time must minimize the visual clutter caused by their introduction. This last aspect is of paramount relevance since several studies (e.g., [150]) point out that an excess of visual information related to VCs is detrimental in terms of awareness, decreasing it instead of increasing it.

The design phase, therefore, aimed, on the first hand, to try to understand which VCs were most suitable (in terms of communicative efficacy) to be included among the tools made available to users. To this end, we initially proposed a series of VCs proposed in the literature, and then we analyzed them with a series of rapid effectiveness and usability studies. This phase led to skimming several VCs used in various works in the literature. To give an example, one of the VCs we discarded is the *frustum* [240]. This VC uses a visual frustum that originates from the device to represent the portion of space framed by the user through his device. The VC is typically represented by a truncated pyramid, displayed with marked transparency to make its presence as unobtrusive as possible (Figure 6.1). In the test phase, this VC was discarded since it does not provide precise information about the object currently observed but only gives an approximate idea of where he is looking at that moment. Another result of this preliminary phase was realizing the need to add a VC that would allow a (more appropriate, in many use cases) exocentric pointing interface, hence the idea to create the SPOI taking as refusal similar VCs used in remote sharing scenarios.

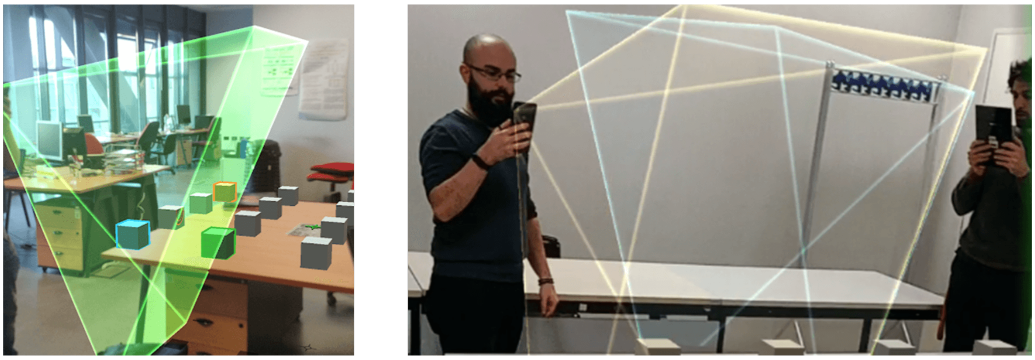


Figure 6.1: The (discarded) frustum VC.

The second step in this prototyping phase was related to analyzing several alternative forms of visual representation of VCs. For the SPOI, we also analyzed different ways to quickly provide users with information regarding the presence and location of SPOIs inserted by other users when and not in the FOV of the user's device. Again, the introduction of these cues is a result of a preliminary analysis of our prototypes.

We stress that, despite our efforts, not all the implemented choices were proven successful. As we have already reported in Section 4.3.3, the way we handled the Highlight (i.e., using the same VC to both indicate the interactable objects to a user and signal to the peers when the user is manipulating a virtual object) was confusing, as shown by the subjective results collected at the end of the experimental phase.

6.2 Discussion

The first general observation we can make is the link between the usability levels of the various applications developed and their effectiveness in terms of educational outcomes. Although, as we mentioned earlier, maximizing usability does not necessarily have positive effects on achieving the final educational goals, it is also true (as we pointed out) that a low level of usability and UX can have adverse effects.

One of the most striking examples of the link between learning effectiveness and (low) usability is probably the firefighter training system described in Section 3.1.7. Indeed, one of the features of this work was the analysis of different interaction interfaces for locomotion management in the VR space. These were a desktop interface based on mouse and keyboard (DVR), one using a teleporting metaphor with an immersive HMD and its controllers (IR), and the third leveraged a treadmill (KAT). The KAT's usability was evaluated as decidedly low, scoring a SUS value barely sufficient (68.3) and decidedly lower than the other two, which obtained optimal values (DVR 85.90, IR 81.7). These values negatively affected the learning process results, measured through the number of tasks completed correctly and timely, where KAT users had significantly worse performances than those who used the other interfaces (83% correct actions versus 100% for the other two groups).

While the Firefighter training program highlights the (negative) effect of usability on expected educational outcomes, in other case studies the positive educational outcomes (extensively discussed in the Chapter 5) are in all cases accompanied by positive results in terms of usability.

In Holo-BLSD, the good SUS score (72.03) was paired with similar values for most of the Nielsen's attributes (learnability 3.63, efficiency 3.93, memorability 3.89, and satisfaction 4.44). It is interesting to note that only a small number of volunteers had previous (limited) experiences with VR (23.07%) and AR (23.07%), and almost none with the AR HMD used (only one user had a previous experience with a wearable AR device). These numbers place greater emphasis on the usability results. Indeed, they suggest that we managed to design and develop a system that allowed users to enjoy educational content (and to learn the conveyed information effectively, as shown by the experimental results of Section 3.1.6) through technologies that were entirely new for them and with minimal time available to adapt and learn to use the system effectively.

It is interesting to highlight the similar results obtained with Asteroid Escape, where, despite the low experience users had with some of the technologies involved (e.g., 75% of volunteers never or rarely used AR-related systems), the SUS scores highlighted excellent usability (76.67 for the Navigator and 74.37 for the Crew roles). Similar results were obtained for the Nielsen's usability scores, which were high for both interfaces, exception made for the possibility to recover from errors (mainly as a result of the Crew frustration when experiencing a chain of crashes of the spaceship against the asteroid tunnels).

ARSCape is another example where users largely inexperienced with AR (75.5% had rarely or never used AR reported excellent usability (in terms of SUS and HARUS scores) and the effectiveness of the proposed VCs in supporting the collaborative tasks.

We firmly believe that one of the relevant elements contributing to these results was the initial training phase we included in Holo-BLSD and Asteroid Escape on using the tools and the interactions proposed in the application. To some extent, the same approach was applied in ARScape and Sustain. In ARScape, before using the proposed VCs to succeed in discovering clues and solving the escape room puzzles, they were analyzed in ad hoc test scenarios preceded by short training sessions. Thus, even if implicitly, users were still introduced to VCs' use incrementally with respect to their complexity. Sustain includes an initial training stage, managed by the facilitator, to explain the game's objectives and mechanics to the players and familiarize them with the UI (through a tutorial conceived as a series of step-by-step instructions on using the game interface).

With this observation, we want to emphasize the importance of providing informational (and possibly interactive) material at the beginning of the experience to allow users to become familiar with (and, ideally, master the use of) the interface. Above all, we stress that this guideline, which should be followed in the development of any SG, becomes, in our view, nearly mandatory with SGs based on technologies unfamiliar or unknown to users. This latter case was indeed the one we faced in Holo-BLSD and Asteroid Escape, which both leveraged the Hololens AR HMD, a device that, at the time, was a novelty for the whole research community.

As another relevant element contributing to achieving the expected SG outcomes, we suggest the carefully designed UX, which ultimately provided a meaningful, pleasant, and relevant experience to users in our works.

In Holo-BLSD, learners reported that immersion in the virtual experience was important for completing the assigned tasks (Section 3.1.6). Subjective evaluation results indicate that the proposed AR-based learning path can stimulate learners' attention at levels similar to those achieved with traditional training. Furthermore, the virtual environment also facilitates learners' comprehension of the activities to be performed compared to the verbal description made by tutors of the standard courses (where many of the elements involved in the procedures are only described, opposite to Holo-BLSD, where they are present in the environment).

In Asteroid Escape, we strived to create a highly immersive application to ensure users would engage with the game's narrative, objectives and, in turn, actively communicate and collaborate, thus strengthening team-work related abilities (i.e., the primary goal of our SG). To achieve this goal, we created a fairly complex setup that was rated as positive by users. As for the UX, our results suggests that the high level of engagement and commitment towards the game objectives (Table 3.15) was conducive to improve their communication skills across the multiple runs of the game (as shown in Fig. 3.24). Although the limited number of volunteers does not allow us to provide solid statistical evidence, we assume that these positive outcomes were also solicited by the sense of immersion and presence users felt during the game.

Regarding ARScape, even if not enough experimental results have been collected to evaluate the quality of the UX, the free comments received from users have been very positive. Our volunteers emphasized the realism of the environments and virtual objects used to solve the proposed puzzles, the level of complexity of the game (comparable to their other experiences with a "real" escape rooms), and their enjoyment in being able to solve an escape room in AR.

In addition to these positive elements, our work allowed us to highlight critical elements that need to be improved. In the context of procedural learning, our users identified as a negative element the rigidity of the application, which by requiring to perform operations in a precise way did not allow users to operate in alternative forms more familiar to them. As a result, users had to adapt their behavior to the system rather than the other way around. In addition, users often complained about the inability to remedy trivial interaction errors, with negative consequences on completing operations. In Asteroid Escape, one of the most negative elements indicated by users was the difficulty of navigating after a crash in particularly complex situations. This issue ultimately required the Crew to solve technical problems of little importance and interest for the overall goal (since Crew communication had to focus on figuring out how to use the commands to get out of the generated impasse, thus it did not involve the Navigator and was not related in any way with the tactical and strategic elements of the game). In ARScape, the experiments allowed us to identify a flaw in the design of the Highlight VC that was used for two different types of communication (one to inform the user of the interactive objects around him, and the other to inform his peers of the manipulations made by the user on a virtual object) creating problems in promptly interpreting the conveyed information.

These indications emphasize the need to follow an iterative design and implementation methodology in which the working prototype is continuously analyzed in terms of usability, UX, and clarity of information and feedback communicated to users. These steps are necessary to identify limitations and errors in the current design as early as possible. However, it must be stressed that despite the care and attention that can be put into these design phases, inevitably, the final version of the application is not always flawless (as demonstrated in our work). One of the main limitations that we have identified in this process is the limited number of users involved in the preliminary usability tests, which does not allow collecting exhaustive information to promptly identify all possible problems. However, we must emphasize the cost in terms of time and human material to be involved in such tests. This problem becomes an issue for many projects where the time and budget available for development are limited. In all of these cases, it is even more critical to include usability and UX good practices as guiding goals throughout the process and from the earliest design stages.

6.3 Conclusions

The main takeaway that can be drawn from our results is that usability matters. From a certain point of view, this may seem a trivial observation, but for all intents and purposes, it is not, especially considering that often, in the literature, this seems to be an element that in many cases did not receive the proper attention. Many works focus on the design of the training path and the instructional scaffolding, which are the main elements that contribute to the effectiveness of the training intervention. Nevertheless, in serious games, usability and UX (and player experience) play an essential role. In particular, since AR technologies are not yet so widespread, usability (or rather the lack thereof) can become a barrier for users, with negative consequences in terms of expected results.

These limitations, reported in the literature [246, 325], have led to the formulation of the following RQ, *RQ3: Usability can be an issue in AR-SGs, can it be softened?*

We believe that by describing the solutions we eventually conceived to solve the usability issues faced during design and development, we can affirm that these issues can indeed be softened through careful, user-centered design. From this point of view, we believe that it is essential to put users in the condition to operate proficiently in the shortest time possible to focus on the activities they have to perform in the game and on the content provided by the application. From our experiments, implementing training sessions dedicated to learning how to use the tools seems to be the most suitable solution to achieve this goal. Another recommendation we feel we can make is to validate design choices as soon as possible with a panel of volunteers and through a mix of measures from both subjective and objective sources.

Chapter 7

Conclusions

Interest in AR for educational purposes is rapidly growing thanks to three main factors: recent technological advancements, the availability of simple development tools, and AR's unique visual and interactive characteristics. In parallel, research exploiting SGs for learning purposes is now paying attention to this novel technology. The many similarities between these two learning tools/methods generated a growing interest into approaching a novel line of research advocated to explore the possibilities and benefits of combing AR and SGs. Researchers believe that AR-SGs can be more captivating, more immersive, better at supporting collaboration, and overall more engaging compared to the use of AR and SG as separate entities. Despite these promising outcomes, we are still in the early days of AR-SG research and several open problems have not been addressed yet.

In this thesis, we pursued the objective of gaining insights on how AR and SG could effectively be combined, addressing educational fields that had been given little attention in previous AR-SG literature. From a methodological standpoint, this research was conducted through the design, development, and evaluation of four AR-SGs (Holo-BLSD, Sustain, Asteroid Escape, and ARScape), which served as demonstrators to address the main problems identified in the literature.

7.1 Research Questions and Contributions

The research described in this thesis was guided by the following research questions:

- *RQ1: Can we establish synergies between AR and SGs in order to develop effective AR-SGs in a variety of learning contexts?*
- *RQ2: Are AR-SGs effective?*
- *RQ3: Usability can be an issue in AR-SGs, can it be softened?*

In the following, we will summarize what are the collected findings that allowed us to answer these RQs, highlighting the specific contributions made by this thesis. For a more detailed discussion of each RQ, the interested reader may refer to the concluding sections of the previous Chapters, namely Section 3.4 (RQ1), Section 5.2 (RQ2), and Section 6.3 (RQ3).

To address RQ1, we identified learning domains and specific topics where AR or SGs had been successfully adopted but mostly individually, as examples combining their strengths were rare or nil. In particular, we selected the following three learning domains (and topics): (i) *procedure learning* (in particular, Basic Life Support and Defibrillation), (ii) *learning about complex systems* (in particular, sustainability), and (iii) *soft-skills learning* (in particular, team-related skills). Therefore, to evaluate possible synergies between AR and SGs (i.e., the creation of an effective "relationship" between these two approaches/tools for learning), we designed and developed four AR-SGs addressing these learning domains/topics (where 2 of them were targeting "soft-skill" learning).

To summarize, the main contributions of this thesis in answering RQ1 are:

- the collection of practical evidence showing that positive synergies can be established between AR and SGs within learning domains which, up to now, were relatively unexplored with these approaches. The proposed AR-SGs approaches enhanced the following features and outcomes:
    - immersion and engagement with the learning experience
    - motivation to play, motivation to learn
    - deeper understanding of the (complex) problem/concept addressed
    - collaboration and communication in multiplayer contexts
    - mutual awareness in co-located collaborative applications
- the proposal of a software framework for the design of procedural learning applications. The framework models procedure activities as nodes organized into a graph, whose edges define dependency requirements between activities. The framework models the learning path into three phases: (i) learning, which provides all instructional material to complete the procedure, (ii) rehearsal, where the learner can practice the procedure through a gamified approach, receiving feedback on the correct/incorrectness of the performed tasks, and (iii) evaluation, where the system tracks users' actions (giving no feedback) and outputs at the end an evaluation form. This framework enables the rapid development of applications addressing procedure learning and is technology and domain agnostics;

- the proposal of game design features and mechanics to develop AR-SGs addressing complex systems learning. Through Sustain, we defined nine high-level game features inspired by sound theoretical frameworks adapted to encompass an AR-enhanced collaborative environment. Despite Sustain specifically targets sustainability awareness, the exploited frameworks and AR-SG approaches are general (and thus transferrable to other learning scenarios). Therefore, by employing or adapting our game features, we expect that other contexts can experience the same benefits found in Sustain.

To answer RQ2, we conducted extensive experimental assessments by exploiting a broad set of data collected from numerous sources encompassing: (i) subjective data from questionnaires (ii) users' behavior observational data (collected and coded from the video recording of play sessions), and (iii) objective measurements collected from the in-game interactions. Data from these sources were first analyzed individually, and then relations among them were searched. This approach provided us with more robust results addressing both the ludic and educational purpose and the relationship between these two elements. According to our results, our contributions in answering RQ2 can be summarized as follows:

- the collection of evidence that our AR-SGs were effective in their ludic purpose. More surprisingly, or we could say less obvious, evidence suggests that the positive ludic outcome influenced the learning one. This relationship mostly stemmed from motivation and commitment to the game outcomes, immersion in the virtual environment, and a general feeling of enjoyment, leading to higher engagement with the learning content. While similar results can be found in the general literature about AR, SGs and AR-SGs [246, 188], we stress that no such results were yet available in the tackled domains, which were mostly unexplored with AR-SGs;
- the assessment of a wide range of learning outcomes, which featured both tangible and intangible concepts/skills. These more intangible skills, primarily associated with the broader category of 21-st century skills, are still being explored in SG-related research and have hardly been explored in AR-SGs. The (positive) outcomes we could detect were:
    - effective support to learning and training on complex procedures
    - adequate comprehension of complex system issues and positive effects on intangible learning outcomes (such as on sustainability awareness and commitment)
    - improvement of different skills associated with collaboration (i.e., communication, negotiation and problem-solving skills)

- the proposal of analysis and assessment methodologies that combine in-game data with data collected from player's behavior observations. Earlier, we mentioned that the game design features proposed could be transferred to other games. Similarly, we believe the same logic applies to the assessment methodologies adopted for analyzing the effectiveness of our AR-SGs. Hence, we believe future practitioners could exploit similar approaches to data collection and analysis in assessing the effectiveness of AR-SGs that present similarities with ours or take as reference our game design features.

RQ3 originated from evidence in the literature reporting compromised learning outcomes caused by inadequate usability levels or poorly designed UX in AR learning applications [246]. We approached this problem developing our AR-SGs through a careful process that placed users' needs at the cornerstone of our design choices. The outcomes of this process were tested through usability studies conducted for each AR-SGs. Users' evaluations were extremely positive, also taking into account that many of our volunteers had little to no experience with AR. The main contribution in answering this RQ were the following:

- the identification of guidelines/suggestions that can be followed to address usability and UX in the design of AR-SGs adequately. They can be summarized as:
    - consider that, although AR is appraised as a mature technology, the diffusion of AR technology is still limited, and therefore, most users are still inexperienced with it. In these cases, practitioners should give nothing for granted, even if AR is administered through devices (like smartphones and tablets) most users are familiar with. The design should include appropriate feedback on *when*, *how*, and *where* action should be performed. To ensure the message is correctly conveyed, feedbacks should leverage stimuli across all (possible) senses (i.e., visual, auditory, and haptic). We stress the importance of explicating *where* digital contents are. At the moment, augmented contents are primarily experienced through windows (i.e., the screen of a mobile device or the limited FOV of HMD). As a result, users might risk not seeing digital content as they must orient their devices in a specific direction. Therefore, it is important to inform the user about the direction to turn to see and interact with the augmented contents.
    - enable users to learn "how" to use the AR-SGs *before* the learning experience starts. In this way, users can get acquainted with AR usage, commands and interaction metaphors of the AR-SGs, thus enabling them to reduce the mental effort to operate the system and focus on the learning activities proposed by the AR-SGs. An effective way we found to

  accomplish this objective was implementing interactive training sessions that can be repeatably delivered before the learning experience starts;

  - validation of design choices through an iterative development process where a running prototype is continuously evaluated. It could be argued that such a process can increase time, money, and effort. However, in the long run, its benefit are clear as they lead to an early identification of usability and UX issues that, ultimately, contribute to the final application success and avoid discovering "too late" issues that require drastic changes in design choices (which, in turns, need to be implemented and tested). Then, it is clear that a low level of usability can bias the experimental results collected, as users are more focused on trying to use the application, rather than on the educational contents proposed. Our experience let us solicit the research community to place adequate care (and time) in carefully designing for (and assessing usability and UX from the early steps of the development process.

- the proposal of novel VC aimed at eliciting users' mutual awareness in co-located collaborative AR experiences, which contribute to improve the overall user experience and the collaborative efforts to accomplish tasks in the shared virtual environment.

7.2 Limitations and Open Problems

As for the limitations of this work, the first we highlight is the adoption of a case study approach (i.e., developing AR-SGs targeting a specific topic). Although this approach allowed us to gather interesting results, their generalization is not always straightforward. Despite that, given that the exploration of AR-SGs in these learning domains is still in its infancy, we think our work can help advance the current state of the art. Indeed, we adopted design choices informed by high-level theoretical frameworks to foster the repeatability of the specific implementation described in this thesis. Therefore, although the data collected leads to evidence in specific cases (or learning topics), we believe similar results can be achieved also in different topics by adopting the same guidelines/frameworks.

Another limitation we wish to acknowledge relates to the reported effectiveness arising from the combination of AR and SG. We tested this relationship as a whole. However, we did not collect evidence that helps us truly understand "if", "which", and "how" specific affordances of either AR or SG and their association accounted for the positive detected results. We stress that answering these questions would have required developing different (yet similar) applications with and without certain features and comparing their outcomes. However, the choice to focus on single applications encompassing both AR and SG features was mainly dictated by time constraints (this Ph.D. program had a fixed duration of three years) and the increased complexity of such evaluation (i.e., enlarging the size of the volunteer panel). Therefore, we could not isolate and entirely understand direct mappings and influences between features of each technology/approach and outcomes. Nevertheless, we favored dedicating the available time to carefully design the proposed AR-SGs, which required numerous iteration cycles to reach the envisioned outcomes.

From the analysis of state of the art in AR-SGs and thanks to issues that emerged throughout the development and assessment of our AR-SGs, we wish to highlight the following open problems that might inform practitioners on topics to pursue in future research.

- When evaluating AR-learning tools, SGs or AR-SGs, the majority of authors focus on the effectiveness of the proposed applications as a whole. Nevertheless, these solutions are the outcomes of a tight relationship of many elements (e.g., game features, interaction patterns and different visualization modalities). Therefore researchers suggest [36, 188] that more evidence should be gathered to better understand which and how individual elements contribute to the pursued outcome.

- In order to assess AR-SGs in detail, multiple data sources must be combined. Unfortunately, these data often involve time-consuming labeling procedures (e.g., the manual video annotation process performed for Asteroid Escape and Sustain). Thus, we believe that recent technological advancements in computer vision and machine learning could enable more automated assessment processes, hence, freeing researchers from the "boring" part of data annotation and focus more on its analysis. Unfortunately, to the best of our knowledge, these systems are not yet available, and we believe they could be of priceless help to the research community.

- In current AR-SGs reviews, collaboration is reported with different inclusion criteria. Although every study reports that collaboration is a positive (and vital) feature of AR-SGs, to the best of our knowledge, there are no reviews capable of providing a complete picture of collaborative AR-SGs, emphasizing which are their affordances and benefits. Therefore, we deem it interesting (and vital) for the community to perform such a review.

- Sound theoretical frameworks have been proposed for the design of SGs [54]. However, to the best of our knowledge, similar results in the field of AR learning are still limited and in the particular case of AR-SGs have not been proposed yet.

- In Chapter 6 we have described several solutions to usability and UX problems that may arise from the novel use of a technology or a digital application.

Nevertheless, when these digital applications are games, there are elements that influence the personal gaming experience at a *behavioral*, *physiological*, and *psychological* level [322]. Such a broader view over the player-game relationship is usually referred to as *player experience* (PX). When games also have an educational purpose, the PX must seamlessly blend with the learning experience. In the research proposed in [268], we explored which methodologies have been proposed for (i) designing educational SGs under the lens of PX and (ii) evaluating the PX in this context. The design frameworks we identified are mostly focused on either the educational or PX element, but rarely these two elements are addressed jointly. Moreover, to evaluate PX, there is a general consensus that mixed methods combining multiple data sources are preferable. However, no guidelines or tools have been proposed yet in the literature.

- Serious pervasive games (SPG) are SG where the virtual boundaries are broken, and reality becomes part of it (i.e., elements in the physical world influence the game) [63]. SPG expands the game's space according to the context where it is played, establishing a learning experience where players' attention should be focused on both the physical domain as the digital realm. Therefore, given this continuum between physical and virtual, AR is one of the technologies employed in creating SPGs. The new generation of mobile devices with marker-less AR capabilities can significantly expand the physical playspace where these activities can occur (i.e., the game is not bound to play areas previously equipped with markers). Despite the vast potentialities of experiencing these games globally (i.e., imagine the AR game Pokemon Go), examples of AR-SPGs are still limited. Therefore, more research is required to understand better how to fully exploit these new technologies in the creation of "serious" activities which take the world as a playground leading to new kinds of immersion, engagement, and eventually learning [229].

7.3 Future Works

We wish to conclude this thesis by briefly describing the activities we are currently involved in, as they mainly originate from limitations or problems we identified through the activities that led to this Ph.D. program conclusion.

First, when tackling procedure learning, volunteers reported a certain "rigidity" in the management of the educational contents. In Holo-BLSD, besides receiveing instructions, learners have no ways to ask questions to the system, or request additional information, as usually happens when students interact with a human trainer. To address this issue, we are planning to leverage adaptive learning practices. Adaptive learning is a new educational approach that allows the personalized delivery of content in the training phase (e.g., communicating it in different ways and at different levels of depth) according to the learner's needs, skill level, and interests. In Holo-BLSD, the adaptability is limited to providing learners with the choice of receiving from scratch all instructions related to the procedures to be learned or training to practice them. Overcoming this limitation requires the development of a more advanced adaptive system. In turn, this adaptive system requires implementing a module in the application that can detect the user's behaviors, progress, and specific requests and analyze this information to adapt the training content delivered. One of the components we consider necessary in this module is the inclusion in the training experience of a virtual instructor capable of establishing a realistic interaction with the user through voice and gestures. In addition to conveying training content, the presence of a virtual instructor allows the learner to turn to him or her for clarification in case of difficulty, request alternative explanations, or repetition of exposed concepts. Such a virtual instructor should be capable of (i) sensing the environment and understanding the current state of the simulation in order to understand how the trainee is conducting his task and, possibly, intervene to help him (ii) understanding user questions and answer in proper ways (iii) delivering emotional rich feedback through body and facial expressions in order to improve the empathy with the learner.

Second, concerning awareness in AR-based collaborative tasks, the novel VCs we propose in Chapter 4 have been only evaluated in a co-located context. Nevertheless, AR collaborative environments can also be exploited by users who are remotely connected. We are therefore committed to understanding if our VCs can be successfully used also in such collaborative contexts. We believe this research is of even greater importance as the current pandemic situation (COVID-19) is pushing forward the establishment of new working practices. Above all, remote work and remote conferences. We believe shared AR spaces will help make collaboration more accessible and more efficient in these new working scenarios, eventually leading to higher quality outcomes.

Finally, we are committed to the exploration of the potentialities offered by learning through making SG. Rather than simply playing them, designing and developing a SG can be envisioned as an educational activity. This approach stems from the constructionist theory [231] applied to games. The general concept behind the idea of learning by making games is that the process of designing and creating games helps students to (i) improve their understanding of the subject matter, which needs to be broken down and analyzed in every detail to allow for the development of contextually appropriate game mechanics, (ii) construct new relationships with knowledge, (iii) express in more depth their ideas and feelings about the subject matter of the game, and (iv) develop collaborative (and creative) problem solving abilities [93, 165]. The possibility of addressing this research topic was offered through our involvement in an activity promoted by Politecnico di Torino. From March to June 2020, we tutored 59 students challenged to design and develop a SG to raise awareness on sustainability-related issues within the university campus. By evaluating the developed SGs and collecting questionnaires from all the participants, we assessed (i) the effectiveness of SG design and development as a learning activity, (ii) its impacts on teamwork, and (iii) the likeability of students to repeat a similar experience. Since students were not required to create AR-SGs, this activity, and its results, have not been discussed in this thesis, but details can be consulted in [69]. However, although not directly addressing AR, some findings can be contextualized to the broader field of SGs and consequently to the more specific one of AR-SGs. Despite the positive results, what mainly emerged were difficulties associated with the development of the SG as the students had no background in making games. From this experience, we recognized that making SG can be a positive and powerful activity for "serious" purposes, but unfortunately, the tools available are inadequate for a non-specialized audience. Therefore new authoring tools are required to expand the target audience of people who can approach the design and development of SG. Finally, considering the rising interest in AR, extensively discussed in this thesis, we believe these tools should also provide the possibilities of creating experiences that exploit the numerous benefits elicited by AR.

Appendix A

Developed AR-SGs Design Building Blocks

In this Appendix we would like to outline the core building blocks on which all of the AR -SGs discussed in this paper are built. These building blocks are taken from Tracy Fullerton's Game Design Workshop [116]. Each AR-SG is described in terms of its Formal and Dramatic elements, game mechanics, and game systems. Below, we first provide a template for the content addressed in each AR-SG. For each element and sub-element, we give a brief description of its meaning. Then the same template will be filled in for each AR-SG.

A.1 Building Blocks Template

Formal Elements

- **Players:** number of players, roles and their interaction patterns.
- **Objectives**: what players are trying to accomplish within the rules of the game.
- **Procedures**: the methods of play and the actions that players can take to achieve the game objectives.
    - Starting Action: how to put the game into play.
    - Progression of Action: ongoing procedures after the starting action.
    - Special Action: available conditional to other elements or game state.
    - Resolving Action: how to bring gameplay to a close.
- **Rules**: they define game objects and define allowable actions by the players.

- **Resources**: assets that can be used to accomplish certain goals, but which are made scarce in the system by the designer.
- **Conflict**: emerges from the players trying to accomplish the goals of the game within its rules and boundaries. Conflict can originate from:
  - Obstacles: physical and/or virtual. They can be surpassed by means of both mental and physical skills.
  - Opponents: other players or non player (i.e., controlled by the game).
  - Dilemmas: dilemma-based choices that players have to make.
- **Boundaries**: what separate the game from everything that is not the game.
- **Outcome**: the element of uncertainty which is measured at the end of the game.

Dramatic Elements

- **Challenge:** the overall challenge of the game, used to establish a meaningful experience and should be balanced with the players' abilities throughout the gaming experience.
- **Premise:** it establishes the action of the game within a setting or metaphor.
- **Characters:** they are the agents through whose actions a drama is told.

Game Mechanics

List of the core game mechanics.

Game Dynamics

List of the core game dynamics.

A.2 Holo-BLSD

Formal Elements

- **Players**
  - 1 player vs the game
  - Multilateral Competition: players compete trying to beat each others' scores

- **Objectives**
  - Practice the BLSD procedure by trying to complete it with the fewest errors and in the shortest time
- **Procedures**
  - Starting Action: after the learning phase, players can decide to start (or repeat) the rehearsal phase (i.e., the Holo-BLSD gamified system)
  - Progression of Action: players execute all the correct actions in a given step of the BLSD procedure
  - Special Action: N/A
  - Resolving Action: the last step of the procedure is correctly completed
- **Rules**
  - Rules in each step of the procedure are defined according to the BLSD protocol
  - Each step must be completed correctly, in the right order and within a given time
- **Resources**
  - Knowledge acquired during the learning phase where all the steps of the procedure are explained in detail
  - Physical CPR manikin
  - Time limit for each step of the procedure
- **Conflict**
  - Obstacles: each step of the procedure has its specific obstacles
  - Opponents: N/A
  - Dilemmas: N/A
- **Boundaries**
  - Physical room in which the AR experience takes place
  - The FOV of the AR-HMD
- **Outcome**
  - Save the unconscious person's life making most points possible

Dramatic Elements

- **Challenge:** correctly apply all the notions acquired in the learning phase to prepare for the final (one time only) evaluation
- **Premise:** players find themselves in a room and there is an unconscious person on the floor. They must do all they can to save his life.
- **Characters**
  - The unconscious person
  - Bystanders for asking the defibrillator
  - The hospital call center operator

Game Mechanics

- Players can move freely in the real and virtual environment by walking in it.
- Players can interact with the different objects through voice, gaze and tap. The actions to perform are specific to each step of the augmented BLSD procedure.

Game Dynamics

- By completing each step of the procedure players receive points for (1) the correctness and (2) velocity of their actions. If the time limit for a given step of the procedure the player looses points for every second.
- To foster players' competitiveness, at the end of the procedure a leaderboard shows the players' achievement in comparison to its peers.
- Players may make mistakes of varying severity, which are treated as either warnings or errors and immediately reported by the system whom explains what exactly has been done wrong. Warnings allow the player to continue in the procedure, Errors make the application start again.
- Before accessing the evaluation phase, players are required to complete a full rehearsal phase with no warnings and no errors.

A.3 Sustain

Formal Elements

- **Players**
  - 3 players vs the game
  - cooperative
- **Objectives**
  - Build a sustainable city trying to reach a population of 10000 citizens
- **Procedures**
  - Starting Action: all players connect and choose their role
  - Progression of Action: each turn is structured into 3 phases: (i) *Planning*, players make plans for the forthcoming turn highlighting which variables they wish to improve, (ii) *Execution*, players select building actions to implement the collectively-agreed plan, and (iii) *Evaluation*, players are confronted with the effects of their actions on the state of the city
  - Special Action: N/A
  - Resolving Action: the game sessions reaches the 6^{th} turn
- **Rules**
  - Plans must be collaboratively defined.
  - Each building action has a cost (Economic & State Variables)
  - All actions must be approved by all players at the end of the turn
- **Resources**
  - 6 turns
  - Number of stars to allocate for the collaborative plan
  - Shared economic budget
  - Maximum of 2 actions per player per turn
  - The available buildings per turn, unique to each role
  - *Global State Variables:* Environment, Happiness and Population
  - *Local Impact Variables:* Housing, Working Places, Energy, Food, Transportation, Leisure and Pollution

- **Conflict**
  - Obstacles: N/A
  - Opponents: N/A
  - Dilemmas: grow population at the expense of their happiness and environmental sustainability
- **Boundaries**
  - Physical table where the augmented content (the city being constructed) can be seen through a tablet
  - The augmented environment features 2 4x4 grid blocks where building can be constructed
- **Outcome**
  - *Global State Variables* at the end of the 6th turn

Dramatic Elements

- **Challenge:**
  - Understand the relation between *Local Impact Variables* and Global State Variables
  - Achieve this understanding by collectively sharing ideas, viewpoints and knowledge
- **Premise:** players are part of the city council which has been appointed to expand the city's population maintaining a sustainable balance between citizens happiness and environmental pollution
- **Characters:** Mayor, Minister of Agriculture, Minister of Energy

Game Mechanics

- *Planning*
  - Players assign stars to *Local Impact Variables* and Global State Variables
- *Execution*
  - Players select buildings and chose where to construct them in one of the 2 4x4 grids

  - Buildings can be destroyed
  - New buildings are unlocked as the game unfolds
  - Each building can have 3 levels: medium, big and high, each with a different cost and impact
  - When players are satisfied with their actions they must explicitly confirm them
  - Players can decide to skip the turn with no actions performed
  - Players choose to agree or disagree with other players' actions

- *Evaluation*
  - A Fuzzy System computes the values of the Global State Variables according to those of the *Local Impact Variables*

Game Dynamics

- The *Planning* phase a turn t is performed with the information received in the *Evaluation* of turn t-1.
- The actions performed in the *Execution* phase are informed based on the collaboratively agreed plan done in the *Planning* phase
- The *Planning*, *Execution*, *Evaluation* cycle is repeated for every turn
- Collaborative dynamics are fostered as to progress actions from all players must be discussed and agreed on

---

A.4 Asteroid Escape

Formal Elements

- **Players**
  - 3 players vs the game
  - cooperative
- **Objectives**
  - Pilot the spaceship out of an asteroid before it implodes

- **Procedures**
    - Starting Action: players complete the tutorial
    - Progression of Action: 2 *crew members* pilot the spaceship supported by the information provided by the *navigator*
    - Special Action: shoot rockets and access the *navigator*'s view from inside the spaceship
    - Resolving Action: exit the asteroid from the only available exit
- **Rules**
    - By hitting the tunnels' walls the spaceship receives damage
    - If the spaceship is attacked by the mining drones it looses fuel
- **Resources**
    - Spaceship's damage level
    - Spaceship's fuel
    - Countdown for the asteroid's implosion
- **Conflict**
    - Obstacles:drones inside tunnels which attack the spaceship and tunnels collapsing due to minor explosions
    - Opponents: N/A
    - Dilemmas: N/A
- **Boundaries**
    - Asteroid's tunnel walls
    - The projected screen depicting the spaceship's interior
    - The FOV of the AR-HMD
    - The spaceship controls available on 2 smartphones possessed by the *crew* members
- **Outcome**
    - Exit the asteroid before the countdown expires

Dramatic Elements

- **Challenge**

  - *Crew*, become skilled in piloting the spaceship learning to coordinate with the other crew member who has different, yet complementary controls
  - *Navigator*, learn to interpret the augmented 3D map and provide effective information to the *crew* piloting the spaceship
  - *Crew & Navigator*, learn to effectively communicate with one another

- **Premise:**the AI controlling drones in a mining asteroid has taken over, activating the asteroid's implosion procedure. You are the only human crew left inside, escape before the asteroid explodes while you are still inside

- **Characters:** the mining drones

Game Mechanics

- The 2 *crew* members have different and complementary controls: one manages the spaceship's roll and pitch and the other yaw and propulsion

- The mining drones attack when the spaceship is sufficiently close. They attach to the spaceship stealing its fuel

- The *crew* members can shoot rockets to the drones.

- The *crew* members can perform a coordinated action to deattach the drones

- Tunnels collapse due to the asteroid's internal minor implosions.

- The *navigator* has a complete 3D representation of the tunnel's map and can explore it in AR

Game Dynamics

- The *crew* members do not have access to the tunnel's map and can only see ahead of them.

- The *navigator* can only see the tunnel's map but cannot see drones and obstacles in front of the *crew* members

- The *crew* members and the *navigator* are allowed (should) communicate complete each others missing pieces of information

A.5 ARScape

Formal Elements

- **Players**
  - 3 players vs the game
  - cooperative
- **Objectives**
  - solve all the puzzles
- **Procedures**
  - Starting Action: all players connect and enter the shared AR environment
  - Progression of Action: solve one puzzle at a time
  - Special Action: N/A
  - Resolving Action: solve the last puzzle
- **Rules**
  - To proceed to the next puzzle the previous one must be completed
- **Resources**
  - Different tools available to solve individual puzzles: torch, keys, skull
  - Time required to complete each puzzle.
- **Conflict**
  - Obstacles: puzzles to solve
  - Opponents: N/A
  - Dilemmas: N/A
- **Boundaries**
  - The physical room in which the AR contents are placed.
  - The virtual world seen through the HHD's screen
- **Outcome**
  - Solve all the puzzles

Dramatic Elements

- **Challenge:** puzzles must be solved collaboratively
- **Premise:** three archaeologists have found an ancient table and must unveil its hidden secret
- **Characters:** N/A

Game Mechanics

- Navigation: players have to (synchronously) position themselves in specific points of the environment
- Gaze: players have to (synchronously) observe specific objects in the environment
- Manipulation: players have to (synchronously) manipulate specific objects in the environment
- Specific puzzles require a combination of the above actions.
- Each puzzle requires at least one action/interaction by all players

Game Dynamics

- To solve each puzzle coordination and communication is required by the players as the game does not progress without without everyone's contribution/cooperation

Appendix B

Personal Contribution to Research Projects

As mentioned in the introduction of this thesis, the term “we” was used instead of “I”, since all the contents which have been presented have benefit from a collaborative effort of both the research group I am affiliated with and external collaborators. However, to clarify how I have personally contributed to each research project/activity described in this thesis, I here present a detailed list of the carried out activities.

- Holo-BLSD (Section 3.1.1):
  - Design of the Application/Experience
  - Development of the Application
  - Experimental Design definition
  - Data collection
- Firefighter Training (Section 3.1.7):
  - Design of the Application/Experience
  - Development of the Application
  - Experimental Design definition
  - Data Analysis
- Sustain (Section 3.2.1):
  - Design of the Application/Experience
  - Development of the Application
  - Experimental Design definition

  - Data collection
  - Data Analysis
- Asteroid Escape (Section 3.3):
  - Design of the Application/Experience
  - Development of the Application
  - Experimental Design definition
  - Data collection
  - Data Analysis
- AR Awareness (& ARScape) (Chapter 4):
  - Design of the Framework
  - Experimental Design definition
  - Data collection
  - Data Analysis

Appendix C

List of the Published Works

This Ph.D. thesis aims to present the main research activities and personal contributions to the scientific community in recent years. These works have been disseminated through the following publications:

- Alysson Diniz Dos Santos, Francesco Strada, Andrea Martina, and Andrea Bottino. "Designing collaborative games for children education on sustainable development". In: 8th International Conference on Intelligent Technologies for Interactive Entertainment. Utrecth, NL. 2016.
- Alysson Diniz dos Santos, Francesco Strada, Andrea Bottino. "The design of an augmented reality collaborative game for sustainable development". In: International Conference on Games and Learning Alliance (GALA). Utrecth, NL. 2016.
- Alysson Diniz dos Santos, Francesco Strada, and Andrea Bottino. "Investigating the Design and Evaluation of Educational Games Under the Perspective of Player Experience". In: International Conference on Games and Learning Alliance (GALA). Lisbona, PT. 2017.
- Andrea Bottino, Pierluigi Ingrassia, Fabrizio Lamberti, Fernando Salvetti, Francesco Strada, and Antony Vitillo. "Holo-BLSD: an Augmented Reality self-directed learning and evaluation system for effective Basic Life Support Defibrillation training". In: 18th International Meeting on Simulation in Healthcare (IMSH). Los Angeles, US. 2018
- Pier Luigi Ingrassia, Giulia Mormando, Eleonora Giudici, Francesco Strada, Fabio Carfagna, Fabrizio Lamberti, and Andrea Bottino. "Augmented reality learning environment for basic life support and defibrillation training: usability study". In: JMIR. JOURNAL OF MEDICAL INTERNET RESEARCH. 22(5). 2019

- Francesco Strada, Andrea Bottino, and Fabrizio Lamberti. “Holo-BLSD – A holographic tool for self-training and self-evaluation of emergency response skills”. In: IEEE TRANSACTIONS ON EMERGING TOPICS IN COMPUTING. 2019.
- Alysson Diniz dos Santos, Francesco Strada and Andrea Bottino. “Approaching sustainability learning via digital serious games”. In: IEEE TRANSACTIONS ON LEARNING TECHNOLOGIES. 12(3), pp. 303-320. 2019
- Filippo Gabriele Pratticò, Francesco Strada, Fabrizio Lamberti, and Andrea Bottino. “Asteroid Escape: A Serious Game to Foster Teamwork Abilities”. In: 40th Annual Conference of the European Association for Computer Graphics (Eurographics). Genova, IT. 2019.
- Fabrizio Corelli, Edoardo Battegazzorre, Francesco Strada, Andrea Bottino, Gian Paolo Cimellaro.“Assessing the Usability of Different Virtual Reality Systems for Firefighter Training”. In: Proceedings of 4th International Conference on Human Computer Interaction Theory and Applications (HUCAPP). Valletta, Malta. 2020
- Edoardo Battegazzorre, Davide Calandra, Francesco Strada, Andrea Bottino, and Fabrizio Lamberti. “Evaluating the suitability of several AR devices and tools for industrial applications”. In: 7th International Conference on Augmented Reality, Virtual Reality and Computer Graphics (AVR). Lecce, IT. 2020
- María Ximena López, Francesco Strada, Andrea Bottino, and Carlo Fabricatore. “Using Multimodal Learning Analytics to Explore Collaboration in a Sustainability Co-located Tabletop Game”. In: Proceedings of 15th European Conference of Game Based Learning (ECGBL). Brighton, UK. 2021
- Sara Cravero, Francesco Strada, Isabella Lami and Andrea Bottino. “Learning sustainability by making games. The experience of a challenge as a novel approach for Education for Sustainable Development”. In: th International Conference on Higher Education Advances (HEAd). Valencia, Spain. 2021.

These works have recently been submitted and are still under review:

- Francesco Strada, Edoardo Battegazzorre, Enrico Ameglio, Simone Bruno Turello, and Andrea Bottino. “Assessing Visual Cues for Improving Awareness in Collaborative Augmented Reality”. In: 24th ACM Conference on Computer-Supported Cooperative Work and Social Computing (CSCW). 2021.

Bibliography

[1] Gary L Adams and Siegfried Engelmann. *Research on Direct Instruction: 25 Years beyond DISTAR*. Eric, 1996.

[2] Adesuwa Vanessa Agbedahin. “Sustainable development, Education for Sustainable Development, and the 2030 Agenda for Sustainable Development: Emergence, efficacy, eminence, and future”. In: *Sustainable Development* 27.4 (2019), pp. 669–680. ISSN: 10991719. DOI: `10.1002/sd.1931`.

[3] Gloria Ajit, TERRY LUCAS, and RINGAH KANYAN. “A Systematic Review of Augmented Reality in STEM Education”. In: *Estudios de economía aplicada* 39.1 (2021), p. 4.

[4] Murat Akçayır and Gökçe Akçayır. “Advantages and challenges associated with augmented reality for education: A systematic review of the literature”. In: *Educational Research Review* 20 (2017), pp. 1–11.

[5] Ismo Alakärppä et al. “Using nature elements in mobile AR for education with children”. In: *Proceedings of the 19th International Conference on human-computer interaction with mobile devices and Services*. 2017, pp. 1–13.

[6] Stephen M Alessi and Stanley R Trollip. *Multimedia for learning: Methods and development*. Allyn & Bacon, 2001.

[7] Cristina Alonso-Fernández et al. “Improving evidence-based assessment of players using serious games”. In: *Telematics and Informatics* 60 (2021), p. 101583.

[8] American Heart Association. *My Cardiac Coach*. 2017. URL: `https://play.google.com/store/apps/details?id=cc.aha.cardiaccoach`.

[9] Margarita Angelidou and Artemis Psaltoglou. “An empirical investigation of social innovation initiatives for sustainable urban development”. In: *Sustainable cities and society* 33 (2017), pp. 113–125.

[10] Sylvester Arnab et al. “The development approach of a pedagogically-driven serious game to support Relationship and Sex Education (RSE) within a classroom setting”. In: *Computers & Education* 69 (2013), pp. 15–30.

[11] Jakob J Assmann and Julia V Gallenkamp. “Communication and leadership trustworthiness in virtual teams: an empirical comparison of the US and China”. In: *2009 42nd Hawaii International Conference on System Sciences.* IEEE. 2009, pp. 1–10.

[12] Dana Atwood-Blaine and Douglas Huffman. “Mobile gaming and student interactions in a science center: the future of gaming in science education”. In: *International journal of science and mathematics education* 15.1 (2017), pp. 45–65.

[13] Steven K Ayer, John I Messner, and Chimay J Anumba. “Augmented Reality Gaming in Sustainable Design Education”. In: *Journal of Architectural Engineering* 22.1 (2016), p. 04015012.

[14] Aida Azadegan, Johann CKH Riedel, and Jannicke Baalsrud Hauge. “Serious games adoption in corporate training”. In: *International conference on serious games development and applications.* Springer. 2012, pp. 74–85.

[15] Ronald Azuma. “Tracking requirements for augmented reality”. In: *Communications of the ACM* 36.7 (1993), pp. 50–51.

[16] Ronald Azuma et al. “Recent advances in augmented reality”. In: *IEEE computer graphics and applications* 21.6 (2001), pp. 34–47.

[17] Ronald T Azuma. “A survey of augmented reality”. In: *Presence: Teleoperators & Virtual Environments* 6.4 (1997), pp. 355–385.

[18] Jorge Luis Bacca Acosta et al. “Augmented reality trends in education: a systematic review of research and applications”. In: *Journal of Educational Technology and Society, 2014, vol. 17, núm. 4, p. 133-149* (2014).

[19] Per Backlund and Maurice Hendrix. “Educational games-are they worth the effort? A literature survey of the effectiveness of serious games”. In: *2013 5th international conference on games and virtual worlds for serious applications (VS-GAMES).* IEEE. 2013, pp. 1–8.

[20] Albert Bandura et al. “Guide for constructing self-efficacy scales”. In: *Self-efficacy beliefs of adolescents* 5.1 (2006), pp. 307–337.

[21] Doreen K Baringer and James C McCroskey. “Immediacy in the classroom: Student immediacy”. In: *Communication education* 49.2 (2000), pp. 178–186.

[22] Alma Gloria Barrera Yañez, Cristina Alonso-Fernandez, and Baltasar Fernandez Manjon. “Review of serious games to educate on gender equality”. In: *Eighth International Conference on Technological Ecosystems for Enhancing Multiculturality.* 2020, pp. 662–668.

[23] B.I. Batuwanthudawa and K.P.N Jayasena. “Real- Time Location based Augmented Reality Advertising Platform”. In: *2020 2nd International Conference on Advancements in Computing (ICAC)*. IEEE, Dec. 2020. DOI: 10.1109/icac51239.2020.9357261. URL: https://doi.org/10.1109/icac51239.2020.9357261.

[24] Katrin Becker, Jim R Parker, and Kevin McNulty. *The guide to computer simulations and games*. John Wiley & Sons, 2012.

[25] Francesco Bellotti et al. “Assessment in and of serious games: an overview”. In: *Advances in human-computer interaction* 2013 (2013).

[26] Emelia J. Benjamin and al. “Heart Disease and Stroke Statistics: A Report From the American Heart Association”. In: *Circulation* 137.12 (2018).

[27] Mark Billinghurst et al. “Asymmetries in collaborative wearable interfaces”. In: *Digest of Papers. Third International Symposium on Wearable Computers*. Ieee. 1999, pp. 133–140.

[28] Mark Billinghurst et al. “Experiments with face-to-face collaborative AR interfaces”. In: *Virtual Reality* 6.3 (2002), pp. 107–121.

[29] Arthur van BILSEN, Geertje BEKEBREDE, and Igor MAYER. “Understanding Complex Adaptive Systems by Playing Games”. In: *Informatics in Education* 9.1 (Apr. 2010), pp. 1–18. DOI: 10.15388/infedu.2010.01. URL: https://doi.org/10.15388/infedu.2010.01.

[30] Gary Bitter and Allen Corral. “The pedagogical potential of augmented reality apps”. In: *International Journal of Engineering Science Invention* 3.10 (2014), pp. 13–17.

[31] Timo Bleeker, Gun Lee, and Mark Billinghurst. “Ego-and Exocentric interaction for mobile AR conferencing”. In: *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. Ieee. 2013, pp. 1–6.

[32] Audrey L. Blewer et al. “Continuous chest compression cardiopulmonary resuscitation training promotes rescuer self-confidence and increased secondary training”. In: *Critical Care Medicine* 40.3 (2012), pp. 787–792.

[33] Ian Bogost. *Persuasive games*. Vol. 5. Cambridge, MA: MIT Press, 2007.

[34] Walter R Boot et al. “The effects of video game playing on attention, memory, and executive control”. In: *Acta psychologica* 129.3 (2008), pp. 387–398.

[35] D. A. Bowman, D. Koller, and L. F. Hodges. “Travel in immersive virtual environments: an evaluation of viewpoint motion control techniques”. In: *Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality*. Mar. 1997, pp. 45–52. DOI: 10.1109/vrais.1997.583043.

[36] Elizabeth A Boyle et al. “An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games”. In: *Computers & Education* 94 (2016), pp. 178–192.

[37] Elizabeth A. Boyle et al. “An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games”. In: *Computers & Education* 94 (2016), pp. 178–192. ISSN: 03601315. DOI: `10.1016/j.compedu.2015.11.003`.

[38] Aysun Bozanta et al. “Effects of serious games on perceived team cohesiveness in a multi-user virtual env.” In: *Comp. in Human Behavior* 59 (2016), pp. 380–388.

[39] Paula Bräuer and Athanasios Mazarakis. “Gamification und Augmented Reality für Lagerprozesse”. In: *Mensch und Computer 2018-Workshopband* (2018).

[40] Susan E Brennan et al. “How conversation is shaped by visual and spoken evidence”. In: *Approaches to studying world-situated language use: Bridging the language-as-product and language-as-action traditions* (2005), pp. 95–129.

[41] Denise M Bressler and Alec M Bodzin. “A mixed methods assessment of students’ flow experiences during a mobile augmented reality science game”. In: *Journal of computer assisted learning* 29.6 (2013), pp. 505–517.

[42] Johannes Breuer and Gary Bente. “Why so serious? On the relation of serious games and learning”. In: *Journal for Computer Game Culture* 4 (2010), pp. 7–24.

[43] John Brooke. “SUS – A quick and dirty usability scale”. In: *Usability Evaluation in Industry* 189 (1996). Isbn: 9780748404605, pp. 4–7.

[44] John Brooke. “SUS – A quick and dirty usability scale”. In: *Usability Evaluation in Industry* 189 (1996). Isbn: 9780748404605, pp. 4–7.

[45] John Brooke. “SUS: A Retrospective”. In: *Journal of Usability Studies* 8.2 (2013), pp. 29–40.

[46] John Brooke. “SUS: a retrospective”. In: *Journal of usability studies* 8.2 (2013), pp. 29–40.

[47] Carolee Buckler and Heather Creech. *Shaping the Future We Want - UN Decade of Education for Sustainable Development (2005-2014) Final Report.* Paris, France: Unesco, 2014, p. 198. ISBN: Isbn 978-92-3-100053-9.

[48] Patrick Buckley and Elaine Doyle. “Gamification and student motivation”. In: *Interactive learning environments* 24.6 (2016), pp. 1162–1175.

[49] Alejandro Calderón and Mercedes Ruiz. “A systematic literature review on serious games evaluation: An application to software project management”. In: *Computers & Education* 87 (2015), pp. 396–422.

[50] Andrés-Marcelo Calle-Bustos et al. “An augmented reality game to support therapeutic education for children with diabetes”. In: *PloS one* 12.9 (2017), e0184645.

[51] Ilaria Caponetto, Jeffrey Earp, and Michela Ott. “Gamification and education: A literature review”. In: *European Conference on Games Based Learning*. Vol. 1. Academic Conferences International Limited. 2014, p. 50.

[52] Peter H Carstensen and Kjeld Schmidt. “Computer supported cooperative work: New challenges to systems design”. In: *In K. Itoh (Ed.), Handbook of Human Factors*. Citeseer. 1999.

[53] Maira B. Carvalho et al. “An activity theory-based model for serious games analysis and conceptual design”. In: *Computers & Education* 87 (Apr. 2015), pp. 166–181. ISSN: 03601315. DOI: `10.1016/j.compedu.2015.03.023`. URL: `http://www.sciencedirect.com/science/article/pii/S0360131515001050`.

[54] Chiara Eva Catalano, Angelo Marco Luccini, and Michela Mortara. “Guidelines for an effective design of serious games”. In: *International Journal of Serious Games* 1.1 (2014).

[55] Hsin-Yi Chang, Ying-Shao Hsu, and Hsin-Kai Wu. “A comparison study of augmented reality versus interactive simulation technology to support student learning of a socio-scientific issue”. In: *Interactive learning environments* 24.6 (2016), pp. 1148–1161.

[56] Hsin-Yi Chang, Hsin-Kai Wu, and Ying-Shao Hsu. “Integrating a mobile augmented reality activity to contextualize student learning of a socioscientific issue”. In: *British Journal of Educational Technology* 44.3 (2013), E95–e99.

[57] Peter Checkland and John Poulter. “Soft Systems Methodology”. In: *Systems Approaches to Managing Change: A Practical Guide*. Ed. by Martin Reynolds and Sue Holwell. London: Springer, 2010, pp. 191–242. DOI: `10.1007/978-1-84882-809-4\_5`. URL: `http://link.springer.com/10.1007/978-1-84882-809-4%5C_5`.

[58] Kun-Hung Cheng and Chin-Chung Tsai. “Affordances of augmented reality in science learning: Suggestions for future research”. In: *Journal of science education and technology* 22.4 (2013), pp. 449–462.

[59] Tosti HC Chiang, Stephen JH Yang, and Gwo-Jen Hwang. “An augmented reality-based mobile learning system to improve students’ learning achievements and motivations in natural science inquiry activities”. In: *Journal of Educational Technology & Society* 17.4 (2014), pp. 352–365.

[60] Jeffrey Chin, Richard Dukes, and William Gamson. “Assessment in simulation and gaming: A review of the last 40 years”. In: *Simulation & Gaming* 40.4 (2009), pp. 553–568.

[61] S. P. Chung et al. “Effects of script-based role play in cardiopulmonary resuscitation team training”. In: *Emergency Medicine Journal* 28.8 (Sept. 2010), pp. 690–694.

[62] Darragh Coakley and Roisin Garvey. “The Great and the Green: Sustainable Development in Serious Games”. In: *9th European Conf. Games Based Learning (ECGBL2015)*. 2015, p. 135.

[63] António Coelho et al. “Serious Pervasive Games”. In: *Frontiers in Computer Science* 2 (2020), p. 30.

[64] Jeff Conklin. *Dialogue Mapping: Building Shared Understanding of Wicked Problems*. New York, NY, USA: John Wiley & Sons, 2005.

[65] Thomas M Connolly et al. “A systematic literature review of empirical evidence on computer games and serious games”. In: *Computers & Education* 59.2 (2012), pp. 661–686.

[66] Fabrizio Corelli et al. “Assessing the Usability of Different Virtual Reality Systems for Firefighter Training.” In: *VISIGRAPP (2: HUCAPP)*. 2020, pp. 146–153.

[67] Kevin Corti. “Games-based Learning; a serious business application”. In: *Informe de PixelLearning* 34.6 (2006), pp. 1–20.

[68] Tim Coughlan et al. “The conceptual framing, design and evaluation of device ecologies for collaborative activities”. In: *International Journal of Human-Computer Studies* 70.10 (Oct. 2012), pp. 765–779. DOI: `10.1016/j.ijhcs.2012.05.008`. URL: `https://doi.org/10.1016/j.ijhcs.2012.05.008`.

[69] Sara Cravero et al. “Learning sustainability by making games. The experience of a challenge as a novel approach for Education for Sustainable Development”. In: *7th International Conference on Higher Education Advances*. 2021.

[70] Chris Crawford. “The art of computer game design”. In: (1984).

[71] Joaquín Cubillo et al. “A learning environment for augmented reality mobile learning”. In: *2014 IEEE Frontiers in Education Conference (FIE) Proceedings*. IEEE. 2014, pp. 1–8.

[72] Sébastien Cuendet et al. "Designing augmented reality for the classroom". In: *Computers & Education* 68 (2013), pp. 557–569.

[73] Arthur Lyon Dahl. "Achievements and gaps in indicators for sustainability". In: *Ecological indicators* 17 (2012), pp. 14–19.

[74] Stéphane D Dandeneau and Mark W Baldwin. "The inhibition of socially rejecting information among people with high versus low self-esteem: The role of attentional bias and the effects of bias reduction training". In: *Journal of Social and Clinical Psychology* 23.4 (2004), pp. 584–603.

[75] Dragoş Datcu, Stephan G Lukosch, and Heide K Lukosch. "Handheld augmented reality for distributed collaborative crime scene investigation". In: *Proceedings of the 19th International Conference on Supporting Group Work*. 2016, pp. 267–276.

[76] Sebastian Deterding. "Situated motivational affordances of game elements: A conceptual model". In: *Gamification: Using game design elements in non-gaming contexts, a workshop at CHI*. Vol. 10. 1979742.1979575. 2011.

[77] Jenni Deveau, Gary Lovcik, and Aaron R Seitz. "Broad-based visual benefits from training with an integrated perceptual-learning video game". In: *Vision Research* 99 (2014), pp. 134–140.

[78] Arindam Dey et al. "A systematic review of 10 Years of Augmented Reality usability studies: 2005 to 2014". In: *Frontiers Robotics AI* 5.37 (2018). ISSN: 22969144. DOI: `10.3389/frobt.2018.00037`.

[79] Ines Di Loreto, Simone Mora, and Monica Divitini. "Collaborative serious games for crisis management: an overview". In: *2012 IEEE 21st International Workshop on Enabling Technologies: Infrastructure for Collaborative Enterprises*. IEEE. 2012, pp. 352–357.

[80] Ángela Di Serio, María Blanca Ibáñez, and Carlos Delgado Kloos. "Impact of an augmented reality system on students' motivation for a visual art course". In: *Computers & Education* 68 (2013), pp. 586–596.

[81] Cambridge Dictionary. *Meaning of effectiveness in English*. URL: `https://dictionary.cambridge.org/dictionary/english/effectiveness` (visited on 06/23/2021).

[82] Phil Diegmann et al. "Benefits of augmented reality in educational environments-a systematic literature review". In: *Benefits* 3.6 (2015), pp. 1542–1556.

[83] Jeremiah Diephuis et al. "Game Changer: Designing Co-Located Games that Utilize Player Proximity". In: *Proceedings of DiGRA 2015: Diversity of Play* (2015), pp. 1–26.

[84] Kody R Dillman et al. "A visual interaction cue framework from video game environments for augmented reality". In: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 2018, pp. 1–12.

[85] Alysson Diniz dos Santos, Francesco Strada, and Andrea Bottino. "Approaching sustainability learning via digital serious games". In: *IEEE Transactions on Learning Technologies* 12.3 (2019), pp. 303–320. ISSN: 19391382. DOI: `10.1109/tlt.2018.2858770`.

[86] Damien Djaouti et al. "Origins of serious games". In: *Serious games and edutainment applications*. Springer, 2011, pp. 25–43.

[87] Mary Jo Dondlinger. "Educational video game design: A review of the literature". In: *Journal of appl. educ. technology* 4.1 (2007), pp. 21–31.

[88] Ralf Dörner et al. *Serious Games*. Springer, 2016.

[89] Ashish Doshi et al. "Use of projector based augmented reality to improve manual spot-welding precision and accuracy for automotive manufacturing". In: *The International Journal of Advanced Manufacturing Technology* 89.5-8 (2017), pp. 1279–1293.

[90] Duke Health. *Duke CPR*. 2015. URL: `https://itunes.apple.com/us/app/duke-cpr/id1022440083?mt=8`.

[91] Riley E Dunlap et al. "New trends in measuring environmental attitudes: measuring endorsement of the new ecological paradigm: a revised NEP scale". In: *Journal of social issues* 56.3 (2000), pp. 425–442.

[92] Matt Dunleavy, Chris Dede, and Rebecca Mitchell. "Affordances and limitations of immersive participatory augmented reality simulations for teaching and learning". In: *Journal of science Education and Technology* 18.1 (2009), pp. 7–22.

[93] Jeffrey Earp. "Game making for learning: A systematic review of the research literature". In: *Proceedings of 8th international conference of education, research and innovation (ICERI2015)*. 2015, pp. 6426–6435.

[94] Alejandro Echeverría et al. "Exploring different technological platforms for supporting co-located collaborative games in the classroom". In: *Computers in Human Behavior* 28.4 (2012), pp. 1170–1177.

[95] Brent D. Jr. Edwards et al. "Achieving and Monitoring Education for Sustainable Development and Global Citizenship: A Systematic Review of the Literature". In: *Sustainability* 12.4 (2020), p. 1383. DOI: `https://doi.org/10.3390/su12041383`. URL: `www.mdpi.com/journal/sustainability`.

[96] Daniel Elford, Simon J Lancaster, and Garth A Jones. "Stereoisomers, Not Stereo Enigmas: A Stereochemistry Escape Activity Incorporating Augmented and Immersive Virtual Reality". In: *Journal of Chemical Education* (2021).

[97] Jason B Ellis et al. "Games for virtual team building". In: *Proceedings of the 7th ACM conference on Designing interactive systems*. Acm. 2008, pp. 295–304.

[98] Kay Emblen-Perry. "Enhancing student engagement in business sustainability through games". In: *International Journal of Sustainability in Higher Education* 19.5 (2018), pp. 858–876. ISSN: 14676370. DOI: `10.1108/ijshe-05-2017-0075`.

[99] Hendrik Engelbrecht, Robert Lindeman, and Simon Hoermann. "A SWOT Analysis of the Field of Virtual Reality for Firefighter Training". In: *Frontiers in Robotics and AI* 6 (2019), p. 101.

[100] Barrett Ens et al. "Revisiting collaboration through mixed reality: The evolution of groupware". In: *International Journal of Human-Computer Studies* 131 (2019), pp. 81–98.

[101] Noel Enyedy et al. "Learning physics through play in an augmented reality environment". In: *International journal of computer-supported collaborative learning* 7.3 (2012), pp. 347–378.

[102] Anabela Estudante and Nicolas Dietrich. "Using Augmented Reality to Stimulate Students and Diffuse Escape Game Activities to Larger Audiences". In: *Journal of Chemical Education* 97.5 (2020), pp. 1368–1374.

[103] Carlo Fabricatore. "Underneath and beyond mechanics: An activity-theoretical perspective on meaning-making in gameplay". In: *Games and Rules: Game Mechanics for the "Magic Circle"*. Ed. by Beat Suter, Mela Kocher, and Renè Bauer. Bielefeld, DE: Transcript Verlag, 2019. Chap. 5, pp. 87–112. ISBN: 978-3837643046. DOI: `10.14361/9783839443040-006`.

[104] Carlo Fabricatore, Dimitar Gyaurov, and Ximena Lopez. "Rethinking serious games design in the age of COVID-19: Setting the focus on wicked problems". In: *Entertainment Computing and Serious Games. ICEC-JCSG 2020. Lecture Notes in Computer Science*. Ed. by Minhua Ma et al. Cham, CH: Springer, 2020, pp. 243–259.

[105] Carlo Fabricatore and Ximena Lopez. "A Model to Identify Affordances for Game-Based Sustainability Learning". In: *Proceedings of the 8th European Conference on Game Based Learning*. Ed. by Carsten Busch. Reading, UK: Academic Conferences and Publishing International Limited, Oct. 2014, pp. 99–109. ISBN: 978-1-910309-55-1. DOI: `10.13140/2.1.5028.0967`.

[106] Carlo Fabricatore and Ximena Lopez. “Game-based learning as a meaning making-driven activity process: a human factors perspective”. In: *Proceedings of the 13th International Conference on Game Based Learning ECGBL 2019*. Ed. by Lars Elbæk et al. Reading, UK: Academic Conferences and Publishing International, 2019, pp. 218–226. ISBN: 978-1-912764-37-2. DOI: `10.34190/gbl.19.177`.

[107] Carlo Fabricatore and Ximena Lopez. “Sustainability Learning through Gaming: An Exploratory Study”. In: *Electronic Journal of e-Learning (EJEL)* 10.2 (2012), pp. 209–222. ISSN: 1479-4403. URL: `http://www.ejel.org/issue/download.html?idArticle=194`.

[108] Carlo Fabricatore and Ximena López. “Education in a complex world: nurturing chaordic agency through game design”. In: *Seeking Understanding*. Brill Sense, 2019, pp. 325–353.

[109] Steven Feiner et al. “Windows on the world: 2D windows for 3D augmented reality”. In: *Proceedings of the 6th annual ACM symposium on User interface software and technology*. 1993, pp. 145–155.

[110] Andrea Ferracani et al. “Natural and virtual environments for the training of emergency medicine personnel”. In: *Univ. Access in the Inf. Society* 14.3 (2015), pp. 351–362. ISSN: 1615-5297.

[111] Stephen M Fiore, Arthur Graesser, and Samuel Greiff. “Collaborative problem-solving education for the twenty-first-century workforce”. In: *Nature Human Behaviour* 2.6 (2018), pp. 367–369.

[112] Jay W Forrester. “Counterintuitive behavior of social systems”. In: *Theory and decision* 2.2 (1971), pp. 109–140.

[113] Piero Fraternali and Sergio Luis Herrera Gonzalez. “An Augmented Reality Game for Energy Awareness”. In: *Computer Vision Systems*. Ed. by Dimitrios Tzovaras et al. Cham: Springer International Publishing, 2019, pp. 629–638. ISBN: 978-3-030-34995-0.

[114] Sara de Freitas and Helen Routledge. “Designing leadership and soft skills in educational games: The e-leadership and soft skills educational games design model (ELESS)”. In: *British Journal of Educational Technology* 44.6 (2013), pp. 951–968.

[115] Fong-Ling Fu, Ya-Ling Wu, and Hsi-Chuan Ho. “An investigation of coopetitive pedagogic design for knowledge creation in Web-based learning”. In: *Computers & Education* 53.3 (2009), pp. 550–562.

[116] Tracy Fullerton. *Game design workshop: a playcentric approach to creating innovative games*. AK Peters/CRC Press, 2019.

[117] Erich Gamma et al. *Design Patterns: Elements of Reusable Object-Oriented Software*. 1st ed. Addison-Wesley Professional, 1994. ISBN: 0201633612. URL: http://www.amazon.com/Design-Patterns-Elements-Reusable-Object-Oriented/dp/0201633612/ref=ntt%5C_at%5C_ep%5C_dpi%5C_1.

[118] Rosemary Garris, Robert Ahlers, and James E. Driskell. "Games, Motivation, and Learning: A Research and Practice Model". In: *Simulation & Gaming* 33.4 (2002), pp. 441–467.

[119] Gartner. *5 Trends Drive the Gartner Hype Cycle for Emerging Technologies, 2020*. 2020. URL: https://www.gartner.com/smarterwithgartner/5-trends-drive-the-gartner-hype-cycle-for-emerging-technologies-2020 (visited on 03/16/2020).

[120] Steffen Gauglitz et al. "World-stabilized annotations and virtual scene navigation for remote collaboration". In: *Proceedings of the 27th annual ACM symposium on User interface software and technology*. 2014, pp. 449–459.

[121] Sofía Ruiz de Gauna et al. "Feedback on the Rate and Depth of Chest Compressions during Cardiopulmonary Resuscitation Using Only Accelerometers". In: *Plos One* 11.3 (2016). Ed. by A.B. Johan Groeneveld, e0150139.

[122] James Paul Gee. *Good video games+ good learning: Collected essays on video games, learning, and literacy*. Peter Lang, 2007.

[123] James Paul Gee. *Situated language and learning: A critique of traditional schooling*. routledge, 2012.

[124] James Paul Gee. *What Video Games Have to Teach Us About Learning and Literacy*. 2nd. New York, NY: Palgrave Macmillan, 2007. ISBN: 1403984530.

[125] Niklas Gericke et al. "The Sustainability Consciousness Questionnaire: The theoretical development and empirical validation of an evaluation instrument for stakeholders working with sustainable development". In: *Sustainable Development* 27.1 (2019), pp. 35–49.

[126] Saket Girotra et al. "Regional Variation in Out-of-Hospital Cardiac Arrest Survival in the United States". In: *Circulation* 133.22 (2016), pp. 2159–2168.

[127] Ian Glover. "Play as you learn: gamification as a technique for motivating learners". In: *Edmedia+ innovate learning*. Association for the Advancement of Computing in Education (AACE). 2013, pp. 1999–2008.

[128] Carina González-González et al. "Design and analysis of collaborative interactions in social educational videogames". In: *Computers in Human Behavior* 31 (2014), pp. 602–611.

[129] Fabio Grandi et al. "A reference framework to combine model-based design and AR to improve social sustainability". In: *Sustainability (Switzerland)* 13.4 (2021), pp. 1–16. ISSN: 20711050. DOI: 10.3390/su13042031.

[130] Audrey Joyce Gray. *Constructivist teaching and learning.* Saskatchewan School Trustees Association, 1997.

[131] Irene Greif. *Computer-supported cooperative work: A book of readings.* Morgan Kaufmann Publishers Inc., 1988.

[132] Robert Greif et al. "European Resuscitation Council Guidelines for Resuscitation 2015: Section 10. Education and implementation of resuscitation". In: *Resuscitation* 95 (2015), pp. 288–301. ISSN: 0300-9572.

[133] Mariluz Guenaga et al. "A serious game to develop and assess teamwork competency". In: *2014 International Symposium on Computers in Education.* 2014, pp. 183–188.

[134] Ezgi Tosik Gün and Bilal Atasoy. "The effects of augmented reality on elementary school students' spatial ability and academic achievement". In: *Egitim ve Bilim* 42.191 (2017).

[135] Anhong Guo et al. "Blocks". In: *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 3.3 (Sept. 2019), pp. 1–24. DOI: `10.1145/3351241`. URL: `https://doi.org/10.1145/3351241`.

[136] Carl Gutwin and Saul Greenberg. "A Descriptive Framework of Workspace Awareness for Real-Time Groupware". In: *Computer Supported Cooperative Work (CSCW)* 11.3-4 (Sept. 2002), pp. 411–446. DOI: `10.1023/a:1021271517844`. URL: `https://doi.org/10.1023/a:1021271517844`.

[137] Dimitar Gyaurov, Carlo Fabricatore, and Ximena Lopez. "CPS-GBL Framework to Design and Evaluate Game Systems Promoting Intrinsically-Motivating Complex Problem-Solving Processes". In: *Proceedings of the 14th International Conference on Game Based Learning ECGBL 2020.* Reading, UK: Academic Conferences and Publishing International, 2020, pp. 220–228.

[138] Robert Jan den Haan and Mascha C. van der Voort. "On evaluating social learning outcomes of serious games to collaboratively address sustainability problems: A literature review". In: *Sustainability* 10.12 (2018), p. 4529. ISSN: 20711050. DOI: `10.3390/su10124529`.

[139] Guillaume Hacques et al. "Exploring to learn and learning to explore". In: *Psychological Research* (2020), pp. 1–13.

[140] Päivi Häkkinen et al. "Multiplayer 3D game in supporting team-building activities in a work organization". In: *12th International Conference on Advanced Learning Technologies.* 2012, pp. 430–432.

[141] Philip Hallinger et al. "A bibliometric review of research on simulations and serious games used in educating for sustainability, 1997–2019". In: *Journal of Cleaner Production* 256 (2020). ISSN: 09596526. DOI: `10.1016/j.jclepro.2020.120358`.

[142] Fatima Yahya Al-Hammadi et al. “Augmented reality in educational games: City of Life (COL) emirati sustainability-edutainment interactive game”. In: *2018 Advances in Science and Engineering Technology International Conferences (ASET)*. 2018, pp. 1–7. DOI: `10.1109/icaset.2018.8376921`.

[143] Jonathan M. Harris. “Basic Principles of Sustainable Development”. In: *Dimensions of Sustainable Development, volume 1*. Ed. by Kamaljit S. Bawa and Reinmar Seidler. Oxford, UK: EOLSS Publishers | UNESCO, 2009, pp. 21–40.

[144] James J Heckman and Tim Kautz. “Hard evidence on soft skills”. In: *Labour economics* 19.4 (2012), pp. 451–464.

[145] Steven Henderson and Steven Feiner. “Exploring the benefits of augmented reality documentation for maintenance and repair”. In: *IEEE transactions on visualization and computer graphics* 17.10 (2010), pp. 1355–1368.

[146] Emi Higashi et al. “Development and evaluation of a corrective feedback system using augmented reality for the high-quality cardiopulmonary resuscitation training”. In: *2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 2017.

[147] Luke Houghton and David Tuffley. “Towards a Methodology of Wicked Problem Exploration through Concept Shifting and Tension Point Analysis”. In: *Systems Research and Behavioral Science* 32.3 (2015), pp. 283–297. ISSN: 10991743. DOI: `10.1002/sres.2223`.

[148] María Blanca Ibáñez et al. “Experimenting with electromagnetism using augmented reality: Impact on flow student experience and educational effectiveness”. In: *Computers & Education* 71 (2014), pp. 1–13.

[149] María-Blanca Ibáñez and Carlos Delgado-Kloos. “Augmented reality for STEM learning: A systematic review”. In: *Computers & Education* 123 (2018), pp. 109–123.

[150] Andrew Irlitti et al. “Conveying spatial awareness cues in xR collaborations”. In: *IEEE transactions on visualization and computer graphics* 25.11 (2019), pp. 3178–3189.

[151] Petra Isenberg et al. “An exploratory study of co-located collaborative visual analytics around a tabletop display”. In: *VAST 10 - IEEE Conference on Visual Analytics Science and Technology 2010, Proceedings*. 2010, pp. 179–186. ISBN: 9781424494866. DOI: `10.1109/vast.2010.5652880`.

[152] Petra Isenberg et al. “An exploratory study of co-located collaborative visual analytics around a tabletop display”. In: *2010 ieee symposium on visual analytics science and technology*. IEEE. 2010, pp. 179–186.

[153] *ISO 9241-400 Ergonomics of human-system interaction – Part 400: Principles and requirements for physical input devices*. 2007.

[154] Michael J. Jacobson and Uri Wilensky. “Complex Systems in Education: Scientific and Educational Importance and Implications for the Learning Sciences”. In: *Journal of the Learning Sciences* 15.1 (2006), pp. 11–34. DOI: `10.1207/s15327809jls1501\_4`.

[155] Michael J. Jacobson and Uri Wilensky. “Complex systems in education: Scientific and educational importance and implications for the learning sciences”. In: *Journal of the Learning Sciences* 15.1 (2006), pp. 11–34. ISSN: 10508406. DOI: `10.1207/s15327809jls1501\_4`.

[156] Shamila Janakiraman, Sunnie Lee Watson, and William R Watson. “Using Game-based Learning to Facilitate Attitude Change for Environmental Sustainability”. In: *Journal of Education for Sustainable Development* 12.2 (Sept. 2018), pp. 176–185. ISSN: 0973-4082. DOI: `10.1177/0973408218783286`.

[157] Joy Jansen. “Piaget’s Cognitive Development Theory”. In: *Encyclopedia of Child Behavior and Development*. Springer US, 2011, pp. 1104–1106. DOI: `10.1007/978-0-387-79061-9\_2164`. URL: `https://doi.org/10.1007/978-0-387-79061-9%5C_2164`.

[158] Daniela Janßen et al. “Towards measuring user experience, activation and task performance in immersive virtual learning environments for students”. In: *International Conference on Immersive Learning*. Springer. 2016, pp. 45–58.

[159] Charlene Jennett et al. “Measuring and defining the experience of immersion in games”. In: *International journal of human-computer studies* 66.9 (2008), pp. 641–661.

[160] Robert Johansen. *Groupware: Computer support for business teams*. The Free Press, 1988.

[161] Eric Johnson. *Boeing Says Augmented Reality Can Make Workers Better, Faster*. 2015.

[162] Scott D Johnson, Steven R Aragon, and Najmuddin Shaik. “Comparative analysis of learner satisfaction and learning outcomes in online and face-to-face learning environments”. In: *Journal of interactive learning research* 11.1 (2000), pp. 29–49.

[163] David H. Jonassen and Lucia Rohrer-Murphy. “Activity theory as a framework for designing constructivist learning environments”. In: *Educational Technology Research & Development* 47.1 (1999), pp. 61–79. ISSN: 10421629. DOI: `10.1007/bf02299477`. URL: `http://www.springerlink.com/index/10.1007/BF02299477`.

[164] Peter H. Jones. “Systemic design principles for complex social systems”. In: *Social systems and design*. Ed. by Gary S. Metcalf. Springer. Tokio: Springer, 2014, pp. 91–128. ISBN: 978-4-431-54478-4. DOI: 10.1007/978-4-431-54478-4. URL: http://link.springer.com/chapter/10.1007/978-4-431-54478-4%5C_4.

[165] Yasmin B Kafai and Quinn Burke. “Constructionist gaming: Understanding the benefits of making games for learning”. In: *Educational psychologist* 50.4 (2015), pp. 313–334.

[166] R. S. Kalawsky. “VRUSE – A computerised diagnostic tool for usability evaluation of virtual/synthetic environment systems”. In: *Applied Ergonomics* 30.1 (1999), pp. 11–25.

[167] Tuomas Kantonen, Charles Woodward, and Neil Katz. “Mixed reality in virtual world teleconferencing”. In: *2010 IEEE Virtual Reality Conference (VR)*. IEEE. 2010, pp. 179–182.

[168] Korina Katsaliaki and Navonil Mustafee. “Edutainment for Sustainable Development: A Survey of Games in the Field”. In: *Simulation and Gaming* 46.6 (2015), pp. 647–672. ISSN: 1552826x. DOI: 10.1177/1046878114552166.

[169] Clinton Keith. *Agile game development with Scrum*. Pearson Education, 2010.

[170] Sebastian Kelle, Roland Klemke, and Marcus Specht. “Effects of Game Design Patterns on Basic Life Support Training Content”. In: *Educational Technology & Society* 1.16 (2013), pp. 275–285.

[171] J. M. Keller. “Development and use of the ARCS model of instructional design”. In: *Journal of Instr. Dev.* 10.3 (1987), pp. 2–10.

[172] Kangsoo Kim et al. “Revisiting Trends in Augmented Reality Research: A Review of the 2nd Decade of ISMAR (2008–2017)”. In: *IEEE Transactions on Visualization and Computer Graphics* 24.11 (Nov. 2018), pp. 2947–2962. DOI: 10.1109/tvcg.2018.2868591. URL: https://doi.org/10.1109/tvcg.2018.2868591.

[173] Seungwon Kim et al. “Evaluating the combination of visual communication cues for HMD-based mixed reality remote collaboration”. In: *Proceedings of the 2019 CHI conference on human factors in computing systems*. 2019, pp. 1–13.

[174] Seungwon Kim et al. “Improving co-presence with augmented visual communication cues for sharing experience through video conference”. In: *2014 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. Ieee. 2014, pp. 83–92.

[175] Michael Kipp et al. “Single-Person and Multi-Party 3D Visualizations for Nonverbal Communication Analysis”. In: *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014, Reykjavik, Iceland, May 26-31, 2014*. Ed. by Nicoletta Calzolari et al. European Language Resources Association (ELRA), 2014, pp. 3393–3397. URL: http://www.lrec-conf.org/proceedings/lrec2014/summaries/160.html.

[176] Eric Klopfer, Scot Osterweil, Katie Salen, et al. “Moving learning games forward”. In: *Cambridge, MA: The Education Arcade* (2009).

[177] Jan Kolkmeier et al. “With a little help from a holographic friend: The openimpress mixed reality telepresence toolkit for remote collaboration systems”. In: *Proceedings of the 24th ACM Symposium on Virtual Reality Software and Technology*. 2018, pp. 1–11.

[178] Raph Koster. *Theory of fun for game design*. " O’Reilly Media, Inc.", 2013.

[179] Aaron Kotranza et al. “Real-time in-situ visual feedback of task performance in mixed environments for learning joint psychomotor-cognitive tasks”. In: *2009 8th IEEE International Symposium on Mixed and Augmented Reality*. Ieee. 2009, pp. 125–134.

[180] George Koutromanos, Alivisos Sofos, and Lucy Avraamidou. “The use of augmented reality games in education: a review of the literature”. In: *Educational Media International* 52.4 (2015), pp. 253–271.

[181] Birgul Kutlu, Aysun Bozanta, and Nuket Nowlan. “Multi-User Virtual Environments and Serious Games for Team Building in Organizations”. In: *Proceedings of the Sixth International Conference on E-Learning in the Workplace (ICELW 2013). New York, USA*. 2013.

[182] Yeram Kwon et al. “HeartiSense: A Novel Approach to Enable Effective Basic Life Support Training Without an Instructor”. In: *Chi ’14*. 2014, pp. 431–434. ISBN: 978-1-4503-2474-8.

[183] Fedwa Laamarti, Mohamad Eid, and Abdulmotaleb El Saddik. “An overview of serious games”. In: *International Journal of Computer Games Technology* 2014 (2014).

[184] Teemu H Laine. “Mobile educational augmented reality games: a systematic literature review and two case studies”. In: *Computers* 7.1 (2018), p. 19.

[185] Teemu H Laine et al. “Science Spots AR: a platform for science learning games with augmented reality”. In: *Educational Technology Research and Development* 64.3 (2016), pp. 507–531.

[186] Joey J Lee and Jessica Hammer. “Gamification in education: What, how, why bother?” In: *Academic exchange quarterly* 15.2 (2011), p. 146.

[187] Kangdon Lee. “Augmented reality in education and training”. In: *TechTrends* 56.2 (2012), pp. 13–21.

[188] Jingya Li et al. “Augmented reality games for learning: A literature review”. In: *International Conference on Distributed, Ambient, and Pervasive Interactions*. Springer. 2017, pp. 612–626.

[189] Fotis Liarokapis and Sara de Freitas. “A Case Study of Augmented Reality Serious Games”. In: *Looking Toward the Future of Technology-Enhanced Education: Ubiquitous Learning and the Digital Native*. Ed. by Martin Ebner and Mandy Schiefner. Hershey, PA: IGI Global, 2010, pp. 178–191. ISBN: 9781615206780. DOI: `10.4018/978-1-61520-678-0.ch010`.

[190] S Jay Liebowitz and Kenneth P De Meuse. “The application of team building”. In: *Human Relations* 35.1 (1982), pp. 1–18.

[191] Heide Lukosch et al. “Building a virtual world for team work improvement”. In: *Frontiers in Gaming Simulation*. Springer, 2014, pp. 60–68.

[192] Heide K. Lukosch et al. “A Scientific Foundation of Simulation Games for the Analysis and Design of Complex Systems”. In: *Simulation & Gaming* 49.3 (May 2018), pp. 279–314. DOI: `10.1177/1046878118768858`. URL: `https://doi.org/10.1177/1046878118768858`.

[193] Stephan Lukosch et al. “Collaboration in augmented reality”. In: *Computer Supported Cooperative Work (CSCW)* 24.6 (2015), pp. 515–525.

[194] Minhua Ma, Andreas Oikonomou, and Lakhmi C Jain. *Serious games and edutainment applications*. Vol. 504. Springer, 2011.

[195] Kaveh Madani, Tyler W. Pierce, and Ali Mirchi. “Serious games on environmental management”. In: *Sustainable Cities and Society* 29 (2017), pp. 1–11. ISSN: 22106707. DOI: `10.1016/j.scs.2016.11.007`.

[196] G Elisabeta Marai, Angus G Forbes, and Andrew Johnson. “Interdisciplinary immersive analytics at the electronic visualization laboratory: Lessons learned and upcoming challenges”. In: *2016 Workshop on Immersive Analytics (IA)*. Ieee. 2016, pp. 54–59.

[197] David Jerner Martin and Kimberly S Loomis. *Building teachers: A constructivist approach to introducing education*. Cengage Learning, 2013.

[198] Sergio Martin et al. “New technology trends in education: Seven years of forecasts and convergence”. In: *Computers & Education* 57.3 (2011), pp. 1893–1906.

[199] Jorge Martín-Gutiérrez et al. “Design and validation of an augmented book for spatial abilities development in engineering students”. In: *Computers & Graphics* 34.1 (2010), pp. 77–91.

[200] Roberto Martinez-Maldonado et al. "Capturing and analyzing verbal and physical collaborative learning interactions at an enriched interactive tabletop". In: *International Journal of Computer-Supported Collaborative Learning* (2013). ISSN: 15561607. DOI: 10.1007/s11412-013-9184-1.

[201] Tristan McCowan. *Higher Education for and beyond the Sustainable Development Goals*. Cham, Switzerland: Palgrave Macmillan, 2019. ISBN: 978-3-030-19596-0.

[202] Reuben R. McDaniel and Dean J. Driebe. "Uncertainty and Surprise: An Introduction". In: *Uncertainty and Surprise in Complex Systems*. Ed. by Reuben R. McDaniel and Dean J. Driebe. Heidelberg: Springer, 2006, pp. 3–11. DOI: 10.1007/10948637_1.

[203] Angela McFarlane, Anne Sparrowhawk, Ysanne Heald, et al. *Report on the educational use of games*. TEEM (Teachers evaluating educational multimedia), Cambridge, 2002.

[204] Joseph Edward McGrath. *Social psychology: A brief introduction*. Holt, Rinehart and Winston, 1964.

[205] Gabriele Meiselwitz and W Sadera. "Investigating the connection between usability and learning outcomes in online learning environments". In: *Journal of Online Learning and Teaching* 4.2 (2008), pp. 234–42.

[206] David R Michael and Sandra L Chen. *Serious games: Games that educate, train, and inform*. Muska & Lipman/Premier-Trade, 2005.

[207] Alice Mitchell and Carol Savill-Smith. "The use of computer and video games for learning". In: *A review of the literature* (2004).

[208] Roxana Moreno and Richard E Mayer. "Learning science in virtual reality multimedia environments: Role of methods and media." In: *Journal of educational psychology* 94.3 (2002), p. 598.

[209] Gareth Morgan and Jean Adams. "Pedagogy first! making web-technologies work for soft skills development in leadership and management education". In: *J. of Int. Learning Res.* 20.2 (2009), pp. 129–155.

[210] José Miguel Mota et al. "Augmented reality mobile app development for all". In: *Computers & Electrical Engineering* 65 (2018), pp. 250–260.

[211] Jens Müller, Roman Rädle, and Harald Reiterer. "Remote Collaboration With Mixed Reality Displays". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, May 2017. DOI: 10.1145/3025453.3025717. URL: https://doi.org/10.1145/3025453.3025717.

[212] Hannes Münchow and Maria Bannert. "Feeling good, learning better? Effectivity of an emotional design procedure in multimedia learning". In: *Educational Psychology* 39.4 (2019), pp. 530–549.

[213] William C. Mundell et al. “Simulation technology for resuscitation training: A systematic review and meta-analysis”. In: *Resuscitation* 84.9 (2013), pp. 1174–1183.

[214] Juan A Munoz-Cristobal et al. “Supporting teacher orchestration in ubiquitous learning environments: A study in primary education”. In: *IEEE Transactions on Learning Technologies* 8.1 (2014), pp. 83–97.

[215] Mathieu Nancel et al. “High-precision pointing on large wall displays using small handheld devices”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2013, pp. 831–840.

[216] Magy Seif El-Nasr, Anders Drachen, and Alessandro Canossa. *Game analytics*. Springer, 2016.

[217] Diep Nguyen and Gerrit Meixner. “A survey of gamified Augmented Reality systems for procedural tasks in industrial settings”. In: *IFAC-PapersOnLine* 53.2 (2020), pp. 10096–10100.

[218] Scott Nicholson. “Peeking behind the locked door: A survey of escape room facilities”. In: (2015).

[219] Birgitte Lund Nielsen, Harald Brandt, and Hakon Swensen. “Augmented Reality in science education–affordances for student learning”. In: *NorDiNa* 12.2 (2016), pp. 157–174.

[220] Jakob Nielsen and Rolf Molich. “Heuristic evaluation of user interfaces”. In: *SIGCHI Conference on Human Factors in Computing Systems*. Academic Press, 1990, pp. 249–256.

[221] Niels Christian Nilsson et al. “Natural walking in virtual reality: A review”. In: *Computers in Entertainment (CIE)* 16.2 (2018), p. 8.

[222] Katsuo A Nishikawa and Joseph Jaeger. “A computer simulation comparing the incentive structures of dictatorships and democracies”. In: *Journal of Political Science Education* 7.2 (2011), pp. 135–142.

[223] Shuo Niu, D Scott McCrickard, and Steve Harrison. “An observational study of simultaneous and sequential interactions in co-located collaboration”. In: *IFIP Conference on Human-Computer Interaction*. Springer. 2017, pp. 163–183.

[224] Michele Notari, Adrian Baumgartner, and Walter Herzog. “Social skills as predictors of communication, performance and quality of collaboration in project-based learning”. In: *Journal of Computer Assisted Learning* 30.2 (2014), pp. 132–147.

[225] Manuela Núñez et al. “Collaborative augmented reality for inorganic chemistry education”. In: *WSEAS international conference. Proceedings. Mathematics and computers in science and engineering*. Vol. 5. Wseas. 2008, pp. 271–277.

[226] J. O’Flaherty and M. Liddy. “The impact of development education and education for sustainable development interventions: a synthesis of the research”. In: *Environmental Education Research* 24.7 (2018), pp. 1031–1049. ISSN: 14695871. DOI: `10.1080/13504622.2017.1392484`.

[227] Diana G Oblinger. “Games and learning.” In: *Educause quarterly* 29.3 (2006), pp. 5–7.

[228] OECD. *Education at a Glance 2021*. 2021. URL: `https://www.oecd-ilibrary.org/content/publication/b35a14e5-en`.

[229] Leif Oppermann and Michaela Slussareff. “Pervasive games”. In: *Entertainment computing and serious games*. Springer, 2016, pp. 475–520.

[230] Ioannis Paliokas and Stella Sylaiou. “The use of serious games in museum visits and exhibitions: A systematic mapping study”. In: *2016 8th International Conference on games and virtual worlds for serious applications (VS-GAMES)*. IEEE. 2016, pp. 1–8.

[231] Seymour Papert and Idit Harel. “Situating constructionism”. In: *Constructionism* 36.2 (1991), pp. 1–11.

[232] Lena Pareto et al. “A teachable-agent-based game affording collaboration and competition: Evaluating math comprehension and motivation”. In: *Educational Technology Research and Development* 60.5 (2012), pp. 723–751.

[233] Nohyoung Park et al. “Projected AR-Based Interactive CPR Simulator”. In: *Virtual, Augmented and Mixed Reality. Systems and Applications*. Ed. by Randall Shumaker. Berlin, 2013, pp. 83–89. ISBN: 978-3-642-39420-1.

[234] Nikolaos Pellas et al. “Augmenting the learning experience in primary and secondary school education: A systematic review of recent trends in augmented reality game-based learning”. In: *Virtual Reality* 23.4 (2019), pp. 329–346.

[235] Marco Aurelio Pellens, Marcelo Da Silva Hounsell, and André Tavares da Silva. “Augmented reality and serious games: A systematic literature mapping”. In: *2017 19th Symposium on Virtual and Augmented Reality (SVR)*. Ieee. 2017, pp. 227–235.

[236] Gonçalo Pereira et al. “Serious games for personal and social learning & ethics: status and trends”. In: *Procedia Computer Science* 15 (2012), pp. 53–65.

[237] Vasco Pereira et al. “Extended reality framework for remote collaborative interactions in virtual environments”. In: *2019 International Conference on Graphics and Interaction (ICGI)*. IEEE. 2019, pp. 17–24.

[238] Catlin Pidel and Philipp Ackermann. “Collaboration in Virtual and Augmented Reality: A Systematic Overview”. In: *Lecture Notes in Computer Science*. Springer International Publishing, 2020, pp. 141–156. DOI: `10.1007/978-3-030-58465-8_10`. URL: `https://doi.org/10.1007/978-3-030-58465-8_10`.

[239] Marcio S Pinho, Doug A Bowman, and Carla M Dal Sasso Freitas. “Cooperative object manipulation in collaborative virtual environments”. In: *Journal of the Brazilian Computer Society* 14.2 (2008), pp. 53–67.

[240] Thammathip Piumsomboon et al. “Exploring enhancements for remote mixed reality collaboration”. In: *SIGGRAPH Asia 2017 Mobile Graphics & Interactive Applications*. 2017, pp. 1–5.

[241] Thammathip Piumsomboon et al. “The Effects of Sharing Awareness Cues in Collaborative Mixed Reality”. In: *Frontiers in Robotics and AI* 6 (Feb. 2019). DOI: `10.3389/frobt.2019.00005`. URL: `https://doi.org/10.3389/frobt.2019.00005`.

[242] Polycular. *ÖkoGotschi goes School*. `http://oekogotschi.at/`. Accessed: 2020-04-30.

[243] Marc Prensky. *Digital game-based learning*. McGraw Hill, 2001.

[244] Marc Prensky. “Digital natives, digital immigrants part 2: Do they really think differently?” In: *On the horizon* (2001).

[245] F. Pretto et al. “Experiences using Augmented Reality Environment for training and evaluating medical students”. In: *2013 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*. 2013, pp. 1–4.

[246] Iulian Radu. “Augmented reality in education: a meta-review and cross-media analysis”. In: *Personal and Ubiquitous Computing* 18.6 (2014), pp. 1533–1543.

[247] Iulian Radu et al. “A Survey of Needs and Features for Augmented Reality Collaborations in Collocated Spaces”. In: *Proceedings of the ACM on Human-Computer Interaction* 5.CSCW1 (Apr. 2021), pp. 1–21. DOI: `10.1145/3449243`. URL: `https://doi.org/10.1145/3449243`.

[248] R Abindra Ratan and Ute Ritterfeld. “Classifying serious games”. In: *Serious games*. Routledge, 2009, pp. 32–46.

[249] Lisa Rebenitsch and Charles Owen. "Review on cybersickness in applications and visual displays". In: *Virtual Reality* 20.2 (June 2016), pp. 101–125. ISSN: 1434-9957. DOI: `10.1007/s10055-016-0285-9`. URL: `https://doi.org/10.1007/s10055-016-0285-9`.

[250] Jun Rekimoto. "Matrix: A realtime object identification and registration method for augmented reality". In: *Proceedings. 3rd Asia Pacific Computer Human Interaction (Cat. No. 98EX110)*. Ieee. 1998, pp. 63–68.

[251] Christian Reuter et al. "Game Design Patterns for Collaborative Player Interactions." In: *DiGRA*. 2014.

[252] Paula Ceccon Ribeiro, Bruno Baere Pederassi Lomba de Araujo, and Alberto Raposo. "ComFiM: a cooperative serious game to encourage the development of communicative skills between children with autism". In: *Brazilian Symposium on Computer Games and Digital Entertainment*. 2014, pp. 148–157.

[253] Daniel C Richardson and Rick Dale. "Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension". In: *Cognitive science* 29.6 (2005), pp. 1045–1060.

[254] Ganit Richter, Daphne R Raban, and Sheizaf Rafaeli. "Studying gamification: the effect of rewards and incentives on motivation". In: *Gamification in education and business*. Springer, 2015, pp. 21–46.

[255] Lloyd P Rieber. "Seriously considering play: Designing interactive learning environments based on the blending of microworlds, simulations, and games". In: *Educational technology research and development* 44.2 (1996), pp. 43–58.

[256] Horst W. J. Rittel and Melvin M. Webber. "Dilemmas in a general theory of planning". In: *Policy Sciences* 4.2 (1973), pp. 155–169. DOI: `10.1007/bf01405730`.

[257] George Ritter et al. "The effect of bystander CPR on survival of out-of-hospital cardiac arrest victims". In: *American Heart Journal* 110.5 (1985), pp. 932–937. ISSN: 0002-8703.

[258] José Bernardo Rocha, Samuel Mascarenhas, and Rui Prada. "Game mechanics for cooperative games". In: *ZON Digital Games 2008* (2008), pp. 72–80.

[259] Geoffrey Rockwell and Kevin Kee. "The leisure of serious games: A dialogue". In: (2011).

[260] Danilo Gasques Rodrigues, Janet G Johnson, and Nadir Weibel. "Real-time guidance for cardiopulmonary resuscitation in Mixed Reality". In: 2018.

[261] Sylvia Rojas-Drummond and Neil Mercer. “Scaffolding the development of effective collaboration and learning”. In: *International journal of educational research* 39.1-2 (2003), pp. 99–111.

[262] Jannick P Rolland, Richard L Holloway, and Henry Fuchs. “Comparison of optical and video see-through, head-mounted displays”. In: *Telemanipulator and Telepresence Technologies*. Vol. 2351. International Society for Optics and Photonics. 1995, pp. 293–307.

[263] Margarida Romero, Mireia Usart, and Michela Ott. “Can serious games contribute to developing and sustaining 21st century skills?” In: *Games and Culture* 10.2 (2015), pp. 148–177.

[264] Jeremy Roschelle and Stephanie D Teasley. “The construction of shared knowledge in collaborative problem solving”. In: *Computer supported collaborative learning*. Springer. 1995, pp. 69–97.

[265] Helen Routledge. *Why games are good for business: How to leverage the power of serious games, gamification and simulations*. Springer, 2016.

[266] Katie Salen and Eric Zimmerman. *Rules of Play: Game Design Fundamentals*. Cambridge, MA: MIT Press, 2003. ISBN: 0262240459.

[267] Fariha Hayat Salman and David R Riley. “Augmented reality crossover gamified design for sustainable engineering education”. In: *2016 Future Technologies Conference (FTC)*. 2016, pp. 1353–1356. DOI: `10.1109/ftc.2016.7821781`.

[268] Alysson Diniz dos Santos, Francesco Strada, and Andrea Bottino. “Investigating the Design and Evaluation of Educational Games Under the Perspective of Player Experience”. In: *International Conference on Games and Learning Alliance*. Springer. 2017, pp. 218–227.

[269] Marc Ericson C Santos et al. “A usability scale for handheld augmented reality”. In: *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*. 2014, pp. 167–176.

[270] Marc Ericson C Santos et al. “Augmented reality as multimedia: the case for situated vocabulary learning”. In: *Research and Practice in Technology Enhanced Learning* 11.1 (2016), pp. 1–23.

[271] Marc Ericson C Santos et al. “Augmented reality learning experiences: Survey of prototype design and evaluation”. In: *IEEE Transactions on learning technologies* 7.1 (2013), pp. 38–56.

[272] Birgit Schmitz, Marcus Specht, and Roland Klemke. “An Analysis of the Educational Potential of Augmented Reality Games for Learning.” In: *mLearn*. 2012, pp. 140–147.

[273] Bertrand Schneider et al. “Leveraging mobile eye-trackers to capture joint visual attention in co-located collaborative learning groups”. In: *International Journal of Computer-Supported Collaborative Learning* 13.3 (2018), pp. 241–261.

[274] Felix D Schönbrodt and Jens B Asendorpf. “Virtual social environments as a tool for psychological assessment: Dynamics of interaction with a virtual spouse.” In: *Psychological assessment* 23.1 (2011), p. 7.

[275] Martijn J Schuemie et al. “Research on presence in virtual reality: A survey”. In: *CyberPsychology & Behavior* 4.2 (2001), pp. 183–201.

[276] Natalie Sebanz and Guenther Knoblich. “Prediction in joint action: What, when, and where”. In: *Topics in Cognitive Science* 1.2 (2009), pp. 353–367.

[277] Federico Semeraro et al. “Kids (learn how to) save lives in the school with the serious game Relive”. In: *Resuscitation* 116 (2017), pp. 27–32.

[278] Federico Semeraro et al. “Motion detection technology as a tool for cardiopulmonary resuscitation (CPR) quality training: A randomised crossover mannequin pilot study”. In: *Resuscitation* 84.4 (2013), pp. 501–507. ISSN: 0300-9572.

[279] Mickael Sereno et al. “Collaborative Work in Augmented Reality: A Survey”. In: *IEEE Transactions on Visualization and Computer Graphics* (2020).

[280] Ángel Serrano-Laguna et al. “A methodology for assessing the effectiveness of serious games and for inferring player learning outcomes”. In: *Multimedia Tools and applications* 77.2 (2018), pp. 2849–2871.

[281] Robert A Simon, Gary N Grubb, and Dennis K Leedom. *Validation of Crew Coordination Training and Evaluation Methods for Army Aviation.* Tech. rep. Dynamics Research Corp. Wilmington MA, 1995.

[282] Mustafa Sirakaya and Ebru Kilic Cakmak. “Effects of augmented reality on student achievement and self-efficacy in vocational education and training”. In: *International journal for research in vocational education and training* 5.1 (2018), pp. 1–18.

[283] E. Skura. “Pros and cons of first aid training?” In: *Canadian Medical Association Journal* 182.12 (2010), E549–e550.

[284] Rasa Smaliukiene. “Stakeholders’ impact on the environmental responsibility: Model design and testing”. In: *Journal of Business Economics and Management* 8.3 (2007), pp. 213–223. DOI: `10.1080/16111699.2007.9636171`. eprint: `https://www.tandfonline.com/doi/pdf/10.1080/16111699.2007.9636171`. URL: `https://www.tandfonline.com/doi/abs/10.1080/16111699.2007.9636171`.

[285] Shamus P Smith et al. “Improving undergraduate soft skills using m-learning and serious games”. In: *Int. Conf. on Teaching, Assessment, and Learning for Engineering*. 2015, pp. 230–235.

[286] Kurt Squire and Henry Jenkins. “Harnessing the power of games in education”. In: *Insight* 3.1 (2003), pp. 5–33.

[287] Marios Stanitsas, Konstantinos Kirytopoulos, and Elise Vareilles. “Facilitating sustainability transition through serious games: A systematic literature review”. In: *Journal of Cleaner Production* 208 (2019), pp. 924–936. ISSN: 09596526. DOI: 10.1016/j.jclepro.2018.10.157. URL: https://hal-mines-albi.archives-ouvertes.fr/hal-01918890/document.

[288] Linda Steg et al. “An integrated framework for encouraging pro-environmental behaviour: The role of values, situational factors and goals”. In: *Journal of Environmental psychology* 38 (2014), pp. 104–115.

[289] Neal Stephenson. *Snow crash*. Bragelonne, 2014.

[290] John D. Sterman. “Learning in and about complex systems”. In: *System Dynamics Review* 10.2-3 (1994), pp. 291–330. DOI: 10.1002/sdr.4260100214. URL: https://doi.org/10.1002/sdr.4260100214.

[291] Stone Meadow Development LLC. *Resuscitate! CPR AED & Choking*. 2010. URL: https://itunes.apple.com/us/app/resuscitate-cpr-aed-choking/id363393502?mt=8.

[292] Tarja Susi, Mikael Johannesson, and Per Backlund. “Serious games: An overview”. In: (2007).

[293] Ivan E Sutherland. “A head-mounted three dimensional display”. In: *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. 1968, pp. 757–764.

[294] Theresa Jean Tanenbaum, Alissa N Antle, and John Robinson. “Three perspectives on behavior change for serious games”. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2013, pp. 3389–3392.

[295] Anthony Tang et al. “Collaborative coupling over tabletop displays”. In: *Proceedings of the SIGCHI conference on Human Factors in computing systems*. 2006, pp. 1181–1190.

[296] Theophilus Teo et al. “Investigating the use of Different Visual Cues to Improve Social Presence within a 360 Mixed Reality Remote Collaboration”. In: *The 17th International Conference on Virtual-Reality Continuum and its Applications in Industry*. ACM, Nov. 2019. DOI: 10.1145/3359997.3365687. URL: https://doi.org/10.1145/3359997.3365687.

[297] Daniella Tilbury. “Rising to the challenge: Education for sustainability in Australia”. In: *Australian Journal of Environmental Education* 20.02 (2004), pp. 103–114.

[298] Daniella Tilbury and David Wortman. *Engaging People in Sustainability.* Gland, Switzerland: Iucn, 2004. ISBN: 2831708230.

[299] Adam Timmis and al. “European Society of Cardiology: Cardiovascular Disease Statistics 2017”. In: *European Heart Journal* 39.7 (2018), pp. 508–579.

[300] Hendrys Fabián Tobar Muñoz et al. “Supporting technology for augmented reality game-based learning”. In: (2017).

[301] Sigmund Tobias et al. “Review of research on computer games.” In: (2011).

[302] Undpi. *Agenda 21: Programme of action for sustainable development, Rio Declaration on Environment and Development, statement of forest principles.* New York, NY: United Nations Department of Public Information, 1994, p. 294. ISBN: 92-1-100509-4. URL: `https://sustainabledevelopment.un.org/outcomedocuments/agenda21`.

[303] Unesco. *United Nations Decade of Education for Sustainable Development (2005-2014): International implementation scheme.* Paris, France: Unesco, 2005.

[304] Richard Van Eck. “Digital game-based learning: It’s not just the digital natives who are restless”. In: *EDUCAUSE review* 41.2 (2006), p. 16.

[305] Sonya E Van Nuland and Kem A Rogers. “The anatomy of E-Learning tools: Does software usability influence learning outcomes?” In: *Anatomical sciences education* 9.4 (2016), pp. 378–390.

[306] Jani Väyrynen et al. “Exploring Head Mounted Display based Augmented Reality for Factory Workers”. In: *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia.* ACM, Nov. 2018. DOI: `10.1145/3282894.3289745`. URL: `https://doi.org/10.1145/3282894.3289745`.

[307] Francisco del Cerro Velázquez and Ginés Morales Méndez. “Augmented reality and mobile devices: A binominal methodological resource for inclusive education (SDG 4). an example in secondary education”. In: *Sustainability (Switzerland)* 10.10 (2018), pp. 1–14. ISSN: 20711050. DOI: `10.3390/su10103446`.

[308] Alice Veldkamp et al. “Escape education: A systematic review on escape rooms in education”. In: (2020).

[309] Marcia L Verduin et al. “Computer simulation games as an adjunct for treatment in male veterans with alcohol use disorder”. In: *Journal of substance abuse treatment* 44.3 (2013), pp. 316–322.

[310] Joost M. Vervoort et al. “Stepping into futures: Exploring the potential of interactive media for participatory scenarios on social-ecological systems”. In: *Futures* 42.6 (2010). Special issue: futures for multiple civilizations, pp. 604–616. ISSN: 0016-3287. DOI: `http://dx.doi.org/10.1016/j.futures.2010.04.031`.

[311] Cordula Vesper et al. “A minimal architecture for joint action”. In: *Neural Networks* 23.8-9 (2010), pp. 998–1003.

[312] Benjamin Volmer et al. “A comparison of predictive spatial augmented reality cues for procedural tasks”. In: *IEEE transactions on visualization and computer graphics* 24.11 (2018), pp. 2846–2856.

[313] Iro Voulgari and Vassilis Komis. “Massively multi-user online games: The emergence of effective collaborative activities for learning”. In: *2008 Second IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning*. IEEE. 2008, pp. 132–134.

[314] Wiebe de Vries et al. “Self-training in the use of automated external defibrillators: The same results for less money”. In: *Resuscitation* 76.1 (2008), pp. 76–82. ISSN: 0300-9572.

[315] Daniel Wagner and Dieter Schmalstieg. “First steps towards handheld augmented reality”. In: *Seventh IEEE International Symposium on Wearable Computers, 2003. Proceedings.* IEEE Computer Society. 2003, pp. 127–127.

[316] Xiangyu Wang and Phillip S Dunston. “Comparative effectiveness of mixed reality-based virtual environments in collaborative design”. In: *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)* 41.3 (2011), pp. 284–296.

[317] Harald Warmelink et al. “AMELIO: Evaluating the Team-Building Potential of a Mixed Reality Escape Room Game”. In: *Extended Abstracts Publication of the Annual Symposium on Computer-Human Interaction in Play.* CHI PLAY ’17 Extended Abstracts. Amsterdam, The Netherlands: Association for Computing Machinery, 2017, pp. 111–123. ISBN: 9781450351119.

[318] Harald Warmelink et al. “AMELIO: Evaluating the team-building potential of a mixed reality escape room game”. In: *Extended abstracts publication of the annual symposium on computer-human interaction in play.* 2017, pp. 111–123.

[319] Voravika Wattanasoontorn et al. “A Kinect-Based System for Cardiopulmonary Resuscitation Simulation: A Pilot Study”. In: *Serious Games Development and Applications.* Ed. by Minhua Ma et al. Berlin, 2013, pp. 51–63. ISBN: 978-3-642-40790-1.

[320] Wced. *Our Common Future.* Oxford, England: Oxford University Press, 1987. ISBN: 019282080x.

[321] Thomas Wells and Steven Houben. "CollabAR–Investigating the Mediating Role of Mobile AR Interfaces on Co-Located Group Collaboration". In: *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 2020, pp. 1–13.

[322] Josef Wiemeyer, Lennart Nacke, Christiane Moser, et al. "Player experience". In: *Serious games*. Springer, 2016, pp. 243–271.

[323] Phillip Wilkinson. "Affective Educational Games: Utilizing Emotions in Game-Based Learning". In: *Games and Virtual Worlds for Serious Applications (VS-GAMES), 2013 5th International Conference on Games and Virtual Worlds for Serious Applications*. Bournemouth University. Ieee, Sept. 2013, pp. 1–8. DOI: 10.1109/vs-games.2013.6624219.

[324] Pieter Wouters, Herre van Oostendorp, and Erik D. van der Spek. "Game Design: The mapping of Cognitive Task Analysis and Game Discourse Analysis in creating effective and entertaining serious games". In: *Proceedings of the ECCE'10, 28th Annual European Conference on Cognitive Ergonomics*. August. Delft: ACM Press, Aug. 2010, pp. 287–293. ISBN: 9781605589466. DOI: 10.1145/1962300.1962358. URL: http://dl.acm.org/citation.cfm?id=1962300.1962358.

[325] Hsin-Kai Wu et al. "Current status, opportunities and challenges of augmented reality in education". In: *Computers & education* 62 (2013), pp. 41–49.

[326] Jason S Wu and Joey J Lee. "Climate change games as tools for education and engagement". In: *Nature Climate Change* 5.5 (2015), pp. 413–418. ISSN: 17586798. DOI: 10.1038/nclimate2566. URL: http://www.climateinteractive.org.

[327] Joyce Yeung et al. "AED training and its impact on skill acquisition, retention and performance – A systematic review of alternative training methods". In: *Resuscitation* 82.6 (2011), pp. 657–664. ISSN: 0300-9572.

[328] Steve Chi-Yin Yuen, Gallayanee Yaoyuneyong, and Erik Johnson. "Augmented reality: An overview and five directions for AR in education". In: *Journal of Educational Technology Development and Exchange (JETDE)* 4.1 (2011), p. 11.

[329] Jose P. Zagal, Rick Jochen, and Idris Hsi. "Collaborative games: Lessons learned from board games". In: *Simulation & Gaming* 37.1 (2006), pp. 24–40. ISSN: 1046-8781. DOI: 10.1177/1046878105282279.

[330] Jia Zhang et al. "The development and evaluation of an augmented reality-based armillary sphere for astronomical observation instruction". In: *Computers & education* 73 (2014), pp. 178–188.

[331] Michael Zyda. “From visual simulation to virtual reality to games”. In: *Computer* 38.9 (2005), pp. 25–32.

This Ph.D. thesis has been typeset by means of the TEX-system facilities. The typesetting engine was pdfLATEX. The document class was `toptesi`, by Claudio Beccari, with option `tipotesi=scudo`. This class is available in every up-to-date and complete TEX-system installation.