Doctoral thesis

Doctoral theses at NTNU, 2021:214

Rabail Tahir

Game-based learning design and evaluation: Towards better understanding and improvement

NTNU
Norwegian University of Science and Technology
Thesis for the Degree of Philosophiae Doctor
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

Rabail Tahir

Game-based learning design and evaluation: Towards better understanding and improvement

Thesis for the Degree of Philosophiae Doctor

Trondheim, June 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Computer Science

NTNU
Norwegian University of
Science and Technology

NTNU
Norwegian University of Science and Technology

Thesis for the Degree of Philosophiae Doctor

Faculty of Information Technology and Electrical Engineering
Department of Computer Science

© Rabail Tahir

ISBN 978-82-326-6522-8 (printed ver.)
ISBN 978-82-326-5979-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

Doctoral theses at NTNU, 2021:214

Printed by NTNU Grafisk senter

‘What is not started will never get finished’

Johann Wolfgang von Goethe

Dedication

To my mother, my greatest strength and my father, my greatest support.

You are the reason for all my success and accomplishments.

This research work is dedicated to my beloved amazing parents (Samina Kausor and Tahir Pervaiz) for their endless and unconditional love, sacrificial care, prayers, support, and showing confidence in me, without which it would have been impossible for me to pursue my studies up to this stage. They always understood me, provided me with the best they could and without complaint, and guided me in all life's tasks to the best of their ability. Their encouragement, faith, and motivation gave me the strength and confidence to pursue this PhD journey and made it possible for me to complete this work. Thank you for believing in me, for always being there when I need you, for giving me the strength to keep going, and for taking part in my journey. You are the reason for all my success and accomplishments.

Abstract

In recent years, Game-based learning (GBL) has proliferated rapidly and is widely used in various fields to improve learning. Educational games can actively engage learners and meet individual learning situations, making them effective educational tools for various learning contexts. The rapid increase in educational games makes it necessary to understand and improve the GBL design and evaluation process and analyze the various aspects that influence the GBL experience. The GBL research presents a wide range of diverse criteria for designing and evaluating educational games used inconsistently across literature, hindering a holistic view. It is important to conceptualize the GBL elements for comprehensive design, analysis, and evaluation. Educational game development is a complex process and introduces challenges related to balancing different GBL aspects that require collaboration among the heterogeneous team of experts. However, GBL research lacks operationalizable approaches for integrating GBL design knowledge into the educational game design process to facilitate collaboration and shared understanding among team members. Educational game researchers are interested in investigating the effectiveness of the GBL approach in different educational settings. Especially the recent Syrian war and COVID-19 crisis have emphasized the potential of educational technology such as learning games to support informal and non-formal learning for the marginalized population of refugee children and online education during the pandemic. Despite the growing research in GBL, there is still a considerable need to further and systematically investigate the effectiveness of learning games to provide empirical evidence on their appropriate design. Identifying the evaluation criteria is complex and time-consuming, making GBL evaluation a critical undertaking. Therefore, appropriate techniques, methods, and principles are required to conduct the evaluation research.

The research undertaken in this thesis investigates how the GBL design and evaluation process can be facilitated to improve the effectiveness of learning games. The research work considers the need for multi-dimensional focus with various interlinked aspects that need to be balanced to facilitate the learning process in GBL. Consequently, the focus of this doctoral work is to build a more comprehensive understanding of GBL by investigating factors that influence the learning experience with games in different domains and learning contexts. Moreover, the research work focuses on facilitating the GBL design team in collaboratively ideating effective design and facilitating stakeholders (such as designers, educators, researchers) to analyze learning games and plan the GBL evaluation studies. The research work also explored the potential of GBL in different learning contexts focusing on evaluating educational games in formal and online education (amid COVID-19) with students and informal learning setup (focusing refugee crisis) with refugee children. Moreover, since the current Human-Computer Interaction (HCI) and GBL research do not consider the peculiar, situated nature of GBL research with refugee children, this research work also focused on understanding methods and guidelines for conducting GBL research with this user group.

To this end, this doctoral research followed the design science research (DSR) methodology to develop a GBL framework and three tools/instruments for the design and evaluation of learning games, answering the identified problems. A mixed methods research design (exploratory-triangulation design) was used in connection with DSR in three cycles focusing on the GBL phenomenon, GBL design process and GBL evaluation process. This research comprises three review studies, the design of artifacts, three case studies, and six empirical studies (GBL evaluation studies and GBL design workshops). The research work focused on two primary activities of DSR, building and evaluating. First, the GBL framework was constructed by combining the existing knowledge then GBL evaluation studies were conducted to understand and validate the framework constructs. The GBL framework is put into practice by utilizing the framework components to design and implement three instruments/tools to design, analyze, and evaluate learning games. The evaluation results from design workshops and GBL evaluation studies (quasi-experiments) along with case studies were used to revise the tools/instruments, validate the framework concepts, and provide further implications/guidelines for improvement.

The resulting contributions provide substantive findings and identify challenges and issues regarding current GBL design and evaluation practices. A holistic conceptual GBL framework has been devised to bootstrap the design and evaluation process. The framework provides the key GBL elements in a detailed hierarchy focusing on a multi-dimensional approach. The research also investigates the GBL learning process and the contributing factors by employing educational games in different learning contexts. Moreover, this doctoral work includes developing and implementing a card-based toolkit for ideating educational game design, facilitating collaboration and completeness in the GBL ideation process. It also provides the knowledge for transforming a theoretical framework into a design toolkit. Furthermore, the GBL evaluation process is facilitated by devising an analysis instrument and an integrated evaluation approach to guide stakeholders. Finally, guidelines are presented for evaluating and designing effective learning games in general and particularly for refugee children. The particular emphasis on this special group is based on empirical evidence from user studies and lessons learned from the author's practical experience from the EduAPP4Syria project. The GBL community of educators, researchers, and practitioners can learn from these guidelines for more inclusive design focusing on user characteristics.

Preface

This thesis is submitted to the Norwegian University of Science and Technology (NTNU) for partial fulfilment of the requirements for the degree of Philosophiae Doctor.

The doctoral work was carried out at the Department of Computer Science, NTNU, Trondheim, Norway. Professor Alf Inge Wang was the main supervisor, and Professor Monica Divitini and Professor Dag Svanæs the co-advisers.

This doctoral work is funded by "EduApp4Syria", an international innovation competition funded by the Norwegian government and coordinated by the Norwegian Agency for Development Cooperation (Norad) in cooperation with NTNU and other partners.

Acknowledgments

First and foremost, I would like to thank Almighty God for blessing me with this life-changing opportunity to pursue the journey towards a PhD and giving me the strength and determination to complete this thesis.

I would like to thank all the people who have contributed in different ways to this research journey. I strongly believe that without their support, guidance, and sincere cooperation, the completion of this achievement would not have been possible. First, I would like to express my deepest gratitude and very special thanks to my supervisor, Alf Inge Wang, for his patient guidance, advice, and encouragement. Alf has always been very supportive, believed in me, gave me the confidence to choose my path and the flexibility to explore options. In fact, Alf has always impressed me by exceeding my expectations. He is a real gentleman and an excellent teacher: very encouraging, understanding, sincere, honest, kind, and generous. He was no less than a perfect supervisor I could ever possibly imagine. I grew up as a researcher and as a person under your guidance. Thank you for the effort you dedicated to me. I will always be very grateful to Alf for making my graduate experience very productive, enjoyable, and memorable.

Special thanks to my co-advisors, Professor Monica Divitini and Professor Dag Svanæs. I was always inspired by your work and found support and advice when needed. I would like to thank Monica for providing me with opportunities to collaborate with fellow PhDs and Master students and for her availability and constructive suggestions, which helped my PhD research work.

I wish to thank all my colleagues and co-authors who directly or indirectly contributed to my work. Many thanks to Ali Shariq Imran and Krenare Pireva Nuci for sharing their knowledge and collaborating with me in several research papers. Working with you has been fundamental, and you have been wonderful collaborators and friends.
A very special thanks to Serena Lee Cultura for being part of the design workshops; this would not be easy without you. You are an amazing person and a wonderful officemate and friend. It was encouraging to be able to talk to you when we shared some common work habits.

I extend my gratitude to all the colleagues from the Department of Computer Science (IDI), the ISSE group, and the EduApp4Syria Project. A special thanks to Nektaria

Kaloudi, Deepika Verma, Sofia Aftab, Zacharoula Papamitsiou, Dimitra Christidou, and Madeleine Lorås for all the engaging and constructive discussions and support, also to Kshitij Sharma, Javier Gómez, and Evangelos Niforatos for being very supportive officemates.

Thanks to all my friends in Norway and Pakistan for always being supportive, helping me survive from all the low moments, forget my PhD struggles, providing moments of joy, and sharing my happiness with me. They made me see things differently and more optimistically.

Last but certainly not least, I want to express my deepest, heartfelt, and most sincere gratitude to my family: Samina Kausor, Tahir Pervaiz, and Muhammad Rohail Tahir. Without you, all of this would not have been possible; each of you built a piece of who I am today. First, to my parents for always standing by my side and giving me strength, confidence, and encouragement to pursue my career. Thanks to my mother, who made me who I am today with her love, prayers, and encouragement throughout my studies, and my father, who supported and guided me in taking the first step towards my career in this field. I especially thank my mother and brother, who encouraged me to take this opportunity and pursue this PhD journey. They always supported my decisions, including moving to Norway, and gave me the strength to accomplish them. Besides, big thanks to my brother for always being there when I need him, for being my support system, and for his love and care.

Table of Contents

List of Figures

List of Tables

Abbreviations

AR Augmented Reality

ARCS Attention, Relevance, Confidence and Satisfaction

CCI Child-Computer Interaction

CLT Cognitive Load Theory

COTS Commercial off-the-Shelf

CS Computer Science

DEGs Digital Educational Games

DGBLL Digital Game-Based Language Learning

DSR Design-Science Research

E-learning Electronic Learning

EEG Electroencephalogram

GBL Game-Based Learning

GDPP Game Design Pedagogic Plan

GEM Game-Based Learning Evaluation Model

GEQ Game experience Questionnaire

GOM Game Object Model

GQM Goal Question Metric

HCI Human-Computer Interaction

HEEG Heuristic Evaluation for Educational Games

ICT Information and Communication Technology

IMI Intrinsic Motivation Inventory

LEAGUÊ Learning, Environment, Affective cognitive reactions, Game factors, Usability, and User

MACF Meaningful learning, ARCS motivation model, Cognitive load, and Flow theory

MDA Mechanics, Dynamics, Aesthetics

mGBL Mobile Game-Based Learning

MST Muslim Society Trondheim

NFE Non-Formal Education

NTNU Norwegian University of Science and Technology

PHEG Playability Heuristic for Educational Games

RCTs Randomized Controlled Trials

RETAIN Relevance, Embedding, Transfer, Adaptation, Immersion and Naturalization

RST Rough Set Theory

TEL Technology Enhanced Learning

USE Usefulness, Satisfaction, Ease of use

UX User Experience

ZPD Zone of Proximal Development

Part I. Synopsis

1 Introduction

This thesis is about game-based learning (GBL) and how the GBL design and evaluation process can be supported for developing effective learning games. This chapter starts with the background and motivation of the thesis. Then the problem statement is presented, and the research methodology is described. Next, the research goal and research questions are defined that express the problem foundation of this thesis. The chapter continues with this doctoral work's research outcomes, presenting the list of publications and contributions. Finally, the structure of the thesis is explained.

1.1 Background, Context, and Motivation

Over the past years, Information and communication technology (ICT) has developed substantially, and it influences almost every aspect of human life (Ratheeswari, 2018). Saariluoma (2005) argued that technology is designed to satisfy human needs and is a part of human activities. Therefore, it should be considered within the context of human experience forming the foundation of human technology. Human technology is a multidisciplinary field with no single theoretical or empirical approach that can be applied to all problems. Moreover, with the emergence of complex information technologies, multidisciplinary and interdisciplinary approaches in human technology research are imperative (Saariluoma, 2005). Today, ICT plays salient roles in education, entertainment, workplaces, business, and many other areas (Ratheeswari, 2018). In this digital era, the use of ICT in education is an essential and effective means for broadening educational opportunities, creating a robust learning environment, and providing students with the opportunities to learn and apply crucial 21st-century skills (Afshari, Bakar et al., 2009, Ratheeswari, 2018). Wasson and Morgan (2013) provided a summary of the field of ICT in education. They highlighted that it is subject to rapid change as it reflects advances in the underlying technology's capabilities. The use of technology for education is not new. The field of learning technology comprises the study and practice of aiding the learning process and improving learners' performances by creating, using, and managing appropriate technological resources (Mishra, Koehler et al., 2009, Rushby and Surry, 2016). Educational technology is developed not only to make education extensively available, but also to improve educational quality (Sampath, 1981). Winn (2002) identified four stages of educational technology research evolution (1) focus on the content, (2) focus on format, (3) focus on interactions, and (4) the latest era of research, focus on the learning environments.

Education is the process of transmitting the experience, i.e., knowledge, skills, and attitudes, to the members of the community. It is synonymous with learning as any sort of acquired experience and is referred to as any act that holds a formative effect on an individual's personality (Sampath, 1981). The educational research literature goes beyond formal education and frequently focuses on alternative systems such as "open systems", "distance learning", and "non-conventional studies". Depending on different learning situations with varying degrees of intentionality, education is found in three main forms (formal, non-formal, and informal), often combining and complementing one another (Melnic and Botez, 2014). After extensive literature review, the CEDEFOP[1] glossary defined formal, non-formal, and informal learning (Tissot, 2004) (Protopsaltis, Pannese et al., 2011). *Formal learning* is the learning that occurs within a structured and organized context that is designated as learning and may lead to formal certification or recognition. *Non-formal learning* is the learning embedded within planned activities that contain important learning elements but are not explicitly designated as learning. Lastly, *informal learning* is the learning that results from everyday life activities related to leisure work or family, also frequently referred to as experiential learning. Dib (1988) investigated formal, non-formal, and informal education concepts and highlighted their advantages, inter-relations, and limitations. Non-formal education (NFE) began as a service for rural populations and undeveloped countries deprived of formal education. However, it has now expanded to serve diverse audiences (Romi and Schmida, 2009). Non-formal education potentially solves many educational problems to meet individuals and society's needs (Dib, 1988). Online education has grown significantly over the past ten years (Li and Irby, 2008). There are many reasons for learners to participate in online learning environments (external requirements or pressures, personal needs or desires). Due to this, the non-formal and informal groups significantly differ from the formal groups because participation is not a requirement instead based on affinity. Most non-formal and informal online learning communities are based on the involvement of relatively independent and autonomous individuals who can easily engage or disengage from the group without personal consequence (Schwier and Seaton, 2013). Many researchers have recognized the importance of informal and non-formal education in addition to the formal education system and see the future of these alternative forms of education as quite promising (Dib, 1988, Gallacher and Feutrie, 2003, Romi and Schmida, 2009, Melnic and Botez, 2014).

Recent incidents such as the Syrian war and the COVID-19 pandemic can further put light on the importance of and need for informal and online education. Millions of Syrian children are out of school because of the conflict, and many have to cope with traumas and high levels of stress, which also affects their learning ability. They have to take refuge

[1] CEDEFOP is one of EU's decentralized agencies that supports development of European Vocational education and training policies and contributes to their implementation. https://www.cedefop.europa.eu/en/about-cedefop

in other countries and often miss out on essential formal education. It is also important to note that even for non-formal education, very few opportunities are available to them (Deane, 2016). The parents are mostly struggling with finding work and fulfilling basic needs and cannot give their children enough time to teach them. The risk is that we end up with a whole generation of children and young people with limited to no literacy. One of the main problems is that these children are not learning to read or write their mother tongue. It will make the future integration in schools difficult for those who want to return to Syria after the war to grow up in their country (Nordhaug, 2016). Thus, there is a need for any education initiative (formal, non-formal, emergency education) that can be useful to fill the knowledge gap of refugee children. As it is essential for their future institutional learning and acceptance among host communities to help refugee children thrive (Culbertson and Constant, 2015, Deane, 2016, Palaiologou, Fountoulaki et al., 2019). The refugee crisis and their increased number attracted the research community's interest to focus on refugee education and identified the need for new knowledge and better understanding for research with this particular group with the challenging background (de Wal Pastoor, 2016, Kaukko, Dunwoodie et al., 2017).

Another situation is the COVID-19 crisis that led to the physical closure of schools and universities, and institutions across the globe had to adopt the online teaching mode leading to accelerated digitalization of teaching, making online education more prevalent (Daniel, 2020, Langford and Damşa, 2020). About 97% of students as part of educational institutions are affected by this situation concerning online teaching (amid COVID-19) in Europe and Central Asia (Bank, 2020). The results from a survey with teachers (concerning COVID-19 online teaching) indicated that most of them conduct their lectures online (either live or using pre-recorded videos), adapting to the new ways of teaching (Langford and Damşa, 2020). Therefore, as educational technology professionals, it is essential to understand the non-formal and informal learning opportunities and use learning technology for these alternative learning setups in addition to supporting formal education (Schwier, 2012).

Technology-enhanced learning (TEL) is not only limited to traditional education systems. Numerous educational institutions have adopted "technology-rich" solutions to foster active learning, expand education beyond the classroom lecture settings, and help learners build knowledge on real-world problems (Trinidad, 2003). Research has reported high availability of smartphones among war-affected Syrian families, which can help reach children with engaging and fun learning supplements. It can help facilitate their continued learning and future reintegration into school (AbuJarour, Krasnova et al., 2016, Narli, 2018, Drolia, Sifaki et al., 2020, Neag, 2020). Similarly, online education platforms and e-learning tools (such as Zoom, Google Classroom) saw considerable demand during the COVID pandemic when businesses and day-to-day activities were shutting down (Azlan, Wong et al., 2020, Radha, Mahalakshmi et al., 2020, Soni, 2020). Many pedagogical theories explain the role of technology in education. The most prominent notion is that digital technology (as a medium or tool) organizes and imparts the learning activities by enabling the communication of content and structure (Laurillard, 2013). Therefore, technology should be actively and meaningfully used for education (Graham, Woodfield et al., 2013). However, similar to traditional learning environments where student engagement and learning performance have been a concern for researchers and educators (McMahon and Portelli, 2004, Ahlfeldt, Mehta et al., 2005, Noel and Liub, 2017), there are many challenges in e-learning as well. The typical constraints include lack of student-teacher interaction, difficulty focusing and understanding major concepts due to lack of active learning, and lack of engagement (Peña-Lévano, 2020).

Research has revealed that students enjoy playing computer games, and they are becoming a significant part of their lives (Pillay, Brownlee et al., 1999). According to Piaget (Piaget, 1962, Piaget, 1964), play is integral to and evolves with the different stages of children's cognitive development. Computer games are played for various reasons ranging from entertainment to educational purposes (Mohamed and Jaafar, 2010a). The use of play and games in a learning context is not a new phenomenon; it has been in practice for quite some time to increase learner motivation and fun aspects (Pillay, Brownlee et al., 1999, Plass, Homer et al., 2015). Educators acknowledged computer games' potential for learning purposes back in the 1980s (Pivec, 2007, Wang, Liu et al., 2015). The games which embody educational objectives are considered to make education more learner-centered by increasing fun and enjoyment, making learning more effective (Malone, 1980, Gee, 2003, Mohamed and Jaafar, 2010a, Wang, Liu et al., 2015). Games have transitioned from recreational practice to educational practice since the technology-based learning environments have attracted significant attention and interest. Researchers believe that in addition to increasing engagement, computer games can facilitate cognitive processes such as lateral thinking and making inferences that are beneficial for educational context (Quinn, 1996, Pillay, Brownlee et al., 1999). Furthermore, findings reveal that playing computer games can provide various cognitive, affective, perceptual, and behavioral outcomes (Connolly, Boyle et al., 2012). According to Koutromanos and Avraamidou (2014), games have a great potential for enhancing engagement, motivation, interaction, and interest by offering a range of opportunities (such as hands-on activities, story-line proving opportunities for role-playing, opportunities for developing skills like critical thinking, debating and constructing arguments, promote collaboration). Therefore, the use of mobile games as a learning tool is compelling in both contexts of a formal and informal learning environment.

The term "game-based learning (GBL)" refers to the use of games intended for some educational or learning purposes and have defined learning outcomes (Prensky, 2003a, Plass, Homer et al., 2015, Sanchez, 2019). According to Hsu, Hung et al. (2013), one of the major research topics in educational technology research, among others (such as multimedia, educational software, e-learning), is educational games. GBL has the potential to solve most of the issues discussed above by facilitating the learning process and increasing learner motivation and engagement (Kiili, 2005a). Several studies have been published on using GBL in the classroom and out of school (Wang and Tahir, 2020). A survey by Kim and Bonk (2006) indicated an expected increase in the use of interactive games in online teaching and learning. The term "game" usually refers to the use of digital games called educational games, serious games, or digital learning. Still, it is not always the case as the non-digital games are also broadly used for learning purposes (Plass, Homer et al., 2015, Sanchez, 2019). However, with the increasing use of digital technology and mobile phones, digital GBL is becoming increasingly popular. GBL has been extensively implemented in various courses and gained wide acceptance over the years (Wallner and Kriglstein, 2011, Alfadhli and Alsumait, 2015). Educational games have become a growing market in the game industry and academia alike. According to Protopsaltis, Pannese et al. (2011), educational games are normally considered as informal learning. Still, they can be a part of formal, non-formal, or informal learning environments if they are approached as learning elements that can be potentially integrated into multiple learning settings. Therefore, GBL has the potential to facilitate both formal and informal learning by promoting student motivation and enhancing the effectiveness of the learning process by adding fun (Prensky, 2002b, Protopsaltis, Pannese et al., 2011). It is a complex phenomenon and practice and includes a large variety of games, subjects, educational contexts (formal and non-formal), and other educational practices (Sanchez, 2019). Games have a high presence in primary education in both non-formal and informal segments of children learning. However, GBL is now also applied in the classroom lecturing to address new ways of ICT-based instructional design and provide learners with the opportunity to learn 21st-century skills and competencies that can influence their behavioral patterns and reflection (Pivec, 2007). Many researchers have investigated the effect of using GBL platforms and found a positive impact on students' and teachers' attitudes, classroom dynamics, student anxiety, and learning performance (Wang and Tahir, 2020). From the learners' perspective, learning games are played for various reasons, such as having fun while learning, experimenting, expressing feelings about conflict situations, and achieving better scores with challenges. However, from a teacher's perspective, learning games are used to reach the new generation with their preferred medium they are engaged with from their childhood (Pivec, 2007). One obstacle for introducing games in a learning environment is the belief that the novelty effect of learning through games will wear out quickly. However, studies show that game-based learning can also improve engagement, motivation, concentration, and perceived learning over time (Wang, 2015). According to de Freitas (2018), educational games have cross-disciplinary nature, and the related research seems to fall into four disciplinary categories: education science (theory and practice using pedagogy and psychology elements), game science (technology-enhanced learning perspective), neuroscience (brain-function), and information science (behavioral modeling and data analytics). However, there is a need to bring together the sub-fields' substantive literature into one distinct perspective.

This prevalent use of games in the young generation's lives has stimulated researchers' interest in investigating GBL tools (Van Eck, 2006, Ebner and Holzinger, 2007, Alaswad and Nadolny, 2015). The Human-Computer Interaction (HCI) researchers are increasingly interested in generating scientific and methodological knowledge concerning interactive systems' design and evaluation (Markopoulos and Bekker, 2003b). Today, technological advances are leading to immense improvements and significantly impact education. However, these developments are followed by considerable new challenges, making it difficult for researchers and educators to follow up and assess technology tools' effectiveness for learning (Alqurashi, 2019). Therefore, researchers should investigate which characteristics of learning technologies help or hinder learning and focus on the practical use of such technology in real contexts (Hsu, Hung et al., 2013, Alqurashi, 2019). Educational technology researchers are most concerned with technology's pedagogical use and its effectiveness in achieving intended learning outcomes. Therefore, it is essential to investigate the use of e-learning technology by students to contribute to e-Learning design and support student engagement (Hsu, Hung et al., 2013). It is important to understand how learning games impact the learner engagement and learning performance and which factors mitigate these effects to gain insights into the GBL process and implications for effective GBL design (Khan, Ahmad et al., 2017, Kadel, Halder et al., 2018, Eltahir, Alsalhi et al., 2021). GBL studies have attracted much attention internationally, and efforts have been made to demonstrate its usefulness (Shi and Shih, 2015). de Freitas (2018) highlighted that the literature concerning the use, design, and efficacy of GBL approaches is fragmented with inconsistent referencing patterns across sub-disciplines. This is because GBL literature has been built gradually in phases and in an ad hoc way spanning different disciplines. Due to GBL's cross-disciplinary nature, there are changing terminologies in different contexts. Researchers target a single viewpoint for studies using multi-methodological approaches; therefore, no distinct perspective has emerged. It is essential to evaluate the value of learning games as all serious games are not successfully used. Research demonstrates that ineffective design of learning games can produce negative results (Carrión, Santórum et al., 2020). The development of GBL applications is a time-consuming, laborious and complex process of game creation and integration with the learning process (Liu, Shaikh et al., 2020). Research has shown that it is important to design an effective learning game; otherwise, the learners simply don't use them (Marconi, Schiavo et al., 2018, Nousiainen, Kangas et al., 2018, Park, Kim et al., 2019, Liu, Shaikh et al., 2020). However, a comprehensive and robust design for learning games is still difficult to achieve because game designers often neglect the learning materials, and educators do not focus on exciting game elements (Shi and Shih, 2015). Studies show a significant increase in students' interest and motivation in learning with this technique.

However, long-term consequences and impact on students should be determined (Liu, Shaikh et al., 2020). Although the research regarding the effectiveness of learning games is overwhelmingly positive, researchers are challenged to investigate the best practices for the GBL approach to meet the learner's interests in different contexts. Therefore, more design studies are needed as GBL enters into this new wave of implementation. GBL researchers and practitioners need to ensure lessons from different disciplines are consolidated into general practice and unified literature to distill the key benefits and harmonize multidisciplinary perspectives. Furthermore, it is important to address the methodological challenges and create a shared terminology for GBL (de Freitas, 2018).

This doctoral work's overall research objective is to investigate how to facilitate the design and evaluation of GBL approaches to improve the effectiveness of learning games. Based on the above topics, this PhD work is positioned among four research fields: Information and Communications Technology, Educational Research, Game Research, and Human-Computer Interaction, as shown in Figure 1.1.

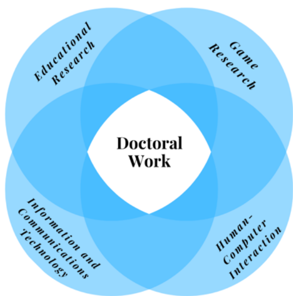


Figure 1.1: Venn diagram with fields of research

The research described in this thesis concerns the use of technological innovation (GBL) for teaching and learning. ICT includes the potential and limitations that technology brings within the education domain. It is linked to the digital and communication transformation in education and the development of innovative technologies for supporting the learning experience. The educational research in this PhD work focuses on evaluating the GBL approaches in the three main educational contexts: formal, non-formal, and informal learning. Gaming research covers the potential of using different types of games to provide powerful learning experiences and understating how game elements generate motivation and flow to facilitate the learning process. This PhD research investigates and compares GBL and non-GBL approaches to understand the impact on learning outcome uncovering GBL phenomenon. Finally, the HCI research in this thesis stresses the design and evaluation of learning games, understanding the tools and approaches that can provide support as well as guidelines to produce effective learning games. In this doctoral work, we conducted three design workshops focusing on leveraging the learning game design process and three quasi-experimental studies focusing on understanding the GBL phenomenon and underlying factors, as well as evaluation methods and guidelines. The potential role of educational technology (especially learning games) in the refugee crisis and COVID-19 directed the GBL evaluations conducted in this thesis to target formal, non-formal, and informal learning. The three quasi-experiments (GBL evaluation studies) in this doctoral work included university setting, online teaching (amid COVID-19 pandemic), and refugee children context, respectively thereby, covering the three educational settings. The previous research identified the need for new knowledge and better understanding for research with the special user group of refugee children because of their challenging background (Kaukko, Dunwoodie et al., 2017). Therefore, refugee children's inclusion in the GBL evaluation study instigated further research focusing on this special group and added to investigating methods and guidelines for evaluating learning games with refugee children.

1.2 Problem Statement

Nowadays, learning games are becoming a powerful educational tool enhancing learning both in and out of the classroom supporting individual learner needs (Protopsaltis, Pannese et al., 2011). Their growing use has attracted both industry and academic research communities (Backlund and Hendrix, 2013, Zeng, Parks et al., 2020). There is ample research focusing on the benefits of using games for learning. However, there are gaps in the literature that need further research to understand GBL as an approach and improve its effectiveness to generalize the use of learning games in different educational settings (Torrente, Moreno-Ger et al., 2009, Hainey, Connolly et al., 2016). Analyzing the work on learning game design and evaluation practices, we discovered a lack of research regarding the following:

- Despite the increasing utilization of learning games and research investigating its use, there *is still a lack of empirical evidence to support GBL as an approach and its validity to generalize its use with different pedagogic models and learning practices* (Torrente, Moreno-Ger et al., 2009, de Freitas and Liarokapis, 2011, Van Staalduinen and de Freitas, 2011, Ariffin and Sulaiman, 2013, Giannakos, 2013, Boyle, Hainey et al., 2016, Hainey, Connolly et al., 2016, Voulgari and Yannakakis, 2019). Additional research is needed to further and more thoroughly explore the learning process with games and identify the influencing factors to understand what makes GBL effective and how it works within particular contexts (Squire, 2006, Dickey, 2007, Nakayama, Yamamoto et al., 2007, Protopsaltis, Pannese et al., 2011, Van Staalduinen and de Freitas, 2011, Giannakos, 2013). It is important to understand the GBL phenomenon and provide insights for the design of learning games to support different learning contexts and physical settings (Jagušt, Botički et al., 2018, Voulgari and Yannakakis, 2019).

- Researchers have identified evaluation as an integral part of applications' success to remove imperfections, increase their effectiveness, and fit their purpose (de Freitas and Oliver, 2006, Gossen, Hempel et al., 2013). Educational game development is a costly and time-consuming process; therefore, there is a demand for continued research in assessing GBL approaches' efficacy and a requirement for identifying principal criteria (de Freitas and Oliver, 2006, de Freitas and Liarokapis, 2011). Although researchers have highlighted different aspects important for GBL, *more research is needed for an overreaching approach to guide evaluation and design iterations* (de Freitas and Liarokapis, 2011, Van Staalduinen and de Freitas, 2011, Oprins, Visschedijk et al., 2015). According to Kebritchi, Hirumi et al. (2010), a cursory literature review indicated that GBL is not always effective. Therefore, to verify the potential of learning games, it is essential to systematically evaluate them (Marciano, de Miranda et al., 2014) but the diverse characteristics of GBL make it a difficult task (Djelil, Sanchez et al., 2014). Previous research has identified evaluation criteria and the evaluation process as main challenges in evaluating educational games (Mohamed and Jaafar, 2010a). According to Dondi and Moretti (2007), identifying criteria is a complex and time-consuming process, and it is also difficult to identify different evaluation processes and the difference between the analytical (single aspect) and global (holistic) evaluation process. Moreover, there are not many approaches available to guide the process of evaluating GBL applications (Becker, 2011). It highlights the need to define the key aspects of educational games that must be focused on to make them effective learning tools (Ak, 2012) to serve as evaluation criteria and guide the learning game evaluation process to improve game design. Researchers (Hays, 2005, All, Nunez Castellar et al., 2014, Vanderhoven, Willems et al., 2015) have highlighted that the research studies evaluating educational games' effectiveness struggled with various methodological issues, and it is essential to recognize and overcome these challenges to improve GBL evaluation research.

- According to research, complex products' design requires understanding multiple aspects (Zahedi, Tessier et al., 2017). It is vital to involve different stakeholders (game designers, developers, educationalists, researchers, and learners) in the design and evaluation process of learning games (de Freitas and Liarokapis, 2011). Previous research indicates that there is no single path to learning game design, and the involvement of different experts in educational game development makes it a difficult task. However, only a few methods specifically support incorporating vital elements for educational game design (Harteveld, 2011, Silva, 2020). Furthermore, there is a need to combine different experts' perspectives to effectively achieve the goal (Brandt and Messeter, 2004, Ahmad, Rahim et al., 2015). It is challenging for the interdisciplinary GBL design team to understand different aspects and their relationships embedded in GBL, influencing learning game design (Ahmad, Rahim et al., 2015, Ávila-Pesántez, Rivera et al., 2017). The stakeholders often have limited knowledge outside their area and different design space interpretations (Theodosiou and Karasavvidis, 2015, Wetzel, Rodden et al., 2017). Game designers focus on creating fun and engaging games but do not have knowledge of teaching materials whereas, educators emphasize learning material but neglect exciting game features (Shi and Shih, 2015). There is a lack of communication between practitioners and researchers and adoption of models and tools in game design practice (Neil, 2012). Furthermore, there *is a lack of research focusing on improving the educational game development process* to help the GBL team make effective learning games that provide a more engaging, rich, and constructive player experience (Paz and Fernandes, 2018). Educational game design is a complex process. One of the main problems in learning game design practice is involving different experts to connect the different aspects to support learning effectively. Therefore, there is a need to explore ways to scaffold team members in the early steps of educational game design (Theodosiou and Karasavvidis, 2015).

Based on the above, this thesis's objective is to support the design and evaluation of GBL approaches to improve the effectiveness of learning games. This doctoral research work is a step toward developing a more holistic understanding of GBL, especially the different aspects that shape this phenomenon and influence interaction with learning games. We postulate that building such an understanding can inform the design and evaluation of effective learning games with embedded practical value for researchers, educators, and designers and guide factors that can influence learning in situ. This knowledge can be used to develop instruments and tools that can guide and scaffold the design and evaluation process. The PhD research also utilizes and stresses the importance of theories and concepts relevant to HCI, learning, game design, and psychology to ground the measures we have selected and applied in our analyses.

1.3 Research Methodology

The doctoral work presented in this thesis followed the design-science research (DSR) (March and Smith, 1995, Hevner, March et al., 2004, Hevner and Chatterjee, 2010) as a research methodology. DSR is used by many researchers in educational technology (Chard, 1999, Cheong, Cheong et al., 2013, Van Biljon, Traxler et al., 2015, Spill and Bruinsma, 2016, Apiola and Sutinen, 2020). It is a research paradigm to provide a solution to practical problems via building and applying artifacts, thereby adding new knowledge and contributing to the body of scientific evidence (Hevner and Chatterjee, 2010). The term "artifact" refers to something that is human-constructed or artificial rather than something that occurs naturally (Simon, 1996). The artifacts in design science either improve the existing solutions or provide the first solution to a problem (Hevner and

Chatterjee, 2010). The research by March and Smith (1995) characterizes four types of artifacts as research output for design science research: *constructs* (vocabulary and symbols of a domain), *models* (abstractions, representations, or propositions expressing relationships among constructs), *methods* (algorithms, guidelines or practices used to perform a task), and *instantiations* (realizations of artifacts, i.e., implemented and prototype systems or tools). Following this classification, the research outcomes (artifacts) of this doctoral work are a model (LEAGUÊ conceptual framework), methods (ten-step process for transforming framework to design cards and integrated LEAGUÊ-GQM evaluation approach), and the instantiations (card-based toolkit, LEAGUÊ analysis instrument, and LEAGUÊ evaluation guide). However, the thesis's main focus is on developing the LEAGUÊ framework and its three instantiations (card-based toolkit, LEAGUÊ analysis instrument, and LEAGUÊ evaluation guide). These research outcomes involved constructing a framework and set of tools to support the GBL design and evaluation process at various educational games' developmental stages. The design of the produced artifacts was grounded in relevant theories and further refined through results from the evaluation studies feeding the design iterations and contributing to the validation of theories and development of new constructs. The design science research highly emphasizes the importance of evaluation (Pries-Heje, Baskerville et al., 2008), as it is crucial to demonstrate the value of the research outcomes (artifacts) using rigorous evaluation methods to prove its relevance for practice (Sonnenberg and Vom Brocke, 2011). All the produced artifacts were evaluated during evaluation studies, where some of the tools went through multiple iterations. The performed evaluation studies facilitated understanding GBL design and evaluation and provided theoretical and practical impact extracting design principles.

This doctoral research used a *mixed methods* research design equated with design science research, employing both qualitative and quantitative research methods. As highlighted by Ågerfalk (2013), a mixed methods approach can potentially play an important role in DSR and can be used to provide deeper insights (Cleven, Gubler et al., 2009, Aramo-Immonen, 2011, Genemo, Miah et al., 2016). We adopted the *exploratory-triangulation design* in mixed method research within DSR cycles (explained in detail in Chapter 4). The design science research in this thesis is combined with mixed method research design. Three cycles of mixed method design were performed within design science research to answer the research questions and generate implications that can inform the GBL design and evaluation process to produce effective learning games. Each cycle started with "qualitative research" by conducting a systematic literature review or comparative analysis that identified the problems and needs using existing literature and theories (answering RQ1) and provided the knowledge for constructing artifacts (see Figure 4.3, DSR rigor cycle). It is followed by "quantitative and qualitative research" that involved the evaluation of built artifacts by conducting user studies (quasi-experiments and design workshops) using mixed methods (answering RQ2-4), see Figure 4.3 (DSR relevance cycle). The first cycle of mixed methods design focused on the GBL phenomenon and interrelated factors targeting the LEAGUÊ framework (RQ2). The second cycle focused on the GBL design process targeting card-based tool (RQ3). Finally, the third cycle focused on the GBL evaluation process targeting the LEAGUÊ-GQM approach (RQ4) (see Figure 4.4 for mixed methods cycles).

The evaluation studies were conducted mainly as *quasi-experimental studies* (GBL evaluations), *design workshops*, and *case studies*. Consistent with the DSR methodology grounded in the activities of "building" artifacts for a specific purpose and "evaluating" how well they perform (March and Smith, 1995), we conducted three quasi-experimental studies, three design workshops, and three case studies. During these activities, the framework and tools developed during multiple iterations were validated on the field. Co-design was used as a strategy in design workshops to pursue collaboration and produce more effective educational outcomes (Myers, Piccolo et al., 2018, Trischler, Pervan et al., 2018). The methods used for collecting quantitative data were *questionnaire, pre and post-test, usability testing, game logs, EEG,* and *game design artifacts*. During design workshops, the users produced game design ideas using ideation sheets. These user-generated materials (game design artifacts) aided the improvement and refinement of the tool and methods employed. The methods used for collecting qualitative data were *semi-structured interviews, observations, video recording, focus group*, and *documents/articles analysis*. All the data has been analyzed respectively based on their type, founding our results in the relevant literature and theory.

1.4 Research Questions

The main objective of this doctoral research is to understand and improve the design and evaluation of learning games for them to be effective, building on the current state-of-the-art in GBL and the gap in literature hindering the holistic view of aspects to be integrated into educational games, and devising instruments and methods to facilitate the process in order to develop games that are both engaging, educative and effective for its users. The main research goal that expresses the problem foundation for my PhD work is as follows:

Research Goal: How can the design and evaluation of game-based learning (GBL) approaches be supported to improve the effectiveness of learning games?

In order to answer the main research goal, the work has been broken down into four research questions as follows:

RQ1. What are the challenges and problems in the current GBL design and evaluation practices?

RQ2. What are the key elements for the GBL phenomenon, and how are they related?

RQ3. Which kind of approaches, tools, and guidelines can be employed to facilitate the *GBL design process* for effective learning games?

RQ4. Which kind of approaches, tools, and guidelines can be employed to facilitate the *GBL evaluation process* for effective learning games?

1.5 Research Outcomes

The research outcome of this doctoral work is based on ten research papers published (or under review) in international peer-reviewed journals and conference proceedings that explored the research questions and added to the contributions. The results reported in these research papers contributed to the body of knowledge in the fields of GBL, HCI, and Educational Technology.

1.5.1 Research Papers

The research questions (RQ1-RQ4) are addressed in the following research papers. The connection and mapping between research papers and research questions are presented in Table 1.1.

P1. **Tahir, Rabail**, and Alf Inge Wang. (2017). "State of the art in game-based learning: Dimensions for evaluating educational games." In *Proceedings of the 2017 European Conference on Games Based Learning (ECGBL)*, Academic Conferences International Limited, pp. 641-650.

P2. **Tahir, Rabail**, and Alf Inge Wang. (2020). "Codifying game-based learning: Development and application of LEAGUÊ framework for learning games." *Electronic Journal of e-Learning* 18, no. 1: 69-87. DOI: 10.34190/EJEL.20.18.1.006

P3. Pireva, Krenare, **Rabail Tahir**, Ali Shariq Imran, and Niraj Chaudhary. (2019). "Evaluating learners' emotional states by monitoring brain waves for comparing game-based learning approach to pen-and-paper." In *Proceedings of the 2019 IEEE Frontiers in Education Conference (FIE)*, IEEE, pp. 1-8. DOI: 10.1109/FIE43999.2019.9097262

P4. Pireva, Krenare, **Tahir, Rabail**, Alf Inge Wang, and Ali Shariq Imran. (2021). "Game-based digital quiz as a tool for improving students' engagement and learning in online lectures". Ready for submission.

P5. **Tahir, Rabail**, and Alf Inge Wang. (2018). "Insights into the design of educational games: Comparative analysis of design models." In *Proceedings of the 2018 Future Technologies Conference (FTC)*, Springer, Cham, pp. 1041-1061.
DOI: https://doi.org/10.1007/978-3-030-02686-8_78

P6. **Tahir, Rabail**, and Alf Inge Wang. (2020). "Transforming a theoretical framework to design cards: LEAGUÊ ideation toolkit for game-based learning design." *Sustainability, Special Issue Design Methodology for Educational Games* 12, no. 20: 8487.
DOI:10.3390/su12208487

P7. **Tahir, Rabail**, and Alf Inge Wang. (2021). "Completeness and collaboration in the early design phase of learning games: Do ideation cards provide scaffolding?". Accepted in the 2021 *International Conference on Human-Computer Interaction (HCII)*.

P8. **Tahir, Rabail**, and Alf Inge Wang. (2019). "Exploring methods and guidelines for child-computer interaction research with refugee children." In *Proceedings of the 2019 International Conference on Human-Computer Interaction*, Springer, Cham, pp. 70-89.
DOI: 10.1007/978-3-030-22636-7_5

P9. **Tahir, Rabail**, and Alf Inge Wang. (2021). "Evaluating the effectiveness of game-based learning for teaching refugee children Arabic using the integrated LEAGUÊ-GQM approach". Ready for submission.

P10. **Tahir, Rabail**, and Alf Inge Wang. (2019). "How to evaluate educational games with refugee children: Methodological aspects and lessons learned from EduApp4syria." In *Proceedings of the 2019 European Conference on Games Based Learning (ECGBL),* Academic Conferences International Limited, pp. 722-730.
DOI: 10.34190/GBL.19.136

Table 1.1: Mapping the connection between research papers and research questions.

	P1	P2	P3	P4	P5	P6	P7	P8	P9	P10
RQ1	•				•			•		
RQ2		•	•	•					•	
RQ3		•	•	•		•	•		•	
RQ4		•	•	•				•	•	•

1.5.2 Research Contributions

The main contributions (encompassed in ten research papers) from this doctoral work to the academic body of research are summarized as follows (see Figure 1.2 for the overall mapping of research questions, research papers, and contributions exhibiting the flowchart of the doctoral research work):

C1: ***Summarize and conceptualize the state of the art in GBL design and evaluation practices and identify existing challenges and issues.*** This doctoral work's first contribution is presented as review studies consisting of two systematic literature review studies (Tahir and Wang, 2017, Tahir and Wang, 2019a) and a comparative analysis study (Tahir and Wang, 2018). The results from these studies provide an overview and substantive insights regarding the design and evaluation practices for GBL. The reviews aimed to show the state of the art depicting ongoing advances and changes in the field of GBL and identify the challenges and potential research gaps that, once addressed, can lead to more meaningful tools and methods for producing effective learning games.

C2: ***Present a conceptual hierarchical framework of six dimensions for comprehensive design and evaluation of GBL applications.*** This doctoral work's second contribution presents a GBL framework (LEAGUÊ) (Tahir and Wang, 2020a) to bridge the gap in the literature regarding core dimensions of GBL for comprehensive design and evaluation. The GBL literature highlighted an inconsistency in terminology, scope, definition, and usage of elements leading to the absence of a holistic view of GBL. This research includes directed content analysis of existing theories on GBL to validate and conceptually extend it, producing a comprehensive framework for GBL. The framework presents a hierarchical structure with four conceptual levels, with six dimensions elaborated through factors, subfactors, and metrics to guide the researchers and designers to create effective learning games.

C3: ***Empirical evidence on the application of GBL approaches in different contexts for improved understanding about the process of learning with educational games and contributing factors.*** The third contribution of this doctoral work is the examination of the GBL phenomenon. It includes the findings from the three GBL evaluation studies (quasi-experiments) (Pireva, Tahir et al., 2019, Pireva, Tahir et al., 2021, Tahir and Wang, 2021b) conducted in this doctoral research. We utilized existing learning games in different domains (sorting algorithm, HCI, language learning) and contexts (formal learning, online learning, and informal learning) to evaluate the effectiveness of GBL in comparison to other instructional approaches. The aim was to better understand how the learning process acquired when using a GBL approach and how the key elements of GBL are related and affect the GBL experience. These studies' findings provide implications to support the design of effective learning games and learners' learning experience.

C4: ***Contribution to the design, implementation, and evaluation of a card-based design toolkit for the ideation phase of educational game design, facilitating multidimensional focus and collaboration in the GBL design process.*** This doctoral work's fourth contribution is the development, evaluation, and refinement of the card-based toolkit to facilitate the GBL design process (Tahir and Wang, 2020b). It includes the ten-step process of transforming the framework into the toolkit that can guide other researchers and designers to develop similar tools and findings from three design workshops detailing the toolkit's strengths and limitations to support GBL design practices. The developed toolkit scaffolds for collaboration and completeness in the early phase of the learning game design process (Tahir and Wang, 2021a).

C5: ***Contribution to the development and application of an analysis instrument and an integrated evaluation approach to support the educational game evaluation process***. The fifth contribution of this doctoral work is framed as the development and application of an analysis instrument (Tahir and Wang, 2020a) and an integrated evaluation approach (Tahir and Wang, 2021b) to assess learning games and to guide the GBL evaluation process. The developed artifacts are based on the LEAGUÊ framework (Tahir and Wang, 2020a), grounding the research. The analysis instrument can be seen as a support tool for any stakeholder (designers/developers, researchers, and intermediates like teachers/parents) who aim to understand the potential of an educational game in a specified environment. The instrument ensures that they take into account the essential factors associated with GBL to know the strengths and areas that need improvement. The proposed integrated evaluation approach can be seen as an essential step towards future empirical research in GBL, providing support for planning and executing educational game evaluation studies. It provides guidance for developing a GBL evaluation plan by establishing goals, defining questions, and identifying measures for the evaluation process.

C6: ***Guidelines for improving the design and evaluation of GBL applications in general and specifically for refugee children***

The sixth and final contribution of this doctoral work elaborates on the best practices in designing and evaluating learning games in general and specifically for refugee children as a special user group. It presents the lessons learned and recommendations that emerged from the field experience in evaluating learning games, analyzing empirical data from evaluation, and the reviewed literature (Tahir and Wang, 2019a, Tahir and Wang, 2019b). These design and evaluation guidelines can help researchers and designers critically reflect on the methodological issues, selection of methods, and design needs to create effective learning games for refugee children. However, most of these findings can be useful for general GBL research as well.

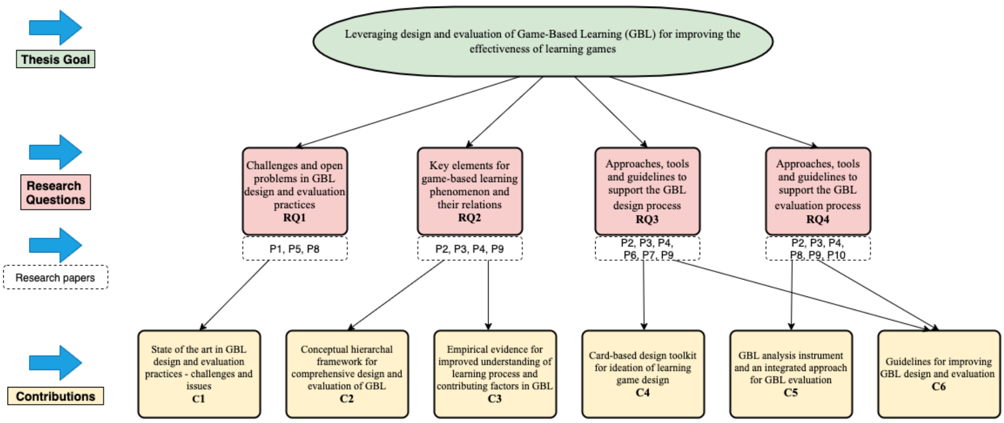


Figure 1.2: Thesis flowchart

1.6 Structure of the Thesis

This thesis is composed of two main parts and is structured as follows:

- **Part I:** includes the introduction to the research work, an overview of the related work and relevant theories, research methodology used, results attained, contributions made, and the conclusion of this thesis. Part I is organized as follows:

Chapter 2. Presents an overview of the related work in GBL, the relevant background theories, and the main elements for educational gaming as the theoretical underpinning of this PhD thesis.
Chapter 3. Describes the overall research approach and methods adopted in this PhD. It includes the research cycles, the research artifacts developed, and an overview of the

evaluation studies (including quasi-experiments, design workshops, and case studies) conducted.

Chapter 4. Presents the results of the research papers included in this thesis.

Chapter 5. Discusses the results and outlines the contributions of the PhD thesis and their relation to the research papers. This chapter also presents the limitations of the research work and evaluation of the contributions with respect to the research questions.

Chapter 6. Concludes the thesis and provides suggestions for future work.

- **Part II:** contains the set of ten research papers (entailing this thesis) in full length.

2 Related Work

This chapter provides an overview of the related work in the game-based learning (GBL) research field, in line with the research questions for this doctoral work. The chapter presents how the previous studies relevant to this thesis have addressed similar topics. It is essential to provide a brief overview of relevant work to ground the choices made in this doctoral research and highlight the research contributions of this thesis.

2.1 Game-Based Learning Definition and Benefits

Game-based learning (GBL) is defined as the type of games with definite educational or learning objectives (Shaffer, Squire et al., 2005, Plass, Homer et al., 2015). It refers to the use of games encompassing educational value and learning outcomes (Tang, Hanneghan et al., 2009). According to Whitton (2012), GBL at a simple level can be defined as the learning enabled by using a game. In contrast to the above definition of GBL, gamification is the use of game elements and principles (such as rewards, scoring, incentive systems, challenges) in a non-game context (Deterding, Dixon et al., 2011, Robson, Plangger et al., 2015) to engage users in a task or activity (Hamari, 2017). It is important to understand this distinction between GBL and gamification. The design of GBL (referred to as *educational games or learning games*) requires balancing the necessity to cover the appropriate subject matter with the desire to emphasize gameplay (Plass, Perlin et al., 2010). Therefore, an educational game's fundamental characteristic is blurring instructional content with game characteristics (Pivec, Dziabenko et al., 2003). As described by Garris, Ahlers et al. (2002), the debriefing process (between the game cycles and attainment of learning outcomes) provides a link between the game experience and learning, and this is also in line with the Kolb's description of an experiential learning process (Kolb, Rubin et al., 1971) that include four stages: doing (concrete experience), reflecting (reflective observation), understanding (abstract conceptualization), and applying (active experimentation).

Prensky (2001) states that six key elements define games. These elements include rules, goals and objectives, challenge/competition/conflict/opposition, outcomes and feedback, representation or story, and interaction. According to Malone (1981a), the four elements of computer games are defined as curiosity, fantasy, challenge, and control. Usually, when we mention GBL, it is assumed that the game is a “digital game.” However, it is not always the case as non-digital games are also used for learning purposes (Plass, Homer et al., 2015, Sanchez, 2019). Nonetheless, digital technology's prevailing use has tremendously increased the popularity and focus on digital GBL. Nowadays, mobile phones have also been widely used for GBL under the label “mobile game-based learning” (Alfadhli and Alsumait, 2015).

Learning games are utilized in education as they can make the overall learning experience more engaging and entertaining for learners (Anastasiadis, Lampropoulos et al., 2018). According to Prensky (2001), fun plays a principal role in creating motivation and relaxation in the learning process. Motivation enables learners to make an effort without displeasure, and relaxation allows them to do things more easily. Therefore, adding a fun element into the learning process makes learning engaging, enjoyable, compelling, and, more importantly, also efficient and effective (Prensky, 2002a, Anastasiadis, Lampropoulos et al., 2018). Prensky (2003a) presented a digital GBL approach to promote motivation by incorporating digital games with curricular contents. He defined the key characteristic of digital GBL as the “coming together” of interactive entertainment and learning through digital games (Prensky, 2003a). According to Anastasiadis, Lampropoulos et al. (2018), digital GBL can potentially enhance learners’ learning experience, improve learning and teaching, and promote active interaction between teachers and students. Moreover, based on their analysis, digital GBL also offers the following benefits: progressive learning through experience, student-centered and feedback-driven learning, social-emotional growth, positively competitive environment, soft skills development, cognitive growth and digital literacy, improved collaboration and communication environment, enhanced decision making, critical thinking, and problem-solving skills, the rewarding feeling of achievement and progression, and high self-esteem and autonomy.

A learning game should be motivating enough to make learners repeat the learning embedded game cycles within the learning game (Pivec, Dziabenko et al., 2003). A learner is expected to acquire desirable behaviors and learning while repeating the cycles (i.e., playing the learning game) based on cognitive and emotional reactions resulting from feedback and interaction with gameplay (Pivec, Dziabenko et al., 2003). Therefore, the most crucial part of the educational process in both GBL and traditional learning is to hold the student’s motivation to learn, thus preventing boredom with the educational material (Alfadhli and Alsumait, 2015).

2.2 Important Aspects of Game-Based Learning

Educational games must be effective to be used for learning or educational purposes (Yusoff, 2010). There is a consensus to a greater degree among most researchers concerning the fact that games can be engaging and instructive. However, there is little agreement regarding GBL's essential aspect for designing and implementing effective educational games (Garris, Ahlers et al., 2002). Benson (2014), based on his research, agrees with Moore's view (Moore and Price, 2009) in concluding that, as of yet, there is no standard theory on designing computer games. All educational gaming theorists do not agree on one best approach to developing and applying learning games (Dedeaux, 2016). The theoretical base of learning games is still evolving. A major problem is the lack of consensus concerning aspects that constitute an effective game (Bedwell, Pavlas et al., 2012).

Researchers have taken different perspectives to assist in developing effective learning games (Yusoff, 2010, Benson, 2014). These perspectives incorporate different research areas, including education, psychology, and computer science, ranging from focusing on pedagogical foundation, intrinsic motivation theory, game design principles to learning theory-based games design. The literature of games for learning goes back to 1970's when the initial discussion started concerning the potential of games as learning simulations (Parry, 1971). Then some successful examples of GBL applications emerged (Sharan and Colodner, 1976). Later, game-based training was discussed from a cognitive psychology context with prominent research by Malone (1981b). In this context, “game for learning” was approached from a cognition and motivation standpoint and described the elements of games that contribute to learners' motivation. The early research of learning games mainly focused on theoretical concepts of individual game-based training platforms. In comparison, the latest 21st-century research concentrates on theoretical and empirical examination of educational games or serious games (Pavlas, 2010). Van Eck (2006) emphasized the need for research explaining why GBL is engaging and what makes it effective, highlighting that several areas account for effectiveness. Below we discuss some of the critical aspects of GBL considered by researchers to describe and advance educational game's understanding and effectiveness.

2.2.1 Education and Pedagogical Foundation

Many supporters of serious games have been researching the best approach to use games for learning resulting in GBL literature embodied in well-established learning principles, models, and theories (Van Eck, 2006). The importance of basing the design of learning games on a pedagogical foundation is recognized by many researchers (Hirumi and Stapleton, 2009, Tang, Hanneghan et al., 2009, Arnab, Berta et al., 2012, Wu, Chiou et al., 2012). Gunter, Kenny et al. (2006) stated that a serious game fails to meet the intended educational goals if its design is not based on well-established learning theories. It is important to integrate the educational effectiveness in the design process from the start.

Learning can be described as the process of acquiring knowledge, skills, behaviors, and attitudes (Salas and Cannon-Bowers, 2001). For an effective learning game, it is essential to evaluate the suitability of the game content with regard to the learning objectives (Dedeaux, 2016). Gee (2003) provides significant academic research on a set of thirty-six learning principles for learning games. These principles are based on his study of complex and self-directed learning undertaken by players as they encounter and master a game. In his research, Gee (2005) supports the use of games (within education) as a learning tool. He argues that games inherently contain learning principles (such as reflection, curiosity, repetition). Therefore, his proposed learning principles can build into good computer games that can transform learning for teachers and students (Gee, 2007).

Many researchers emphasize the need to base the educational game design on recognized learning theories. However, to support this argument, limited examples are found in the literature. A review by Wu, Hsiao et al. (2012) found that only a few GBL studies are established on the use of recognized learning theory which mostly favored constructivism and humanism compared to other theories. The most commonly referenced learning theories in GBL are cognitivism, constructivism, behaviorism, and humanism (Benson, 2014). According to constructivism, learners process or absorb new knowledge by linking it to their existing knowledge. In this way, they effectively construct their own mental models of this new knowledge (Bruner, 1966). Cognitivism is related to the mental processes that take place within the human mind. Piaget's research on cognitive developmental stages is considered the most well-known work of cognitivism (Piaget, 1976, Benson, 2014). According to humanism, learners have a willingness and natural desire to learn; hence, learning should concentrate on the learner instead of the process (Huitt, 2009). Behaviorism views humans' minds as a black box, and it can be stimulated to generate a response. An observable change can be reinforced in a learner's behavior either positively through some reward or through punishment, depending on the required response (Pritchard, 2017). The theories relevant to the research work in this thesis are described in detail in Chapter 3.

Many researchers have primarily focused on education/learning aspects in GBL. de Freitas and Oliver (2006) proposed a four-dimensional framework. It focuses on pedagogy by considering four dimensions collectively to help tutors evaluate the potential of employing simulation/GBL in practice. Connolly, Stansfield et al. (2009) described an evaluation framework that focuses on the pedagogical aspect, introducing attributes to measure the GBL environment with attention on the learner and learning. Another evaluation framework proposed by Wang, Liu et al. (2015) emphasized the learning perspective concerning learning results, learner motivations, and learner experience.

2.2.2 The Game Design Perspective

Many proponents of Digital GBL have spoken about the risks of “academizing” games, which Prensky (2002a) mentions as “sucking the fun out” (Van Eck, 2006). Many researchers argue that this shortcoming is due to academicians designing educational games who have insufficient understanding of the art, culture, and science of game design. Therefore, the resulting products are sometimes educationally sound but drearily weak as games (Van Eck, 2006). Mildner, Stamer et al. (2015) stated that for effective learning games, it is not enough just to emphasize good learning content, but it is also essential to consider engaging game design for developing learning games that can teach and motivate.

Papert (1998) argues that game designers have a better perspective on the nature of learning compared to curriculum designers. Prensky takes on a similar view in describing Gee's learning principles (Gee, 2003) as jargon-filled that must have been intended for an academic audience. Game designers would find it difficult to understand these principles thinking they will suck the fun out (Prensky, 2003b). Prensky adapted the learning principles by Gee from a design perspective focusing on video game players learn from what practices in the game design. In Prensky's view, the work by Gee (2003) does not discuss much about the field of game design. Therefore, he extracted game design rules from the project “400 Project” by (Barwood and Falstein, 2002, Kreimeier, 2002) and analyzed relative equivalents in Gee's learning principles (Gee, 2003) to focus on using game design principles for learning games.

Further, Prensky (2003a) offered a range of principles that game designers should follow to produce good games. Other researchers focusing on game design include MacKenty (2006), who offered advice on what should be included in an educational game. According to MacKenty's views (Benson, 2014), an educational game has to be a good game first; only then can it be a good learning game. It implies that a thorough understanding of game design is required as a foundation for designing educational games. Plass, Homer et al. (2015) described some fundamental game design elements (such as game mechanics, narrative, incentives, visual aesthetics, musical score, content and skills, and learning objectives) used in learning games for understanding the engagement in GBL. According to their research, game design patterns are comparatively more preferable than design principles or guidelines for their relatively abstract level. It is because they can be customized or localized for application in specific projects (Plass, Homer et al., 2015).

Several researchers have proposed a framework or model focusing on educational game design. Mitgutsch and Alvarado (2012) proposed a framework that structures different game design elements to analyze the formal, conceptual design of serious games. The framework provides recommendations on how to shape serious games' assessment in terms of design. Chorianopoulos and Giannakos (2014) presented design principles for serious games in mathematics. Shi and Shih (2015) focused on game design aspects proposing 11 game factors for GBL design. These factors described a thinking process to design and evaluate educational games using game elements.

Wilson, Bedwell et al. (2009) investigated the specific game attributes that influence learning outcome. It provided a starting point for further research focusing on game attributes (such as challenge, game fiction, feedback) and impact on learning outcomes

(Pavlas, 2010). In line with attribute-based research, Giannakos (2013) focused on the effect of attitudes in educational games on learners' performance.

2.2.3 Pedagogy and Game Design Integration

The learning games developed on strong pedagogical foundations incorporating sound educational practices will meet the educational goals. However, they will lack fun and engagement, thus failing to meet user expectations. On the other hand, learning games developed with game designers dominating the GBL design process may be entertaining. However, they will fail to apply essential pedagogical principles thus, lacking vital knowledge. According to Gunter, Kenny et al. (2006), merely pouring the learning content in an ad hoc manner in the game will not result in effective learning just because the content is present inside a game that might motivate learners to learn.
Hirumi and Stapleton (2009) suggested applying pedagogy during the development process for enhancing GBL. Therefore, a systematic process that integrates instructional design tasks with the game development process for educational game design can optimize GBL. Van Eck (2006) highlighted that in order to maximize the learning potential of educational games, it is crucial to integrate games into the learning process. The answer to effective learning games lies in finding a synergy between engagement or game and pedagogy in GBL, not in privileging one arena over the other (Van Eck, 2006). Further, Van Eck (2007) highlighted that the goal of building learning games meeting the educational goals without losing the engagement part of the game is still not achieved. One of the main problems in integrating pedagogy and entertainment in GBL is to bring the learning content into the game world (Van Eck, 2007). The ability to effectively integrate the content is the key to creating good educational games that can produce noticeable learning outcomes along with overall value (Seeney and Routledge, 2009). Van Eck (2007) proposed four principles of learning in games that can guide GBL designers in developing new learning games.

Some researchers have proposed educational game design and evaluation frameworks that combine focus on learning and game design as two critical aspects of GBL. Some of these frameworks include educational game design framework by Ibrahim and Jaafar (2009), a framework for the analysis and design of educational games by Aleven, Myers et al. (2010), an adaptive digital GBL framework proposed by Tan, Ling et al. (2007), a RETAIN model presented by Zhang, Fan et al. (2010), a GBL evaluation model (GEM) by Oprins, Visschedijk et al. (2015) and a Game object model (GOM) proposed by (Amory, 2007). Rooney (2012) introduced the concept of fidelity in his framework for serious game design in addition to play and pedagogy and explored the challenges involved in balancing these three elements. The framework proposed by El-Sattar and Hussein (2016) described learning as an activity system (with learners as subjects, the task as an object, and the educational game as a tool). It introduced a game design pedagogic plan (GDPP) to facilitate learning in games by providing guidance on game design and pedagogy to balance learning with game attributes.

2.2.4 Motivational Perspective

The educational game research was closely tied with the notion of motivation since Malone's work on intrinsically motivating games (Malone, 1980). According to his theory, three categories: challenge, fantasy, and curiosity creates intrinsically motivating computer games. Another popular work on motivation is the ARCS model by Keller (1983). This model includes a synthesis of motivational concepts and theories clustered into four categories: attention, relevance, confidence, and satisfaction (ARCS). Both these works are significantly used and cited in GBL research as the basis for educational game design (Habgood, 2005, Dickey, 2007, Kenny and Gunter, 2007, Kuo, 2007, Derbali and Frasson, 2010a, Liu and Chu, 2010, Boyle, Connolly et al., 2011, Shute and Ke, 2012, Belanich, Orvis et al., 2013, Wang and Tahir, 2020). Motivation is vital for educational games as there is a positive relationship between motivation and learning (Bixler, 2006). However, some researchers state that it is unlikely that a high level of engagement with games will transfer to educational contexts (Hoffman and Nadelson, 2010). According to Plass, Homer et al. (2015), motivation is the most frequently cited characteristic of games. It is because entertainment games have long shown to motivate learners through various game features of motivational nature.

When viewing GBL from a motivational perspective, the ability of learning games to motivate and engage players by offering an enjoyable experience that they want to continue is emphasized (Ryan, Rigby et al., 2006, Zusho, Anthony et al., 2014, Plass, Homer et al., 2015). Besides Malone and Keller's work, several efforts have been made to identify elements contributing to motivation and engagement in games (Loftus and Loftus, 1983, Squire, 2013). Some integrated models were developed for motivation in games. Garris, Ahlers et al. (2002) adapted the input-process-output model to motivation for instructional games and learning. Ryan, Rigby et al. (2006) applied their self-determination theory (SDT) (Ryan and Deci, 2000) for exploring motivation for gameplay.

2.2.5 Flow Experience

In line with the past research trend related to motivation, the topic of flow drew considerable attention and interest in educational game research. Flow theory was initially introduced by Csikszentmihalyi (1990), which served as the basis for the research focusing on flow experience in educational games. The flow theory is explained in detail in Chapter 3, Section 3.1. Many researchers focused on flow aspects for evaluating player enjoyment in games. Flow has a particular value in the area of games as it maps well with immersion experienced by the player during the gameplay (Dunwell, Lameras et al., 2014). Sweetser and Wyeth (2005) proposed the GameFlow model by mapping the elements of flow to elements in the game literature to provide criteria for enjoyment in games.

Similarly, the learning game design is also aimed at creating an experience that is so interesting that it keeps the players' attention intensity and as long as possible (Kiili, de Freitas et al., 2012). Pavlas (2010) proposed a play experience scale for GBL focusing on flow and found that play and in-game performance were key predictors. Kiili (2005a) introduced an experiential gaming model based on the four stages of experiential learning (Kolb, 2014) to facilitate the flow experience. The model serves as a link between game design and educational theory. However, the framework does not offer a complete game design. EGameFlow, proposed by Fu, Su et al. (2009), is a scale for assessing the level of enjoyment delivered by e-learning games. EGameFlow is based on the GameFlow model (Sweetser and Wyeth, 2005) to provide a more rigorous scale for assessing e-learning games' user enjoyment. Kiili, Lainema et al. (2014) presented a flow framework to analyze educational games' overall playing experience through dimensions of flow. It is interesting to note that most of the work concerning flow within GBL research revolves around the original flow dimensions by Csikszentmihalyi (1990).

2.2.6 User Experience or Usability for GBL Design and Evaluation

Besides the aspects already discussed above, few researchers focused on either user experience (UX) or usability as an important aspect for educational game design and evaluation (Nagalingam and Ibrahim, 2015, Petri and von Wangenheim, 2016). According to Nagalingam and Ibrahim (2015), it is important to evaluate the UX of educational games to ensure that effective learning games are produced. It makes it essential to identify the suitable UX elements. According to de Lima, de Lima Salgado et al. (2015), it is important to compare educational games' user experience with mainstream games. It will help understand GBL better and identify important attributes that help improve the UX of educational games. Usually, educational game research has been predominantly concerned about evaluating the learning effects for successful learning games (Virvou and Katsionis, 2008, Yusoff, 2010). However, usability is considered important as it affect educational effectiveness (Mayes and Fowler, 1999, Squires, 1999, Virvou and Katsionis, 2008). Moreover, Markopoulos and Bekker (2003a) pointed out that usability may be crucial for enjoyment and learning. Despite this, not much attention has been given to usability in GBL (Hersh and Leporini, 2013).

User Experience (UX) is a branch of human-computer interaction (HCI) focusing on the interaction between users and products. According to ISO 9241-210 definition, UX is a person's responses and perceptions resulting from the use or anticipated use of a system, service, or product (Standardization, 2008, Law, Roto et al., 2009, Mirnig, Meschtscherjakov et al., 2015). UX plays a vital role in this era of evolving digital games to identify suitable variables to evaluate the educational game design (Nagalingam and Ibrahim, 2015). UX is a broader concept that incorporates usability and other components (Shiratuddin and Zaibon, 2011, Law and Sun, 2012). According to Nagalingam and Ibrahim (2015), the UX aspect covers not only fun, entertainment, and challenge but also social aspects such as students' culture, background, and the content of the game. It also includes time factors, believing that the user's response towards the product might change with time.

On the other hand, some of the most prominent work in usability is ISO 9241-11 standard, ISO/IEC 9126, and Nielsen's heuristics (Paz and Pow-Sang, 2014, Shafiq and Khan, 2018). According to the ISO 9241-11 standard (ISO, 1998), usability is the extent to which the product can be used with efficiency, effectiveness, and satisfaction to achieve specified goals by specified users in a specified context of use. Whereas ISO/IEC 9126 (Commission, 2001) defined usability as the capability of the product (when used under specified conditions) to be learned, understood, used, and attractive to the user. On the other hand, Nielsen presented ten usability heuristics for the user interface (Nielsen, 1994, Nielsen, 2005). The above describe models and heuristics are also largely used and adapted for GBL as a basis to understand the factors that affect the usability in educational games (Federoff, 2002, Sim, MacFarlane et al., 2005, Sim, MacFarlane et al., 2006, Nousiainen, 2009, Diah, Ismail et al., 2010, Lu, Chang et al., 2011, Krouska, Troussas et al., 2019). Some researchers (Carroll, 2004) suggested extending the usability concept to include fun. However, others (Sim, MacFarlane et al., 2005) regard fun and usability as two completely separate constructs and stick with the traditional definition of usability, which is also the approach embraced in this thesis. Moreover, the concept of pedagogical usability emerged while focusing on educational medias' requirement to support learning in addition to effective and easy use (Nokelainen, 2004, Shield and Kukulska-Hulme, 2006, Kukulska-Hulme, 2007, Hersh and Leporini, 2013, Laurillard, 2013). Sim, Horton et al. (2004) emphasized that interfaces for education should meet the standard usability requirements. However, usable interfaces must also be intuitive and should not distract users from accomplishing their goals. Usability is even more important for educational games since their user interface is typically more complex as it has to support learning and fun. That is why instructional efficiency also becomes important (Yacci, Anne et al., 2004, Virvou and Katsionis, 2008). However, some researchers (Diah, Ismail et al., 2010, Ismail, Diah et al., 2011, Soewardi and Perdana, 2019) have demonstrated that standard usability metrics can be effectively used for educational games.

Only a few GBL researchers have focused explicitly on UX in educational games. Shiratuddin and Zaibon (2011) emphasized on UX of GBL. They presented heuristics focusing on learning content, game usability, mobility component, and gameplay. Law and Sun (2012) proposed an evaluation framework of UX for adaptive educational games focusing on usability, gaming experience, adaptivity, and learning experience. Barbosa, Rego et al. (2015) also emphasized UX by presenting a heuristic focusing on usability and game experience for evaluating educational games. de Lima, de Lima Salgado et al. (2015) proposed evaluating UX using a game experience questionnaire (GEQ) and an adaptation of the Intrinsic Motivation Inventory (IMI). Nagalingam and Ibrahim (2015) explored the UX elements for the design and evaluation of educational games. They highlighted the need for a single complete framework incorporating all essential elements for UX of educational games. Moreover, Law and Sun (2012) emphasized that although a number of evaluation methods are available for UX, a viable analytic framework is missing that can facilitate a more profound understanding of factors (and their dynamics) affecting UX in the GBL context.

Moreover, the existing research shows that interface design and attention to usability can lead to better educational games (Shafiq and Khan, 2018). Therefore, usability is a vital aspect of GBL applications' effectiveness and usefulness (Thomas, Schott et al., 2004). Some GBL researchers focused on usability and presented heuristics for evaluating educational games' usability (Omar and Jaafar, 2010, Mohamed, Yusoff et al., 2012, Barbosa, Rego et al., 2015). These researchers incorporated concepts of learning, gameplay, interface, and enjoyment within heuristics for evaluating GBL usability. Thomas, Schott et al. (2004) presented usability guidelines for mobile learning games based on literature related to games and learning, game design theory, interviews with learning game developers, and game analyses. They reviewed elements that contributed to effective learning games. Yue and Zin (2009) proposed six usability evaluation constructs for the design of history educational games.

2.2.7 Player Characteristics

Another aspect in GBL research is the focus on player characteristics which showed progression in the 21st century (Pavlas, 2010). In the context of GBL, good learning games should be within the zone of proximal development (ZPD) of players (Plass, Homer et al., 2015). The concept of “zone of proximal development” is proposed by Vygotsky (1978). He thought that play is a primary factor in children's development and has a vital role in creating ZPD for children. The “zone” in ZPD represents what can be done by a learner with support from another individual. In contrast, the “core” in ZPD represents the actions that a learner can perform without any support (Borthick, Jones et al., 2003). This notion concentrates on how learners can most effectively learn and is mostly studied in development and education contexts (Dunn and Lantolf, 1998). However, as pointed out in (Pavlas, 2010), there are interesting similarities between ZPD and flow when an external collaborator is viewed as a game used by the learner. In this way, in educational games, ZPD becomes relevant to the concept of “flow state.” However, ZPD is not fundamental to flow theory but can be used to examine and expand the understanding of flow (Pavlas, 2010). Kiili (2005a) proposed linking ZPD with the skill-challenge balance to extend the flow zone and improve educational game design. Moreover, Piaget's theory (Piaget, 1964) also emphasizes the individual differences by presenting four cognitive developmental stages of children (described in detail in Chapter 3, Section 3.2). It has been used by researchers for understanding children experiences in games (Kamii and DeVries, 1980, Schifter, Ketelhut et al., 2011, Bjorklund and Causey, 2017).

The GBL research focusing on player characteristics examines players' response to game challenges (Cowley, Charles et al., 2006). Moreover, the research also investigates players' characteristics (such as game exposure and self-efficacy) that affect educational games' efficacy (Orvis, Horn et al., 2006). Some researchers do not directly include the aspect of player characteristics in their models but focused on explaining the GBL process from the viewpoint of player experiences, such as the experiential gaming model by Kiili (2005a) focuses on player's skill level and challenges. Few researchers include the learner aspect in their proposed GBL models. The four-dimensional model (de Freitas and Oliver, 2006) focuses on learner specifications connected with the other three dimensions. The game-based learning evaluation model (GEM) (Oprins, Visschedijk et al., 2015) includes the learner aspect as a part of a broader concept incorporating personal features, learning indicators, and learning outcomes. However, the authors only describe learning indicators in detail. The framework by El-Sattar and Hussein (2016) mentions the learner aspect as the subject of an activity system focusing on learning. The GBL guidelines proposed by Alfadhli and Alsumait (2015) emphasize child requirements. Tan, Ling et al. (2007) regarded the learner aspect as a vital issue for GBL design. They included it in their proposed framework focusing on the pedagogical perspective. The framework emphasizes that designers and educators should design and recommend suitable games for learners considering their cognitive development, psychological needs, and learning behavior. Ariffin and Sulaiman (2013) found, based on their study results, that learner's motivation to learn is influenced by the learner's background, which affects their performance. They thus highlighted the need to integrate the learner's background parameters in educational games.

2.3 Use and Effectiveness of Game-based Learning in Different Contexts

Although traditional practices for education remain in use, GBL has extensively been implemented in various courses such as computer science, language learning, psychology, mathematics, and pedagogy (Breuer and Bente, 2010, Alfadhli and Alsumait, 2015). Traditional games have long been a part of the human learning experience in informal or formal settings. Nowadays, serious games, including educational games, are receiving attention from the game industry and academic research (Susi, Johanesson et al., 2007, Michaud and Alvarez, 2008, Ritterfeld, Cody et al., 2009, Protopsaltis, Pannese et al., 2011). Although educational games are now increasingly being accepted as a learning tool, this increased acceptance still revolves around the debate about what makes them effective and how they should be used (Protopsaltis, Pannese et al., 2011). de Freitas (2006) highlighted that a key research challenge is to make intellectually appropriate, enriching, and challenging educational games with their integration into the learning process. Moreover, Squire (2006) emphasized that it is essential to explain how particular GBL approaches work in particular contexts using instructional theory approaches (Squire, 2006, Squire, 2007).

Educational games can be used in formal, non-formal, or informal learning environments (defined in Chapter 1, Section 1.1) based on the structure of the context and intention to learn (Colardyn and Bjornavold, 2005, Protopsaltis, Pannese et al., 2011). Voulgari and Yannakakis (2019) suggested strengthening the link between informal, non-formal, and formal learning practices. It could benefit not only formal education but also the effectiveness and access to non-formal and informal practices. According to a review of GBL literature, educators adopt three ways for integrating educational games into the learning process: students building educational games, educators or developers building games to teach students, or integrating commercial off-the-shelf (COTS) games into classrooms (Van Eck, 2006). Voulgari and Yannakakis (2019) examined the use of games in non-formal and informal science learning practices using a case study. They found that despite the diversity in settings, structure, format, and target users of these practices, there was a convergence in some themes. These themes included pedagogical approaches, the importance of fun, and objectives of the practices. Some other issues they found included required resources for practical implementation, gender representation, and parents' role. Binzak, Anderson et al. (2016) examined gameplay across formal and informal learning contexts. They found similarities between gameplay in both contexts and interesting observations concerning integrating video games into formal learning settings. They also highlighted the need for further research in this area. On the other hand, Yelland (2003) investigated primary-school-age children's experiences with computer games for mathematical understandings in formal and informal learning contexts. The study results highlighted significant differences in the use of computer games in-school and after-school. They suggested that after-school (informal) context not only facilitated learning but offered opportunities for new and dynamic ways of interaction.

Some researchers performed studies to explore the impact of using learning games in formal, non-formal, and informal learning settings. Hainey, Connolly et al. (2016) reviewed quality empirical studies related to GBL application in primary education focusing on affective and motivational outcomes, behavioral change, knowledge acquisition, cognitive skills, and content understanding. The results showed that GBL is used in various subjects to teach children at the primary education level. However, more randomized controlled trials (RCT) studies comparing GBL to traditional viable teaching approaches for primary education should be performed to ascertain the usefulness of GBL. A review by Hussein, Ow et al. (2019) examined the effects of using educational games at the elementary education level for teaching science. It showed that GBL has promising potential, specifically for content understanding. Moreover, GBL was also found as an effective approach for postgraduates to teach them research skills. The findings are based on the evaluation of gameplay of 127 university students, which demonstrated positive game experience and transfer of intended learning outcomes (Abbott, 2019). GBL's effectiveness in higher education is also investigated by Ariffin and Sulaiman (2013), focusing on higher education students' knowledge change in three different treatments. The result showed that playing a game with integrated cultural aspects may increase students' knowledge acquisition. Moreover, the mean score with the educational game was highest among all three treatment groups. Koutromanos, Sofos et al. (2015) reviewed the use of augmented reality (AR) games in education for formal and informal learning contexts. They found evidence for positive student learning outcomes. A study by Milovanovic, Minovic et al. (2009) aimed to find some empirical evidence for the effectiveness of educational games for teaching and found positive results. Qian and Clark (2016) reviewed GBL for 21st-century skills and found that it might be effective for students' skills. However, only 37% of the empirical findings reported the effect sizes. Koutromanos and Avraamidou (2014) conducted a review study to explore the use of games in formal and informal learning settings and identified the ways in which mobile games were used in different settings.

Moreover, researchers have focused on different factors for investigating the effectiveness of educational games. Vandercruysse, Vandewaetere et al. (2012) in their review, investigated educational games' learning effects to understand the conditions that make a game effective for learning. Milovanovic, Minovic et al. (2009) investigated the effects of different teaching approaches concerning individual differences in cognitive styles. The results showed that cognitive style had some effect on the effectiveness of GBL. Koutromanos and Avraamidou (2014) investigated the impact of using mobile games on students' achievements, attitudes, and learning. Papastergiou (2009) examined the motivational appeal, learning effectiveness, and potential gender differences of learning games for computer science concepts for Greek high school education. They found that GBL can be used as a motivational and effective learning environment irrespective of students' gender. Protopsaltis, Pannese et al. (2011) investigated the relationship between serious games and learning contexts (formal and informal). They highlighted the challenges related to the learning aspect, gaming aspect, and implementation and technological details that educational game designers and developers must face when designing games for formal and informal learning.

Researchers have highlighted several issues in GBL research concerning the effectiveness and use of educational games. Many researchers have highlighted the dearth of empirical evidence in the GBL literature to support the validity of the GBL approach (Connolly, Stansfield et al., 2007, Hainey, 2010). Hussein, Ow et al. (2019) highlighted the lack of empirical evidence. They further pointed out that there are different opinions and mixed results concerning GBL benefits on students' academic achievements, skills, and motivation. The finding of their review suggested the need for additional research. They suggested researching different learning modes, compare GBL to traditional teaching methods, and conduct more RCTs. Hainey, Connolly et al. (2016), in addition to RCTs, identified a lack of and need for longitudinal studies. Similarly, Ariffin and Sulaiman (2013) also highlighted the lack of empirical evidence for the effectiveness of GBL as a learning tool. On the other hand, some researchers highlighted issues specific to particular contexts, domains, or users. Gasteiger, Obersteiner et al. (2015) highlighted a lack of empirical research on the effectiveness of games for children's early numeracy development in formal and informal learning. They also identified the need for more research with systematic evaluation. Voulgari and Yannakakis (2019) highlighted that there is still limited research focusing on the use of games for non-formal learning practices. The review by Koutromanos and Avraamidou (2014) highlighted a lack of research concerning the use of mobile games in the formal learning contexts and specifically the combination of the formal and informal contexts. Most studies in the review focused on secondary education context, highlighting a gap in literature focusing on younger students. Moreover, Voulgari and Yannakakis (2019) identified some implications of context-related issues on designing and implementing games for non-formal and informal science learning practices. These include settings, diversity, time constraints of practices, and goals.

It is vital to properly evaluate educational games and obtain concrete empirical evidence in different learning scenarios to generalize results to affirm their potential (O'Neil, Wainess et al., 2005). The empirical evidence is fragmented as educational game research focuses on different game types, tasks, age groups and also has methodological flaws. Although the lack of empirical evidence for GBL effectiveness is not new, the technological advancements and growing popularity reinforce the need for generating more empirical evidence.

2.4 Educational Game Design and Evaluation

In this section, the existing models and frameworks for educational game design and evaluation are briefly reviewed, summarizing the previous work on GBL design and evaluation. As highlight by Wasson (2007), a fundamental challenge for technology-enhanced learning (TEL) environments (such as learning games) is how to design them and understand their use considering complex interconnections between various factors such as pedagogical and technological issues. Therefore, it is important to detail the existing frameworks for gaining deeper insights into GBL and mark areas for further improvement to define this doctoral work's contributions. This existing work also served as a theoretical basis for the development of the LEAGUÊ framework proposed in this PhD thesis. The details concerning strengths and weaknesses of existing models and identified challenges in GBL design and evaluation literature are covered in P1 (Tahir and Wang, 2017) and P5 (Tahir and Wang, 2018). Table 2.1, 2.2, and 2.3 presents a summary of existing work.

The existing models, frameworks, and guidelines primarily focusing on GBL evaluation are presented in Table 2.1. The majority of the evaluation-focused frameworks focus on pedagogical aspects such as pedagogic considerations, learning experience, learning results, educational property, learning content, topic coverage, and pedagogical issues. Many frameworks concentrated on evaluating either flow, motivation, engagement, enjoyment, immersion, or fun aspects in learning games. Moreover, usability or UX aspects were also emphasized by evaluation-focused frameworks. Most of the framework incorporating usability or UX aspects also included pedagogical, game design, and playability, in addition to focusing on the interface. The game design aspects focused on by these frameworks included, among others, gameplay, game design, and game experience. However, hardly a few frameworks or models incorporated learner characteristics for GBL evaluation. Most of these frameworks and models aim to help guide learning games' assessment or evaluation focusing on different aspects, stakeholders, and purposes.

Table 2.1: Existing evaluation-focused GBL models and frameworks

Evaluation-focused frameworks/ models	Description
EGameFlow	This scale is for evaluating user enjoyment of e-learning games. It can be useful to measure subjective opinions in large amounts to determine the learning game's strengths concerning students' viewpoints about enjoyment. It aimed at offering an economical and easy evaluation method to survey learners about the used learning game (Fu, Su et al., 2009).
Evaluation Framework for Assessing Games	This framework presents evaluation criteria focused on the quality aspect in assessing and selecting learning games. The framework aims to support the self-evaluation of learning games and emphasize the importance of quality. The criteria are based on two previous e-learning and distance learning initiatives that are modified and extended for educational games. The main dimensions are specified in order, allowing for analysis in terms of presence or absence. They also facilitate quantitative analysis for evaluation purposes. The criteria are identified for different areas, including pedagogical and context criteria, technical criteria, and content criteria. It can support teachers, professionals, and trainers in adopting learning games for teaching by assisting them in selecting the most suitable games to support students' development considering changing educational

Evaluation-focused frameworks/ models	Description
	requirements. However, validity or application of evaluation criteria is not provided (Dondi and Moretti, 2007).
Evaluation Framework for Effective GBL	This framework aims at GBL evaluation with a focus on the pedagogical perspective. It identified that GBL could potentially be evaluated in terms of learner performance, learner motivation, learner perceptions, learner preferences, GBL environment, and collaboration. Depending on what is to be evaluated, the framework categories can be viewed as a collective whole, not necessarily in isolation. It can be used in the formative evaluation to inform design when embedding a GBL application into curricula and in summative evaluation by pointing to individual analytical measurements for evaluation (Connolly, Stansfield et al., 2009).
Evaluation Framework for GBL	This framework guides the GBL evaluation from the learning perspective. The framework provides detailed measurements for learner experience, learner motivations, and learning results to guide the evaluation. Some guidelines are provided for performing a GBL evaluation using the framework to assist researchers (Wang, Liu et al., 2015).
Evaluation Framework of UX	This framework is for the evaluation of User experience (UX) for adaptive digital educational games (DEGs). The framework is proposed to support an in-depth understanding of the dynamics of factors affecting UX in DEGs. Different approaches are integrated to construct this multi-perspective framework for evaluating the UX of adaptive DEGs. It focuses on gaming experience, learning experience, adaptivity, and usability. Further, it applies activity theory for investigating UX data (Law and Sun, 2012).
Flow Framework	This framework describes the dimensions of flow experience that can be used to analyze the overall quality of educational games' playing experience. The principles are based on associative, situative, and cognitive learning theories with a focus on flow and feedback principles. The framework does not focus on the pedagogical aspect but the experience of flow in learning games. However, learning is presented as one of the flow consequences but not further discussed. It can be used for studying game-based learning experiences (Kiili, Lainema et al., 2014).
Four-Dimensional Framework	This framework aims to help tutors evaluate the potential of using games and simulation-based learning in their practice. The framework does not provide a prescriptive approach but a touchstone for understanding. It assists practitioners to critically think about how simulations and games are embedded in their class plans. Evaluators and researchers can use it to develop metrics for analyzing existing educational games and simulations. It focuses on pedagogically specific factors: learner specification, pedagogic considerations, context, and mode of representation (de Freitas and Oliver, 2006).
Framework for Serious Game Design Evaluation	This framework presents a hybrid methodology for serious game design evaluation providing criteria for evaluation based on MACF (meaningful learning, ARCS motivation model, cognitive load, and flow theory). The most suitable MACF criteria design is selected through rough set theory (RST) using fuzzy Delphi, AHP, and

Evaluation-focused frameworks/ models	Description
	TOPSIS. The importance of each criterion was initially established using expert selection, and then RST was used. The most important MACF criteria were determined based on the selection process results as playfulness, skills, attention, and personalized. This framework provides a system evaluation model and design criteria for multimedia game design educators for selecting the most suitable MACF characteristics from the range of evaluation criteria. The MACF design criteria selection is a multicriteria problem. It includes uncertainty, fuzziness, and subjectivity in the evaluation process. Different alternatives must be considered when evaluating game design criteria factors. It makes the decision evaluation method necessary to reinforce MACF designs' decision evaluation quality. The decision-making evaluation model of MACF can help game design educators suggest teaching strategies (Su, Chen et al., 2013).
Game Scale to Evaluate Educational Computer Games	This scale is for evaluating the quality of educational computer games focusing on learning and enjoyment characteristics. Teachers can use it to identify the quality of games to select good games for use in classes. The scale is developed by defining a model based on the input-process-outcome game model and Kolb experience learning cycle. However, only the initial structure of the scale is presented (Ak, 2012).
Game-Based Learning Evaluation Model (GEM)	This evaluation model focuses on measuring the effectiveness of serious games in a practical way. It provides the methodology and indicators that should be measured in serious games validation research. The model focuses on design and learning indicators, learning outcomes, personal features, and environmental influences. The selection of indicators choices should be made based on the research. The use of generic learning and design indicators allows using the GEM model for evaluating multiple games. Such evaluation provides insight into the reasons for serious games' effectiveness, helping designers improve the game. Some challenges exist in GEM concerning the quality of game design indicators scale that needs further attention (Oprins, Visschedijk et al., 2015).
Guidelines for Evaluating Games	These guidelines are for evaluating games to identify promising games for teaching computer science (CS). It helps to determine if computer science educational games are engaging and promising for students based on topics taught, easy to install, engaging, and time to use the game. The developed evaluation criteria are specific for evaluating CS games. These guidelines are intended to help developers produce effective games for teaching CS. Moreover, they can help teachers choose effective CS games to help engage students with CS learning topics (Gibson and Bell, 2013).
Heuristic Evaluation for Educational Games (HEEG)	These heuristics are for evaluating educational games in terms of game experience and usability. They are developed based on existing heuristic evaluation models. These heuristics be used to quickly identify specific problems of usability, design, enjoyment, and gameplay. They provide points for evaluating games to improve general game quality and reduce the risks of conceptual flaws in creating instructional games. These heuristics can be applied during

Evaluation-focused frameworks/ models	Description
	game development allowing earlier detection of problems and after the game is developed to check if it meets the requirements (Barbosa, Rego et al., 2015).
Heuristics Evaluation Strategy	This evaluation strategy is proposed to specifically evaluate mobile game-based learning (mGBL), focusing on game usability, gameplay, mobility, and learning content. These heuristics aim to provide a checklist for evaluating mGBL applications' effectiveness. Each component represents issues to be considered for mGBL evaluation. These heuristics are developed based on playability heuristics for mobile games and adding learning content (Zaibon and Shiratuddin, 2010).
Methodology for Interface Evaluation	This methodology is for heuristics-based usability evaluation and describes usability factors that can be used to evaluate educational games' interface. The factors include interface issues, multimedia issues, pedagogical issues, and playability issues. These are compiled from various studies. This methodology's focus is on formative evaluation and is intended for educational game evaluation during the development process. It includes a set of questions for different evaluator types expecting to optimize resource usage based on the online evaluation tool (Omar and Jaafar, 2010).
Playability Heuristic for Educational Games (PHEG)	These heuristics guide expert evaluation for finding usability problems in educational computer games. PHEG is proposed for potential quantitative usability analysis of educational games. It helps to find usability problems focusing on the interface, educational elements, content, playability, and multimedia. It can be useful for game developers to get feedback from evaluators for identifying usability problems while still in the development phase of the educational game (Mohamed, Yusoff et al., 2012).
Quality Evaluation Model	This ISO quality model is for evaluating mobile games. The model is based on ISO/IEC 25010:2011, focusing on reliability, efficiency, and maintainability. It contains suitable procedures, metrics, and measures for these three policies. Although the model was developed to ensure the quality of an educational game, the identified policies and measures are rather general for use with any mobile game or software product. The model does not consider the pedagogical aspect in evaluation (Alhuhud and Altamimi, 2016).
Quality Evaluation Standard	This framework is for identifying quality evaluation elements and provides evaluation standards for serious educational games. It is designed to evaluate the quality of technical and non-technical elements of educational games. It provides metrics for each element for a comprehensive evaluation. The framework can assist users in selecting an educational game and provide standardized quality criteria to developers for producing high-quality educational games. It is based on the international standard ISO/IEC 9126 and the Korean standard TTAS.KO-11.0078. However, the last two steps for framework development to establish a complete quality standard are not yet conducted (Yoon and Park, 2013).

Evaluation-focused frameworks/ models	Description
Serious Game Design Assessment Framework	This framework is for the assessment of game design in serious games. It provides a structure with different design elements to study the formal conceptual design underlying a serious game for a holistic assessment concerning their implicit and explicit purposes. Six essential components are identified: game purpose, mechanics, content, fiction and narrative, framing, and aesthetics and graphics. It should not be considered a definite objective measurement instrument but recommendations regarding structuring serious games' assessment concerning their design to provide grounds for critical discussions. The frameworks' idea is focused more on purpose-based game systems considering that the purpose of the game is the driving factor functioning as the main influence over game design elements. However, the elements can be assessed and discussed in a flexible order depending on the game and criticism perspective (Mitgutsch and Alvarado, 2012).
Usability Evaluation Constructs	This evaluation method presents six evaluation constructs for usability evaluation for history educational games. The constructs are interface, mechanics, gameplay, playability, feedback, and immersion. These are proposed for usability evaluation based on previous educational game design studies, integrating the game design features and pedagogical elements. However, questionnaires or detailed heuristics for these constructs are not provided (Yue and Zin, 2009).

The existing models, frameworks, and guidelines primarily focusing on GBL design are presented in Table 2.2. Although these frameworks or models mainly aim at guiding learning game design, some of these also intend to support GBL evaluation. The design-focused frameworks also highly emphasize the pedagogical aspect focusing on learning objectives, experiential learning, instructional design, relevance, embedding, transfer, learning content, and knowledge enhancement. However, after the pedagogical aspect, the most emphasized perspective in design-focused frameworks/ models is the game design. The game design aspect in these frameworks/models focuses on game requirements, game mechanics, game dynamics, game aesthetics, game design, fidelity, gameplay, and design principles. Moreover, some frameworks focus on designing flow experience, immersion, play, or enjoyment in learning games. However, only a few design-focused frameworks/ models concentrated on usability or UX.

Table 2.2: Existing design-focused GBL models and frameworks

Design-focused frameworks/ models	Description
A Framework for Serious Educational Game Design	It is a nested model of six elements: identity, immersion, interaction, increased complexity, informed teaching, and instructional. The framework elements are derived from studies on the design and development of games from grade 5 to graduate level and grounded in theory and research within education, instructional technology,

Design-focused frameworks/ models	Description
	psychology, and learning sciences. According to this framework, educational games contain these six elements that come into view in the order of magnitude, starting from the element identity and ending at instructional (Annetta, 2010).
A Theoretical Framework for Serious Game Design	This triadic theoretical framework consists of the elements of pedagogy, play, and fidelity for the design of serious games. The author points out that the inherent inconsistencies between pedagogy, game design, and fidelity make it challenging to balance these elements during a serious game design process and integrate them into one coherent framework (Rooney, 2012).
Adaptive Digital Game-Based Learning Framework	This framework presents essential components and features for designing GBL environments based on four existing models (design framework for edutainment environment, adopted interaction cycle for games, engaging multimedia design model for children, and game object model. Based on the analysis, the developed framework focuses on the learners and the game design. The framework also highlights some important features such as challenge, goals, story, and objectives not included as part of the framework (Tan, Ling et al., 2007).
Design Principals for Serious Game	These design principles identified four main principles based on previous literature to design serious video games for Math. They can also support researchers in evaluating learning effectiveness. The design principles include hero and narrative, familiar interactions, trial and error, and collaboration. The proposed principles provide a starting point for researchers to extend and apply them in other domains (Chorianopoulos and Giannakos, 2014).
e-VITA Framework for SGs	This framework focuses on three critical dimensions, including technical verification, user experience, and pedagogical aspects (learning outcome). It argues that with respect to development and evaluation, educational games should have three critical dimensions to be effective (1) it should be easy-to-use and technically sound; (2) it should be an engaging and fun game; and (3) it should be an effective learning instrument producing desired learning outcomes. In order to improve motivation and learning, all three dimensions should be targeted. The failure to meet any one dimension could compromise the effectiveness of serious games (Pappa and Pannese, 2010).
Educational Games Design Framework	This framework was developed for educational game design for higher education by comparing few available frameworks and recommends the required criteria from pedagogy and game design viewpoint. The idea behind this framework is to combine three factors that include pedagogy, game design, and learning content modeling into the educational game design. The focus of the game design is on multimodality and usability. Similarly, the focus of the pedagogical factor is learning outcomes and motivation theory. The factors of fun, problem-solving, and syllabus matching are also highlighted (Ibrahim and Jaafar, 2009).
EGameDesign	These design guidelines focus on enjoyment and knowledge enhancement for an effective educational game design. They intend

Design-focused frameworks/ models	Description
	to make the design and evaluation of web-based educational games more effective and less complicated, focusing on players' flows to enhance knowledge. The guidelines are based on four dimensions framework and emphasized the six levels of knowledge in Bloom's taxonomy to infer game task arrangement. The guidelines focus on the game goal and game style considerations, game interface considerations, and game task arrangement (Yu, Fu et al., 2009).
Experiential Gaming Model	The experiential gaming model is developed based on the idea of integrating experiential learning theory, flow theory, and game design. The model emphasizes the importance of clear goals, providing immediate feedback, and matching challenges to players' skill levels (Kiili, 2005a).
Framework for the analysis and design of educational games	This framework is developed based on existing components, including a method for specifying the educational objectives, principles for instructional design supported by empirical research in learning sciences, and a framework for linking game dynamics, mechanics, and aesthetics. The framework directs the levels which are essentials for an educational game to be effective. It discusses the three components: learning objectives, MDA (mechanics, dynamics, aesthetics), and instructional principles highlighting the support they can provide to game designers from the analytical angle (Aleven, Myers et al., 2010).
Game Factors and Game-Based Learning Design Model	This model underlines the fact that prior models are designed based on specific game genres, making them difficult to use when the target game genre is different from the default game genres applied in research. Therefore, this framework presents macro-level design concepts comprising of 11 key factors for the game design. The factors include game goals, game fantasy, game mechanism, game value, narrative, interaction, challenges, freedom, sociality, sensation, and mystery. The authors verify the usability of the model and performance of identified factors for designing educational games by analyzing two applications (Shi and Shih, 2015).
Game-Based Learning Guidelines	These GBL guidelines identify any possible issues in the game design, learning objectives, user interface, and child requirements. These guidelines primarily target GBLs' pre-production phase but can also be employed in the post-production phases (Alfadhli and Alsumait, 2015).
Level Up	The goal of Level Up is to build new modes to design and evaluate future game-based learning systems. The framework uses methods from intelligent tutoring system literature and maps empirical learning curves on student game-log data. The author hypothesized that the framework would increase educational games' production speed, increase the quality, and offer scientific evaluation of the games' educational content. The framework stresses the data-driven analysis of learning experiences through educational data mining, statistical techniques, and visualizations applied to game logs. However, the complete framework is not presented (Eagle, 2009).
RETAIN Model	The RETAIN model consists of six elements (relevance, embedding, transfer, adaptation, immersion, and naturalization). The model is

Design-focused frameworks/ models	Description
	constructed on instructional design principles and describes the notorious concepts between instructional design and game. It provides a common framework for educators and game designers by comprehending the effective integration of game and learning content to even them out (Zhang, Fan et al., 2010).
Usability guidelines for mobile educational games	These usability principles provide a conceptual overview of 'good practice' for designing and evaluating mobile educational games. The design principles are presented based on literature related to learning and games, interviews with educational game developers, game, game analyses, and game design theory. Five main categories are identified: adaptation, goals, challenge and mastery, context, and community and collaboration (Thomas, Schott et al., 2004).

The existing models, frameworks, and guidelines primarily focusing on GBL development are presented in Table 2.3. Only three frameworks specifically focused on educational game development. However, it is interesting to note that only one model actually presented GBL development phases and related activities, components, and deliverables for each phase. The rest only provides essential components and a mechanism to balance different components to support GBL development and equally useful for design and evaluation. All three frameworks/ models emphasized the pedagogical aspect. The game design aspect is emphasized by two frameworks focusing on the game definition, narrative, and gameplay. Moreover, flow, usability, and learner were each emphasized by only one of the three frameworks.

Table 2.3: Existing development-focused GBL models and frameworks

Development-focused frameworks/ models	Description
A novel framework for design and development of serious game	This framework is based on the study of serious games, activity theory, and multimedia learning to balance games and learning activities for the design and evaluation of serious games. In order to facilitate the balance between learning and game, a new pedagogical concept, "game design pedagogic plan (GDPP), is provided. The framework inherits the main components of activity theory (subject and object) and extends this using GDPP to facilitate learning in games. The pedagogic model maps to object and rules, the learner component reflects the subject, and the learning context maps to the community and divisions of labor (El-Sattar and Hussein, 2016).
Game Object Model II (GOM II)	This framework provides a theoretical basis to support the development of educational games. It also provides a mechanism to support the evaluation of computer game use in classroom settings from an educational perspective (learning versus instruction). The framework attempts to produce a dialectic between pedagogical and game elements. It includes abstract interfaces (components that promote educational objectives) and concrete interfaces (components that realize such objectives). The interfaces are prioritized from most

Development-focused frameworks/ models	Description
	to least important but without justification for this prioritizing. The interfaces include game space, visualization space, elements space, actors space, problem space objects, and social space objects (Amory, 2007).
mGBL engineering model	This model presents phases, activities, components, and deliverables for the mobile GBL development. It comprises two layers. The first layer (inner layer) presents three general phases (pre-production, production, and post-production). The second layer presents the components that must be included in each respective phase. The three general phases are sequentially executed, starting from pre-production and ending with post-production. The model is based on the GBL concepts and focuses on intertwining gaming and learning (Shiratuddin and Zaibon, 2011).

3 Theoretical Basis

This chapter presents the theories relevant to the research conducted in this thesis, grounding the doctoral research work. These theories define the approach adopted by the LEAGUÊ ideation toolkit in design workshops and are used as a lens for addressing and elaborating the concepts during GBL evaluation studies.

3.1 Flow Theory

Flow theory is one of the principal theories linked with engagement that provides a meaningful framework representing new qualities of experience, also known as optimal experience (Guo, Xiao et al., 2016). It is also equally relevant for describing the learning process and user experience (Vann and Tawfik, 2020). Flow theory introduces the flow state described by Csikzentmihalyi (1975) through studies on human experience involving activities such as rock climbing, dance, and chess. Flow is a mental state in which a person is entirely absorbed and engrossed in an activity that they lose the sense of time. In the flow, state self-consciousness can fade, and nothing else seems to matter during that time. It is referred to as the optimal experience (Csikszentmihalyi, 1991). During this mental state of flow, a person's engagement is often high and intense concentration on the task. Previous research has identified that flow positively impacts learning (Skadberg and Kimmel, 2004). Therefore, it should be considered when designing learning games.

The flow state is defined to have eight dimensions (Csikszentmihalyi, 1991) for an optimal flow performance. It includes *clear goals, immediate feedback, level of challenges well suited to personal skills, merging of action and awareness, focused concentration on the task, sense of potential control, loss of self-consciousness, and an altered sense of time*. The concept of rewarding or autotelic experience has been seen as the ninth dimension. When people feel it, they mention having experienced at least one or all eight dimensions. Therefore, an activity that produces a combination of these dimensions creates a deep sense of enjoyment that is rewarding and worthwhile. A person is willing to do such activity for its own sake without being concerned with external incentives. The flow theory (Csikszentmihalyi, 1991) has been applied in various domains, including educational game design (Kiili, 2005a, Amory, 2007). Further research showed that flow dimensions could be divided into three groups (see Figure 3.1): Flow antecedents, flow state, and flow consequences (Finneran and Zhang, 2005, Kiili, 2005b, Kiili, de Freitas et al., 2012).

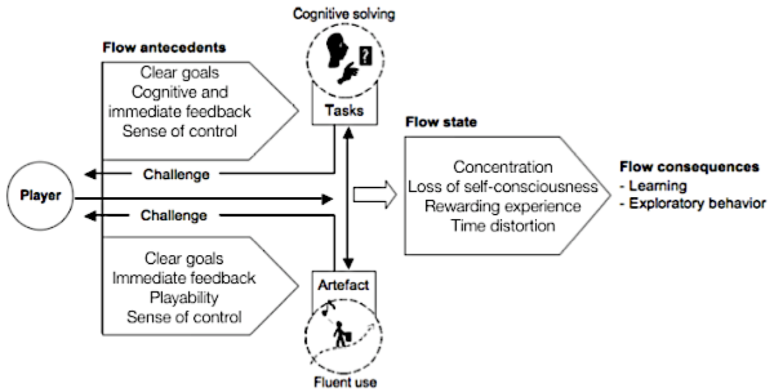


Figure 3.1: The flow framework (Kiili, de Freitas et al., 2012)

The flow antecedents contribute to the flow state. They are the factors (primarily based on original flow dimensions) that should be considered in designing educational games. The flow state is more abstract that contains dimensions that describe the flow experience feelings. The flow consequences are the positive attitudes, learning, and exploratory behavior. Kiili, Lainema et al. (2014) further extended the research and presented a flow framework for educational games to analyze the quality of educational games. The framework divided the flow elements into two groups: flow antecedents and flow state. They introduced five mind lenses to relate the flow dimensions to learning processes.

The challenge-skill pairing is an essential concept of experience. The three-channel model of flow can explain the relationship between the task challenge and the person's skill (Csikzentmihalyi, 1975, Csikszentmihalyi, 1990). According to Csikszentmihalyi (1991), the flow state is possibly preserved if the challenge level (difficulty) of the task increase to match the individual's developing skills as they proceed through a task (see, Figure 3.2).

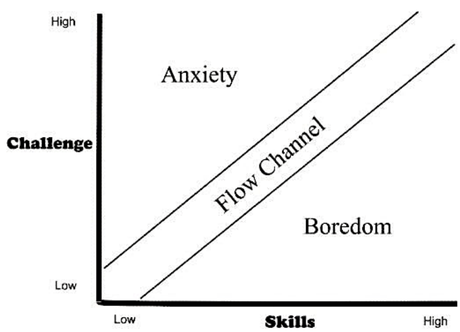


Figure 3.2: The three-channel model of flow (Csikszentmihalyi, 1990)

If the skills of the individual are not comparable to the challenges provided by the tasks, this results in anxiety. Although the tasks within the anxiety channel might be of interest, they are so challenging that the individual loses the motivation to persist. In contrast, if an individual's skills exceed the task's challenge, this results in boredom. In this channel, the individual disengages quickly from the activity and is no longer interested in the task. Therefore, the main goal is to fall between these channels and appropriately balance and maintain skill and challenge levels over time to realize flow. The individual should be challenged without being overwhelmed. Therefore, it can be said that flow occurs in the space between boredom and anxiety. Moreover, anxiety and boredom are the negative experiences that motivate an individual to strive for a state of flow.

Kiili (2005b) emphasized that it is possible to extend the flow channel by providing guidance to the players or the possibility of collaboratively solving the problems. Thus the original model was extended by adding the zone of proximal development (Vygotsky, 1980). It refers to the difference between what a player can do without and with help. The three-channel flow model plays a vital role in representing how the flow process acquire through a single activity that can provide insights into the learning activities (Pearce, Ainley et al., 2005).

In this doctoral work, flow dimensions have been incorporated into the developed framework. They also present a theoretical lens through which to view learners' experiences with GBL approaches for learning. In line with flow theory, an educational game is designed to keep learners engaged and interested in the learning activities by providing game tasks with varying levels of difficulty to impart learning. The flow theory was used to understand how the learning process acquires with the GBL approach by investigating the impact on learners' emotional states. The values of stress, engagement, interest, and relaxation provided useful insight when analyzed through the lens of flow theory. It also provides a basis to understand the cognitive load associated with task difficulty and the learner's cognitive skills in the GBL evaluation study conducted in this thesis. The following section discusses the cognitive theory.

3.2 Cognitive Theory

Cognitive theory is an approach to psychology that explains human behavior by understanding the thought processes and how learning is processed in the brain. This section emphasizes two predominant cognitive theories: Piaget's cognitive development theory and the cognitive load theory (CLT). These theories are briefly discussed to form a conceptual basis of learning. The theories worked as theoretical underpinnings and provided a useful theoretical lens to analyze the GBL evaluation studies' research findings in this doctoral work. Piaget's theory is incorporated in the GBL framework developed in this thesis. It is also used as a theoretical basis for analyzing the data from GBL evaluation with refugee children.

Piaget's cognitive development theory (Piaget, 1964) proposed that mental development takes place through certain stages in life, and the concept formation follows a consistent pattern. According to the theory, a child must pass through and experience a series of clearly definable stages in a set sequence. Piaget identified four stages in cognitive development: sensorimotor stage (infancy), preoperational stage (toddler and early childhood), concrete operational stage (elementary and early adolescence), and formal operational stage (adolescence and adulthood). Concreate behaviors are organized through assimilation and accommodation into patterns of behavior. These patterns are eventually internalized to turn into schemata or abstract models (Semple, 2000). According to Piaget's theory of accommodation (Piaget, 1976), if the learner receives new information (learning), then this information is easily absorbed into the memory. On the contrary, according to Piaget's theory of assimilation, if the newly received information contradicts the learners' existing information, then time is needed to adjust before this second information can be assimilated into the learners' memory (Atherton, 2005, McLeod, 2007). Although his ideas have been criticized, Piaget's work greatly influenced educational thinking and selection and development of educational software (Semple, 2000). Piaget's theory has been applied in game research to understand the play and its effects on cognitive development (Kamii and DeVries, 1980, Linaza, 1984, Weisskirch, 2003, Ojose, 2008). Moreover, researchers have used Piaget's theory as a lens or foundation in educational game research to understand GBL focusing on cognitive development, different operational levels of students, game design, children participation, presence, and learning in educational games (Corbeil and Laveault, 2011, Schifter, Ketelhut et al., 2011, Ni and Yu, 2015). Not many researchers directly incorporated Piaget's theory within their proposed GBL frameworks. Christinaki, Vidakis et al. (2014) extracted six vital elements using a framework for designing their educational games for teaching emotion identification skills to children with autism. The framework is grounded on the integration of Piaget's cognitive model and Kolb's experiential learning model. Tan, Ling et al. (2007) emphasized the importance of investigating and considering cognitive development for GBL design. They referred to Piaget's theory of cognitive development when explaining differences in children thinking with regard to their age as an example to recommend suitable games.

On the other hand, *cognitive load theory (CLT)* looks into the limitations and capacity of human memory in processing the received information (Sweller, 1988). It infers that humans have limited working memory. It is difficult for them to process the information that exceeds the amount limit at which they are overwhelmed (Paas, Renkl et al., 2004). Therefore, the representation of content and level of complexity can influence the amount of information a human brain can absorb (Yusoff, 2010). Effective cognitive processing predicts learning since a person has only a limited number of resources that are required to process the information (Mayer and Moreno, 2003, Paas and Ayres, 2014). There are three categories of cognitive load (Sweller, 1988): intrinsic load, extraneous load, and germane load. *Intrinsic load* is related to the task and its main features that must be processed, describing the active processing (keeping visual and verbal representations) within working memory. *Extraneous load* depends on the type of instructional design techniques or representation, including elements not related to learning, but learners need to process them (Korbach, Brünken et al., 2017). Lastly, *germane load* comprises information consolidation and involves scheme construction for enduring knowledge. This information is included in long-term memory (Sweller, Van Merrienboer et al., 1998, Paas, Renkl et al., 2003, Van Merrienboer and Ayres, 2005). It is important to note that for learning to occur, the total cognitive load must not exceed the available resources in working memory as CLT elements are additive. Therefore, it is also imposed by the effective instructional design of learning materials (Paas, Renkl et al., 2003).

According to Kiili, de Freitas et al. (2012), educational games must be designed in a way that all possible resources are available for processing the main task rather than for using game controls. Bad playability requires the player to sacrifice attention and thus cognitive resources for inappropriate activity. Extraneous load is particularly important in design as it can be manipulated directly by designers (Van Merrienboer and Ayres, 2005) by improving usability. The interface with poor usability increases the extraneous cognitive load obstructing meaningful learning. However, the learning task should impose a required germane load for knowledge construction, and all playing should not be effortless (Sweller, Van Merrienboer et al., 1998, Kiili, de Freitas et al., 2012). The germane cognitive load is important for learner engagement in the learning activity (Schrader and Bastiaens, 2012). Moreover, the complexity and amount of information that the learner is required to process combined with the simultaneous actions of motor and cognitive activities are an example of extraneous and intrinsic factors. These factors can overload the learner's processing resources during learning with games (Lim, Nonis et al., 2006, Kalyuga and Plass, 2009, Schrader and Bastiaens, 2012). Intrinsic load depends on the complexity of the game task in relation to the learner's expertise (Schrader and Bastiaens, 2012).

3.3 Co-Design

Co-design is creating new realms of collective creativity and is altering the landscape of design practice (Sanders and Stappers, 2008). Co-design can refer to arranging open innovation processes where people from different groups share and combine their knowledge and ideas. Moreover, it can also involve users in the design process as participants (Chesbrough, 2003). According to Burkett (2012), co-design means collaborative design. It involves collaborating and designing with different people that will deliver, engage, or use the product. Therefore, co-design is a method to actively engage several people involved in the product (such as users or stakeholders) in its design. Co-design builds on the tradition of participatory design (Schuler and Namioka, 1993). It comprises various approaches such as research-oriented (for example, applied ethnography), design-oriented (for example, using generative tools), or user involvement focused. The approaches involving user involvement range from usability testing (where designers/ researchers move toward users) to participatory design (where users move toward designers/ researchers) (Sanders and Stappers, 2008). Steen (2013) emphasizes that co-design is a process of collaborative design thinking. It can be understood as a joint inquiry and an imagination process in which different people together explore and define problems and develop and evaluate solutions. The participants share their experiences, discuss their roles, and negotiate their interests to bring positive change together.

Two imperative co-design strategies are considered when stakeholders are incorporated in the design process: informant design and cooperative inquiry (de Jans, Van Geit et al., 2017). Informant design is of more interest concerning the work involved in this thesis. It supports stakeholders' involvement in the design in ways to maximize the value of their contributions (Scaife, Rogers et al., 1997). Different stakeholders contribute specific information based on their skills and knowledge that shape the design at various phases (de Jans, Van Geit et al., 2017). Co-designing in a multidisciplinary team offers benefits such as shared knowledge, improves idea generation, provides access to users' experiences, increases speed to market, and produces better quality products (Roser and Samson, 2009, Steen, Manschot et al., 2011). A heterogeneous team offers different approaches and points of view to solve a problem (Dibitonto, Tazzi et al., 2017). Moreover, as stakeholders in the co-design process are professionals from different fields, it is essential to create a shared understanding. The team members must be able to integrate their knowledge bases (the knowledge can be of diverse content and structure) for sharing a common language in design sessions (Bittner and Leimeister, 2014, Dibitonto, Tazzi et al., 2017). On the other hand, the presence of stakeholders in the design process adds complexity (Prieto, Rodríguez-Triana et al., 2019). Some of the generic co-design challenges include the following: need to facilitate communication, decision making, social inclusion, manage power relationships, and encourage certain stakeholders (e.g., students) to challenge others' group ideas (e.g., their teachers) (Lee, 2008, Alvarez, Martinez-Maldonado et al., 2020).

Different co-design techniques tailored to the data-rich educational technologies' design are emerging. Researchers have proposed various co-design tools. Such as questionnaires to interrogate different stakeholders (teachers, researchers, and developers) regarding their particular views (related to data, learning, and technology) and comment on others' perspectives (Prieto, Rodríguez-Triana et al., 2019). Document replay enactments to facilitate teachers with different data representations to provide early feedback (Holstein, McLaren et al., 2019). Moreover, many researchers proposed design games and card-based tools for co-design (Kwiatkowska, Szóstek et al., 2014, Vaajakallio and Mattelmäki, 2014). The card-based co-design tools are proposed in different domains, including learning analytics (Alvarez, Martinez-Maldonado et al., 2020), IoT design (Dibitonto, Tazzi et al., 2017), digital services for the bus context (Hildén, Ojala et al., 2017), and games (Huang, Ng et al., 2020). Few researchers have emphasized the values of co-design in the development of educational games. It can be used as means of additional ideas over other methods (such as expert consultation or analysis), leading to creating more effective games (All, Van Looy et al., 2013). According to Khaled and Vasalou (2014), participatory design methods can improve domain expertise and procedurality in serious games. It can support the existing knowledge of players, making it possible to co-design serious games involving users.

Researchers have employed different techniques for co-designing educational games. de Jans, Van Geit et al. (2017) utilized a methodological co-design framework to improve stakeholder involvement for designing mini-games. Tobar-Muñoz, Baldiris et al. (2016) proposed a co-designing approach to creating augmented reality GBL involving teachers and designers. Penuel, Roschelle et al. (2007) defined co-design as a highly-facilitated process in which developers, teachers, and researchers work together (as a team) to design, realize, and evaluate educational innovation. Jessen, Mirkovic et al. (2018) explored new ways of using participatory design methods for co-designing gameful health apps. Díaz, Paredes et al. (2012) used embodied narratives technique to co-design social games with children. da Costa, Rebelo et al. (2017) described a co-design process based on a user-centered design approach in defining the concepts of a civic educational game.

In this doctoral work, co-design practices have been used in the design workshops to drive the ideation phase of educational game design. The objective is to use a co-design process where GBL stakeholders are brainstorming, ideating, and designing educational games for different learning domains, game genres, technology, and users. A card-based design toolkit was employed as a playful approach for designing learning games in a workshop format.

In addition to the theories described above, it is important to note that the investigation work that led to the development and evaluation of the LEAGUÊ framework was grounded in additional theories. Such as experiential learning, ARCS model, constructivism, and most importantly, usability and HCI guidelines. These theories are situated in existing frameworks and literature in GBL (see Chapter 2) and incorporated in our study (P2).

4 Research Methodology

This chapter presents the research methodology adopted for the research work described in this thesis. First, the design science research (DSR) followed as the research approach for this doctoral work is described. Next, the mixed methods design adopted as the research strategy (connected with DSR) is explained, and the methodological choices are presented. Finally, the research activities conducted in this doctoral work are described detailing DSR and mixed methods cycles, evaluation studies, and methods used for data collection and analysis to address the research questions.

4.1 Design Science Research Approach

The methodological approach adopted in this doctoral work is design science research (DSR) (March and Smith, 1995, Hevner, March et al., 2004, Hevner and Chatterjee, 2010). DSR is adopted considering the research context of this thesis, which investigates the game-based learning (GBL) phenomenon and facilitates the GBL design and evaluation process to improve learning games' effectiveness (RQ1-RQ4). The design science approach meets the aim of this research work. It provides theoretical tools to study and understand a particular domain, together with processes to build artifacts to improve an environment (Simon, 1996, Simon, 2019). Furthermore, the focus of DSR on rapid iterations between the artifact's construction and their evaluation (Hevner and Chatterjee, 2010) makes a suitable approach for the investigation of the research questions RQ2-RQ4. The doctoral work unfolded by performing review studies to understand the GBL domain and identify needs, challenges, and research opportunities, which served as input for building design artifacts with subsequent iterations and evaluating them to address the research problems. The doctoral work adds to the knowledge base by validating and extending the GBL concepts and presenting guidelines and design recommendations.

Hevner (Hevner, March et al., 2004, Hevner, 2007) describes how to perform DSR by explaining three coupled cycles (relevance, rigor, design) of activities (see Figure 4.1). The existence of each of these three cycles is a must in design science research. The relevance cycle input the requirements (to design cycle) and involves designing and conducting the field studies for exploratory or evaluation purposes. The rigor cycle includes both a continuous process of informing (design cycle) about the relevant theories for grounding the design work and adding to the knowledge base by validating and extending those theories, thus maintaining an innovative approach and enabling research contributions. Lastly, the design cycle is the primary cycle where actual design work is carried out. It continuously iterates between building the artifacts and evaluating them.

While the design cycle is independent in actual research execution, it is essential to understand its dependencies on the other two cycles. With the relevance cycle, it is fed with requirements and returns back artifacts for the field testing. With the rigor cycle, it is fed with theories and returns theoretical knowledge. Section 4.3.1 explains the three DSR cycles for this doctoral work and the research activities conducted in each cycle.

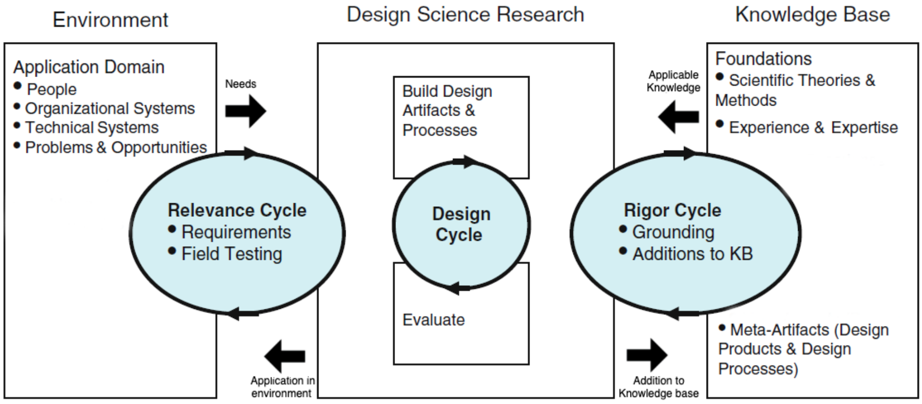


Figure 4.1: The design science cycles, figure adapted from (Hevner, 2007)

Hevner et al. (Hevner, March et al., 2004, Hevner and Chatterjee, 2010) also provide seven guidelines for conducting and evaluating DSR: design as an artifact, problem relevance, design evaluation, research contribution, research rigor, design as a search process, and communication of research. These guidelines are followed for the DSR presented in this thesis. March and Smith (1995) presented a framework for design and natural science research. According to March and Smith (1995), there are two research activities in design science: build and evaluate, the first two columns in Figure 4.2. In comparison, the last two columns are for natural science research. Moreover, the research output or artifacts of DSR can take several forms, including constructs, models, methods, and instantiations. The research work carried out in this thesis focus on the activities of design science: building and evaluating, and the main research outcomes of the thesis are a model (LEAGUÊ framework), methods (Ten-step process for transforming framework into design cards, and LEAGUÊ-GQM approach), and the instantiations (GBL ideation toolkit and GBL analysis instrument). Although the conducted research attempts to theorize and justify where possible by determining how and why the developed artifact worked or did not work in the applied environment, this thesis's main focus is on DSR activities. Figure 4.2 shows the research framework by March and Smith (1995) and the dots indicate the positioning of research work in this thesis.

Research Activities

Research Outputs		Build	Evaluate	Theorize	Justify
	Constructs				
	Model	●	●		
	Method	●	●		
	Instantiation	●	●		

Figure 4.2: DSR activities and outputs in this doctoral work, the framework adopted from (March and Smith, 1995)

Some researchers (Venable, 2006, Gregor and Jones, 2007) have distinguished process artifacts from product artifacts. The former are technologies (anything from diagrams and tools to software) used by people to accomplish a task, and the latter are procedures (methods, approaches) that guide or tell someone how (what to do) to accomplish a task. This thesis focuses on both product (model and instantiations) and process artifacts (methods).

The key purpose of the evaluation activity in DSR is to demonstrate the utility of the developed artifact, i.e., to determine whether or how well it achieves its purpose. The second purpose of DSR evaluation (from the design theory viewpoint) is to validate or enhance the design theory (Venable, Pries-Heje et al., 2012). Peffers, Tuunanen et al. (2007) divided the evaluation into two activities: demonstration and evaluation. The demonstration is a light-weight evaluation used to demonstrate that the developed artifact achieves its purpose at least in one context, mainly to prove that the idea works. The demonstration proves that the artifact solves one or more instances of the problem, as Peffers demonstrates the use of DSRM with four case studies (Peffers, Tuunanen et al., 2007). On the other hand, proper evaluation is extensive and more formal, requiring the knowledge of appropriate metrics and analysis techniques and including empirical evidence or logical proof (Peffers, Tuunanen et al., 2007). According to Sonnenberg and Vom Brocke (2011), a demonstration can be conducted using methods such as demonstration with the prototype, experiment with the prototype, experiment with the system, benchmarking, surveys, expert interview, and focus group. In comparison, an evaluation activity is typically conducted by applying the following methods: field experiment, case study, survey, focus group, or expert interview. The evaluation activity results might stimulate further iterations (concerning the same or adapted problem statement) through the process of DSR. This doctoral research used demonstration for examining process artifacts (the ten-step process for transforming framework into design cards and the LEAGUÊ-GQM approach) and evaluation for examining product artifacts (LEAGUÊ framework and GBL ideation toolkit). The evaluation of the LEAGUÊ framework followed the second purpose of DSR evaluation, i.e., to validate or enhance the developed GBL phenomenon.

4.2 Research Strategy and Methodological Choice

In this thesis, design-science is combined with mixed methods research to obtain completeness, confirmation, and compensation. Mixed methods research combines quantitative and qualitative research methods in the same research inquiry (Venkatesh, Brown et al., 2013). As highlighted by Cleven, Gubler et al. (2009), a mixed methods approach can be used in DSR to provide deeper insights. In this thesis, mixed methods research complements DSR to fulfill our research objectives (defined in Chapter 1) because of its potential to understand and explain complex phenomena (Mingers, 2001, Cao, Crews et al., 2006). This methodological choice has recently attracted much attention in educational technology research (for a review, see (Hung, Yang et al., 2018, Bond, Buntins et al., 2020).

This doctoral research work followed the exploratory-triangulation design for mixed methods research as described by Kwok (2012). This mixed methods design combines the benefits of exploratory and triangulation design, characterized by an initial qualitative phase, followed by a combination of quantitative and qualitative phase of data collection and analysis. The exploratory phase is useful to explore a phenomenon, and triangulation design helps obtain a broader, more complete picture by collecting both types of data, quantitative and qualitative, to converge and enhance the validity of findings (Creswell and Plano Clark, 2011). Three cycles of exploratory-triangulation mixed methods design are performed in this doctoral work: first focusing on *GBL phenomenon*, second on *GBL design process,* and third on *GBL evaluation process*. The mixed methods design is integrated within the three DSR cycles. Each cycle of mixed-method research starts with a review study that constitutes the exploratory phase of exploratory-triangulation design in mixed methods research and represents the rigor cycle of DSR. Next, the artifacts developed in the design cycle of DSR are evaluated using evaluation studies. The quasi-experiments and design workshops constitute the triangulation phase (collect quantitative and qualitative data) of exploratory-triangulation design and represent the relevance cycle of DSR. A mix of qualitative and quantitative methods accounts for user studies' unpredictability (Rogers, Connelly et al., 2007). To implement the main research strategy, we adopted several methods in this doctoral research. During the research work, observations, focus groups, questionnaires, and video recordings were usually employed to evaluate the design artifacts, the process, and the perceived user experience during design workshops. Whereas pre-post-tests, field notes, interviews, questionnaires, EEG, observation checklist, and game logs were the primary means for collecting data in quasi-experiments for evaluating GBL approaches with the users. Table 4.1 presents an overview of all methods used in user studies.

Table 4.1: Quantitative and qualitative methods used

Quantitative		**Qualitative**	
Data collection	**Data analysis**	**Data collection**	**Data analysis**
– Pre and Post Test – Questionnaire – Usability test – Game logs – Artifacts – EEG – Checklist	– Descriptive statistics – Pearson correlation – Pearson-product moment correlation – Wilcoxon signed-rank test – Spearman correlation – Mann–Whitney test	– Semi-structured interviews – Observations – Video recording – Focus Group – Documents/articles	– Categorization of their main elements – Directed content analysis – Quasi-formal comparison technique – Ground theory approach

4.3 Research Activities

This section details the different research activities performed during the doctoral research work that contributed to the three cycles of the DSR methodology. There are two main types of research activities conducted in this thesis: *review studies* and *evaluation studies*. During the course of this doctoral work, a total of three review studies were conducted (including two systematic literature reviews and one comparative analysis), and nine evaluation studies were conducted (including three quasi-experimental studies (GBL evaluations), three design workshops, and three case studies). Table 4.2 summarizes the research activities. The quasi-experiments investigated the GBL phenomenon, learning process, and contributing factors by focusing on evaluating GBL approaches' effectiveness in three different learning setups: *formal learning, non-formal learning,* and *informal learning*. The design workshops explored the scaffolding provided by the card-based toolkit for the GBL design process. The case studies examined the learning game evaluation process, including two case studies on learning game analysis and one project analysis (focusing on empirical evaluations of learning games within the project EduApp4Syria). The evaluation studies are described in detail in Section 4.3.3. The next sections explain the sequence and contribution of these research activities in DSR and mixed method research cycles.

Table 4.2: Summary of research activities

Research activities	Description	Papers	Research questions
Review study 1	Systematic literature review on state of the art in GBL	P1	RQ1
Review study 2	Comparative analysis of educational game design models.	P5	RQ1
Review study 3	Systematic literature review on methods and guidelines for child-computer interaction (CCI) research with refugee children.	P8	RQ1
Quasi-experiment 1	Investigate the GBL approach's learning process (focusing on LEAGUÊ framework concepts) by monitoring brain waves and comparing it to traditional pen-and-paper learning.	P3	RQ2,3,4
Quasi-experiment 2	Investigate the GBL approach's impact (focusing on LEAGUÊ framework concepts) for improving students' engagement and learning in online education.	P4	RQ2,3,4
Quasi-experiment 3	Evaluate the GBL approach's effectiveness (focusing on LEAGUÊ framework concepts) for Arabic reading skills of migrant refugee children in an informal learning setup.	P9	RQ2,3,4
Design workshop 1	Employ card-based toolkit for ideating learning game design with Computer science and engineering students in university	P6, P7	RQ3
Design workshop 2	Employ revised card-based toolkit for ideating learning game design with Technology Enhanced Learning (TEL) researchers in a doctoral summer school	P6, P7	RQ3
Design workshop 3	Employ revised card-based toolkit for ideating learning game design with master students in a game development course	P6, P7	RQ3
Case study 1	Empathy game evaluation with LEAGUÊ analysis instrument carried out by the game developer.	P2	RQ4
Case study 2	VR game evaluation with revised LEAGUÊ analysis instrument carried out by the game developer in two iterations using player feedback	_*	RQ4
Case study 3	Examining field studies in the project "EduApp4Syria" to investigate the evaluation methods used and practical, ethical, and methodological challenges in conducting GBL evaluation research with refugee children.	P10	RQ4

* The study is presented in the master's thesis (Karlstrøm and Markussen, 2020)

4.3.1 Design Science Research Cycles

This section describes the activities during the progress of the research concurrently unfolding intra and inter rigor, design, and relevance cycles of DSR. This design science research developed a GBL framework (LEAGUÊ) for learning games design and evaluation (reported in P2) through directed content analysis of existing knowledge. This framework was used to design and develop three instruments/tools: a card-based toolkit

(P6), an analysis instrument (P2), and an evaluation guide for the LEAGUÊ-GQM approach (P9). In addition, a ten-step process was devised to transform a framework into design cards building on the experience of designing the toolkit and the domain knowledge acquired through literature. The artifacts are evaluated using quasi-experimental studies, design workshops, and case studies, and the results are used to validate and extend the GBL knowledge base and refine the artifacts. Figure 4.3 presents the complete picture of research activities and integration of mixed-method research within three DSR cycles. The DSR cycles are described below whereas, the mixed-methods cycles are explained in the next section.

- **Rigor Cycle**

The work started by conducting a review study on the respective topic (P1, P5, P8), each time before we designed artifacts and conducted the user studies. These review studies represented the starting point to understand the GBL domain and identify the needs, challenges, and research opportunities that build a solid knowledge foundation and theoretical background to support and drive the subsequent research. The findings from review studies provided significant input and emphasized the need to create meaningful artifacts (framework, tools, methods) to solve the identified problems. Therefore, throughout the process, GBL domain knowledge acquired through systematic literature review studies, comparative analysis, and the literature on HCI, psychology, and co-design (Chapter 2 & 3) informed the design work.

Also, in the final iterations of the work, the outcomes of DSR contribute to the knowledge base as part of the rigor cycle. The produced artifacts, design recommendations, gained experiences, valuable insights, and knowledge from evaluation studies improve state of the art and validate and enhance GBL concepts and domain knowledge. Research outcomes are reported in ten research papers (Chapter 5), published (or ready for submission) in academic conferences and journals from which research contributions (Chapter 6) emerged, thereby closing the rigor cycle loop and providing tangible contributions to the field of research.

- **Design Cycle**

The central course of doctoral research concentrated on building and evaluating the artifacts (LEAGUÊ framework, analysis instrument, ideation toolkit, ten-step transformation process, and integrated evaluation approach and guide) targeting the GBL design and evaluation phases. The activities iterated between turning early design ideas into new prototypes and consequent evaluations, often leading to design iterations. The evaluation of the artifact (process and product) can be conducted by selecting appropriate methods from a range of evaluation opportunities (Pries-Heje, Baskerville et al., 2008). The process artifacts were evaluated by demonstration, i.e., applying the developed artifact in a relevant and suitable context to prove the idea. The ten-step process was demonstrated by using the steps to transform the LEAGUÊ framework into the GBL ideation toolkit (reported in P6). Similarly, the LEAGUÊ-GQM approach was demonstrated by developing an evaluation plan for conducting quasi-experiment 3 (reported in P9). However, the product artifacts (model and instantiations) went through proper evaluation. The analysis instrument was examined using two case studies (case study 1&2) whereas, the LEAGUÊ framework and GBL ideation toolkit were evaluated in user studies (quasi-experiments and design workshops, respectively) with subsequent iterations. Other activities involved understanding and investigating the methods and guidelines for conducting learning game evaluation with refugee children, connecting and extending the evaluation phase with this special user group. Table 4.3 presents an overview of the developed artifacts in relation to the papers and research questions that describe the work. Building all the artifacts involved only material development; software and hardware technology were not involved.

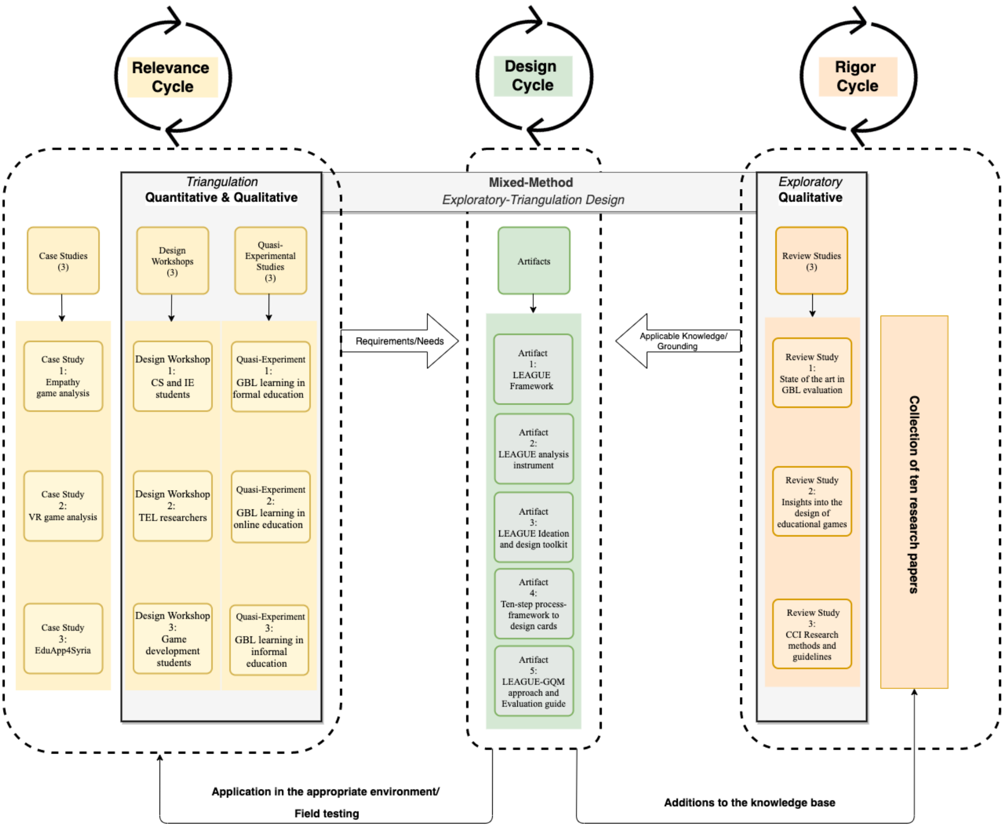


Figure 4.3: Research activities and mixed-method strategy in DSR cycles

Table 4.3: List of artifacts built

ID	Name	Type	Released	Paper	Research question
Artifact 1	LEAGUÊ framework	Product	Autumn 2017	P2	RQ2
Artifact 2	LEAGUÊ analysis instrument	Product	Autumn 2018	P2	RQ4
Artifact 3	LEAGUÊ ideation toolkit	Product	Spring 2019	P6	RQ3
Artifact 4	Ten step process- transform framework to design cards	Process	Spring 2019	P6	RQ3
Artifact 5	LEAGUÊ -GQM approach and evaluation guide	Process	Spring 2018	P9	RQ4

- **Relevance Cycle**

The evaluation of the artifacts (from the design cycle) in the application domain often recurred to new evaluation studies to keep the design process updated with new requirements. The subsequent iterations both validated and improved the artifacts, building on the experience and insights gained from evaluation studies. The design workshops and GBL evaluations with participants facilitated discussions, triggering a better understanding of the domain and leading to refinements and design recommendations. Furthermore, the evaluation studies also contributed to valuable insights by facilitating reflection on employed methods bringing new perspectives and guidelines, and improved knowledge resulted as an outcome.

4.3.2 Mixed Methods: Exploratory-Triangulation Design Cycles

This section describes the three cycles of exploratory-triangulation mixed methods design performed in this doctoral work for investigating the research questions. As previously presented in Figure 4.3, mixed methods design is equated with the three DSR cycles. We will present each of the three exploratory-triangulation design cycles of mixed method research, explaining each cycle's specific activities and their relevance to the research questions. Figure 4.4 presents the three mixed-method cycles.
Overall, the exploratory phase of the mixed methods design constitutes the qualitative studies comprising three review studies and qualitative content analysis. The review studies together answer RQ1 by presenting state of the art in GBL and identifying the existing issues, needs, and opportunities in GBL design and evaluation practices. Similarly, overall, the design's triangulation phase constitutes the combined quantitative and qualitative studies comprising all user studies (quasi-experiments and design workshops).

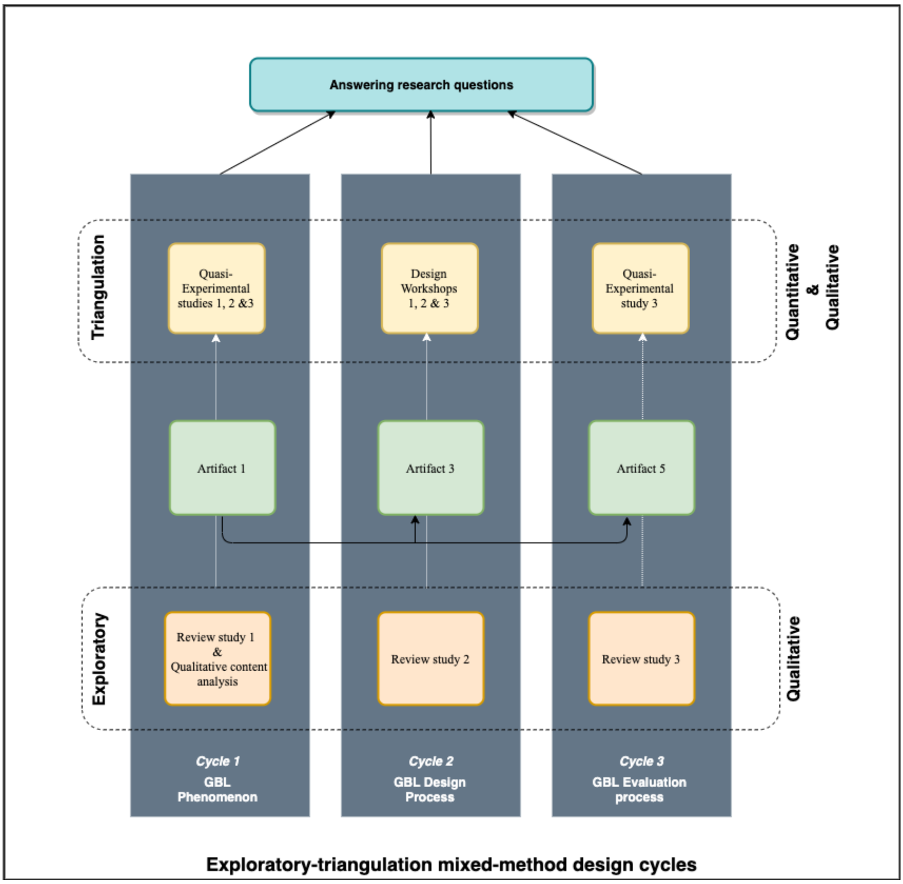


Figure 4.4: Exploratory-triangulation design cycles in mixed methods research

Each of the three complete cycles investigates one research question (RQ2-RQ4), focusing on the *GBL phenomenon*, *GBL design process,* and *GBL evaluation process*. The exploratory-triangulation design was selected as a mixed methods strategy because it fits our research objective.

- **Cycle 1: GBL Phenomenon**

The first cycle of mixed method design investigated the GBL phenomenon finding the answer to research question RQ2. The initial qualitative exploratory phase began with review study 1, in which we performed a systematic literature review on state of the art in GBL (reported in P1). We aimed to systematically summarize the current trends, existing design and evaluation approaches, and GBL evaluation criteria. Our findings highlighted wide diversity and inconsistent use of GBL elements across the studies and emphasized the need for a comprehensive framework for GBL design and evaluation.

The review study's findings provided input for the qualitative content analysis that resulted in the LEAGUÊ framework (reported in P2), thus completing the exploratory phase. The developed LEAGUÊ framework was examined in the triangulation phase by conducting three GBL evaluation studies (quasi-experiments) in different contexts: formal, non-formal, and informal learning setups (reported in P3, P4, and P9), investigating the relationships among different GBL elements of the framework. The framework became a salient input in quasi-experiments regarding 1) the choice of GBL-related concepts and metrics, 2) questionnaire development, and 3) validation of relations and patterns.
The adopted strategy supported our research objective in that the exploratory design allowed us to reveal in-depth information about the GBL phenomenon then use quantitative methods to test some relationships among the constructs in the developed GBL framework. The triangulation design helped make the conclusion more convincing by comparing the data from different sources that confirm and compensate each other (Kwok, 2012).

- **Cycle 2: GBL Design Process**

The second cycle of mixed method exploratory-triangulation design explored the GBL design process addressing RQ3. This cycle began with review study 2, in which we performed a comparative analysis of educational game design models (reported in P5). The objective was to understand GBL design by analyzing essential attributes in different existing educational game design models, highlighting their strengths and weaknesses, and identifying challenges that require further research. The findings identified the lack of tool support and the absence of concrete guidance for model application. Furthermore, it emphasized the need for empirical validation and independent evaluations. The exploratory phase results directed the transformation of the LEAGUÊ framework (from cycle 1) into a card-based GBL ideation toolkit and a ten-step process for converting framework to design cards (reported in P6). The developed toolkit was evaluated in the triangulation phase by conducting three learning game design workshops (reported in P6 and P7). The findings detail the toolkit's strengths and limitations to support the early design phase of educational game development and scaffolding for completeness and collaboration in a multidisciplinary GBL team.

- **Cycle 3: GBL Evaluation**

The third and last cycle focused on methods, tools, and guidelines to facilitate the GBL evaluation process contributing to RQ4. The third GBL evaluation study/quasi-experiment aimed to investigate the GBL approach in an informal learning setup. Correspondingly, an NTNU partnered project, "EduApp4Syria," aimed at a similar goal of developing language learning games for Syrian refugee children. These two corresponding goals, the Syrian crisis (Yazgan, Utku et al., 2015) and the role educational technology can play (Drolia, Sifaki et al., 2020) directed this cycle's focus and the doctoral research work to include refugee children as the target group. We conducted the third quasi-experimental study as part of the “EduApp4syria” project. This special user group's inclusion consequently stimulated research in CCI (children-computer interaction) and GBL research targeted at refugee children. It was essential to get the required knowledge to conduct the third GBL evaluation study (reported in P9) and investigate the GBL approach for informal learning with this user group. Moreover, review study 1 (P1) also emphasized this research need and identified that current GBL evaluation approaches do not consider children's requirements and needs. Also, the majority of educational game evaluation studies do not use existing frameworks or guidelines/approaches.

This cycle began with review study 3, in which we performed a systematic literature review on available research guidelines and methods for CCI research with refugee children (reported in P8). The aim was to summarize the available literature that can guide researchers in the CCI community for conducting evaluation studies (focusing on learning games in this context) with this user group and identify the need for additional research to address specific requirements of this group. The finding of this review study served as a starting point presenting research guidelines and evaluation methods for conducting quasi-experiment 3 and also directed the need to develop the LEAGUÊ-GQM approach and evaluation guide (based on the LEAGUÊ framework) for creating a GBL evaluation plan. The developed approach was employed in the triangulation phase for collecting qualitative and quantitative data. The experience and insights from the evaluation study 3 (reported in P9) provided data regarding guidelines and methods employed for GBL evaluation research with refugee children (reported in P10) in addition to the investigation concerning the effectiveness of the GBL approach for informal learning. Thus, achieving the aim of this cycle contributing to facilitate the GBL evaluation process.

4.3.3 Evaluation Studies

This section gives an overview of the evaluation studies designed and conducted in this doctoral work to understand the GBL domain, familiarise with the design and evaluation process and evaluate artifacts produced during the design cycle. The evaluation studies in this PhD research are composed of three main investigation methods: Quasi-experiments (GBL evaluations), design workshops, and case studies. A total of three quasi-experimental studies, three design workshops, and three case studies were conducted during the course of this PhD. These evaluation studies were used both to validate the developed artifacts and to inform the design iteration. Insights concerning possible improvements were acquired from the experience and results of the evaluation studies. Below, the design and implementation of the three types of evaluation studies is described elaborating on the purpose, setting, design, participants, and the applied approaches for data collection.

- **Quasi-Experimental Studies**

The *first quasi-experimental study* investigated the learning process with a GBL approach compared to a traditional pen-and-paper approach (reported in P3). Figure 4.5 illustrates a picture from the experiment.

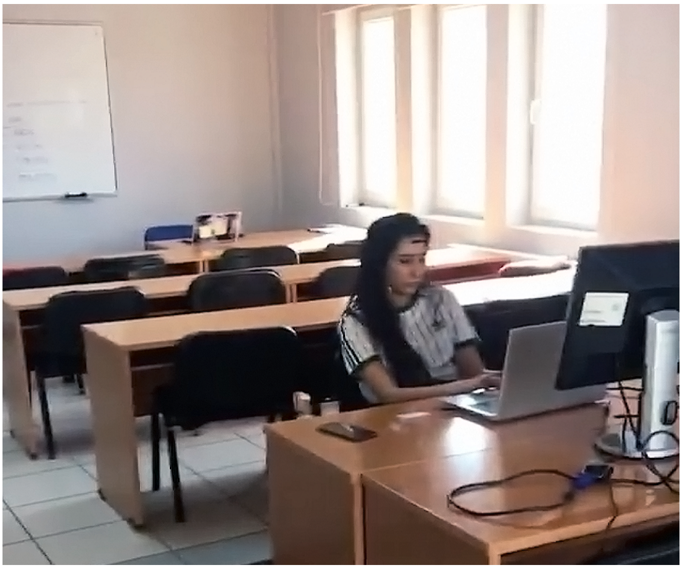


Figure 4.5: First GBL evaluation study in a formal learning setup

This experiment's main objective was to evaluate students' emotional states and learning outcomes on an algorithmic problem (related to a sorting algorithm) when provided with a learning game compared to using printed notes for the same scenario to investigate the GBL phenomenon. The experiment was organized and conducted in the university with two different groups (one using the GBL approach and the other using the pen-and-paper approach). An insertion sort problem was selected for this experiment, and a learning game called "Sort Attack" was used. The experiment consisted of three parts: pre-test, problem-solving exercise, and post-test. A clear set of instructions were provided to both groups concerning the experiment. The sample consisted of twenty-two first-year undergraduate computer science students 18-25 years old (7 females and 14 males) who did not have any formal course (as part of their degree) on the algorithm subject. The students were divided into two groups (each with 11 students). The participants were asked for written consent for data collection and were informed about the research objective, data collection process, and voluntary participation in the experiment. The data was collected from different sources using quantitative and qualitative methods such as EEG headset, questionnaire, pre/post-test, and observation. The data was triangulated and cross-referenced to warrant our interpretations. The evaluation's objective was to test the

GBL phenomenon's hypothesis in the developed framework to validate the framework concepts and enable refinement and enhancement. The analysis focused on how the learning process acquired (in terms of learning outcomes and emotional states) while gaining new knowledge through GBL compared to a traditional learning approach. The analysis examined the impact of affective-cognitive reactions on the learning gains using a GBL approach and identified some issues in the employed game for further improvement to build a more robust and engaging GBL platform.

The *second quasi-experimental study* investigated the potential of using a GBL approach for improving student engagement and their learning curve in online education (reported in P4). Figure 4.6 illustrates a picture from the experiment. This study's main objective was to evaluate using a GBL tool (Kahoot!) in the online Human-Computer Interaction (HCI) course amid the COVID-19 pandemic. There was a need to assess GBL technology's effectiveness in the online learning process and investigate its effects on students' engagement, motivation levels, and learning gains. This study was organized and conducted in online lectures during the quarantine period. Two digital quiz platforms were used in this experiment: Kahoot! and Google Form Quiz. The experiment was conducted in two iterations with different procedures and focus. The first iteration focused on students' engagement and interactivity with the two employed quiz platforms, and the second attempted to evaluate the learning curve using Kahoot! platform. In the first experiment, the participants were divided into two groups (approximately 150 students in each). The control group used the Google Quiz (non-game-based digital quiz), and the experimental group used Kahoot! (game-based digital quiz) for the in-lecture quiz. The quiz was implemented in the middle of the lecture (using the specified platform), and at the end of the online class, students were asked to fill the survey. For the second iteration, the Kahoot! platform was used for quizzes (pre/post-test) every two weeks with standard online teaching in-between weeks for a period of four months (March 2020 to June 2020). This experiment started with a pre-test followed by the online instruction and ended with a post-test in order to measure the learning curve. The sample for this study was students (18 -25 years old) taking the online HCI course. A total of 261 (72 Female / 189 Male) students participated in the first iteration and, 243 (75 female and 168 male) students participated in the second experiment. Both quantitative and qualitative data were collected using digital quiz platforms from pre/post-test, survey questionnaires, and student feedback. In addition, the instructor also made observations during the online classes concerning student's behavior and participation. This evaluation study's objective was also two-fold: to assess the GBL approach's effectiveness in online education and to study, validate and enhance the developed framework dimensions within this experiment. The analysis focused on student engagement, learning curves, game factors, and usability when using game-based digital quizzes and examined their interrelation. The analysis also identified the factors that contributed most to the online learning experience with the GBL approach and suggested key points giving new directions for effective use of the GBL approach in online learning and factors that can improve GBL applications' effectiveness.

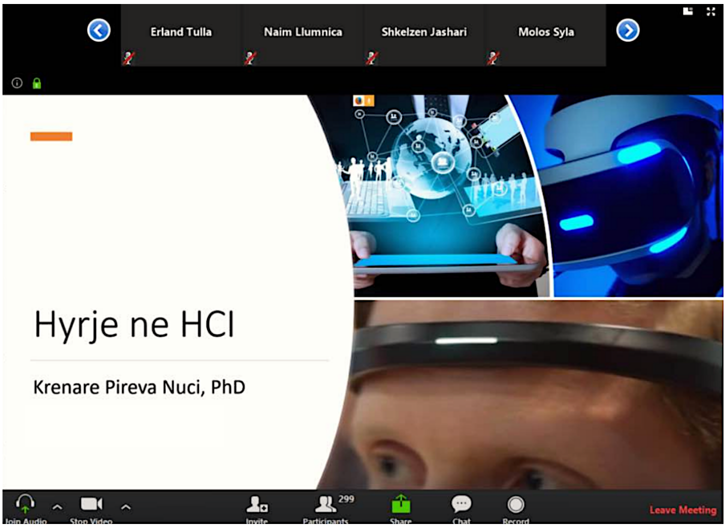


Figure 4.6: Second GBL evaluation study in online learning setup

The *third quasi-experimental study* investigated the potential of a GBL approach in an informal learning environment. Figure 4.7 illustrates a screenshot of the game screen from the experiment. The user study's main objective was to evaluate a language learning game, "Feed the Monster," for improving the Arabic reading skills of migrant refugee children (reported in P9). The quasi-experiment was organized and conducted as an out-of-school activity in a largely informal setting with a one-group pre-post-test design. No control group was used considering the context of refugee children where no other means of education is available for comparison. The experiment consisted of two parts (a playtest session and one-week play at home) and was one-week-long. The playtest session was 40-55 minutes in duration and started with a pre-test; next, the gameplay session began where children played the game for 20-25 minutes, and subsequently, a short follow-up interview was conducted at the end. Afterward, the mobile devices (with the installed game) were handed to parents for children to play the game at home for one week (at least 20 minutes daily). The user study was conducted in spring 2018 from April 7th to May 6th in Trondheim, Norway. The sample was thirty children from a migrant refugee background (aged 5–10 years old) who could speak Arabic but could not read and write in Arabic. A total of 14 girls (mean age: 7.14, SD: 1.875) and 16 boys (mean age: 7.125, SD: 1.746) participated in the study. The parents were contacted to obtain consent for the data collection. The children were also asked for their consent, and their participation in the study was completely voluntary. The study collected quantitative and qualitative data using pre/post-test, observation checklist, game logs, questionnaire, interview, and notes. As defined by Venable, Pries-Heje et al. (2012), the objective of the evaluation was to validate the GBL concepts in the developed framework, investigate the relationship between different GBL elements, and further enhance the GBL applications by presenting design recommendations. The analysis focused on the learning gain with the GBL approach. Moreover, it investigated the effect of user characteristics (age group, learning modality preferences, and mobile usage experience) on learning, usability, and gameplay performance, exploring the GBL phenomenon and the relation between different factors. The analysis also identified issues faced by children when playing with the learning game, feeding the subsequent game iterations and providing guidelines for designing effective learning games.

Figure 4.7: Third GBL evaluation study in informal learning setup

- **Design Workshops**

The design workshops (reported in P6 and P7) were organized and conducted in three different contexts: as a research study (with computer science and engineering students), in a doctoral summer school (with researchers in technology-enhanced learning (TEL), and in a graduate course "Game development" with students. The typical setting for the design workshops was a workshop session (approximately two-hour duration) where participants work in teams (of four to six members) for a collaborative design process. Design workshops' objective was to generate and develop a design idea for a learning game specific to a learning domain, game genre, and target users. Two organizers facilitated the design workshops to lead the teams through the ideation session by presenting the design activities. The workshops started with a short introduction to GBL concepts and the toolkit description. Subsequently, teams start ideating the learning game design using the toolkit, completing five design activities. Figure 4.8 presents a picture from the design workshop.

The workshop participants were usually master students and researchers aged between 25 and 40 and had no to moderate learning game design experience. The sample consisted of 34 participants in total, 21 males and 13 females. Two design workshops were conducted during spring 2019 and one in autumn 2019. The participants were explained about the study's research objective, informed that their participation is voluntary, and asked to sign a consent form.

The primary means to collect the data were a questionnaire, focus group, observations, and pictures of the produced game design artifacts. In addition, video recordings of few teams were collected during some of the design workshops. The role of the workshop organizers was to present the design activities and introduce the concepts of GBL. During the ideation phase, their main task was to observe the teams' work and make notes without intervening. Occasionally, the participants asked for some clarifications, and the organizers provided the required help. The collected data was analyzed using both quantitative and qualitative methods. The focus of the analysis was two-fold. First, to validate the employed toolkit's value and utility in terms of fun, satisfaction, understandability, and usefulness (reported in P6). Second, to investigate scaffolding for completeness and collaboration in the early phase of the GBL design process. The analysis enabled the refinement of the design toolkit and the process adopted, thus closing the design cycle. The analysis identified issues such as inappropriate terminology, confusing guidelines, unclear purpose of the tools at use, poor timing for the activities that fed the subsequent iteration of the design cycle.

Figure 4.8: A design workshop with CS and IE students, part of the studies included in P7

- **Case Studies**

The *first case study* examined the developed LEAGUÊ analysis instrument (reported in P2) by employing it to analyze an empathy game (see Figure 4.9). This online game is for primary school children (8-14 years old) to teach empathy by making stories (about different characters using personality traits) to develop a strategy to achieve a goal. The game is suitable for interactive and collaborative learning and can be played with friends in school or at home. The game developer used the analysis instrument to analyze the beta version of this game and highlighted the strengths, weaknesses, and required improvements in terms of essential GBL components needed to instill desired learning through the game. The feedback was collected from the game developer concerning the effectiveness of the analysis instrument to identify the loopholes and make improvements in the design. The feedback also identified the issues to enable the refinement of the instrument.

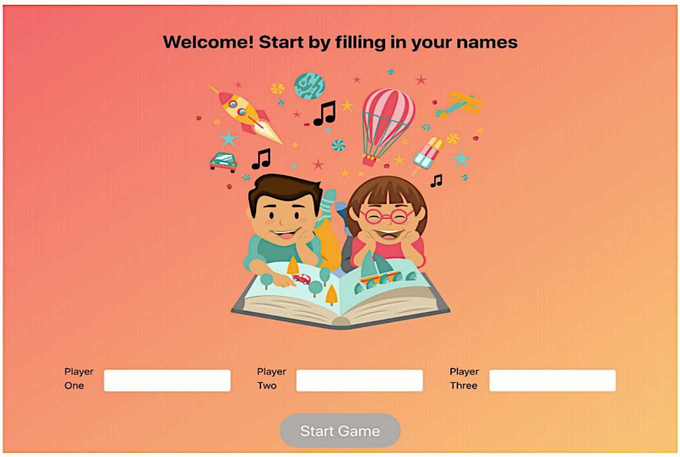


Figure 4.9: Empathy game in the first case study

The *second case study* employed the developed analysis instrument to analyze a VR game (see Figure 4.10) in two iterations on different game versions, reported in (Karlstrøm and Markussen, 2020). The game is a collaborative VR game called "Two Mars" to motivate young adults to pursue a career in STEM by playing a Martian colonist. The first version of the game was evaluated using the LEAGUÊ analysis instrument, which identified some significant flaws in the game to be useful in an educational context. Moreover, it helped identify new requirements for a scenario-based Virtual Reality game focusing on solving day-to-day problems encountered by an early Martian colonist to motivate and engage upper secondary students to learn about the colonization of Mars. The second iteration of the implemented game was again evaluated with the analysis instrument. The two-game developers carried out the evaluations with the analysis instrument in both iterations. In this case study, the developers conducted a user study to observe users playing the game and gather their feedback before self-evaluating the game with the analysis instrument. The improved game findings showed great potential to be used as a supplement to both formal and non-formal learning and as a training tool for soft skills (critical thinking, problem-solving, and collaboration). The feedback was collected from the game developers concerning the instrument's effectiveness in identifying the loopholes in the game design and suggesting improvements. The responses were important for the further refinement and design iteration of the analysis instrument.

Figure 4.10: VR game in the second case study

The *third case study* reflects on the practical experience and field studies carried out in the project "EduApp4Syria" (Nordhaug, 2016) to investigate the applicability of methods used in the context of GBL evaluation with refugee children and the methodological and practical challenges in researching with this user group (reported in P10). Figure 4.11 presents an overview of the multi-stage EduApp4Syria competition. Throughout this project, several evaluation studies illustrate the application and assessment of evaluation methods in various phases of the educational game development life cycle and methodological, ethical, and practical challenges common to GBL evaluation and those unique to refugee children's user group. The methods used to collect the data were project documents, observation notes, interview responses, and practical experience and reflections during evaluation studies. The analysis is presented in terms of the lessons learned and the experience gained from the involvement in research with refugee children. The findings highlight the issues that need to be taken into account when evaluating educational games with refugee children. At the same time, most of the findings are also useful for general research with children.

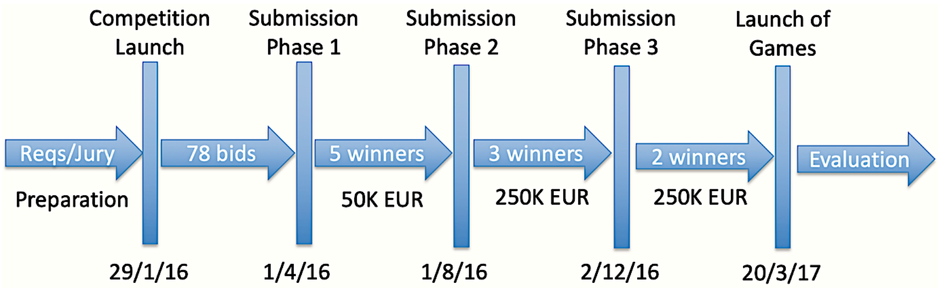


Figure 4.11: EduApp4Syria Project (Nordhaug, 2016) in the third case study

4.3.4 Data Collection and Analysis

This section details the data collection and data analysis method used in papers included in this thesis. Table 4.4 presents a summary of the methodology applied in this doctoral work, detailing the paper's aim, research activity, the type of data collected, the main instruments used for collecting data, and the data analysis methods used in each paper included in this thesis.

Table 4.4: Summary of data collection and analysis

Paper and aim		Research activity	Selected data	Data collection instrument	Data analysis
P 1:	State of the art in GBL evaluation	Systematic Literature Review (SLR)	–58 articles	-	–Categorization of main elements
P 2:	Development and Application of LEAGUÊ Framework	Qualitative content analysis	–58 articles on GBL frameworks, evaluation studies, and reviews – Analysis form from empathy game	–Analysis instrument (three questionnaires/ forms)	–Directed content analysis
P 3:	Investigate differences in learners' emotional states and learning gain in GBL vs. pen-and-paper approach	Quasi-experiment al study	–EEG data from 22 computer science students. –22 pre-and post-test responses –22 emotional state questionnaire responses –Fieldnotes	–EEG –Pre and Post Test –Questionnaire –Observation	–Descriptive statistics –Pearson correlation

	Paper and aim	Research activity	Selected data	Data collection instrument	Data analysis
P 4:	Investigate the effect of the game-based digital quiz on students' engagement, interactivity, and learning gain in online lectures	Quasi-experimental study	–257 survey (usability, technology, game experience, and context) responses from computer science students –243 pre-and post-test responses –257 feedback responses on two open-ended questions	–Survey questionnaire (usability, technology, game experience, and context) –Pre and Post Test –Students' feedback questionnaire –Observation	–Descriptive statistics –Pearson-product moment correlation –Ground theory approach
P 5:	Insights into the design of educational games	Comparative analysis (using analytical framework/ framework analysis)	–15 existing educational design models/guidelines	-	–Quasi-formal comparison technique
P 6:	Design and development of LEAGUÊ ideation toolkit	–Co-design –Design Workshop	–34 Participants' experience questionnaire responses –Fieldnotes –Feedback from focus group sessions	–Questionnaire –Field notes from observations –Focus Group	–Descriptive statistics –Ground theory approach
P 7:	Investigate the scaffolding for completeness and collaboration in the early design phase of learning games with ideation cards	Design Workshop	–Artifacts (games design ideas) from 7 teams –Videos from 3 teams	–Artifacts –Video recording	–Descriptive statistics –Pearson correlation –Ground theory approach
P 8:	Explore Methods and Guidelines for Child-Computer Interaction Research with refugee children	Systematic Literature Review (SLR)	–55 articles	-	–Categorization of main elements
P 9:	Investigate the effect of GBL approach on	Quasi-experimental study	–Game data logs from 30 children	– Pre and Post Test – Questionnaire	–Wilcoxon signed-rank test

Paper and aim	Research activity	Selected data	Data collection instrument	Data analysis
learning gain of refugee children in informal setup – Investigate differences between different age groups (younger and older children) – Investigate the potential association between user demographics, learning, usability and gameplay performance		–VARK questionnaire responses for 30 children – Usability checklist data for 30 children –30 pre-and post-test responses –Demographic questionnaire responses for 30 children –Fieldnotes –Interview responses for 30 children	– Observation checklist – Field notes from observations – Game logs – Usability test – Interview	–Spearman correlation –Mann–Whitney test –Ground theory approach
P 10: Methodological and ethical guidelines to evaluate educational games with refugee children	Case study EduApp4Syria project	–Reflections as notes from the three empirical studies conducted.	–Semi-structured interviews –Field notes from – Observations	– Case analysis/ Categorization of main elements

5 Results

This chapter presents a summary of the papers that document the results of the conducted research. The articles presented in this section are reprinted in full length (following publishing permissions) in Part II of this thesis.

5.1 Overview of the Research Papers

Most of the research work conducted as part of this thesis is published (or under review) in peer-reviewed conference proceedings and journals and therefore accepted as a significant contribution by other researchers in the field. The publications included in this thesis comprise three journal papers and seven conference papers. The summary of research papers is presented in the subsequent sections of this chapter and include the following parts:

- Title of the research paper
- Authors' names and contributions
- Publication Venue (where the paper is published)
- Abstract of the research paper
- A short description of the main findings of the paper and relation to research questions.

5.2 Paper 1

Title: State of the art in Game-Based Learning: Dimensions for evaluating educational games

Authors' names: Rabail Tahir and Alf Inge Wang.

Contributions: Tahir led the research and paper writing and was the main author. The screening of the papers included in the literature review was performed mainly by Tahir. She also collected and analyzed all the articles. Consensus meetings with Wang approved and validated each step of the analysis. Wang provided general supervision for research and paper writing.

Published in: Proceedings of 11th European Conference on Games Based Learning (ECGBL 2017). Graz, Austria. 5-6 October 2017.

Abstract: An increased use of educational games makes it essential to verify these tools for a sound impact by evaluating them from multiple dimensions. This paper presents a Systematic Literature Review (SLR) on state of the art in Game-based Learning (GBL) evaluation. Our research examines the current trends and evaluation practices based on data drawn from a search in four open databases and a manual search of 4 journal proceedings. The paper begins with the context for our study, followed by a depiction of the analysis grid used to generate a database of existing literature and methodology adopted to conduct a systematic review of this literature. From the initial sample of 1929 articles, a total of 58 relevant articles were identified and further examined for the extent of research carried out in GBL evaluation, highlighting the research topics, type of resources, the highly cited papers, and the existing evaluation approaches and criteria used for evaluation of GBL. It then analyses the selected studies for outlining the dimensions for evaluating educational games. The findings of this paper provide insights for researchers and evaluators into current trends, evaluation practices, and multiple dimensions for which an educational game must be evaluated.

Description of main findings and relation to research questions: This paper is a systematic literature review of a state-of-the-art GBL evaluation. It provides grounding and identifies gaps in the GBL literature, thus supporting the subsequent works' rationalization. The results show an increasing trend in GBL research, with most studies from 2009 to 2015 focusing on an evaluation approach followed by design-focused GBL approaches. The findings shed light on a shift in research topics from design to evaluation. However, the studies examined focused on either one or two specific aspects of GBL, and very few focused on overall evaluation specifying all the dimensions essential for GBL evaluation. It highlighted the need for a more comprehensive GBL framework. Moreover, the current GBL evaluation approaches do not cater to children's needs; only one design-focused approach (out of all the reviewed studies) reflected on children requirements. The majority of the studies for educational game evaluations employed some pre-defined (ad hoc) criteria for evaluation and few used general guidelines/approaches, overlooking existing GBL evaluation frameworks. Furthermore, the studies examined revealed a wide diversity of elements used for GBL evaluation. They highlighted the inconsistency in the evaluation elements' definition (dimensions, factors/sub-factors, and metrics) across the studies. The most extensively used dimensions for GBL evaluation are learning, game factors, and usability.

The paper investigates RQ1 to show the state of the art on how educational games are evaluated and the open challenges that exist. The paper provided valuable information that prompted the development of the LEAGUÊ framework in P2.

5.3 Paper 2

Title: Codifying Game-Based Learning: Development and Application of LEAGUÊ Framework for Learning games

Authors' names: Rabail Tahir and Alf Inge Wang.

Contributions: Tahir was the main author and led the research and paper writing. She performed the data analysis and also designed the study and data collection. Wang provided general supervision for the research and the paper writing.

Published in: Electronic Journal of e-Learning, Volume 18, Issue 1, Page no 69-87, 2020.

> *Abstract:* Educational games are now seen as effective learning tools. However, there is a gap in the literature regarding the core dimensions of Game-based learning (GBL) for comprehensive design, analysis, and evaluation due to inconsistent use of elements. GBL literature reports an extensive diversity of elements used to design and evaluate GBL without any categorization of micro and macro-level elements. Hardly any studies systematically decompose these aspects to derivate factors/sub-factors, obstructing identifying any clear pattern. The problem is not the scarcity of GBL research but inconsistency in terminology, scope, definition, and usage of elements leading to the absence of a holistic view of GBL for effective design and evaluation. This study bridges the gap by outlining terminology and scope with four conceptual levels and then systematically categorizing GBL elements by scope, definition, and usage. The methodology used is directed content analysis of GBL literature collected through a previous systematic literature review. Dimensionalization of GBL and further decomposition into factor/sub-factors based on theoretical constructs has resulted in a consistent and clear pattern delineating the structure of the educational game design, analysis, and evaluation. Further codifying metrics and mapping the relationship among GBL dimensions deduce into a conceptual framework that facilitates greater insight into the process of learning with educational games, where to focus, and what to evaluate. The LEAGUÊ framework can be applied for the design, analysis, and evaluation of educational games. The framework is put in practice by utilizing the framework components to develop three items: LEAGUÊ analysis instrument, LEAGUÊ ideation, and design toolkit, and LEAGUÊ evaluation guide that can assist educational game designers, researchers, and evaluators. This paper exemplifies the analysis of learning games using the LEAGUÊ analysis instrument with one case study as an example.

Main findings and relation to research questions: This study expands on GBL design and evaluation literature to overcome the shortcomings in current research (inconsistency in terminology, scope, definition, and usage of elements) by conducting a directed content analysis. The analysis results are translated into the conceptual hierarchical framework LEAGUÊ, which shows that GBL's multidimensionality requires evaluating several aspects referred to as core dimensions, including Learning, Game Factors, Affective-Cognitive Reactions, Usability, User, and Environment. Each dimension focuses on certain factors and sub-factors that constitute that aspect, and metrics are required to assess them. The framework presents 22 factors, 74 sub-factors, and five metrics categories. The dimensions of GBL are related to each other. It is essential to assess the relations presented as a high abstraction of LEAGUÊ for more significant insights into educational games. The framework provides a detailed GBL picture that will guide researchers, evaluators, designers, and developers of educational games. The proposed framework is built on components grounded in theory. Each component has a strong basis for formation that is supported by theoretical constructs in the GBL literature. The framework is put into practice by developing three tools to support GBL practitioners and researchers. The framework can be applied to design, analyze, and evaluate educational games using the LEAGUÊ analysis instrument, ideation and design toolkit, and evaluation guide.

The paper investigates RQ2 and RQ4 and contributes to RQ3. The paper presents a hierarchical framework emphasizing the multidimensionality of the GBL approach. It identifies the core dimensions, factors/sub-factors, and their interrelation to assist the design and evaluation process addressing RQ2. This paper also presents an analysis instrument (developed by utilizing the LEAGUÊ framework components) and its application to analyze an empathy learning game addressing RQ4. The paper served as the primary foundation of investigation for fundamental aspects of the GBL phenomenon. The results from this paper provided the theoretical framework that directed the future research design for RQ2, RQ3, and RQ4.

5.4 Paper 3

Title: Evaluating learners' emotional states by monitoring brain waves for comparing game-based learning approach to pen-and-paper

Authors' names: Krenare Pireva Nuci, Rabail Tahir, Ali Shariq Imran, Niraj Chaudhary.

Contributions: Pireva led the research and coordinated the paper writing. All the co-authors contributed to the research. Tahir designed the study (experiment design and used instruments) included in the paper, and Pireva supervised data collection and performed the data analysis. Besides, Tahir also assisted in the discussion of data analysis results.

Tahir and Pireva also contributed to writing the paper. Imran provided general supervision of the research and paper writing, and Chaudhary guided the data collection using EEG headsets.

Published in: Proceedings of the 49th Annual Frontiers in Education (FIE), Ohio, USA. October 16-19, 2019.

> *Abstract:* A new interest in the use of game factors while acquiring new knowledge has emerged, and a number of researchers are investigating the effectiveness of the game-based approach in education systems. Recent research in game-based learning suggests that this approach imparts learning by involving learners in the learning process. The game factors generate affective-cognitive reactions that absorb users in playing the game and positively influence learning. This paper offers a comparison of the learning processes between the game-based learning and pen-and-paper approaches. In this paper, the analysis of both learning approaches is realized through a brain-controlled technology, using the Emotiv EEG Tech headset, by analyzing the stress, excitement, relaxation, focus, interest, and engagement that the learner is experiencing while going through both approaches.

Main findings and relation to research questions: This paper investigates how the learning process acquire while gaining new knowledge through Game-based learning (GBL) in terms of emotional states and learning outcomes compared to a traditional pen-and-paper approach. This study successfully used EEG data to analyze the emotional states during GBL activity to obtain insights on how the learning process acquire concerning the GBL phenomenon. The *first* interesting feature of the results is that the GBL approach positively facilitates the learning process through emotional states such as excitement, engagement and interest in contrast to the traditional pen-and-paper learning activity. *Second*, the game-based learners experienced higher progress than the pen-and-paper group, especially in the practical and conceptual questions, even though extra attention and explanation were provided to pen-and-paper participants. *Third*, when comparing the number of tasks solved by the participants of both groups, the game-based learners achieved more than double the number of tasks in the same time as the pen-and-paper group. However, when analyzing the emotional state variables, there was no significant difference. One reason could be that the game lacked captivating design and engaging features, which could enhance affective-cognitive reactions. So further research would build upon these findings by using a more robust and engaging gaming platform.

The paper investigates RQ2 and contributes to RQ3 and RQ4. It investigates game elements' role in generating affective reactions (positive attitudes) and their relation to learning outcomes during students' GBL experience compared to a traditional learning approach. This paper's findings and experience also provided valuable information concerning the use of the EEG method for GBL evaluation and design guidelines for effective learning games.

5.5 Paper 4

Title: Game-based digital quiz as a tool for improving students' engagement and learning in online lecture

Authors' names and contributions: Krenare Pireva Nuci, Rabail Tahir, Alf Inge Wang, and Ali Shariq Imran.

Contributions: All the co-authors contributed to the research. Pireva, the first author, led the research and coordinated the paper writing. Tahir contributed to design the study (experiment design and study dimensions) and developed instruments included in the paper with her knowledge of Game-based learning (GBL). Pireva supervised data collection and performed the data analysis, whereas Tahir assisted in the results of the qualitative data analysis and methodology. Tahir and Wang also contributed to writing the introduction, related work, and discussion sections of the paper. Imran provided general supervision of the research and the paper writing.

Published in: Ready for submission.

Abstract: Distance teaching and learning are gaining popularity, especially amidst the COVID-19 crisis at the beginning of this year. Several schools, colleges, and universities across the globe, as a result, have adopted the online mode of teaching. While the businesses and day-to-day activities were shutting down, eLearning tools and online education platforms saw considerable demand past few months. Many institutions with digital infrastructure in place and prior distance teaching experience had a smoother transition from on-campus classes and lecturing to online teaching and learning. In contrast, for many, the transition involved many challenges, including keeping students' motivation, interaction, and interest alive, apart from adapting to the tools and technologies. This paper reports on students' engagement and motivation levels and the learning curve in the Human-Computer Interaction course using a game-based digital quiz tool during the COVID-19 crisis. The study investigates the effect of in-lecture quizzes in online classes concerning learning gains over four months. Two key motivation parameters (students' engagement and interaction) are compared and analyzed on two quiz platforms during the quarantine experience. The platforms include Kahoot! and Google Form Quiz. The results indicated a significant increase in students' engagement and interaction levels in lectures with systematic in-lecture

quizzes. Further, the results show that the learning curve is steeper when using in-lecture quizzes (with 73%) in contrast to classes in-lecture quizzes are not used (57.5%).

Main findings and relation to research questions: This paper investigates the impact of a GBL approach on students' engagement, interactivity, and learning performance in the specific context of online classes amid the COVID-19 crisis. The paper reports on the findings of a four-month investigation on using a game-based digital quiz in online classes of the Human-Computer Interaction (HCI) course during the quarantine period. The prominent results of the study concerning the effects of the GBL approach on student's learning experience include: (1) systematic use of game-based online quizzes impacts the students' engagement and motivation; (2) systematic use of game-based online quizzes impacts class dynamics and is linked with higher interactivity (among professors and students) in online classes; and (3) it also impacts the students' exam performance, the findings showed that students performed significantly better in the subjects with systematic game-based in-lecture quizzes.

The paper contributes to RQ2, providing valuable information regarding the impact of the GBL approach in online lectures, improving our knowledge about the GBL phenomenon. It addresses the importance of game components and interactivity on students' GBL experience in online lectures and the impact of using systematic game-based quizzes on students' engagement and learning performance. The findings and experience from this paper also contribute to RQ3 and RQ4. The paper gave insights into the role of digital quizzes as a data collection method for GBL evaluation in addition to their use as potential learning tools. The paper also provides recommendations for designing effective GBL experiences in online classes.

5.6 Paper 5

Title: Insights into the design of educational games: Comparative analysis of design models

Authors' names: Rabail Tahir and Alf Inge Wang.

Contributions: Tahir led the research and paper writing and was the main author. Tahir performed the comparative analysis, and Wang approved and validated each step of the analysis. Wang provided general supervision for research and paper writing.

Published in: Proceedings of the Future Technologies Conference (FTC). Vancouver, Canada. 13-14 November 2018.

Abstract: The study reports on ongoing research that intends to identify and validate the core dimensions for Game-based learning (GBL) and further explore the shift in dimensional focus between different phases of an educational game development life cycle: pre-production (design), production (development) and post-production (testing and maintenance). Hence, this paper presents the initial work focusing on the design phase by presenting a comparative analysis of educational game design models using GBL attributes, validity, and framework attributes as an analytical lens. The main objective is to analyze the fundamental GBL attributes in existing design models, identify the common attributes that demonstrate their importance for the design phase, and highlight any need for further research in attribute validation and framework improvement. This study also highlights the strengths and weaknesses of existing design frameworks. The results of the analysis underline learning/pedagogical aspects and game factors as the essential attributes for the design phase of educational games. Comparative analysis also guides researchers/practitioners to better understand GBL through various properties of different existing design models. It highlights the open problems such as lack of tool support, empirical validation, independent evaluations, adaptability, and absence of concrete guidance for application to make more informed judgments.

Description of main findings and relation to research questions: This paper presents the comparative analysis of design models/frameworks for educational games using GBL attributes, validity, and framework attributes as an analytical lens. The paper examines the strengths and weaknesses of existing frameworks focusing on common attributes important for the design of educational games. The analysis revealed learning/pedagogy (Learning objective, instructional design, learning content and knowledge enhancement/outcome) and game factors (mechanics, dynamics, narrative, aesthetics, goals) as the essential attributes for the design phase, followed by affective reactions (flow, enjoyment, immersion). Only a few emphasized usability (user interface), user (learner requirements) and environment (including technical and context-related aspects). The analysis brings to attention some open problems in existing design models (such as lack of tool support, empirical validation, independent evaluations, adaptability and absence of concrete guidance for application), which obstruct their application in the practical design process of learning games. Only a few elements, such as learning curve, flow antecedents, and some game design factors, are empirically validated. It is vital to address the lack of tool support, adaptability and concrete guidance for the practical application of framework concepts in the design process of educational game development to facilitate the framework-based educational game design and collaboration between industry and research.

The paper addresses RQ1. It focuses on the GBL attributes in existing educational game design models and highlights the open problems of attribute validation and framework improvement (explicitly pointing to the lack of tool support and guidance for application). This paper's findings guided the need for further research on the GBL design process and led to the development of the LEAGUÊ toolkit in P6.

5.7 Paper 6

Title: Transforming a theoretical framework to design cards: LEAGUÊ ideation toolkit for game-based learning design

Authors' names and contributions: Rabail Tahir and Alf Inge Wang.

Contributions: Tahir designed and created the toolkit and led the research and paper writing. Tahir also designed and conducted user studies and performed the data collection and analysis. Wang participated in the data collection during the third user study and provided general supervision for the toolkit design, research, and paper writing.

Published in: Special Issue Design Methodology for Educational Games, Sustainability journal, Volume 12, issue 20, article number 8487, 2020.

> *Abstract:* Educational game design is a complex process demanding multidimensional focus in a heterogeneous team to balance multiple aspects. The existing Game-based learning (GBL) frameworks detail the required knowledge but are hard to use in design practice. Conversely, card-based design tools are a lightweight approach used to assist the early design phase. While several game design cards exist, none is specific for informing GBL knowledge. There is a lack of operationalizable approaches for designing learning games that integrate research-based GBL knowledge into the actual ideation process. This paper presents a card-based GBL ideation toolkit to reduce the complexity of applying a framework. The toolkit introduces key GBL concepts in the design process as a tangible reference point to facilitate multidimensional focus, supporting idea generation, critical reflection, and creating a shared understanding in the collaborative design process. The paper describes a ten-step process of transforming the LEAGUÊ framework into the LEAGUÊ toolkit (GBL ideation cards), presents the evaluation of the toolkit with design workshop participants, and design lessons detailing strengths and limitations to support GBL design practices.

Main findings and relation to research questions: This paper introduces the LEAGUÊ ideation toolkit: a card-based ideation toolkit for learning game design. The paper presents a ten-step process for transforming a framework into an ideation toolkit. The LEAGUÊ framework described in P3 is translated into a card-based GBL ideation toolkit, addressing the lack of operationalizable approaches for designing learning games. For an overview of how it connects with the LEAGUÊ framework, see (Tahir and Wang, 2020b). The developed toolkit provides tool support for applying theoretical knowledge from the GBL framework in design practice to support the learning game design team in the GBL design process. The toolkit artifacts and the five-step ideation process were evaluated by employing in design workshops, leading to satisfactory results in terms of creative ideas, participants' experience (perceived usefulness, understandability, level of fun, and satisfaction), and support and guidance provided. The results from three design workshops discussed the design lessons by highlighting the developed toolkit's strengths and limitations. The five main successful aspects were: easy to use in practice; stimulate brainstorming and creative thinking; creative elements in the toolkit generate fun; playfully guide the design process and inform and encapsulate theoretical concepts.

The paper investigates RQ3, addressing the lack of operationalizable approaches and integrating research-based GBL knowledge into the actual design process by transforming the GBL framework into the card-based toolkit to support the ideation phase of the GBL design process.

5.8 Paper 7

Title: Completeness and collaboration in the early design phase of learning games: Do ideation cards provide scaffolding?

Authors' names and contributions: Rabail Tahir and Alf Inge Wang.

Contributions: Tahir wrote the paper, designed and conducted the study, and was the main author. Tahir also collected and analyzed the data. Wang participated in the user study and provided general supervision for the research and paper writing.

Published in: Accepted in the 23rd International Conference on Human-Computer Interaction (HCI International 2021), Washington, D.C., USA. 24-29 July 2021.

> *Abstract:* Game-based learning (GBL) has proliferated rapidly in recent years, with both industry and academic research communities calling for collaborative work practices in the educational game design process. There is a need to address all the key GBL aspects and create a shared understanding among team members. Design cards have the potential to improve idea generation and communication between stakeholders. However, potential scaffolding for completeness (focusing

on all key GBL dimensions) and collaboration (working together to produce something) in the learning game design process is not explored. Therefore, in this paper, we investigate how this design approach can scaffold for collaboration and completeness in the early phase of the learning game design process using a card-based GBL ideation toolkit in design workshops. Seven teams were analyzed using design artifacts and video recordings of the workshop session. The results are encouraging in terms of the applicability of ideation cards in the GBL design process to scaffold completeness and collaboration.

Main findings and relation to research questions: This paper evaluates the scaffolding for completeness and collaboration in the ideation process of learning game design ideas generated with the LEAGUÊ ideation toolkit. Team collaboration and focus on all the key dimensions of GBL are critical factors for a learning game's design process. Quantitative and qualitative data collected highlighted the potential of the toolkit to facilitate multidimensional focus and collaboration. One interesting finding is that all the teams focused on the six key GBL aspects (to some extent) in their overall produced game idea, highlighting the potential for scaffolding a multidimensional focus. The typical pattern identified in idea generation was: *First*, the primary card "game" was used, then "learning," followed by the "environment" of the game. Further, the teams typically used a card that addressed the "reaction" that the learning game intended to generate, followed by target "users," and lastly, the "usability" aspect. *Second*, analysis of the play session's video recording resulted in six themes that characterize interaction in the GBL ideation process when using the card-based tool. These six themes aggregated to identify the three central features (points of interaction) of the card-based toolkit that scaffolds collaboration: Physical point of interaction, mental point of interaction, and social point of interaction.

The paper investigates RQ3, addressing GBL's early design phase, highlighting ideation cards' potential to facilitate completeness (multidimensional focus) and collaboration in ideating learning game design and understanding the contributing factors.

5.9 Paper 8

Title: Exploring Methods and Guidelines for Child-Computer Interaction Research with Refugee Children

Authors' names and contributions: Rabail Tahir and Alf Inge Wang.

Contributions: Tahir led the research and paper writing. Tahir was the main author and was responsible for screening, collecting, and analyzing all the articles included in the literature review. Wang approved and validated each step of the analysis and provided general supervision for the research and paper writing.

Published in: Proceedings of the 21st International Conference on Human-Computer Interaction (HCI International 2019), Orlando, Florida, USA. 26-31 July 2019.

Abstract: There are many guidelines and methods for doing Child-Computer Interaction (CCI) research, but very few focus specifically on refugee children with a challenging background. The complex situations and multiple changes refugee children undergo, including community, culture, schooling, friendships, language, war, displacement, physical violence, and even identity, make them different from children who are not refugees. They suffer learning disabilities, mental health issues, poor physical health, trust issues, and overall developmental disabilities. As there are a large number of refugee children in the world who are displaced and out of school, it is crucial to help these children using available technology and assess the effectiveness of the use of technology. This paper presents a literature study on available research guidelines and methods for CCI. The literature has been reviewed for guidelines and evaluation methods, starting from more general research with children, moving to more specific research with refugee children, and finally to identify gaps, present common grounds and directions for research with this specific population. The results from 55 articles reveal that although guidelines and methods for research with children can be used for refugee children, special attention and additional guidelines are needed to address this group's specific needs. Further, the review reveals a lack of CCI research and research methods for refugee children. Most adapted/new children-friendly research methods are not fully employed in research with refugee children. The results of this review could serve as a starting point for researchers entering the CCI field to work with refugee children.

Main findings and relation to research questions: This paper is a Systematic Literature Review (SLR) consisting of 55 articles that emphasized research methods and guidelines for CCI research with children (with or without refugee background). The SLR identified three categories of research methods: Preferred, General, and Specific methods. The preferred and general methods to a large extent are similar for research with children with and without refugee background; however, more variation is found for specific methods. Two categories were identified for research guidelines: general (similar) and specific (different) guidelines. One interesting finding is that there are differences in the details, even for the general guidelines. Additional issues must be taken into account for conducting research with refugee children, revealing the need to adopt guidelines for research with specific emphasis on refugee children's context. The guidelines were further grouped into ethical, practical, and methodological. Only three new or adapted research methods were found in the review specifically for refugee children, but some preferred and specific methods were used with this population. There is a gap in the literature regarding the focus on methodological guidelines for the specific group of refugee children, which is in line with the scarcity of research on the effectiveness of research methods with this user group.

The paper investigates RQ1 and contributes to RQ4. It summarizes the available research methods and guidelines for Child-Computer Interaction (CCI) research with children (with and without refugee background). The paper highlights a gap in CCI literature concerning a lack of methodological guidelines and focus on the effectiveness of evaluation methods for conducting research with refugee children. This paper's findings guided the research on GBL evaluation with refugee children (reported in P9 and P10) in this doctoral work. We aimed for the third GBL evaluation study (quasi-experiment 3, see Chapter 4, Section 4.3, Table 4.2) in this thesis to focus on the informal learning environment, which led us to select this target group. The current refugee crisis and increasing focus of research on using educational technology to help refugee children in their education and well-being directed our research on GBL evaluation for informal learning to emphasize refugee children. Therefore, this paper presents a starting point to guide further research in conducting learning game evaluation with this specific population.

5.10 Paper 9

Title: Evaluating the effectiveness of game-based learning for teaching refugee children Arabic using the integrated LEAGUÊ-GQM approach

Authors' names and contributions: Rabail Tahir and Alf Inge Wang.

Contributions: Tahir led the research and paper writing. She developed the approach, designed and conducted the study, collected and analyzed the data, wrote the paper, and was the main author. Wang contributed to data collection, participated in the user study, and provided general supervision for the research and paper writing.

Published in: Ready for submission.

Abstract: Game-based learning (GBL) is a well-established research area that continues to receive interest and attention from researchers and practitioners alike. GBL is widely utilized in various domains and settings and is growing. However, there is still a lack of empirical evidence concerning its effectiveness, making GBL evaluation a critical undertaking. This paper proposes an integrated (LEAGUÊ-GQM) approach for planning and executing GBL evaluation studies

and presents its application by evaluating the effectiveness of a GBL approach to improve the Arabic reading skills of migrant refugee children in an informal learning setup. The study further focuses on how children's age group, learning modality preference, and prior mobile experience affect their learning gain, usability, and gameplay performance. A quasi-experiment with a one-group pretest-posttest design was used, and 30 children (5-10 years old) with migrant refugee backgrounds participated in this evaluation study. The results suggest that GBL can improve the Arabic reading skills of migrant refugee children at home, showing statistically significant improvement in their reading assessment score. The results also outline a clear impact of children's age groups on their learning gain, usability score, and total levels played. The younger children (5-6 years) outperformed the older children (7-10 years) in learning gain, whereas older children have fewer usability issues than younger. In terms of learning modality preference and prior mobile experience, both had a statistically significant effect related to usability and gameplay performance parameters. However, no effect was found on learning gain. Based on the findings, some design recommendations are suggested for researchers and designers of educational games for more inclusive design focusing on user characteristics.

Main findings and relation to research questions: This paper introduces an integrated (LEAGUÊ-GQM) approach and evaluation guide (for creating an evaluation plan) to facilitate the learning game evaluation process and investigates the potential of a GBL approach for teaching migrant refugee children (Arabic reading skills) in an informal learning setup. The integrated approach was put into practice by employing it to create the evaluation plan for the presented language learning game evaluation with refugee children. The evaluation study focused on 1) the learning gain with this GBL intervention, 2) examined the effects of user characteristics (age-group, learning modality preferences, mobile usage experience) on children's learning, usability, and gameplay with the GBL approach, and 3) usability issues faced by children. The study's main findings suggest that the GBL approach is effective for the informal learning setup of teaching refugee children Arabic reading skills. The study also found that children's age-group is associated with their learning gain, usability score, and total levels played. Moreover, children's learning modality preferences impact their usability score, total levels, and total time played, but it does not impact their learning gain. Similarly, mobile usage experience is related to the children's usability score, the total number of sessions played, and score in a play session, but no significant relationship was found with learning gain. Furthermore, the findings identified the issues faced by children when playing with the employed learning game related to three categories: learnability, interface, and satisfaction. The paper presented recommendations for designing effective learning games for refugee children focusing on the following: age-appropriateness, adaptivity, multimodal learning, intriguing storyline, meaningful feedback, adequate help and tutorial, customizability, pedagogic feedback, empathy and connection, and engaging tasks.

The paper investigates RQ2 and RQ4 and contributes to RQ3. The paper provides valuable information concerning the potential of the GBL approach for teaching refugee children in an informal learning setup. The evaluation study focused on the user characteristics and effect on learning gain, usability, and gameplay performance, thus improving our knowledge about the GBL phenomenon and addressing RQ2. The paper also proposed an integrated LEAGUÊ-GQM approach that guides creating an evaluation plan for conducting GBL evaluation studies addressing RQ4. The evaluation study results in this paper implied recommendations for designing effective learning games contributing to RQ3.

5.11 Paper 10

Title: How to Evaluate Educational Games with Refugee Children: Methodological Aspects and Lessons Learned from EduApp4syria

Authors' names and contributions: Rabail Tahir and Alf Inge Wang.

Contributions: Tahir led the research and paper writing. Tahir also collected and analyzed the data. Wang contributed to writing the paper and provided general supervision for the research and paper writing.

Published in: Proceedings of the 13th European Conference on Games Based Learning (ECGBL 2019), Odense, Denmark. 3-4 Oct, 2019.

Abstract: Educational game evaluation is a multidimensional and complex phenomenon. The growing interest in game-based learning (GBL) results in an increasing need to evaluate this approach's effects, which requires appropriate methods, techniques, and principles that the GBL community can apply. This paper reflects on the methodological aspects of evaluating educational games with refugee children drawing on practical experience and evaluation studies conducted in the EduApp4Syria project. The paper gives an overview of the project and presents three field studies conducted, including the GBL evaluation methods used in the context of refugee children. The methods used included quasi-experimental design, mixed-method approach, observation with/without a checklist, questionnaires, interviews, pre/post-test (using EGRA), screen recording, game-logs, and expert evaluation. The evaluations illustrate the application and assessment of these methods. This paper presents the findings and pitfalls related to the applicability of evaluation methods in various phases of the

game development life cycle and methodological and practical challenges in conducting research and eliciting data in the context of evaluating educational games with refugee children. This article provides an up-to-date examination of both methodological challenges common to GBL evaluation and those unique to the user group of refugee children. Thus, culminating in guidance for researchers on methods and critical issues that need to be considered when designing research studies involving educational games and children. The paper assists researchers to critically reflect on these methodological issues and methods they use as they will have implications on the data obtained.

Main findings and relation to research questions: This paper put forth the findings and pitfalls related to the methodological and practical aspects of evaluating educational games with refugee children based on pragmatic experience and reflections made during evaluation studies conducted in the EduApp4Syria project. Most of these findings are also useful for general research with children. The successful evaluation of a learning game starts with selecting a set of factors or dimensions based on the evaluation objective. There is not just one construct, but several interrelated dimensions shape the GBL experience. Different methods might be beneficial for exploring different objectives and at different development lifecycle stages. *First*, regarding the applicability of methods in different development lifecycle stages, it is worth noticing that expert evaluations are more feasible in the early phases of development to achieve a multidisciplinary approach from the early concept phase, which is essential for educational games. Whereas methods such as experiments, pre/post-test, observation, screen recordings, and game logs are mostly useful in later phases, providing an opportunity for impact evaluation with user data. The game logs can be particularly useful for monitoring long-term engagement and learning, especially for playing in a real context. Moreover, methods such as questionnaires, interviews, and user testing can be employed in multiple phases with slight variations in details. Observations without a checklist are beneficial in the earlier phases of development when the objective is more exploratory, and it helps uncover problems and provide insights into the children's experiences. In later phases, it is more effective to employ observation with a checklist to be more focused and concrete. *Second*, when devising an appropriate methodology for evaluation, several factors must be taken into consideration: such as consent, gaining access and privacy, language barrier, cultural issues, need for a translator, learner related issues (verbalization, previous knowledge, personality, technology experiences, attention span, learning disabilities, age and gender of the children), parents involvement, child-friendly interactions, environment and setup of research, and the effect of information provided on research participants.

The paper investigates RQ4, addressing the GBL evaluation process. The paper focuses on the applicability of evaluation methods in different phases of the educational game development life cycle and presenting methodological, practical, and ethical challenges and guidelines for conducting research and eliciting data for educational games evaluation with refugee children.

6 Contributions

This chapter elaborates and discusses the research contributions of this doctoral work and their implications. The chapter is structured in accordance with the six main contributions as follows:

C1. Summarize and conceptualize state of the art in Game-based Learning (GBL) design and evaluation practices and identify existing challenges and issues.
C2. Present a conceptual hierarchical framework of six dimensions for comprehensive design and evaluation of GBL applications.
C3. Empirical evidence on the application of GBL approaches in different contexts for improved understanding about the process of learning with educational games and contributing factors.
C4. Contribution to the design, implementation, and evaluation of a card-based design toolkit for the ideation phase of educational game design, facilitating multidimensional focus and collaboration in the GBL design process.
C5. Contribution to the development and application of an analysis instrument and an integrated evaluation approach to support the educational game evaluation process.
C6. Guidelines for improving the design and evaluation of GBL applications in general and specifically for refugee children.

The *first* contribution maps out, categorizes, and appraises the existing GBL literature identifying gaps and unsolved problems in the educational game design and evaluation research to commission further primary research. The *second* contribution proposes a comprehensive GBL framework (LEAGUÊ) for educational game design and evaluation, addressing the inconsistency and need for core GBL dimensions in the existing literature. The *third* contribution provides empirical evidence on GBL effectiveness in different learning settings and improved understanding of the learning process and different interrelated aspects in GBL. Moreover, it also provides some design implications for effective learning games from the application of GBL approaches in different contexts. The *fourth* contribution includes the work connected to the creation of the LEAGUÊ design toolkit (card-based tool) for the ideation phase of the GBL design process. It addresses the lack of operationalizable approaches for the learning game design practice. The *fifth* contribution relates to the development of the LEAGUÊ analysis instrument and the integrated LEAGUÊ-GQM approach to support the GBL evaluation process. Finally, the *sixth* contribution presents guidelines and implications that emerged from the knowledge gained from evaluation studies and retrospective analysis of the research outcomes. The guidelines aim to improve GBL design and evaluation for developing effective learning games. Table 6.1 summarizes the six contributions presenting the papers' additions to each contribution.

Table 6.1: Mapping the connection between contributions and research papers

	C1	C2	C 3	C4	C5	C6
P1	•					
P2		•		•	•	•
P3			•			•
P4			•			•
P5	•					
P6				•		•
P7				•		•
P8	•					•
P9			•		•	•
P10						•

6.1 C1. Summarize and Conceptualize State of the Art in GBL Design and Evaluation Practices and Identify Existing Challenges and Issues.

The *first contribution* of this doctoral work provides an assessment and contextualization of up-to-date knowledge of the GBL domain focusing on design, evaluation, and research method and guidelines to identify challenges and support more effective educational game research. It comprises the review studies (Tahir and Wang, 2017, Tahir and Wang, 2018, Tahir and Wang, 2019a), consisting of two systematic literature reviews (SLRs) (P1, P8) and one comparative analysis (P5). The first two review studies (P1, P5) produce substantive findings regarding educational game design and evaluation practices in existing GBL literature. The aim was to present insights into current trends, GBL evaluation approaches and employed evaluation criteria, different dimensions for GBL evaluation, design models, and essential attributes for educational game design. To better understand GBL through various properties of existing approaches and models in the research community. The third review (P8) provides an overview of evaluation methods and guidelines in Child-Computer Interaction (CCI) for conducting evaluation research with children with or without refugee background. It was important because no such guidelines were available in GBL literature. Also, it was important to assess GBL technology's effectiveness with the special group of refugee children as they were our target users for the third quasi-experiment (see Chapter 4, Section 4.3.3). The reviews' outcomes depict state of the art showing ongoing advances and changes in the GBL domain and identify problems and potential research gaps. It was essential for setting the stage for further research.

6.1.1 State of the Art Design and Evaluation Practices for GBL

The GBL design and evaluation practices and evaluation methods and guidelines in existing literature for conducting research with children with or without refugee background (that can guide GBL evaluation with refugee children) are summarized below.

- **GBL Evaluation Practices**

The *first study* (P1) reported on GBL research trends, existing GBL evaluation practices, and dimensions for evaluating GBL. The findings indicated an increasing GBL research trend, with most research focusing on evaluation approaches followed by design approaches. Other research topics included educational game evaluation studies and review studies. The review studies increased over the past few years, and there was a shift from design-focused research to evaluation-focused research topics. A total of 19 evaluation approaches were identified from the GBL literature review comprising frameworks, models, scale, and heuristics or guidelines. Most approaches centered around one or two specific aspects for evaluating GBL, such flow, GBL user experience and user enjoyment. Every few approaches (such as the four-dimensional framework, framework for serious game design evaluation, and serious game design assessment framework) focused on GBL evaluation considering multiple aspects deemed essential. The review highlighted three categories for criteria used in educational game evaluation studies: evaluation approach (framework/model/guidelines), predefined (ad hoc), and not specified. The majority of educational game evaluation studies did not employ any existing GBL evaluation frameworks or models. Most studies used some predefined criteria in an ad hoc manner, just outlining evaluation goals. The selection basis of factors and measures used for the GBL evaluation was not defined in these studies. A few used general guidelines/approaches (not specific for GBL), such as Nielsen's heuristics and USE (usefulness, satisfaction, and ease of use) scale for measuring usability and flow theory and taxonomy of intrinsic motivations for assessing engagement. Only one study (Tseloudi and Tsiatsos, 2015) from the selected review articles used an existing GBL evaluation approach (EGameFlow scale) for measuring enjoyment in an online educational game. The review identified a large number of dimensions (thirty-seven) for GBL design and evaluation dispersed across GBL literature. However, learning is the most extensively used aspect in evaluating educational games, followed by usability. Game factors such as game mechanics, game design and game story were also emphasized in multiple studies. Only one study (Alfadhli and Alsumait, 2015) proposed GBL design guidelines focusing on the needs of children.

- **GBL Design Practices**

The *second study* (P5) reported on GBL attributes used in existing educational game design models, validation of these attributes (theoretically grounded and empirically sound), and comparison of existing models using analytical lenses. The majority of analyzed models emphasized two attributes: *learning* (focusing on learning objective, learning content, instructional design, and knowledge enhancement) and *game* (focusing on goals, mechanics, narrative, dynamics, and aesthetics), followed by reactions such as flow, enjoyment, immersion. Most frameworks are grounded in literature and relevant theories for a pedagogical base and game design principles. The pedagogical theories used include Experiential learning theory, Bloom's taxonomy, Instructional design principles, Piaget’s schemes, Vygotsky's zones of proximal development, and Gagne's events of instruction. Moreover, some frameworks are also based on ARCS model and Flow theory. Only two of the analyzed frameworks had empirical evidence of their validity. However, the authors themselves conducted the evaluations. The elements validated by empirical studies were only three: learning curve, flow antecedents, and game factors. Moreover, the studied models differ in terms of involvement of stakeholders and assessment. The design models are applied in various learning domains such as math, geography, computer science, language, culture, and history. However, most models only emphasized abstract principles limited to high-level concepts rather than concrete procedural guidance to structure educational games' design process. Most of the design models are general for any audience and educational game design. The framework attribute “adaptability in use” is addressed by only two models. They emphasized that the framework should be employed depending on the game characteristics and scope and provided the opportunity for adaptation by offering macro elements adaptable to different game genres. According to the comparative analysis, most of the analyzed frameworks only focused on the design stage. However, three models emphasized use for evaluation or analysis as well. However, not many practical applications of the frameworks are available where they have been used.

- **Evaluation Methods and Guidelines for Research with Children with or without Refugee Background**

The *third study* (P8) presents a starting point to guide researchers and evaluators in the CCI community in conducting GBL evaluations with children with or without refugee background. For evaluating the effectiveness of the GBL approach for informal education, refugee children were selected as the target user group (considering the Syrian crisis and the role educational technology can play (Menashy and Zakharia, 2017). It instigated the need to explore the evaluation methods and guidelines in CCI literature to carry out the educational game evaluation study with this group. The main objective of conducting this Systematic Literature Review (SLR) was to find useful information concerning research with refugee children to allow us to perform our GBL evaluation study. This review helped us analyze guidelines and methods used in research with refugee children to select methods and guidelines to guide the third GBL evaluation (quasi-experiment) study (presented in P9). The findings reported on the available guidelines and evaluation methods for research with children (with and without refugee background) highlighted the specific guidelines and methods for the refugee context, and finally identified gaps and presented directions for research with this specific population. CCI literature highlights several issues (such as cognitive load, verbalization, attention span, skills, nature, and gender differences) that led to developing new/adapted methods for research with children. The issues in the case of refugee children stretch beyond the general issues in research with children (Strekalova and Hoot, 2008). The additional issues are their harsh experiences such as war, dislocation, stress, violence, poverty, discrimination, the difference in culture, language barrier, and loss of family members. These experiences result in learning disabilities, physical and mental health issues, distrust, insecurity, and access issues. Therefore, it is impossible to ignore these issues when conducting research, representing them as a special target group. To exemplify, the issues related to distrust, dislocation, and low literacy required research guidelines to be more flexible. Such as obtaining oral consent (Candappa and Ahmad, 2007), contacting a trusted member of their community to approach them for building trust, and gaining access to local authority social worker following the law (Hopkins, 2008) in case of unaccompanied or separated children. On the other hand, little has been contributed by researchers in reviewed literature regarding research methods. However, many have highlighted the need for such efforts (Block, Warr et al., 2013, Due, Riggs et al., 2014). The research with refugee children on educational game evaluation is still in infancy. Most of the specific methods used with refugee children came from social science literature. The review identified three categories of research methods: preferred, general and specific methods. Specific methods are different for children with and without refugee backgrounds. Moreover, two categories were found for research guidelines: general (similar) and specific (different) guidelines. However, there are some differences in research details for refugee background, even for general guidelines, that must be taken into account. The guidelines were further introduced in three groups: ethical, practical, and methodological. There is a need to adapt guidelines for research with refugee children taking into account the additional ethical, practical, and methodological parameters to ensure the research of introducing educational technology includes a sound understanding of its users. Only three new or adapted research methods were found focusing on refugee children. However, some preferred and specific methods used with this population are highlighted that can guide researchers.

6.1.2 Identified Challenges and Issues in GBL Design and Evaluation Practices

The *first review* improved the understanding of the GBL domain and highlighted the following major issues in GBL research: 1) Most GBL models, frameworks, and studies aimed at exploring GBL through any single aspect (or two), leading to the absence of a holistic view of GBL. 2) Usage of a wide diversity of aspects in GBL literature to design and evaluate educational games, hindering the identification of core aspects essential for an effective GBL approach. 3) Only a few studies systematically break down the high-level GBL aspects based on their theoretical construct, obstructing GBL aspects' hierarchical decomposition in terms of scope. 4) There is inconsistency in the definition (the theoretical meaning of aspect), usage (aspects selected for GBL assessment), scope (High-level and low-level), and terminology (terms used to refer to aspects at a certain scope, e.g., dimensions, criteria, factors) of aspects in GBL literature. The inconsistent use of GBL aspects does not allow the proper categorization of these elements, hindering identifying a clear pattern. There is no discipline between macro and micro-level elements. Some studies use feedback in a broader scope (main aspect) for evaluation. In contrast, other studies use feedback as a sub aspect to achieve usability. 5) Existing GBL evaluation approaches do not cater to children needs. 6) Majority of educational game evaluation studies do not use existing GBL models to devise their evaluation plan. Instead, they mostly use some pre-defined and ad hoc criteria for evaluation, or few follow the general guidelines/approaches. The findings of the review highlighted the need for a comprehensive GBL framework and led to the development of the LEAGUÊ framework presented in P2.

The *second review* improved the understanding of existing design practices in GBL, highlighted their strengths and weaknesses, and identified the following major problems in GBL design research. 1) The existing educational game design models do not offer tool support. 2) There is a lack of adaptability and concrete guidance for the practical application of framework concepts in the educational game design process. 3) There is a scarcity of empirical validation and independent evaluations of design models by other researchers or designers. 4) There is a lack of practical application of design models in the industry for educational game design practice. 5) To get useful insights from the industry, collaboration between industry and research is important. 6) Most of the frameworks do not provide any information on assessment approach, method or stakeholders that are required to participate in the assessment. The findings of the review highlighted the need for an operationalizable approach to assist the GBL design process in practice and led to the development of the LEAGUÊ ideation toolkit presented in P6.

The *third review* was directed to CCI due to the scarcity of educational games evaluation research with refugee children within GBL literature. The review identified some gaps in current CCI literature concerning methods and guidelines for research with refugee children. 1) There is a lack of research concerning the effectiveness of general evaluation methods for research with refugee children and focus on new or adapted research methods. 2) Most of the child-friendly evaluation methods are not utilized in the research in the refugee context. 3) There is a gap in the literature concerning methodological guidelines for research with the refugee group in line with the scarcity of research on methods' effectiveness. The findings of the review can be used as a starting point for research with refugee children. However, the review highlighted the need for more specific implications on methods and guidelines for conducting educational game evaluation research with refugee children that led to the case study research presented in P10.

The first contribution can provide useful information to the GBL community (educational game researchers, designers, developers) concerning the state of the art in GBL. It can help understand and contrast the alternative GBL design and evaluation approaches for selecting an appropriate one and guidelines available for conducting research. The educational game designers can also learn from the essential attributes and criteria highlighted in GBL research. At the same time, the highlighted problems and issues can guide the research community for future research directions. A part of this contribution can also serve as a foundation for novice researchers particularly interested in conducting learning game evaluations with the specific population of refugee children by presenting evaluation methods and guidelines that might also help guide the adaption of the research process.

6.2 C2. Present a Conceptual Hierarchical Framework of Six Dimensions for Comprehensive Evaluation of GBL Applications.

The *second contribution* presents a conceptual GBL framework for comprehensive design, analysis, and evaluation of learning games (Tahir and Wang, 2020a), see Figure 6.1. The framework was proposed to bridge the literature gap (identified in C1, P1) regarding core dimensions of GBL for comprehensive design and evaluation due to inconsistency in the use and comparable importance of aspects leading to the absence of a holistic view of GBL.

The framework was developed in an iterative process through directed content analysis (Hsieh and Shannon, 2005) of GBL literature comprising data extracted from articles in the systematic literature review in P1 (Tahir and Wang, 2017). The key motivation was to validate and conceptually extend the existing GBL research and theory. By analyzing, interpreting and organizing the numerous aspects used for GBL design and evaluation literature to categorize the core aspects systematically. According to the viewpoint of Loh, Sheng et al. (2015), the core aspects optimal for GBL design and evaluation must be focused on specifically GBL literature and not an integration of different sub-fields to properly assess, measure, and improve educational games. The existing GBL frameworks aided in the categorization process and were instrumental in ensuring no omissions.

The proposed LEAGUÊ framework for GBL defines and introduces four conceptual hierarchical levels concerning scope: dimensions, factors, subfactors, and metrics, to remove the identified inconsistently in terminology, scope, definition, and usage in the existing literature. Hierarchy is important to define aspects for a specific application domain (Kececi and Abran, 2001).

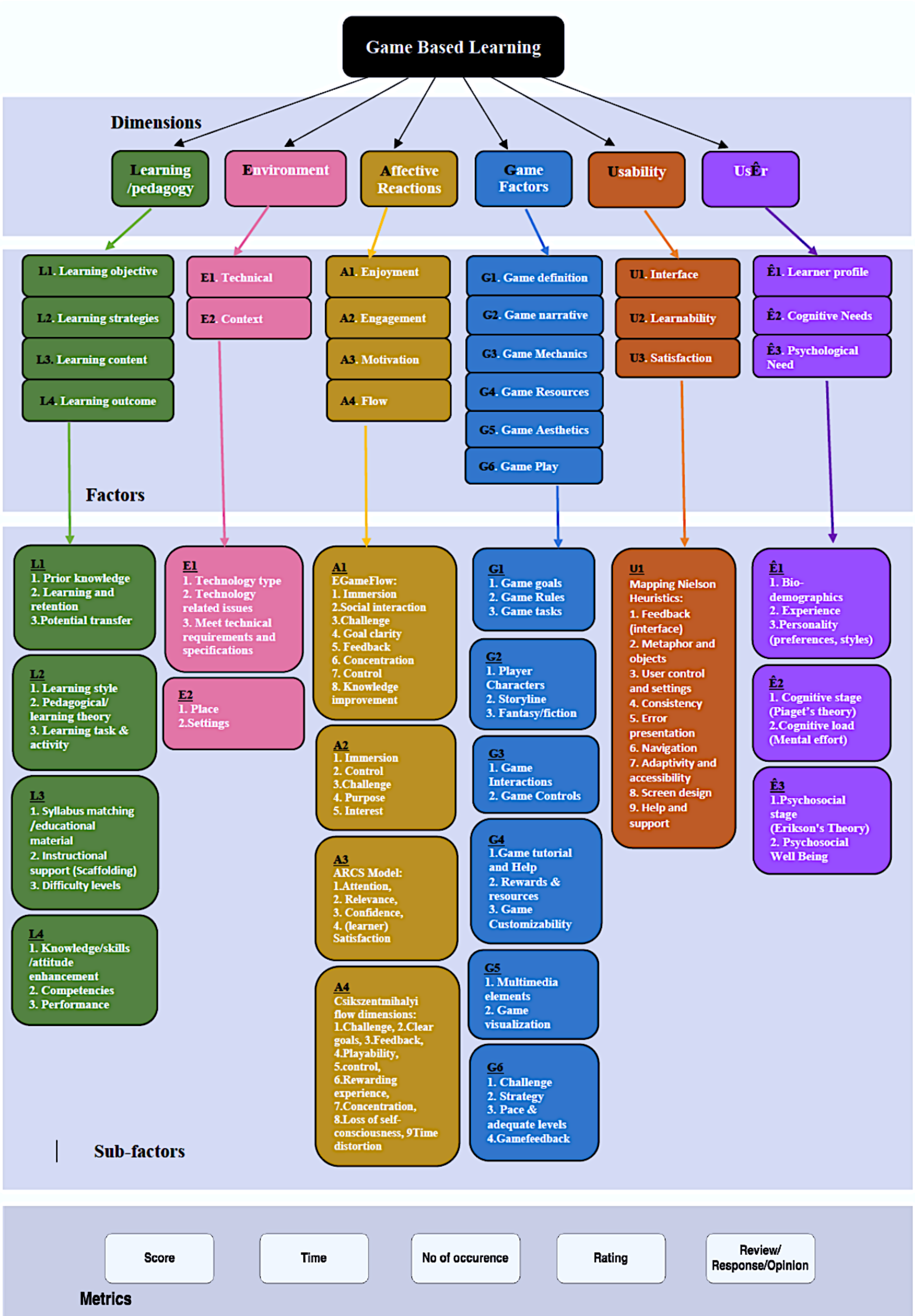


Figure 6.1: LEAGUÊ framework for GBL (hierarchical structure and components)

The framework's hierarchical levels are defined as follows: "*Dimension*" refers to a broader concept yet isolated within its type and not a composition of different aspects. Dimensions represent the main aspects of GBL, and each dimension signifies one specific aspect. "*Factors*" are intermediate-level concepts presenting elements essential for attaining a specific dimension. "*Sub-factors*" are further categorized elements constituting the factor. Lastly, "*Metric*" represents the lowest level in the hierarchy and is the measurement to assess a factor or sub-factor using objective or subjective data.

The conceptual framework identified six dimensions (learning, game factors, affective-cognitive reactions, usability, user, and environment) as key constituents of GBL design and evaluation. These dimensions are systematically categorized into factors and further sub-factors based on their theoretical construct allowing a hierarchical decomposition (see Figure 6.1). Sub-factors are easier to quantify and are mostly devised by mapping and integration conceptual elements from well-developed and widely accepted theories or models. Mostly in areas where researchers had a consensus in the literature. For example, the sub-factors of motivation (A3, Figure 6.1) are integrated from the ARCS (attention, relevance, confidence, and satisfaction) motivation model (Keller, 1987, Su, Chen et al., 2013). Similarly, the sub-factors of flow (A4, Figure 6.1) are incorporated from the flow framework (Kiili, Lainema et al., 2014). The sub-factors of context (E2, Figure 6.1) are adopted from the four-dimensional framework (de Freitas and Oliver, 2006). For a complete description, see P2. Factors where existing theories/models were not available for integration, the sub-factors were proposed following the procedure described in P2 (Tahir and Wang, 2020a). The proposed framework categorized the identified metrics into five types to guide the key metrics types used in GBL evaluation to be utilized and adapted for different evaluation studies depending on the selected evaluation dimensions and factors/sub-factors. The first three metrics for objective data and the last two for subjective (see Figure 6.1, bottom). The framework presented ten fundamental relations to show the GBL phenomenon's high-level abstraction (see Figure 6.2).

The proposed LEAGUÊ framework aims to identify and organize key GBL dimensions essential for producing an effective learning game and present GBL as a complex phenomenon requiring a multidisciplinary approach. The framework establishes a classification schema presenting different dimensions and summarizing the GBL knowledge that can be used as the selection criteria for focusing valuable concepts and establishing a connection between multiple dimensions. A basic vocabulary is provided to facilitate the application and use of framework components in multiple ways. The framework can be applied in the process of design, analysis, and evaluation of GBL. The dimensions in GBL (depending on the evaluation objective) might be considered in isolation (picking and selecting components) for a specific evaluation study. However, GBL dimensions are linked to each other in terms of cause and effect. They can be viewed as a collective whole to understand the process and help in design and analysis. The framework is put in practice by utilizing the framework components to develop three items: LEAGUÊ analysis instrument, LEAGUÊ ideation, and design toolkit, and LEAGUÊ evaluation guide for analysis, design, and evaluation of learning games, respectively. The framework components are validated by using and evaluating the developed instruments in different contexts.

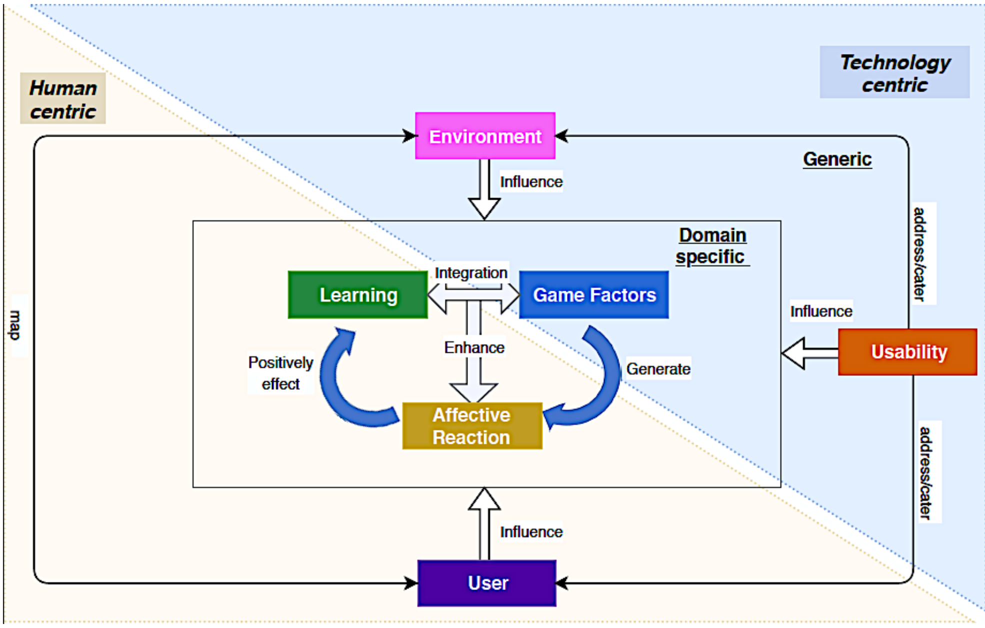


Figure 6.2: LEAGUÊ framework (Multidimensional approach)

The second contribution describes a GBL conceptual framework to support the comprehensive design and evaluation of learning games. The framework systematically categorizes the GBL concepts, providing valuable insights into the GBL phenomenon and the multidimensionality of this approach. The framework can support GBL researchers and stakeholders to understand the relationships between the multiple dimensions embedded in GBL. The framework components can act as a springboard to guide further research that other researchers can utilize and develop. This GBL framework differs from most existing ones, focusing on some specific aspects, making them difficult to use in practice. It is challenging when the objective is complete design and analysis of an effective learning game and not just focusing on few individual aspects of it. Moreover, it is also complicated when the learning game's target genre differs from the default game genre used in research (Shi and Shih, 2015). The specificities of LEAGUÊ with other frameworks can be highlighted by the comprehensiveness of the framework. It details the individual parts to allow analysis in terms of presence/absence and provide an overall picture and interconnection between the core dimensions. It can be equally useful for analytical and holistic evaluation providing a theoretical model essential for unifying all the different aspects of GBL. Furthermore, dimensions presented in LEAGUÊ are higher-level concepts and not restricted by the game genre.

6.3 C3. Empirical Evidence on the Application of GBL Approaches in Different Contexts for Improved Understanding about the Process of Learning with Educational Games and Contributing Factors.

The *third contribution* of this doctoral work comprises the GBL evaluation studies (quasi-experiments) focusing on evaluating GBL approaches' effectiveness in three different contexts: university, online lecture, and informal learning at home. These studies investigate the learning process with GBL, the contributing factors and their interrelation. This contribution provides insights on the emotional states and learning outcome with a GBL approach compared to the traditional learning approach to obtain insights on how the learning process acquire concerning the GBL phenomenon presented in P3 (Pireva, Tahir et al., 2019). The impact on engagement, learning gain, interactivity, and usability with the GBL approach in online lectures (compared to the non-game-based approach) is investigated in P4 (Pireva, Tahir et al., 2021). Lastly, the role of user characteristics in influencing learning gain, gameplay performance, and usability experience with the GBL approach in an informal learning setup is reported in P9 (Tahir and Wang, 2021b). Overall, based on the three quasi-experimental studies, GBL approach is effective for learning in formal, non-formal, and informal learning setups. When using the GBL approach, students performed better in practical concepts in university setup; students' engagement and motivation, learning gain, and interactivity between students and teachers increased in online lectures; and refugee children had a significant increase in reading assessment scores in informal learning setup at home. However, some issues were highlighted, and guidelines and implications (covered in C6) were presented for improving the effectiveness of learning games. Below, each GBL evaluation is discussed.

Game-based learners (students using the learning game) experienced better progress in practical questions than traditional learners (students using lecture notes, i.e., pen-and-paper), although extra attention and explanation were provided to traditional learners (details presented in P3). Game-based learners reached more complex tasks to solve and were also more engaged, interested, focused and relaxed during the learning activity. However, they were less excited compared to pen-and-paper learners. The finding provided insights into the learning process concerning the GBL phenomenon. The learners had the highest focus, interest, and excitement at the beginning of the game, and they experienced maximum relaxation immediately after they started the game. In line with the flow theory (Csikszentmihalyi, 2000, Csikszentmihalyi, 2014), the stress started to increase as the complexity of the game's tasks increased, but a simultaneous increase was also observed in the engagement, after which the stress started to decrease. Moreover, the value of learners' relaxation was also less at this point, which means that the game-based learner skill level was not too high to fall into boredom. The start of a decrease in stress level showed that the game challenge level was not too high for the learner to fall into anxiety. According to the flow theory, the flow emerges in the space between anxiety and boredom. Game design can keep the player in a flow state by increasing the game's challenge level at the same pace as the player's skill level increases to maximize their impact that results in learning. Therefore, from the analyses of the complete data set, it can be concluded based on the values of stress and engagement that the game-based approach can generate affective reactions (flow) that can positively impact the learning. On the contrary, although traditional learners experienced interest and excitement at the beginning of the learning activity, their engagement and focus were overshadowed by stress as the learning activity progressed. The engagement decreased towards the end in this group of learners when tasks became more complex. Another significant difference between the game-based and the pen-and-paper learners was the average number of solved tasks. When comparing the number of tasks solved by the participants of both groups, the game-based learners achieved more than double the number of tasks (reaching complex tasks) with almost no help at the same time as the pen and paper group that required additional explanation. It is a particularly promising line of research because the game, with its interactive and multimedia elements, can facilitate learning and reduce the cognitive load in complex tasks that reduce stress and increase engagement in GBL. The cognitive load can be reduced by presenting information in chunks (Thalmann, Souza et al., 2019) and using interactive visualization (Khalil, Paas et al., 2005).

Due to COVID-19, lectures shifted to the online mode that introduced the challenges concerning keeping students engaged and active virtually. The potential role of learning games made it relevant to evaluate the GBL approach's effectiveness in this context of distance learning during quarantine and understand the GBL learning process. The results further improved our understanding of GBL and contributing factors also the specific impact on online learning and teaching with this approach (details presented in P4). Online learners had a better experience in terms of fun, engagement, and usability with the GBL approach (using game-based digital quiz) compared to the non-game-based approach (non-game-based digital quiz). The categories that contributed to the effectiveness of the GBL approach in online learning were the following: *game factors* (such as competition, time pressure), *engagement*, and *interactivity* (not only within technology but interactivity between students and teachers). It is an interesting finding as it expanded the study dimensions of the learning process with GBL to include and emphasize interactivity and class dynamics for GBL in online learning. It is also interesting to note that this factor was equally important for students and teachers. Both positively experienced it with the GBL approach in an online learning setup. Students preferred the GBL approach owing to better interactivity in addition to game experience and engagement. Thus, one of GBL's strengths is effectively increasing student-teacher interactivity and class dynamics (Wang and Tahir, 2020). The elements that increased cooperation and interactivity between teachers and students were easy and fun teaching and learning (game-like) and being able to assess and evaluate the performance (feedback on answers, whether correct or incorrect). Moreover, results emphasized increasing the simplicity, accessibility, ease of use, and efficiency of GBL applications to improve learning games' usability in online learning. The learning performance of students increased in online lectures with the GBL approach as compared to without GBL. Anecdotally, in general, it is fascinating that students start to complain when they have systematic in-lecture quizzes. However, they felt motivated to have systematic in-lecture quizzes once the professor selects game-based platforms to perform the quizzes.

The recent Syrian crisis (Hamilton and Moore, 2003, Yazgan, Utku et al., 2015, Thomas, 2016) and an initiative project “EduApp4syria” (aimed at developing language learning games for Syrian children) partnered by NTNU provided us the opportunity to investigate GBL for informal learning at home with the special group of migrant refugee children. This was carried out to explore the potential of this approach and the effects of user characteristics on the GBL experience. The results showed that the GBL approach effectively increased the learning gain of children in the context of informal learning (details presented in P9). Children had a significant increase in their scores after playing with the language learning game for one week at home. The increase in learning gain with GBL is not something new, and other researchers have also found similar results focusing on formal teaching setup (Salah, Abdennadher et al., 2016, Azizt and Subiyanto, 2018, Kenali, Yusoff et al., 2019). Some interesting findings were the impact of user characteristics on GBL experience of learning gain, usability, and gameplay performance. In particular, the findings suggest that the children's age can be associated with their learning gain, usability, and gameplay performance with GBL. Thus, a learning game needs to adapt to each child individually for improved effectiveness (Andersen, 2012, Vandewaetere, Cornillie et al., 2013). The game's educational material and the difficulty level are important. They should not be too easy for older children, so they start feeling bored. The learning game should offer children different challenges as the perceived challenge is the strongest motivator for children (Greenberg, Sherry et al., 2010). However, the challenges must be balanced with their skill level in line with flow theory (Csikszentmihalyi, 2000, Csikszentmihalyi, 2014).

The need for adaptivity in the learning game is emphasized, so the game should increase the difficulty level in relation to player skills (Peirce, Conlan et al., 2008, Streicher and Smeddinck, 2016, Plass and Pawar, 2020). The older children had better usability score than the younger children. It is interesting to note that this difference is not linked to their higher mobile usage experience in the conducted study but to age-related factors. The highlighted usability issues were related to cognitive and motor skills. According to Piaget's cognitive development theory (Huitt and Hummel, 2003) these skills are not fully developed in younger children. Moreover, the older children played more game levels than the younger. It is also linked to the nature of older children being competitive (Greenberg, Sherry et al., 2010), and they want to complete all levels to win the game. The children's learning modality preferences are not associated with their learning gain debunking the theory that presenting material in preferred modality can improve learning (Lodge, Hansen et al., 2016). However, the use of different media may afford different instructional methods and improve the learning process (Moreno, 2006). On the other hand, learning modality preferences do impact usability and gameplay performance. Children’s preference for the read/write learning mode positively affected their usability score. The preference for visual, read/write, and auditory learning modalities affected the total time played. The preference for read/write and auditory modalities affected the total game levels completed by children. The employed game focused on language learning and had more audio and text to teach the sound and written form of alphabets and words. Therefore, children with a higher preference for read/write, video, and auditory modalities played the game more than those with a higher preference for kinesthetics. Some children prefer playing more physical games, and they are difficult to keep engaged and focused. Therefore, for GBL to be effective for informal learning specifically targeting children, it is recommended to provide different learning activities incorporating multiple modalities (Alkhasawneh, Mrayyan et al., 2008, Ward, Paul et al., 2017). Moreover, as opposed to the modality-specific learning style theory, multimodal learning is supported by theoretical development (Aslaksen, Haga et al., 2020).

Children's mobile usage experience also correlates to their usability and gameplay performance but does not impact learning. Mobile usage expertise and years of use positively affect the usability score of first-time play. The children who require less help and have previously used mobile for some years are already familiar with many of the interface characteristics shared in the current game (Salanova, Grau et al., 2000). Moreover, the number of years of mobile use is also positively correlated to the score in the gameplay session. Therefore, experience parameters are important for effective technology (Orvis, Horn et al., 2008, Belay, McCrickard et al., 2016). Another interesting factor was the relationship between dependency in mobile usage and gameplay performance. The children with higher dependency (need someone else to play with them) played a greater number of sessions. These children are mostly playing with family (or friends), and this social interaction drives the additional engagement. Research has shown that availability of adults and co-participation in play can increase the children's duration of play (Siraj-Blatchford, 2009, Pursi and Lipponen, 2018).

The third contribution provides useful insights into GBL's effectiveness in different contexts and the factors affecting the GBL learning experiences. This contribution improves the understanding of the GBL phenomenon. It can be useful for researchers, designs and educators interested in researching, designing or employing GBL applications. The presented findings concerning GBL dimensions and interrelated factors rationalize and improve the interpretation of GBL design space that is important to increase the effectiveness of this approach in different setups. Educational game stakeholders should consider the existence of these interrelated factors and realize their impact on learning with the GBL approach.

6.4 C4. Contribution to the Design, Implementation, and Evaluation of a Card-Based Design Toolkit for the Ideation Phase of Learning Game Design, Facilitating Multidimensional Focus and Collaboration in the GBL Design Process.

The *fourth contribution* comprises new knowledge about how the GBL framework's theoretical concepts (from P2) can inform the creation of a card-based GBL ideation and design toolkit. This contribution presents the design, implementation and evaluation of the GBL design toolkit and focuses on facilitating the learning game design practice (Tahir and Wang, 2020b). It also explores the scaffolding such an approach can provide for multidimensional focus and collaboration in the GBL design process (Tahir and Wang, 2021a). This contribution presents a ten-step process for transforming a theoretical framework into design cards and an ideation toolkit to support the GBL design practice (presented in P6). It further investigates completeness and collaboration in the GBL design process facilitated by the proposed card-based approach (described in P7).

The theoretical frameworks detail the required knowledge but are hard to use in design practice. They lack tool-support and guidance for the practical application (issues identified in C1, reported in P5). Therefore, to address this lack of tool support and operationalizable approaches for the educational game design process, a lightweight approach was required to integrate the research-based theoretical GBL knowledge into the actual ideation process and hands-on learning game design practice to assist the early design phase of GBL. As a result, a card-based toolkit was developed to reduce the complexity of framework application and the introduction of key GBL concepts in the design process to bridge the gap between theory and practice. Although several game design cards are available, none is specific for GBL. Therefore, they cannot inform the required design knowledge to reinforce essential GBL concepts and facilitate the required multi-dimensional focus. Consequently, to make the GBL design knowledge easily accessible in the early design process, the proposed GBL ideation toolkit was developed following the Hornecker's proposition (Hornecker, 2010). The LEAGUÊ framework (presented in P2) was transformed into an ideation and design toolkit. The framework's theoretical concepts were converted into easy questions that designers can relate to, which can be introduced in the design process. Another limitation of existing card-based tools is that most researchers have not articulated the design knowledge embedded in them (Deng, Antle et al., 2014). Moreover, except Mueller, Gibbs et al. (2014), none of the others explicitly detail the steps of transforming the framework into a design tool. Not many well-defined processes exist, and the five-stage process described by Mueller, Gibbs et al. (2014) is not validated beyond their work. Therefore, their process was used as a starting point and further adapted and extended (based on experience gained) to validate and extend the prior work. Our extended process of transforming the framework into ideation cards consisted of the following ten steps (for detailed description, see P6): define goals/objectives, establish target boundaries, scrutinize framework to extract concepts, decide the type of cards, formulate the content, reduce items, define rules/process, visualize, gather feedback, refine and improve.

The LEAGUÊ ideation toolkit aims to make the theoretical knowledge about designing learning games easily accessible to GBL design teams facilitating the design process and providing inspiration. The main objective of the developed toolkit was to (i) inform GBL design knowledge, (ii) support the collaborative design process, (iii) facilitate brainstorming, (iv) reflection, and (v) provide guidance for GBL ideation. The developed toolkit introduces four card decks (primary cards, trigger cards, reflection cards, and custom cards) informing GBL concepts (see Figure 6.3).

Figure 6.3: LEAGUÊ card types

Further, it presents five design activities with ideation sheets (idea generation, idea development, idea refinement, idea illustration, and idea documentation), supporting the ideation process; a board with a playbook; a log sheet; and a workshop technique (played in a group of four to six players) for learning game design ideation. For a detailed description, see P6. The toolkit was developed during several design iterations. It was empirically evaluated in three workshop sessions to investigate its effectiveness in informing and guiding the ideation of learning games design in practice. The feedback gathered enabled the refinement of the toolkit and the process. Also, it strengthened the argument for the transformation of the framework into design cards. The toolkit was useful for the GBL early design process concerning participants' experience (perceived usefulness, understandability, level of fun and satisfaction) and achieving toolkits' objectives. In a short time, teams could ideate, develop, refine, illustrate, and document their educational game design ideas using the toolkit artifacts. The results also revealed the elements that obstructed the learning game design concerning workshop format and working with cards that informed design iterations and future directions.

Although design cards can improve idea generation and communication between stakeholders, the potential scaffolding for completeness and collaboration in learning game design is not explored. Completeness refers to focusing on all key GBL dimensions, and collaboration refers to working together to produce something. The investigation explicitly focusing on the scaffolding for multidimensional focus and collaboration provided by this design approach (ideation cards or card-based toolkit) is described in P7. These two elements are essential for GBL design practice. They can be used to learn about the GBL design process as a collaborative design activity engaging various stakeholders. The previous research showed that participants' subjective opinion is not enough to evaluate design cards (Sintoris, Mavrommati et al., 2018). Therefore, the produced design artifacts and video recordings from the design workshop sessions were the primary data sources for investigation. Thus, the study also provides insights into the potential of these data collection approaches for analyzing the GBL ideation process. The results are encouraging in terms of the applicability of ideation cards in the GBL design process to scaffold completeness and collaboration and also to reflect on factors and design decisions in the employed card deck/activities that advance these key outcomes. Considering the multidisciplinary nature of GBL (Winn and Heeter, 2006, de Freitas, 2018), it is essential to focus on key GBL dimensions (targeting relevant concepts within each dimension) for effective GBL design. The card-based tool can facilitate the teams to address key GBL dimensions in the produced game design ideas. The different categories (color-coded for easy searchability) can support achieving multidimensionality. Since cards act as tangible idea containers, by converting the key GBL dimensions into different card categories, they act as design building blocks that team members can use to develop and complete their design ideas from multiple angles (achieving multidimensionality in design). The activity format can remind the team to revisit the design decisions and improve their idea by working on the missing aspects. Moreover, breaking the ideation task into different activities supports completeness. Each new activity puts things into perspective, providing an opportunity to revisit the design decisions and further add or modify them if needed. The time restrictions in the design can limit the use of cards and thus ignoring some concepts. However, this can be controlled by changing time limits in real-life settings. The results highlighted that a card-based tool provides three central features that scaffolds collaboration. It includes physical point of interaction (tangible props), social point of interaction (social setting) and mental point of interaction (common goal/task). These three interaction points (see Figure 6.4) effectively instigate and foster collaboration among team members in the design process. Moreover, these three points have mutually beneficial relationships that together support the collaborative design process. Since each card focused on one specific GBL element, it provided a comprehensive description of that element (using definition, examples or images). It made it easier for all stakeholders (from different areas) to understand the concept. Moreover, it also made it easy for team members to use that tangible information to further extend and explain their ideas to other team members.

The fourth contribution provides a toolkit that can be a resource for the GBL team (including GBL researchers, students, practitioners in the industry, or anyone interested in generating ideas for learning game design) to support the hands-on learning game design practice. It can be used to improve the collaborative design process and guide the team members. The LEAGUÊ toolkit can function as both a practitioner tool and a research instrument to further the GBL design domain. Researchers in other domains can also learn from transforming the framework's theoretical knowledge into a lightweight card-based tool. The proposed ten-step process can guide other researchers to do a similar task of converting a framework into design cards. The developed card-based toolkit provides a tangible reference point to facilitate multi-dimensional focus, support idea generation, facilitate critical reflection, and create a shared understanding in the collaborative design process. The toolkit supports co-design as a strategy to involve different stakeholders in the ideation phase of learning game design. The group-based activities, the use of specific cards designed to inform GBL knowledge, spark creativity and reflection, and the physicality of combining toolkit artifacts to produce game design ideas helped develop a shared understanding despite the diversities in participants' backgrounds and skills. It was intended to be a GBL-specific yet generic tool that can be used to design educational games for diverse learning domains and game genres, supporting GBL designers in the initial ideation phase. This contribution also provides valuable insights into how a card-based tool can scaffold multi-dimensional focus and collaboration, highlighting the contributing factors. The developed card-based tool also has a strong potential of being a framework for analyzing the GBL's collaborative ideation process of multidisciplinary teams. Researchers in other domains can also learn from this approach to investigate collaborative design activities.

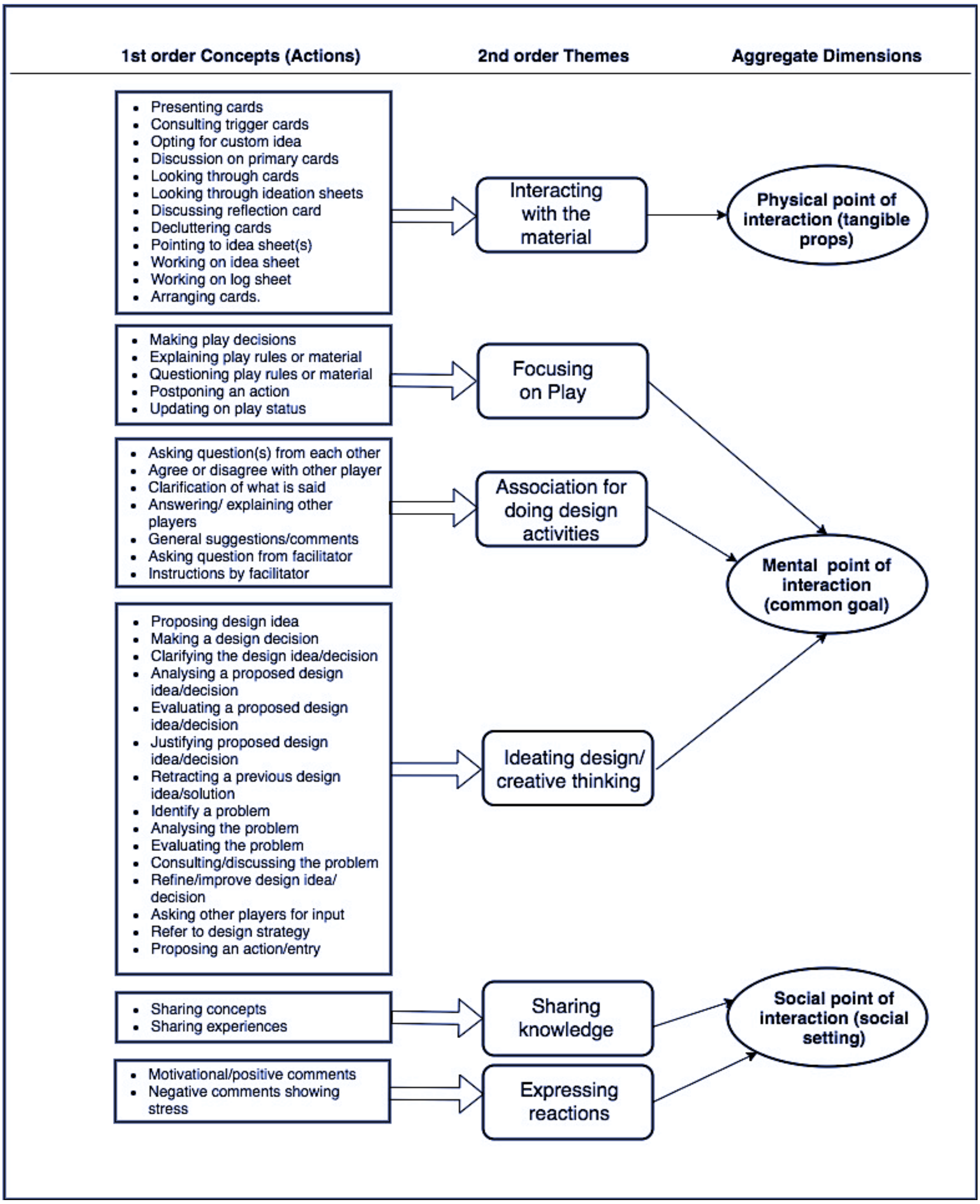


Figure 6.4: Card-based tools' features scaffolding for collaboration

6.5 C5. Contribution to the Development and Application of an Analysis Instrument and an Integrated Evaluation Approach to Support the Educational Game Evaluation Process.

The *fifth contribution* of the thesis includes the development and application of a GBL analysis instrument (Tahir and Wang, 2020a) and an integrated GBL evaluation approach (Tahir and Wang, 2021b) based on the LEAGUÊ framework (presented in P2). This contribution focuses on facilitating the GBL evaluation process with the developed instrument and approach to assess and evaluate educational games. The findings from learning game analysis using the developed LEAGUÊ instrument (presented in P2) and the quasi-experimental (GBL evaluation) study using the proposed integrated LEAGUÊ-GQM approach (presented in P9) provide design guidelines for producing effective educational games (covered in C6). However, the artifacts (instrument and approach) themselves are oriented into supporting the GBL evaluation process.

As highlighted by Baehr (2005), there is a distinction between assessment and evaluation: the prior provides insights, strengths, areas for improvement and feedback on performance, whereas the latter determines whether the quality and standard were met and measured performance was a success or failure. The two processes are complementary and necessary to occur at separate times and settings through different roles. Although aspects important for GBL are present in literature, researchers have highlighted the need for an overarching approach to guide evaluation and design iterations (de Freitas and Liarokapis, 2011, Van Staalduinen and de Freitas, 2011, Oprins, Visschedijk et al., 2015). Therefore, to facilitate the GBL evaluation process, two artifacts were developed for supporting assessment (with GBL analysis instrument) and evaluation (with integrated GBL evaluation approach).

According to Marciano, de Miranda et al. (2014), it is essential to systematically assess learning games to verify their potential. The diverse characteristics of GBL make it a difficult task (Djelil, Sanchez et al., 2014). The developed LEAGUÊ analysis instrument facilitates the systematic GBL assessment to understand the potential of educational games in a specified environment by knowing the strengths and areas in which it could improve. It is often recommended to carry out analysis early because it is easy to make changes and improvements at an early stage of development as they get expensive later. Therefore, it is useful to carry out an analysis before actual evaluation. Researchers have highlighted the importance of lightweight instruments (such as heuristics) as valuable inspection techniques or tools, commonly used for formative evaluation of computer games in design and evaluation phases of development (Mohamed and Jaafar, 2010b). However, most heuristics or guidelines do not allow for in-depth analysis and reflection. In comparison, the LEAGUÊ analysis instrument probes questions to gain more significant insights and instigate reflective thinking. The proposed instrument comprises three parts: a primary analysis form, a secondary form (to reinforce in-depth analysis), and a reflection form. Four steps are outlined for using the LEAGUÊ analysis instrument for assessing learning games to facilitate understanding by evaluators (stakeholders from different backgrounds and areas of expertise). The primary form is presented in Figure 6.5. For a more detailed description, see P2.

Learning	Environment	Affective-Cognitive Reactions	Game Factors	Usability	UsÊr
L1. What are the learning objectives of the game?	**E1**. What technical aspects are required for the game to work and best support learning?	**A1**. How can the game provide enjoyment to the users?	**G1.** What are the game objectives to integrate learning?	**U1.** How interface of the game is made easy to use for target users?	**Ê1.** What are the attributes of the target users of the game?
L2. Which learning strategies are being used to enable learning through the game?	**E2**. What is the context for playing the game for learning?	**A2**. How can the game engage the users?	**G2**. What narrative is used to make game compelling and integrate learning?	**U2**. How does the game provide easy learnability to its target users?	**Ê2.** Which cognitive needs (of target users) are considered in the game?
L3. What learning content is being used in the game for target users?		**A3**. How can the game motivate the users?	**G3.** What mechanics are used to make game compelling and support learning?	**U3.** How does the game provide satisfaction to its target users?	**Ê3.** Which Psychological needs (of target users) are considered in the game?
L4. What learning outcome(s) can be acquired from the game?		**A4.** How can the game generate flow?	**G4.** What resources are provided to the users to function effectively that also support learning?		
			G5. What aesthetics are used to make game compelling for target users?		
			G6. What game play is used to make game compelling for target users and support learning?		
Strength of the analyzed game					
Weakness of the analyzed game					

Figure 6.5: LEAGUÊ analysis instrument (primary form)

The learning game is analyzed based on whether it contains the elements deemed important for game-based learning by splitting the subject matter into its fundamental components. The factors laid out in the *primary form* help analyze the learning game in terms of individual elements essential for an effective game-based learning approach and highlight any weak or neglected areas in the game, presenting the game's overall strength and weakness. The *secondary form* supports an in-depth analysis of each element of the primary form. It facilitates the thinking process to give concrete answers to questions in the primary form. The *reflection form* instigates critical analysis and design trade-offs to highlight revisions or improvements needed in the game. The developed analysis instrument is applied in two case studies to illustrate the instrument's application and test its efficacy for analyzing learning games (see Chapter 4, Section 4.3.3). A pilot study was carried out with the first version of the instrument to inform improvements before the two actual case studies were performed. The improvements included the refinement of questions, introducing secondary form, adding strength/weakness, and required improvements sections. Both case studies were carried out by game developers (with modifications in the application process) to guarantee effective use by stakeholders. The first case study on an online empathy game is described in P2, and the second case study on a VR game (for STEM and space) is reported in (Karlstrøm and Markussen, 2020). The second case study included a more extensive evaluation where the LEAGUÊ instrument was used in two iterations (using feedback from user study) for improving the game versions. The feedback on the usefulness and effectiveness of the LEAGUÊ instrument was collected from the game developers for both case studies that enabled the design iterations.

One of the problems identified in C1 was the use of pre-defined and ad hoc criteria in GBL evaluation studies. Previous research has identified that the evaluation criteria and evaluation process are the main challenges in evaluating educational games (Mohamed and Jaafar, 2010a). According to Dondi and Moretti (2007), it is difficult to classify different evaluation processes (analytical or single aspect and global or holistic). Moreover, there are not many approaches available to guide the GBL evaluation process (Becker, 2011). Evaluation is an integral part of applications' success to remove imperfections, increase their effectiveness, and fit their purpose (de Freitas and Oliver, 2006, Gossen, Hempel et al., 2013). However, identifying evaluation criteria is a complex and time-consuming process (Dondi and Moretti, 2007). Research has highlighted the need to define the key aspects that can serve as evaluation criteria and guide the GBL evaluation process to improve learning games (Ak, 2012). The proposed integrated *LEAGUÊ-GQM approach* guides the planning and execution of GBL evaluation studies. It is a simple and effective approach that integrates the LEAGUÊ framework and GQM approach to presents a three-step parallel process with an evaluation guide. It can be used for creating an evaluation plan for GBL evaluation studies. The components of the LEAGUÊ framework (presented in P2) are used to develop an evaluation guide (see Table 6.2) presenting GBL evaluation criteria with three measurement levels of the GQM approach. The LEAGUÊ evaluation guide facilitates establishing GBL evaluation goals, defining evaluation questions, and identifying measures for the evaluation process by picking and selecting elements from the guide. The complete description is presented in P9. According to GBL evaluation literature (Calderón and Ruiz, 2015, Tahir and Wang, 2017), educational games are evaluated at different development stages. Depending on the evaluation goal, different characteristics are assessed by selecting different criteria. The integrated approach can be used to identify the critical GBL evaluation criteria and prioritize with respect to evaluation goals for planning GBL evaluation to verify that game has satisfied its specified objectives. The main novelty of the presented approach is its simplicity and the GBL-specific guidance. The proposed integrated approach is employed to plan a GBL evaluation study investigating a language learning game's potential for teaching Arabic reading skills to migrant refugee children (presented in P9).

Table 6.2: LEAGUÊ-GQM evaluation guide

<table>
<tr><th>Conceptual level
(objects/goals)</th><th colspan="3">Operational level
(assessment/questions)</th><th>Measurement level
subjective objective/ measures)</th></tr>
<tr><td>Plan evaluation goal choosing dimensions</td><td colspan="3">Plan evaluation questions choosing factors/sub-factors, relations, and select data sources</td><td>Plan evaluation measures and select analysis methods choosing metric types</td></tr>
<tr><td>Dimensions</td><td>Factors</td><td>Sub-factors</td><td>Relations</td><td>Metrics types</td></tr>
<tr><td rowspan="4">Learning</td><td>Learning Objectives</td><td>Prior knowledge, learning, and retention, Potential transfer</td><td rowspan="10">- Learning & Game Factors
- Game Factors & ACR
- ACR & Learning
- (Integration of gameplay and learning) & ACR
- Usability & (Learning, Game factors, ACR)
- Usability & Environment
- Usability & User
- User & Environment
- User & (Learning, Game factors, ACR)
- Environment & (Learning, Game factors, ACR)</td><td rowspan="6">Objective:
- Scores
- Time
- Number of occurrences
- Rating
Subjective:
- Reviews/ responses/ opinions</td></tr>
<tr><td>Learning Strategies</td><td>Learning style, learning theory or model, learning task/activity</td></tr>
<tr><td>Learning Content</td><td>Educational material, instructional support, difficulty levels</td></tr>
<tr><td>Learning Outcome</td><td>Knowledge/skills/attitudes enhancement competencies, performance</td></tr>
<tr><td rowspan="6">Game Factors</td><td>Game Definition</td><td>Game goals, Game Rules, Game tasks</td></tr>
<tr><td>Game Narrative</td><td>Player Characters, storyline, fantasy/fiction</td></tr>
<tr><td>Game Mechanics</td><td>Game Interactions, Game Controls</td><td rowspan="6">Analysis Methods
Qualitative analysis methods:
- Measures of central tendency (mean, median, quartile, mode)
- Measures of dispersion (standard deviation, range)
- Measures of dependency (Pearson correlation coefficient, Spearman rho)
- Graphical visualization (histograms,</td></tr>
<tr><td>Game Resources</td><td>Game tutorial and Help, Rewards & resources, Game Customizability</td></tr>
<tr><td>Game Aesthetics</td><td>Multimedia elements, Game visualization</td></tr>
<tr><td>Game Play</td><td>Challenge, Strategy, Pace &
adequate levels, Game feedback</td></tr>
<tr><td rowspan="2">Affective-Cognitive Reactions (ACR)</td><td>Enjoyment</td><td>Immersion, social interaction, Challenge, Goal clarity, Feedback, Concentration, Control, Knowledge improvement</td><td rowspan="2">Data sources
- Questionnaires
- Observation with checklist
- Observation or field notes
- Interviews
- Pre/post-test</td></tr>
<tr><td>Engagement</td><td>Immersion, Control, Challenge, Purpose, Interest</td></tr>
</table>

	Motivation	Attention, Relevance, Confidence, (learner) Satisfaction	- Screen recording - Game-logs - Expert evaluation - Video recording - Challenges, tasks, or exercises - Learner diaries/reports	frequency diagrams, line charts, box plots, scatter plots, pie charts) - Hypothesis testing (t-test, Mann-Whitney, F-test, Z-test, Wilcoxon matched-pairs test, ANOVA, Kruskal-Wallis, ANCOVA) Qualitative analysis methods - Content analysis - Grounded theory approach
	Flow	Challenge, Clear goals, Feedback, Playability, control, Rewarding experience, Concentration, Loss of self-consciousness, Time distortion		
Usability	Interface	Feedback (interface), Metaphor and Objects, User control and settings, Consistency, Error presentation, Navigation, Adaptivity and Accessibility, Screen design, help, and support		
	Learnability			
	Satisfaction			
User	Learner Profile	Biodemographic, Experience, Personality (preferences, styles)		
	Cognitive Needs	Cognitive stage (Piaget's theory), Cognitive load (Mental effort)		
	Psychological Needs	Psychosocial Stage (Erikson's Theory), Psychosocial Well Being		
Environment	Technical Aspects	Technology type, Technology related issues, Meet technical requirements and specifications		
	Context	Place, Settings		

The fifth contribution can be a resource for GBL stakeholders interested in assessing and evaluating an educational game. The proposed instrument can be used as a support tool for GBL researchers, designers/developers, and intermediates like teachers/parents to ensure that key issues and essential GBL concepts are considered, facilitating learning game assessment practice. Educational game researchers can use this instrument to learn more about the different elements used in the learning games and their relationships to gain experience from both successful and failed game concepts. Educators or parents can use the analysis instrument for analyzing the learning game to assess the potential and develop trust and conviction for justification to use the game as an efficient tool or not. LEAGUÊ-GQM, on the other hand, is a guiding approach to define the study dimensions and develop a GBL evaluation plan. It can be useful for novice GBL evaluators for planning and designing a learning game evaluation study both for single and holistic GBL evaluation by selecting elements from the evaluation guide.

6.6 C6. Guidelines for Improving the Design and Evaluation of GBL Applications in General and Specifically for Refugee Children.

The *sixth contribution* provides an overview of the lessons learned and guidelines concerning GBL design and evaluation. These guidelines emerged from the knowledge gained from conducting GBL evaluation studies, design workshops, case studies, and comparative and retrospective analysis of the outcomes. Based on all the evaluation studies (see Chapter 4, Section 4.3.3) conducted in the Design Science Research (DSR) cycles of this doctoral work (Figure 4.3), guidelines and recommendations emerged for general GBL design and evaluation and specifically for refugee children.

6.6.1 Guidelines for Improving GBL Design and Evaluation

This section presents the guidelines for designing and evaluating effective learning games extracted from research work conducted in this thesis.

- **Guidelines for GBL Design**

The guidelines and recommendations for GBL design are discussed supported by P2, P3, P4, P9. To facilitate understanding and repetition, seven themes are created, consolidating the implications from the papers:

- **Captivating game features**: An educational game needs captivating design and engaging features to enhance affective-cognitive reactions and keep users engaged for a longer time. A fundamental motivational need is to capture and sustain the learner's attention. Therefore, game material must be appealing to users and presented in different ways (Bixler, 2006). If the learning game lacks good game features, learners will be less excited to use the GBL approach in the traditional learning environment, as highlighted in the first GBL evaluation study. Game features (such as time pressure, competition, interactivity) enhance the game experience and keep users interested, as highlighted in the second GBL evaluation study. Moreover, fantasies can be useful in creating motivational learning environments if carefully chosen to attract the target users (Malone, 1981b). The use of fantasies and providing ties to past experiences can enhance learner's satisfaction and boost self-esteem (Bixler, 2006). In the third GBL evaluation study, some children did not connect with the story, resulting in low satisfaction.

- **Good usability**: It is vital that the learning games must have good usability. Simplicity, efficiency, accessibility, and ease of use can increase the usability of GBL applications. Especially for online lectures, the GBL approach should be simple, efficient, easy and user-friendly, have an interface that provides quick access, and no delays in using the learning game. Moreover, it is essential to pay particular attention to designing the educational game interface to offer greater learnability and user satisfaction. Poor usability can increase the extraneous cognitive load in multiple ways. For example, poor navigation requires the extra effort of learners to find the relevant information. Similarly, using unfamiliar terms with respect to learners' mental models or inconsistent design requires additional effort by learners to understand the material (Schmidt, Earnshaw et al., 2020). It is also evident from the present study where younger children had difficulty in understanding the unfamiliar icons.

– **Challenge-skill balance:** It is crucial to create a balance between game challenges and player skills and knowledge (Malone, 1981b, Sweetser and Wyeth, 2005). The game's challenge level should increase in accordance with the increasing skill level of the player to maximize the impact that results in learning. The game should be able to detect the skill level of the player with increasing complexity and provide required help and feedback. The first GBL evaluation study results showed that players' engagement started to increase with increasing challenges. However, towards the end of the learning activity (as the tasks became even more complicated), they started to lose interest and, consequently, relaxation. It is also essential to balance task challenges according to the player's skill level to control the intrinsic cognitive load in a learning game (Schrader and Bastiaens, 2012).

– **In-game assessment options (learning performance):** The educational games should provide player assessment options within the game (Ismail, Ahmad et al., 2019). According to the second GBL evaluation study results, both online learners and teachers liked the GBL application's assessment option and considered it useful. Learners found it useful to evaluate and challenge themselves. In contrast, teachers used the assessment results as input for re-planning and redirected the lectures to focus on questions where students showed low performance. According to Viberg, Wasson et al. (2021), the use of multi-channel data (data logs and self-assessment generated data using instruments to measure learners' motivation) can provide a deeper understanding of learners learning process and provide support.

– **Social interaction.** Educational games should offer elements for social interaction. In the present study, learners preferred and liked the GBL application, incorporating elements that increase cooperation and competition among students and interaction between teachers and students. Interactive teaching and learning are easy and more fun, making learning more attractive and vigilant. Previous research indicated that co-participation and adults' availability could increase children's play duration (Siraj-Blatchford, 2009, Singer, Nederend et al., 2014, Pursi and Lipponen, 2018). It was also observed in the third GBL evaluation study where children with higher mobile usage dependency played more sessions. These children need someone else to play with them. Therefore, because these children are mostly playing with family or friends, the additional engagement may be driven by social interaction.

– **Context.** Context of use should be taken into account when designing learning games. It was observed in the three quasi-experiments that the context had an impact on the GBL experience of participants directly or indirectly. In the first experiment, students learned practical questions more effectively using the game in the university setting than theoretical questions that had better results with the traditional learning method. In the second experiment, the online class had difficulty using “google quiz” because of some technical issues. Therefore, participants did not want to proceed with this platform and chose “Kahoot!” due to its efficient use in the large online class without any technical problems. In the third experiment, the children used the game for one week at home, and from the parents’ feedback, it was found that children played the game longer when they did not have other options to play or were feeling bored. Whereas, in playtest sessions, most children lost interest after 15-20 minutes.

– **Cognitive load.** The learning games must be designed in a way to reduce the unnecessary cognitive load on learners' working memory capacity (Schrader and Bastiaens, 2012). It is crucial to reduce the amount and complexity of information provided by the game that needs to be processed along with the simultaneous actions requiring cognitive and motor activities (Kalyuga and Plass, 2009, Whitton, 2009, Schrader and Bastiaens, 2012). Researchers suggest that designing appropriate support within the game can compensate for the cognitive load associated with its use (Aleven, Stahl et al., 2003). Extraneous cognitive load can be reduced by creating highly usable games allowing the mental resources to focus on germane cognitive load for schemas construction (Sweller, Van Merrienboer et al., 1998, Schmidt, Earnshaw et al., 2020). Furthermore, designing appropriate game tasks and learning activities in accordance with the age and skills of target players can reduce the intrinsic cognitive load in learning games as it depends on the given tasks' complexity in relation to the level of expertise of the player (Schrader and Bastiaens, 2012). Therefore, when designing to immerse players in the game, the human memory's cognitive constraints should be kept in mind (Kiili, de Freitas et al., 2012). The educational game designers should consider the cognitive load theory and the multimedia learning issues to keep a balance as a too rich game environment can overload players mind by increasing incidental process that may disturb learning (Mayer and Moreno, 2003, Kiili, 2005b, Kiili, de Freitas et al., 2012). If both game tasks and the game's use are complex, they may detract player's attention (Pearce and Howard, 2004). Therefore, ideally, a learning game should be transparent with good usability to allow the player to concentrate on higher-order tasks.

- **Guidelines for GBL Design Process**

This section provides guidelines for improving the GBL design process based on lessons learned from the GBL design workshops conducted in this doctoral work covered in P6 and P7. To facilitate understanding and repetition, five themes are created, consolidating the implications from the papers:

– **Multi-dimensional focus in GBL design:** It is essential to incorporate all the key GBL dimensions in the educational game ideation process to create a concrete design idea to build on (de Lope, Medina-Medina et al., 2017, Zahedi, Tessier et al., 2017). The research in this doctoral work found that card-based tools such as ideation cards can be effective in scaffolding the multi-dimensional focus in the GBL design process. Furthermore, it is also possible to identify the design patterns for the GBL design process (exploring the use of these multiple dimensions in different game design projects) that could be useful to guide the process to other novice GBL designers (initial research is presented in P7). A successful pattern can guide the GBL community regarding the best practice to tackle multiple GBL aspects for efficiency and effectiveness in the learning game design practice. Such an approach is expected to reduce the design effort and assist in establishing the right balance between different GBL aspects (Kelle, Klemke et al., 2011).

– **Simplify and guide the complex design process:** It is effective to use structured design activities (each with a different goal) that systematically break down the creative process into individual steps that are easier to understand and operate by the GBL team. Moreover, card-based approaches can be used to support team members to carry out the individual tasks (Mueller, Gibbs et al., 2014, Mora, Gianni et al., 2017). The research in this doctoral work showed that cards simplified the identification of essential GBL concepts in the ideation process and helped multidisciplinary team members recognize that several concepts must be combined in different design activities to make an effective learning game. The team members shuffled through the cards to cover the important concepts in their learning game design. Therefore, the use of cards and design activities together can offer a structured path that guides proceeding with the GBL design process by giving clear directions and order by providing steps (design activities) and building blocks (cards). Different card types are successful for supporting each design activity individually and introducing specific new elements, keeping each activity simple, and introducing newness in different activities.

- **Improve collaboration among team members:** The GBL design process is complex and requires several professionals (from different areas of expertise) to work together towards a common end-product (Tran and Biddle, 2008). Therefore, it is crucial to improve collaboration among GBL team members as it is an important factor in determining and maintaining effectiveness in design (Maldonado, Lee et al., 2007). Researchers have used tools and methods such as board games, game mechanics (Pernin, Mariais et al., 2014), design games and cards (Brandt and Messeter, 2004) to facilitate collaboration among team members in different fields. The research in this doctoral work focused on card-based tools and found them effective for collaboration among GBL team members. *Six themes were identified* that characterize interaction in the GBL ideation process when using a card-based design toolkit: *interacting with the material, focusing on play, an association for doing design activities, ideating design/creative thinking, sharing knowledge, and expressing reactions*. It is important to focus on three points of interaction: physical, mental, and social, to increase team collaboration in the GBL design process (for more description, see P7).

- **Stimulate brainstorming and creative thinking:** Designing involves envisioning, making and rethinking (Vaajakallio and Mattelmäki, 2014). Therefore, it is important to stimulate brainstorming and creative thinking in the GBL design process. According to modern research, it is possible to use methods, tools, and techniques for idea generation that can facilitate creativity. However, different techniques are needed for innovation in general, and no single technique provides the ultimate solution (Kultima and Paavilainen, 2007). The research focusing card-based design toolkit in this doctoral work found trigger cards (cards providing hints and example ideas, see P6) useful to kick start brainstorming and stimulate creative thinking. They provided the existing ideas that helped generate new ones. Some teams would select a trigger card with team discussion to elaborate on the idea that led to creative discussion resulting in combining different trigger cards with a custom card (blank cards to write new ideas) to generate a new idea. Moreover, it was also found that the toolkit's creative elements generate fun in the GBL design process. During the conducted workshops in this doctoral work, the fun element in the GBL ideation process was led by the creativity involved in the design activities. The design activities which required more creativity were considered more fun (for more description, see P6).

- **Inform and encapsulate theoretical GBL design concepts in design practice:** The team members designing learning games are experts from different domains (such as game designers, usability experts, and educators) and often have limited knowledge beyond their expertise (Ahmad, Rahim et al., 2015). It is challenging for multidisciplinary team members to retain and balance different aspects of GBL. Therefore, it is important to adopt ways to make available the multidimensional GBL design concepts within the design process to facilitate team members in sharing knowledge (Wetzel, Rodden et al., 2017, Silva, 2020). It is especially important for novices to understand the overall GBL design space. From the design workshops in this thesis, cards were useful for informing and encapsulating theoretical GBL design concepts (for more description, see P6 and P7), also highlighted by other researchers (Kwok, Harrison et al., 2017). The majority of the team members thought that they considered elements they might have overlooked otherwise, and the information on the cards was useful. The cards' information acts as a quick reminder for designers to the related knowledge/experience, which helps them focus on "all GBL aspects" during ideation process to develop a more concrete design. The design workshop participants praised the potential of the toolkit for informing learning game design concepts.

- **Guidelines for GBL Evaluation**

The implications on GBL evaluation presented are supported by P3, P4, P9, P10. During these studies, knowledge was gathered in understanding the aspects influencing the effectiveness of GBL design and evaluation considering different contexts. The guidelines and implications from the papers are consolidated in five themes below:

- **Real-time data for investigating the GBL learning process**: Electrophysiological measures such as Electroencephalogram (EEG) can be used for an in-depth analysis of the learning process and player's motivation with the GBL approach (Derbali and Frasson, 2010b, Derbali and Frasson, 2010a, Giannakos, Sharma et al., 2019). It is useful for real-time tracking of students learning and provides comprehensive data to measure the player's learning and game experience during the GBL activity. In the present study (P3), the recorded real-time data gave insights into the changing emotional states from start to end with respect to the complexity of learning activity, providing the opportunity to understand the learning and game experience in terms of stress and engagement, shedding light on cognitive load and flow theory. Such investigation can be useful for creating more challenging and enjoyable educational games by triggering dynamic difficulty adjustment concerning the game's response to the player's experience (Stein, Yotam et al., 2018). Previous research mostly focused on using questionnaires to study flow experience (Kiili, 2005b, Kiili, de Freitas et al., 2012). However, the first GBL evaluation study (P3) in this thesis emphasizes the potential of using EEG data to study flow dimensions (especially challenge-skill balance) in educational games. Some researchers (Minovic and Milovanovic, 2013) have focused on real-time learning analytics tools for real-time tracking of students learning experience. It can enable educators to influence the learning process (e.g., by giving hints or interacting with them) to improve students' learning outcomes. Game designers can use such techniques and embed them in the game to manipulate the students' learning path to enhance knowledge adoption. Moreover, capturing multimodal data coming from physiological sensing improves the accurate prediction of learning performance (Giannakos, Sharma et al., 2019).

– **Use of mixed methods:** It is important to combine qualitative and quantitative measures for GBL evaluation (Kato, Cole et al., 2008, de Freitas, 2018) and mixed methods are plausible for merging the findings from the two complementary research paradigms (Pauline-Graf and Mandel, 2019). The efficacy of educational games is hard to measure due to its cross-disciplinarity nature, and methodology is the key to establishing the lines of the discipline. It requires mixed methods studies (combining qualitative methods from education with experiments and other approaches such as neurological studies) to provide the level of detail that can support effective GBL design for improved learner experience. The GBL evaluation studies in this thesis employed mixed methods that offered a variety of benefits, including triangulation of data, more comprehensive interpretation of the phenomena giving completeness to the analysis, increase the validity of findings, and compensate for the weakness of using a single method where one method allows for discoveries in data, and the other provides depth, enhancing the analysis. Similar findings are also highlighted by other researchers (Caracelli and Greene, 1997, Johnson and Onwuegbuzie, 2004, Steinkuehler, King et al., 2011).

– **Use of in-game assessment data for GBL evaluation:** As discussed earlier, the learning game should provide assessment options. Such data is useful for the formative assessment of users' learning outcomes with the GBL approach. Therefore, if the learning game provides assessment reports, the game itself can be used as a GBL evaluation tool. In-game learning analytics serves two purposes. *First*, it provides learners' results (progress report) that determine educational games' effectiveness for learning and highlight issues such as required additional support, improvements in learning tasks, or educational resources. *Second*, such assessment is embedded in the game and therefore conducted non-invasively without disrupting the GBL experience (Bellotti, Kapralos et al., 2013, Snow, Likens et al., 2016). Therefore, it is an effective way to monitor and foster the

GBL learning process (Steiner, Kickmeier-Rus et al., 2015, Shute, Wang et al., 2016).

– **Use of game logs data:** Game logs data analysis is not something new in the game user research. The game logs can be used as consistent measurement and analysis techniques to record players' behavior in the educational games for exploring multiple measures of learning with GBL. Such a technique can help boost the methodological rigor of the educational game design field (Harpstead, Myers et al., 2013). The third evaluation study used game data logs for evaluating children's gameplay performance at home and provided useful data. Therefore, game logs can be useful and effective for the long-term evaluation of user game performance measures. Game logs have a highly descriptive nature can also carry contextual information (such as start and end of sessions and levels, levels completed, win or lose, other information concerning initial conditions of levels that might be relevant) in addition to basic actions. This information provides raw data facilitating grain analysis concerning overall performance or focusing on particular levels (Harpstead, Myers et al., 2013). It is also possible to extract gaming patterns from game log data that can help improve the GBL design and make them more effective and interesting by intelligently improving game difficulty, thus increasing flow (Wang, Wang et al., 2015).

– **Systematic evaluation of learning outcomes in GBL**: Systematic evaluation of GBL learning outcomes provides more useful information for evaluators and educators compared to a single pre/post-test. In the first GBL evaluation study (P3), a single pre-post-test was employed. Although the learning gain was calculated, there were some uncertainties when comparing results for the two groups (such as the effect of selection bias). Similarly, in the first part of the second GBL evaluation study (P4), only one in-lecture quiz was used, which lead to a lack of confidence in results (could have been affected because some students retook the quiz, or one group is more knowledgeable). A pre-post-test design is more effective than only a post-test as it provides a learning gain. However, measuring the learning curve using systematic assessment of GBL learning outcome with multiple iterations is useful to measure the GBL approach's long-term impact and improved results, which counters the testing threat because results would be more sensitive to any variability in the outcomes (reported in P4: second part of the second GBL evaluation study). This systematic evaluation and comparison of learning outcomes increases confidence and validity of data and improves the understanding of learning with GBL highlighting the strengths and weaknesses GBL approach concerning learning material (such as it might be more effective for practical vs. theoretical knowledge). Moreover, the strategy of having a systematic evaluation for the GBL approach within formal or non-formal education can also be a useful resource for educators as input for replanning and revising the teaching plans based on results to have a more significant impact on students' performance. However, depending on the evaluation goal, it is crucial to understand the potential risk of contamination by reintroducing treatment if participants become aware of the researchers' expectations (Bell, 2010). Nevertheless, multiple observation points will help to control for testing, maturation, and regression as they tend to level off or diminish over time (Privitera and Delzell, 2019).

6.6.2 Guidelines for Improving GBL Design and Evaluation for Refugee Children

The third GBL evaluation study focusing on informal learning with refugee children, considering the Syrian crisis and role of educational technology (Yazgan, Utku et al., 2015, Drolia, Sifaki et al., 2020), and the lack of GBL research focusing on this population led to the further research targeting this particular user group. The practical experience of working with this group, research outcomes, and examined literature resulted in important knowledge that can guide other researchers for GBL design and evaluation with refugee children in this research community. However, most of these guidelines are equally relevant for the general GBL research.

The implications and design recommendations for effective learning game design (based on the third evaluation study on using the GBL approach for teaching refugee children in an informal learning setup) are presented in P9. The design recommendations are summarized in the following ten themes: *age-appropriate design, customizability, adaptivity, meaningful feedback, adequate help and tutorial, pedagogic feedback for reinforcement learning, engaging game tasks and learning activities, empathy and connection, effective use of multimedia for multimodal learning and intriguing storyline*. For a detailed description, see P9.
According to Hill (1997), it is important to provide the details of methods employed in research-based publications providing feedback and assessments on the satisfaction of particular techniques (Fargas-Malet, McSherry et al., 2010). The evaluation methods and guidelines for research with children with or without refugee background are reported in P8. The list of evaluation methods is presented in three categories: *preferred, general, and specific* methods (see Table 2 in P8), and the list of guidelines is also presented in three categories: *ethical, practical,* and *methodological* (see Table 3 in P8). Although these are not in the specific context of educational game evaluation, they can serve as a starting point for the research community. More specific guidelines for educational game evaluation with refugee children are covered in P10 (cases study 3), reporting on methodological aspects and lessons learned from field studies (learning game evaluations) conducted with refugee children in the EduApp4syria project (see Chapter 4, Section 4.3.3). The field studies illustrate the application and assessment of different

evaluation methods in various phases of the educational game development life cycle. The methods used included *quasi-experimental design, mixed methods approach, observation with/without a checklist, questionnaires, interviews, pre/post-test (using EGRA), screen recording, game-logs, and expert evaluation*. Table 6.3 presents the educational game development phases and the applicability of these methods (for more details, see P10). Researchers (Hays, 2005, All, Nunez Castellar et al., 2014, Vanderhoven, Willems et al., 2015) have highlighted that the research studies evaluating educational games' effectiveness struggled with various methodological issues. It is essential to recognize and overcome these challenges to improve GBL evaluation research. Therefore, methodological challenges and practical and ethical considerations were identified in the learning game evaluations conducted with refugee children (cases study 3). The guidelines focused on several factors that must be considered when devising an appropriate methodology for GBL evaluation with refugee children. It includes *learner related issues (learning disabilities, previous knowledge, technology experiences, personality, attention span, verbalization, age and gender of children), language barrier, parents' involvement, environment and setup of research, cultural issues, need for a translator, child-friendly interactions, and the effect of information provided on research participants.* For complete details, see P10.

Table 6.3: Educational game development phases and applicability of methods

Evaluation Methods	**Concept (Phase 1)**	**Pre-Production (Phase 2)**	**Production (Phase 3)**	**Post-Production (Launch of game)**
Expert evaluation using checklist criteria	X	X	X	
User Testing		X	X	X
Interview		X	X	X
Observation without checklist		X		
Observation with checklist			X	X
Questionnaire		X		X
Pre/Post Test				X
Screen/Video Recording		X	X	X
Game logs				X
Quasi-experiment				X

The last contribution provides guidelines useful to researchers and practitioners interested in GBL design and evaluation in different fields and contexts, and especially those interested in GBL research with the special user group of refugee children or developing learning games targeting this population. The novice researchers in this field can learn from the experience (design and evaluation guidelines and methods) and find the most appropriate way to apply them to diminish the drawbacks as far as possible and maximize the benefits. Moreover, the guidelines can encourage researchers to critically reflect on the methodological and practical issues since they will have implications for the data produced.

7 Evaluation

This chapter presents the evaluation of this doctoral work's contributions (presented in Chapter 6) with reference to the research questions. In addition, limitations of the research work and validity threats are discussed.

7.1 Evaluation of Research Questions

This section provides the evaluation of the main research goal and each of the four research questions in terms of answers provided by the contributions.

7.1.1 Research Goal: How can the design and evaluation of game-based learning (GBL) approaches be supported to improve the effectiveness of learning games?

The main research goal is answered by Contributions 1-6. Using review studies, the problems and issues in existing design and evaluation practices for GBL were identified, which led to the development of a suite of artifacts for facilitating GBL design and evaluation: LEAGUÊ framework, card-based ideation toolkit, analysis instrument, integrated evaluation approach, and ten-step transformation process from framework to design cards. Each artifact assists a different activity focusing on the design, analysis, and evaluation of learning games. The artifacts have demonstrated their usefulness in supporting the GBL design and evaluation process through design workshops, quasi-experimental (GBL evaluation) studies, and case studies, which resulted in implications and guidelines for evaluating and designing effective learning games.

7.1.2 Research Question 1: What are the challenges and problems in the current game-based learning design and evaluation practices?

This first research question is answered by contribution C1. C1 identified a set of challenges and issues in the existing GBL literature regarding the design and evaluation practices, focusing on the state of the art, existing design and evaluation approaches and models, attributes for design and evaluation, and employed evaluation methods and guidelines. Emphasis is also placed on examining which aspects or attributes are considered most important for the design and evaluation of learning games and to what extent they are validated and employed in practice. The main issues identified in GBL literature were as follows: a wide diversity of aspects used for educational game design and evaluation, inconsistency (in the definition, use, and terminology) of GBL aspects, no systematic breakdown of high-level GBL aspects, and absence of a holistic view of GBL. The main challenges identified for GBL design practice included: lack of tool support and guidance for applying framework concepts in the design practice, lack of empirical evidence for validation and practical application of frameworks in the design process, and scarcity of information on assessment approach, method or stakeholders. The main challenges identified for GBL evaluation practice included: lack of usage of existing GBL evaluation models for conducting educational game evaluations, lack of approaches addressing children's needs, use of ad hoc criteria for evaluating learning games. One of the identified challenges in GBL evaluation is the lack of research focusing on children and the need for conducting GBL evaluation studies with refugee children as a user group. This need caused a further investigation on the effectiveness of evaluation methods and guidelines for researching with children with or without refugee background.

7.1.3 Research Question 2: What are the key elements for the game-based learning phenomenon, and how are they related?

The answer to the second research question is provided by the contributions C2 and C3. The six-dimensional LEAGUÊ framework for GBL is described in detail in C2. The framework provides a comprehensive hierarchical structure outlining the six core GBL aspects in four conceptual levels (dimensions, factors, sub-factors, and metrics). It facilitates greater insight into the process of learning with educational games, where to focus, and what to evaluate. The identified key elements of GBL were found to be related to six dimensions: The learning aspects of using GBL (Learning), the environment where GBL takes place (Environment), the learner's emotional and cognitive responses from using GBL (Affective-cognitive reactions), factors related to game design and gameplay of GBL (Game factors), the ease-of-use to perform defined tasks with GBL product (Usability), and the type of characteristics of the learners using GBL (UsÊr). Each dimension has a set of factors (22 in total) and sub-factors (74 in total) that are mostly devised based on theories and other frameworks grounded in GBL literature. The dimensions are linked to each other in terms of cause and effect, and these relations are crucial for more significant insights into the GBL process. The framework identified ten key relations between core GBL dimensions to provide a detailed picture of the GBL phenomenon focusing on a multidisciplinary approach. Furthermore, five metric types are outlined to facilitate the evaluation of GBL elements. C3 explores and employs the framework components in GBL evaluation studies (quasi-experiments) to improve the understanding of the GBL phenomenon and interrelated elements. The evaluation studies investigated GBL approaches' effectiveness in three different learning setups (formal, non-formal, and informal) to validate and enhance the proposed GBL elements. The studies' findings provided significant insights into the learning process with GBL and contributing factors focusing on the impact on emotional states, learning gain, engagement, interactivity, usability experience, the user characteristic, and gameplay performance.

7.1.4 Research Question 3: Which kind of approaches, tools, and guidelines can be employed to facilitate the GBL *design* process for effective learning games?

The third research question is answered by the contributions C4 and C6. A ten-step process for transforming a framework to ideation cards is proposed, and a card-based design tool (LEAGUÊ ideation toolkit) to support the early design process of learning games is described in C4. The toolkit is evaluated during design workshops with students and researchers. C4 also investigates how this card-based toolkit approach can scaffold collaboration and completeness in the early phase of the GBL design process. Based on the experience gained from the design workshops and GBL evaluation studies, a set of guidelines are presented in C6 to improve the effectiveness of learning game design and the GBL design process. The guidelines for effective GBL design are presented in seven themes: captivating game features, good usability, challenge-skill balance, in-game assessment options, social interaction, context, and cognitive load. The guidelines for the GBL design process are summarized in five themes: multi-dimensional focus in GBL design, simplify and guide the complex design process, improve collaboration among team members, stimulate brainstorming and creative thinking, and inform and encapsulate theoretical GBL design concepts in the design practice. Moreover, based on the GBL evaluation research with refugee children, additional design recommendations are presented for designing effective learning games for this user group. However, these specific guidelines are equally relevant for the GBL design in general.

7.1.5 Research Question 4: Which kind of approaches, tools, and guidelines can be employed to facilitate the GBL *evaluation* process for effective learning games?

This fourth and final research question is answered by the contributions C5 and C6. C5 introduces a GBL analysis tool (LEAGUÊ analysis instrument) for assessing learning games by knowing the strengths and areas of improvement to understand their potential in a specified environment. It is found that learning games are evaluated at different development stages with different purposes and identifying evaluation criteria is a time-consuming and complicated process. An integrated GBL evaluation approach (LEAGUÊ-GQM approach and guide) is proposed to guide planning and creating a learning game evaluation plan by outlining the critical GBL evaluation criteria in three levels and facilitate both analytical (single aspect) and global (holistic) GBL evaluation process. Based on evaluation studies conducted in this doctoral work and the literature, guidelines are presented in C6 to improve the GBL evaluation process in general and additionally for the specific group of refugee children as target users for producing effective learning games. The general guidelines for GBL evaluation are summarized in five themes: real-time data for investigating the GBL learning process, use of mixed methods, use of in-game assessment data for GBL evaluation, use of game logs data, and systematic evaluation of learning outcomes in GBL. The specific guidelines for GBL evaluation with refugee children focused on the methodological, practical, and ethical challenges that must be considered when conducting learning game evaluation with this user group. Moreover, C6 also described the applicability of various evaluation methods and their effectiveness in different phases of the educational game development life cycle.

7.2 Evaluation of the Research Approach

This section describes the limitations of the research approach employed in this doctoral work. The validity and reliability issues are also discussed.

7.2.1 Limitations

This thesis has some limitations. *First*, the review studies have some limitations concerning the choice of databases and search strings used for selecting articles. Some important work might have been missed.

Second, the GBL evaluation studies (quasi-experiments) focused on three different learning settings: formal, non-formal, and informal. Therefore, the conditions (such as target population, employed learning game, and learning domain) were not comparable. The diverse conditions in which these studies took place had to be taken into account during analysis and when generalizing the findings cross-case to provide a robust interpretation of the phenomenon. This heterogeneity added complexity and limited the possibilities of comparing results across studies for generalization. However, it was possible to generalize some findings related to the GBL experience evident in all settings. Such as *challenge-skill balance (flow theory), game components increasing engagement and improving the experience, increase in learning outcomes, need for good usability, the effect of context, and need for managing the cognitive load*. Although the results reported in these studies should apply to GBL usage for teaching and learning various subjects, the results might be limited to the particular learning games employed in the study or learning games exhibiting similar features and not being exactly transferable to other platforms. It is also possible that similar analyses (especially for migrant refugee children) with other user groups might produce different results in terms of age differences. Moreover, the first experiment used a learning game prototype (sort attack). No other and better game was available for that specific learning domain (insertion sort). Developing a new game from scratch would take too much time and be out of the research scope of this thesis. Although this specific study's focus (quasi-experiment 1, in P3) was not usability, the lack of captivating design could have affected participants' emotional states during the GBL experience. Focusing on a more robust and engaging gaming platform may have provided stronger and significant results. Nonetheless, overall results gave a picture of the learning process in GBL. Furthermore, the data collection in the GBL evaluation studies was comprehensive and extensive, requiring time for collection and analysis. Consequently, some data might have received less attention or not included because of limited time and resources. For example, for the first and third quasi-experiments, data was also collected through screen recording. However, the video analysis was not included in the papers (P3 and P9). Moreover, the data for quasi-experiment three was extensive, covering many demographic variables and other characteristics from log files. However, only a fraction was analyzed in P9.

Third, the design workshops evaluated the toolkit only with researchers and students and did not include design practitioners and learning game experts. Therefore, results do not represent the overall design community. Another limitation is that the employed toolkit is not representative of all GBL ideation cards. Therefore, the results are only generalizable to ideation cards presenting similiter features to the proposed toolkit or providing enough knowledge of GBL concepts. There were some issues with time management and workshop organization in the second workshop due to the summer school management's unexpected matters on the workshop's scheduled date. These issues affected the understandability of a few activities (especially the first activity) and working with the cards (presented in P6). However, results are explained, keeping an account of these issues and how they can be mitigated by simple modification in the workshop technique. Due to time-restricted activities in design workshops, participants could not use all the cards in a limited time. Thus, a clear use pattern could not be identified in the second activity (described in P7). Lastly, since the author/researcher predominantly led the design workshops, no conclusions can be made about the supervision level needed when other researchers use the toolkit. However, the second facilitator was different in the third design workshop. Moreover, the facilitators were involved with different design activities at different levels in the three workshops, indicating that the knowledge is in the toolkit and not the person introducing them. It suggests that the toolkit can also be used in settings with other researchers. The quality of the generated educational game design ideas was not evaluated or ranked. Therefore, conclusions cannot be drawn in terms of their novelty. Although the quality of generated ideas is important, this research's main focus was on understanding the toolkit's scaffolding. The ultimate goal was not to produce a functional learning game since that would require evaluating fully developed and deployed ideas, a task beyond this thesis' scope. Therefore, the efficacy of ideas developed by the teams in design workshops was not tested. The focus was to evaluate if the necessary GBL concepts and dimensions were included in the educational game design ideas. Moreover, due to the short duration of the design workshops, it was only possible to assess short ideation iterations. Long-term dynamics and effects connected to the iterative design process might have led to additional insights and understandings.

Fourth, the case studies might have limitations with regard to including expert feedback as they were conducted to demonstrate the use of built artifacts or knowledge in the relevant application domain and not formal evaluations. Although expert evaluations might have provided a different perspective, the fact that case studies involved assessment by external designers and developers provides strong evidence for validity.

In addition, some limitations due to data collection methods used might apply. One such limitation is related to the use of EEG headsets. EEG requires proper adjustment to collect data. The researcher's limited experience with EEG made it difficult to initially gather good quality data. Some of the data had to be discarded due to weak signals. However, mixed methods were used to neutralize the bias and weaknesses linked with a single method. Lastly, the research work presented in this thesis is still in its early stages. The developed research artifacts (framework, toolkit, instrument, and approach) need to be more exhaustively validated with more complex educational games, in different contexts, and with greater sample sizes. Although all the framework components are grounded in GBL literature and existing theories, not all could be investigated in the research work conducted in this thesis due to the comprehensive nature of the framework and time limitations. Especially, psychosocial needs and psychosocial well-being need to be studied thoroughly in educational games.

7.2.2 Internal Validity

Internal validity concerns the establishment of cause-and-effect relationships in the study and the extent to which the experimental design can control the extraneous variables to eliminate and control alternative explanations for the findings (Gall, Gall et al., 1999). Extraneous variables influence and weaken the internal validity by inferring the causal relationships of the variables being examined. Researchers must be aware of the aspects that might threaten the study's validity and take necessary actions to control them.

During the *design workshops* (especially the one arranged by summer school), it was difficult to control some variables. Sometimes, it was required to slightly deviate from the agreed procedure (usual activities' duration) due to unpredictable events and time constraints. The workshop participants had no or little GBL design experience. It was useful to explore the card-based toolkit's support and ensuring the validity of data not influenced by their experience and knowledge. It also allowed us to examine the cards' usefulness for early-career GBL designers but not senior designers. There was no control group in design workshops to compare the GBL design ideas without using a toolkit and assess the intervention's effect. A control group could be used employing some other approach (such as a checklist or framework). However, the aim was to be close to a real-life setting and demonstrate the effectiveness in designers' practice. For this research study, design workshops were conducted instead of using the toolkit in the designer's day-to-day practice in a game studio with professionals as it was practically difficult to achieve. However, the previous work (Flanagan, Belman et al., 2007, Hornecker, 2010, Lucero and Arrasvuori, 2010, Mueller, Gibbs et al., 2014) suggests that design workshops are a way to approximate design practice. It offers a similar environment with team-based design exercises and a time-constrained format, similar to the environment to which designers are exposed. In design workshops, the instruments and testing used were performed in a consistent manner in all workshops conducted without any changes. There were almost no dropouts in the design workshop sessions. Only one person who volunteered to join in the first workshop but arrived late did not participate as the teams were already formed and had started the design activities. However, this did not affect the results. Care was taken to inform the participants about the data collection process which was also required because the consent forms were collected from them.

During the *quasi-experimental (GBL evaluation) studies*, there were few relevant internal validity threats. However, steps were taken in the studies to preserve their internal validity depending on the relevant threats. All participants in each GBL evaluation study followed the study's basic requirements and procedure and received the same treatment. The observers had the same attitude and used the same materials. However, it was challenging to control few variables in some studies due to unpredictable events (such as external involvements, participants' age, and technology issues). Thus, there were slight variations. For instance, it was challenging to gather access to refugee children for the third quasi-experimental study. The experiment was conducted in an organization Muslim Society Trondheim (MST), and it was sometimes difficult to fully control the setup without interruptions (for detailed description, see P10). However, the lack of control was traded for realism. Participants age sometimes required to finish the play session early if they did not want to continue. However, this was important and provided useful insights to understand the engagement with the GBL approach and not something that should be controlled. Moreover, it was not possible to control the children's game usage at home. However, it did not affect the study results as game data logs were used to monitor the game usage and performance. Moreover, in the second quasi-experiment, some technical difficulties were encountered in using the Google Quiz platform with many students, due to which some changes were made in the experiment design. However, it did not affect the results as the focus of the study was on the GBL platform, and we were able to gather enough data for the learning curve to ensure validity.

Although the random selection was not used to choose study samples, steps were taken to ensure that participants are selected in a manner that they are representative of the population. Moreover, the selection was based on defined criteria, and the gender distribution was fairly equal, thus reducing the effect of selection bias (Sharma, 2017). The participants in all experimental studies had no or little knowledge of the learning content being taught by the employed GBL approach. Further, pre-and-post tests were used to ensure that learning outcomes were derived from the GBL experience during the study and not coming from previous knowledge. There was no control group in the second (only in the 2nd part of the study) and third quasi-experiment to compare the learning gain. However, in the first and second experiments (1st part of the study) that used a control group, both groups were comparable concerning age, gender distribution, education level, and study area. Also, there were no differences in how the GBL and non-GBL approaches were used. There was no control group in the second experiment due to technical issues with the Google Quiz platform. However, to mitigate this, the results were repeated in multiple iterations and compared with lectures without using the GBL approach. The reason for no control group in the third experiment was to realize the real settings of the context of use (informal learning). There are no other means like traditional education available for refugee children. However, to mitigate this, a matched-pairs study was performed with one group pre-test–post-test design to investigate the difference in value (pre and post-intervention) from same subjects using Wilcoxon matched-pairs test (MacFarland and Yates, 2016).

Regarding pre-and-post tests used to measure the learning gain in studies conducted, there might be a chance that the pre-test could have affected the outcome of the post-test. Learners might have had better performance in the post-tests merely due to practice. However, care was taken to decrease these chances in each study. In the first quasi-experiment, the problem's complexity and the array size for insertion sort were kept fixed. However, the order of appearance of the questions and the array values for sorting were changed for pre-and-post tests. In the second experiment, for measuring the learning curve, the scenario was repeated in five iterations with a large sample size. It counters the testing threat because results would be more sensitive to any variability in the outcomes. Also, the results were compared for online lectures with and without the use of the GBL approach. In the third experiment, the post-test was conducted after a one-week gap and was not exactly the same, countering the testing threat. Although the same pattern was followed in the post-test and the questions involved the same ten alphabets, there were changes in order and arrangement of the task and some content.

Similar to design workshops, there were no major dropouts in the quasi-experimental studies, only very few cases in the first experiment in which the EEG signals were not successfully collected or were not good quality. These cases were removed from the data analysis process and therefore did not affect the results. Again, to prevent the possible dropout, care was taken to inform the participants about the data collection process. It was also important as the consent forms needed to be collected from participants and, in the third quasi-experimental study, from the parents or guardians. Furthermore, researchers particularly check for non-filled questionnaires and pre-and-post tests after data collection.

7.2.3 Construct Validity

Construct validity concerns if the methods employed measure what they intend to be measuring (Robson and McCartan, 2016). It questions whether correct operational measures have been adopted for the concepts being researched and if the sampling particulars can be supported as measures of general constructs (Shadish, Cook et al., 2002). More specifically, it implies that first, researchers must understand how measures and constructs behave and relate with each other based on relevant theory. After that, the challenge is to give evidence through examination that the measures or constructs behave in the same way in practice.

For construct validity, it is essential to ensure that the indicators are developed carefully based on relevant existing knowledge. For that reason, this doctoral work demonstrates good construct validity as the proposed GBL framework (presented in P2) is built on components grounded in theory extracted from existing GBL literature. Each framework component has a strong basis for construction and is supported by theoretical constructs in GBL literature, not merely based on suspicion. The toolkit and instruments used in evaluation studies were developed based on framework components, thus increasing the construct validity with each study (as proof of validity for C2). Therefore, the selected data sources are strong indicators (there are theoretical grounds and statistical evidence) of GBL.

The statements in the developed tools and instruments, as well as questionnaires and the interviews used in evaluation studies, were inspected and agreed upon by at least two researchers. It ensured that words were clear, precise, and objective and conveyed the same meaning as intended. Collecting data from multiple sources increases construct validity (Flick, 1992). The research work in this thesis adopted a mixed methods design. Data was collected using qualitative and quantitative sources (such as questionnaires, observations, interviews, game artifacts, video recordings, EEG, game logs), including objective and subjective measures. For a complete description, see Chapter 4, Table 4.1 and 4.4. Therefore, the collected data were cross-checked for triangulation, gaining more significant insights, and eliminating the researcher bias, thus increasing the construct validity (Anderson, Bachman et al., 1991, Flick, 1992, Nightingale, 2009). During the research design and data collection period, efforts were made to avoid subjective judgments. Results were discussed among at least two researchers before formalizing theories and constructs to enhance the construct validity.

During the quasi-experiments and design workshops, the researchers were in close contact with the participants, which helped establish the construct validity. It enabled to convey the chosen meaning of constructs, control the operationalization of a construct, and emphasize the theoretical view of constructs relations. The total number of participants in design workshops was 34. Participants in quasi-experiments were as follows: 22 in the first experiment, 261 in the second experiment, and 30 in the third. All design workshops and quasi-experiment employed mixed methods using qualitative and quantitative data. Although the sample size of the data in conducted studies allowed for some statistical analysis, it was emphasized that confirmation in the qualitative data was always obtained before formalizing the results. However, for the design workshops, 34 participants formed seven teams. The size of data samples concerning team's data is therefore limited for usage pattern recognition across teams and for statistical analysis, which weakens the conclusion validity. Each of the evaluation studies was of short duration. Therefore, it is not possible to prove the long-term effects of the GBL approach and the developed toolkit. Moreover, constructs like psychosocial well-being could not be evaluated in GBL evaluation studies.

7.2.4 External Validity

External validity concerns the generalization of the study findings to other settings. It refers to the degree to which causal relationship holds for variations in settings, persons, treatments, and outcomes both that were and were not in the experiment (Shadish, Cook et al., 2002). In simple words, external validity focuses if the study conclusion can be generalized to other people in other places at other times.

Several actions were followed to improve the external validity of the research work in this thesis. The evaluation studies were conducted with variations in places (setup), times, and users. Thus, the external validity became stronger as the logic of the studies was replicated across different contexts. The three quasi-experimental (GBL evaluation) studies focused on exploring the GBL approach's effectiveness in different contexts: formal, distance, and informal learning. Similarly, the design workshops were also carried out in three contexts: research study, summer school, and game development course. The participants were different in each study. For design workshops, participants were mostly students and researchers with no or little GBL design experience, which means everyone can use the toolkit. It is important to note that participants were mostly from computer science (CS) backgrounds (some were from game development). A few were from other disciplines, such as electrical engineering and mathematics. The design workshops were not repeated with the same participants. However, the difference in participants' knowledge related to educational game design added to the card-based toolkit's external validity. On the other hand, the quasi-experimental studies did not have the same population, so the results cannot be compared across studies. However, the different contexts (formal, online and informal learning) demanded using different user groups, which justified not having the same sample. For example, refugee children were selected for the context of informal education as they were most relevant for this type of learning where no other means of traditional education are available. Similarly, for distance learning, students taking online courses following the Covid-19 pandemic were selected.

Moreover, the GBL applications employed in the three studies were also different related to the learning domain and context. The first experiment employed a sorting algorithm game, the second employed a game-based digital quiz focusing on Human-computer interaction (HCI), and the third used an Arabic language learning game. However, all experimental studies' objective was to evaluate GBL's effectiveness in their context of application focusing on the contributing factors related to GBL concepts in the proposed framework (presented in P2). Moreover, the fact that studies took place in Norway and Kosovo, and Italy increases the external validity as demographic variables might affect participants' behavior. Furthermore, there were variations in users and their demographics, further improving the external validity of GBL concepts with different user groups. The first two experiments involved university-level students and the third experiment involved children with a refugee background.

For each study, the research's scope and boundaries were defined, and the findings were explained in detail, as well as the target users and the study dimensions. Field notes were taken in each study to gather as much detail as possible. When conducting the questionnaires and interviews, raw data was collected. However, the credibility of the data depended on participants. The age gap between design workshop participants and the researchers was not that evident, possibly strengthening trust.

Lastly, comparing the evidence with the existing literature in the GBL domain and the proposed framework components helped to clearly outline and generalize the contributions within GBL design and evaluation research.

7.2.5 Reliability

Reliability refers to the trustworthiness of the research study procedures, i.e., the repeatability and consistency of the process. It is the extent to which the results would be consistent if other researchers repeat the research process under comparable conditions (Johnson and Christensen, 2019). Therefore, it is essential to have transparency about the research process to achieve reliability.

To overcome the researcher bias in this research work, two or more researchers were involved in the data collection and analysis process. It is also important to note that different researchers led the work for different quasi-experimental studies. The procedure and findings of each study are described in detail. Researchers communicated about the methodological decisions, findings, and analysis on a continuous basis. Also, they kept notes about the decisions made to explain and justify the choices throughout the process. Moreover, the first two case studies for game analysis were carried out by external game designers/ developers. It suggests that the reliability is satisfactory.

8 Conclusion and Future Work

This doctoral work has focused on understanding and leveraging the game-based learning (GBL) design and evaluation practices for improving the effectiveness of learning games.

The main research methodology adopted was design science research (DSR), combined with mixed methods approach using exploratory-triangulation design within DSR cycles, employing qualitative and quantitative research methods. Three review studies were performed to identify the issues and challenges in current GBL design and evaluation practices. The knowledge from the review studies and existing theories inspired building and evaluating the research outcomes (artifacts) developed in this doctoral work. The artifacts built include a GBL framework and set of tools (card-based toolkit, analysis instrument, and an evaluation guide) to support the GBL design and evaluation process at various developmental stages of learning games. The developed artifacts were evaluated and refined in evaluation studies, including three quasi-experimental (GBL evaluations) studies, three design workshops, and three case studies. The work is grounded in the literature on human-computer interaction (HCI), learning, game design and psychology. The doctoral work resulted in ten research papers published (or ready for submission) in peer-reviewed journals and conference proceedings that explored the research questions and provided six research contributions to the field. These contributions (presented in Chapter 6) answered the main research goal, "*How can the design and evaluation of game-based learning (GBL) approaches be supported to improve the effectiveness of learning games?*" by addressing the four research questions (evaluated in Chapter 7).

Below, the doctoral research findings are summarized in conclusion, and future work is delineated.

8.1 Conclusion

The research in this doctoral work began with systematically performing review studies that provided state of the art, identified issues in GBL evaluation and design practices and presented methods and guidelines for conducting GBL evaluation research with refugee children (*covering the first contribution*). An increasing number of papers show that GBL design and evaluation research is gaining momentum, and existing approaches focus on different aspects of GBL covering various learning domains. More specifically, the findings of the three review studies focused on the trends in GBL evaluation, research topics, current practices in GBL evaluation, dimensions for evaluating GBL, attributes for designing GBL in existing models, validation of these attributes, comparison of existing design models highlighting strength and weaknesses and the existing guidelines and evaluation methods for research with children (with and without refugee background). The findings showed that models and frameworks are available for GBL design and evaluation, where most focus on one or two main aspects (dimensions). The most extensively used dimensions in literature for GBL evaluation are learning, usability, and game factors. The existing design models focus most on the learning/pedagogy and game factors followed by reactions such as flow, enjoyment, and immersion. However, the user interface, learner requirements, technical and context-related aspects were less emphasized by the analyzed design models for educational games. Most of the existing models and frameworks are grounded in learning and game design theories focusing on ARCS models and flow theory, among others. Moreover, the last review study focused on research with refugee children as this was the target user group for the third GBL evaluation study (see Chapter 4, Section 4.3.2 and 4.3.3). The findings identified three categories of research methods: preferred, general, and specific methods, and three categories of guidelines: ethical, practical, and methodological for research with children (with and without refugee background). The review identified a need for additional ethical, practical and methodological parameters for research with refugee children to ensure that the results of introducing educational technology include a good understanding of its users.

The issues and challenges identified from the review studies directed the further research (artifact development and evaluation) conducted in this thesis. The major challenges and problems identified in the current GBL design and evaluation practices through review studies include the following: 1) Inconsistency in a wide diversity of aspects for GBL design and evaluation, hindering a clear pattern of what is important. 2) Absence of a holistic view of GBL. 3) Use of ad hoc criteria for GBL evaluation. 4) Lack of focus on children needs. 5) Scarcity of empirical evidence for validation of GBL frameworks and models. 6) Absence of tool support for GBL design practice. 7) Lack of guidance for practical application of framework concepts in the design process. 8) Scarcity of educational games evaluation research with refugee children within GBL literature. 9) Lack of research on the effectiveness of evaluation methods (general and child-friendly) for research with refugee children. 10) The need for additional research guidelines or adaption to cater to the refugee context's specific needs. 11) Lack of focus on methodological guidelines for the specific group of refugee children. The identified issues collectively address *the first research question (RQ1)*.

The lack of holistic GBL view and inconsistency concerning GBL aspects in literature highlighted the need for developing a comprehensive framework for GBL design and evaluation (*covering the second contribution*). A conceptual hierarchical framework LEAGUÊ is presented through directed content analysis of GBL literature. The framework shows that GBL's multi-dimensional nature requires focusing on six aspects referred to as core dimensions. The dimensions presented in the framework are higher-level concepts and not restricted by the learning domain, game genre, or technology. It includes learning, environment, affective-cognitive reactions, game factors, usability, and user. Each dimension comprises certain factors and further sub-factors; the framework presents 22 factors and 74 sub-factors. Moreover, metrics are outlined in five categories to assess them. The GBL dimensions are interrelated, and it is crucial to assess these relations for more significant insights into educational games. The framework presents a detailed holistic picture of GBL, providing a theoretical model essential for unifying all the different aspects of GBL to guide educational game design and evaluation. The proposed framework is built on components grounded in theory, and each component has a strong basis for formation that is supported by theoretical constructs in GBL literature.

To address the scarcity of empirical evidence concerning GBL effectiveness and framework validations, the doctoral research work focused on exploring the framework components in three GBL evaluation studies (quasi-experiments) to validate and enhance it further. The GBL evaluation studies focused on three different contexts: university, online lecture amid COVID-19, and informal learning at home for refugee children. The GBL evaluation studies focused on investigating the GBL approaches' effectiveness in different learning settings to improve the understanding of the GBL phenomenon, learning process, and interrelated factors (*covering the third contribution).* The findings in the quasi-experimental studies provided empirical evidence on GBL's effectiveness for learning in formal, non-formal, and informal learning setups. Furthermore, they provided significant insights into the learning process with GBL and the possible interrelation and role and impact of emotional states, learning gain, interactivity, usability, user characteristic, and gameplay performance when using a GBL approach compared with others. The students performed better on practical concepts and had a higher engagement, interest, focus, and relaxation when using the GBL approach compared to traditional learning in the university setting. GBL participants solved double the number of tasks at the same time even though extra attention and explanation were provided to traditional learning participants. The students' learning gain, engagement and motivation, and interactivity (between students and teachers) increased with systematic use of the game-based learning platform in online lectures. Online students preferred the GBL platform for the game experience, usability, technology, and context of use (quarantine online learning) compared to the non-game-based platform. The students felt motivated to be part of systematic in-lecture quizzes with game-based platforms and had higher learning gains. They claimed that they had more fun due to the game components (such as competition, bonus points, and music). Both student and instructor noticed and liked the increased interactivity with the use of the GBL platform in the online lectures. Lastly, refugee children had a significant increase in reading assessment scores when using the language learning game in an informal learning setup at home. The user characteristics (age group, learning modality preferences, and mobile usage experience) significantly impacted refugee children's GBL experience concerning their learning gain, gameplay performance, and usability experience. Refugee children's age affects their learning gain, usability, and gameplay performance with GBL. However, children's learning modality preferences and mobile usage experience do not affect their learning gain but impact their usability and gameplay performance parameters with GBL. The proposed LEAGUÊ framework and the empirical evidence (concerning GBL effectiveness, learning process, and interrelated factors) of three GBL evaluation studies collectively address *the second research question (RQ2)*.

Moreover, to support GBL design and evaluation practice, the proposed LEAGUÊ framework is put into practice by developing three tools based on the framework components: card-based ideation and design toolkit, an analysis instrument, and an integrated evaluation approach and guide.

The lack of tool support and guidance for framework application in GBL design practice highlighted the need for an operationalizable approach to assist the educational game design process. It led to the development of the LEAGUÊ ideation toolkit and its evaluation concerning participants' experience, toolkit objectives, refinements, and scaffolding for multi-dimensional focus and collaboration in the GBL design process (*covering the fourth contribution*). To facilitate the effective transfer of theoretical knowledge (framework concepts) to GBL design practice, the proposed LEAGUÊ framework was transformed into a card-based GBL ideation toolkit to support the ideation phase of educational game design. A ten-step process is presented for transforming any framework to design cards to guide other researchers. The developed toolkit comprises a set of four card deck types (Primary, Trigger, Custom, and Reflection cards) that entail GBL design concepts and a workshop technique with five structured design activities that provide step-by-step guidance for the ideation process. The findings showed the potential of the toolkit in informing and guiding educational game design in practice. Moreover, the research findings also highlighted the usefulness of a card-based tool for scaffolding completeness and collaboration. The toolkit features that most contributed to scaffold completeness were different card categories and different tasks (design activities) in addition to the general characteristics of cards as tangible idea containers. The toolkit features that most contributed to collaboration were tangible props, common goals, and social setting. Furthermore, based on the experience and research findings from design workshops and GBL evaluation studies conducted in this doctoral work, design guidelines are identified to facilitate the GBL design practice (*covering the design part of the sixth contribution*). Seven themes were presented concerning implications for designing effective learning games and five themes for the GBL design process. Moreover, in addition, design recommendations were presented for designing effective learning games focusing on GBL evaluation with refugee children. However, these additional guidelines might be equally relevant for GBL design in general. The proposed ideation toolkit, investigation of card-based tool for scaffolding the design process, and the presented design guidelines collectively address *the third research question (RQ3).*

The problem concerning the use of ad hoc criteria for GBL evaluation and GBL's diverse characteristics that make identifying evaluation criteria complex and time-consuming process highlighted the need for an overreaching approach to guide the GBL evaluation process. It led to the development of the LEAGUÊ analysis instrument and integrated LEAGUÊ-GQM evaluation approach and their application in case studies to demonstrate their utility (*covering the fifth contribution*). There is a distinction between different evaluation processes (assessment vs. formal evaluation). LEAGUÊ analysis instrument was developed to facilitate the systematic assessment of learning games to verify their potential and suitability in their designated purpose and application context. The analysis instrument highlights the strengths, weaknesses, and particular challenges in the learning game with regard to essential GBL components required to embed the desired learning through the game into effective practice. The instrument can be used by the educational game stakeholders (designers, developers, researchers, and intermediates like educators or parents) to identify the loopholes and refine the design (both early-stage and beta versions of the game) to improve learning games. The instrument's application was exemplified with two case studies on an online empathy game and a VR game, externally conducted by the game designers and developers. Moreover, an integrated LEAGUÊ-GQM approach was proposed to guide developing a GBL evaluation plan for determining the effectiveness of learning games by ensuring that the quality and performance were met. It presents a three-step parallel process with an evaluation guide providing GBL-specific guidance (listing essential GBL evaluation components) to facilitate defining evaluation goals, formulating questions and data sources, and identifying measures and analysis methods. The approach is equally useful for analytic (single aspect) or holistic evaluation of learning games as it allows picking and selecting elements from the guide based on evaluation rationale. The proposed approach is applied to plan a GBL evaluation study with migrant refugee children. The user group of refugee children was selected for informal learning with GBL due to the Syrian crisis and affiliation with the EduApp4Syria project. The lack of focus on children needs and the scarcity of educational games evaluation research with refugee children within GBL literature instigated further focus on this user group. Furthermore, the lack of research on the effectiveness of evaluation methods and methodological guidelines for research with refugee children led to a side research line dedicated to exploring GBL evaluation research's methodological aspects with refugee children. The third case study was performed on field studies (learning game evaluations) conducted within the EduApp4Syria project that provided specific implications and lessons concerning evaluation methods and guidelines for conducting educational game evaluation with this user group. Finally, the experience gained and research outcome from the GBL evaluation studies and case studies conducted in this doctoral work deduced into evaluation guidelines to facilitate the GBL evaluation practice (*covering the evaluation part of the sixth contribution*). Five themes were presented concerning guidelines for GBL evaluation in general to improve the effectiveness of learning games. Moreover, implications were presented for the effectiveness of evaluation methods in different phases of the educational game development life cycle and methodological, practical, and ethical guidelines for specific research with refugee children. The proposed analysis instrument, integrated evaluation approach, and the evaluation guidelines collectively address *the fourth research question (RQ4)*.

8.2 Future Work

The following research directions are proposed for future work that could be investigated as an extension to the doctoral research presented in this thesis.

- **Further evaluation studies for validation and generalization:** The research artifacts developed in this thesis are still in the early stages, and additional work is required for thorough validation. Future research will focus on designing experiments with control groups, randomization, and large sample sizes to address the validity issues. Moreover, the future work will also focus on generalizing results across different learning contexts, user groups, and more complex educational games.
  Additional GBL evaluation studies will be planned to investigate all components and, specifically, relations between dimensions in the proposed GBL framework. The analysis instrument and integrated evaluation approach and guide will be formally evaluated by comparing the results with other existing instruments or approaches. Further studies will involve design workshops led by external organizers for independent evaluation and with teams with different experience levels to validate the toolkit's usefulness for both early-career and senior GBL designers. The quality of ideas generated with the toolkit will also be evaluated and ranked in future studies to determine their novelty. An interesting possibility is to use a control group employing some other approach (such as a checklist, framework) to compare results (generated learning game ideas) with and without the proposed GBL toolkit.

- **Involving experts and stakeholders (such as educators):** Future work will involve GBL experts and stakeholders such as teachers in the design and evaluation process. One possibility would be to investigate the usefulness of the proposed analysis instrument with teachers, parents, and researchers for different purposes. Teachers will assess the potential of educational games for justification of its use as an efficient tool in classrooms. Parents will use the instrument to assess the learning game to be convinced of the game's positive effects and ascertain its effectiveness for selecting it for their children. GBL researchers will use the instrument to analyze different

successful and failed learning games to learn more about the different elements used in the games and their relationships. It will help validate and improve the instrument concepts, explore the usefulness of the analysis instrument for different stakeholders, and propose more detailed guidelines for designing effective learning games.
Furthermore, the future work will focus on conducting design workshops with multidisciplinary teams of experts (involving educators, usability experts, designers, and developers). It will be interesting to compare the video recording for expert and novice teams to generalize results for collaboration and scaffolding provided by the card-based toolkit for multidisciplinary GBL teams. Moreover, expert evaluations will be conducted focusing on evaluating the developed learning game ideas to provide valuable insights into the quality of generated ideas in terms of completeness and multi-dimensional focus.

- **Refining and extending the toolkit to include prototyping:** Some enhancements are planned in the proposed card-based toolkit (such as improving reflection cards with defined criteria and examples, introducing accessories to make cards more searchable). It will further reduce the complexity and increase the ease of use. Future research will focus on extending the toolkit to cover the prototyping phase, where teams develop learning games that can be evaluated. One possibility is to use a programming environment (such as scratch) to develop prototypes of the generated learning game ideas that can be evaluated. Future plans could also include integrating the developed GBL toolkit in a course (such as game development) and changing the design activity rules from time-bound to finishing cards. It will be interesting to investigate detailed and long-term use of the developed toolkit where teams will have the opportunity to complete all primary cards without time limitation. Future work could also involve specializing the toolkit for specific application domains, e.g., security games or health games, and internationalizing the toolkit by translating it into other languages.

- **Exploring other measures to gain deeper understanding:** Further research will involve extending the design workshops' study dimensions to include creative thinking (creativity), which is also essential for the early design phase of learning games. Moreover, the analysis instrument and integrated evaluation approach will be evaluated in more detail using a questionnaire to investigate utility, value, and ease of use. Furthermore, one interesting possibility for GBL evaluation is investigating other user characteristics such as gender and cultural aspects to compare learning games' effects.

- **Using multi-model data for in-depth analysis:** Game-based learning is a complex process associated with many aspects (such as cognitive load, engagement, attention). Therefore, future work will involve using multimodal data to capture the experience with multiple data streams complementing each other for a complete picture of the GBL learning experience. It will be interesting to use eye-tracking, physiological data, and game logs to identify the physiological features, gaze patterns, and specific game log parameters that best predict learning and flow. This type of multi-model data will be useful to guide real-time adaptive scaffolding in learning games.

- **Collaboration between academia and Industry**: A collaboration between industry and research is essential for thorough insights. Therefore, future research will involve industry practitioners to evaluate toolkit in industry-based projects with real-life constraints and extended periods. It will bring useful insights from the industry, and research will go beyond the academic researchers.

- **Automating GBL design and evaluation:** Future research will focus on automating or partially automating GBL evaluation using the proposed framework and game data logs. The future work will also focus on developing a web-based ideation and evaluation tool that will facilitate the educational game design and evaluation process during different phases of the development lifecycle and help professional and game companies working with game-based learning.

- **Identifying GBL design patterns:** Future work will focus on identifying GBL design patterns in the ideation process that can result in effective and efficient learning game designs to further help GBL designers in the learning game design process. The research will focus on investigating team dynamics with professional GBL design teams consisting of multidisciplinary experts and track design decisions to identify design patterns leading to effective educational games.

- **Long-term effects of the GBL approach:** Future studies could obtain more in-depth insights from a longitudinal collection of engagement and psychosocial data for educational games in different contexts. One possibility would be to explore, in detail, the relationship between user characteristics and affective reactions, such as focusing on the relation between parent-children play that increases engagement. Furthermore, the research could focus on exploring the effects of GBL on the psychosocial well-being of refugee children.

References

Abbott, D. (2019). "Game-based learning for postgraduates: an empirical study of an educational game to teach research skills." Higher Education Pedagogies **4**(1): 80-104.

AbuJarour, S., H. Krasnova, H. Wenninger, J. Fedorowicz, S. Olbrich, C.-W. Tan and C. Urquhart (2016). "Leveraging Technology for Refugee Integration: How Can We Help?".

Afshari, M., K. A. Bakar, W. S. Luan, B. A. Samah and F. S. Fooi (2009). "Factors affecting teachers' use of information and communication technology." Online Submission **2**(1): 77-104.

Ågerfalk, P. J. (2013). Embracing diversity through mixed methods research, Taylor & Francis.

Ahlfeldt, S., S. Mehta and T. Sellnow (2005). "Measurement and analysis of student engagement in university classes where varying levels of PBL methods of instruction are in use." Higher Education Research & Development **24**(1): 5-20.

Ahmad, M., L. A. Rahim and N. I. Arshad (2015). Towards an Effective Modelling and Development of Educational Games with Subject-Matter: A Multi-Domain Framework. 2015 5th International Conference on IT Convergence and Security (ICITCS), IEEE.

Ak, O. (2012). "A game scale to evaluate educational computer games." Procedia-Social and Behavioral Sciences **46**: 2477-2481.

Alaswad, Z. and L. Nadolny (2015). "Designing for game-based learning: The effective integration of technology to support learning." Journal of Educational Technology Systems **43**(4): 389-402.

Aleven, V., E. Myers, M. Easterday and A. Ogan (2010). Toward a framework for the analysis and design of educational games. 2010 third IEEE international conference on digital game and intelligent toy enhanced learning, IEEE.

Aleven, V., E. Stahl, S. Schworm, F. Fischer and R. Wallace (2003). "Help seeking and help design in interactive learning environments." Review of educational research **73**(3): 277-320.

Alfadhli, S. and A. Alsumait (2015). Game-based learning guidelines: Designing for Learning and Fun. 2015 International Conference on Computational Science and Computational Intelligence (CSCI), IEEE.

Alhuhud, G. and W. Altamimi (2016). "Quality evaluation of mobile game: Miftah Alfasaha." Mobile Information Systems **2016**.

Alkhasawneh, I. M., M. T. Mrayyan, C. Docherty, S. Alashram and H. Y. Yousef (2008). "Problem-based learning (PBL): assessing students’ learning preferences using VARK." Nurse education today **28**(5): 572-579.

All, A., E. P. Nunez Castellar and J. Van Looy (2014). Defining best practices for assessing the effectiveness of digital game-based learning. 64rd International Communication Association (ICA) Annual Conference.

All, A., J. Van Looy and E. P. N. Castellar (2013). "An evaluation of the added value of co-design in the development of an educational game for road safety." International Journal of Game-Based Learning (IJGBL) **3**(1): 1-17.

Alqurashi, E. (2019). Technology tools for teaching and learning in real time. Educational technology and resources for synchronous learning in higher education, IGI Global**:** 255-278.

Alvarez, C. P., R. Martinez-Maldonado and S. B. Shum (2020). LA-DECK: A card-based learning analytics co-design tool. Proceedings of the tenth international conference on learning analytics & knowledge.

Amory, A. (2007). "Game object model version II: a theoretical framework for educational game development." Educational Technology Research and Development **55**(1): 51-77.

Anastasiadis, T., G. Lampropoulos and K. Siakas (2018). "Digital Game-based Learning and Serious Games in Education." International Journal of Advances in Scientific Research and Engineering **4**(12): 139-144.

Andersen, E. (2012). Optimizing adaptivity in educational games. Proceedings of the International Conference on the Foundations of Digital Games.

Anderson, N. J., L. Bachman, K. Perkins and A. Cohen (1991). "An exploratory study into the construct validity of a reading comprehension test: Triangulation of data sources." Language testing **8**(1): 41-66.

Annetta, L. A. (2010). "The "I's" have it: A framework for serious educational game design." Review of General Psychology **14**(2): 105-113.

Apiola, M. and E. Sutinen (2020). "Design science research for learning software engineering and computational thinking: Four cases." Computer Applications in Engineering Education.

Aramo-Immonen, H. (2011). Mixed methods research design. World Summit on Knowledge Society, Springer.

Ariffin, M. M. and S. Sulaiman (2013). Evaluating game-based learning (GBL) effectiveness for higher education (HE). 2013 International Conference on Advanced Computer Science Applications and Technologies, IEEE.

Arnab, S., R. Berta, J. Earp, S. de Freitas, M. Popescu, M. Romero, I. Stanescu and M. Usart (2012). "Framing the adoption of serious games in formal education." Electronic Journal of e-learning **10**(2): 159-171.

Aslaksen, K., M. Haga, H. Sigmundsson and H. Lorås (2020). Evidence for a Common Multi-Modal Learning Style in Young Adults? A Psychometric Investigation of Two Modality-Specific Learning Style Inventories. Frontiers in Education, Frontiers.

Atherton, J. (2005). "Learning and teaching: Assimilation and accommodation." UK: Available: http://www. learningandteaching. info/learning/assimacc. htm.

Ávila-Pesántez, D., L. A. Rivera and M. S. Alban (2017). "Approaches for serious game design: A systematic literature review." The ASEE Computers in Education (CoED) Journal **8**(3).

Azizt, A. and S. Subiyanto (2018). Digital Game-Based Learning in Arabic Language Learning and its effects on Students' Academic Performance. International Conference on Indonesian Technical Vocational Education and Association (APTEKINDO 2018), Atlantis Press.

Azlan, C. A., J. H. D. Wong, L. K. Tan, M. S. N. A. Huri, N. M. Ung, V. Pallath, C. P. L. Tan, C. H. Yeong and K. H. Ng (2020). "Teaching and learning of postgraduate medical physics using Internet-based e-learning during the COVID-19 pandemic–A case study from Malaysia." Physica Medica **80**: 10-16.

Backlund, P. and M. Hendrix (2013). Educational games-are they worth the effort? A literature survey of the effectiveness of serious games. 2013 5th international conference on games and virtual worlds for serious applications (VS-GAMES), IEEE.

Baehr, M. (2005). "Distinctions between assessment and evaluation." Program Assessment Handbook **7**(1): 231-234.

Bank, W. (2020). The COVID-19 Crisis Response: Supporting Tertiary Education for Continuity, Adaptation, and Innovation, World Bank.

Barbosa, M. B., A. B. Rego and I. de Medeiros (2015). HEEG: Heuristic evaluation for educational games. XIV Brazilian Symposium of Games and Digital Entertainment. Teresina-PI, Brazil.

Barwood, H. and N. Falstein (2002). More of the 400: Discovering design rules. Lecture at Game Developers Conference.

Becker, K. (2011). "The magic bullet: A tool for assessing and evaluating learning potential in games." International Journal of Game-Based Learning (IJGBL) **1**(1): 19-31.

Bedwell, W. L., D. Pavlas, K. Heyne, E. H. Lazzara and E. Salas (2012). "Toward a taxonomy linking game attributes to learning: An empirical study." Simulation & Gaming **43**(6): 729-760.

Belanich, J., K. L. Orvis and D. E. Sibley (2013). "PC-based game features that influence instruction and learner motivation." Military Psychology **25**(3): 206-217.

Belay, E. G., D. S. McCrickard and S. A. Besufekad (2016). Mobile user interaction development for low-literacy trends and recurrent design problems: A perspective from designers in developing country. International conference on cross-cultural design, Springer.

Bell, B. A. (2010). "Pretest-posttest design." Encyclopedia of research design: 1087-1092.

Bellotti, F., B. Kapralos, K. Lee, P. Moreno-Ger and R. Berta (2013). "Assessment in and of serious games: an overview." Advances in human-computer interaction **2013**.

Benson, R. M. (2014). Games based learning, University of Warwick.

Binzak, V., C. G. Anderson, J.-D. A. Kumar and M. Berland (2016). "Comparing gameplay across formal and informal contexts." Proceedings of the 1st DiGRA and FDG: 1-2.

Bittner, E. A. C. and J. M. Leimeister (2014). "Creating shared understanding in heterogeneous work groups: Why it matters and how to achieve it." Journal of management information systems **31**(1): 111-144.

Bixler, B. (2006). "Motivation and its relationship to the design of educational games." NMC. Cleveland, Ohio. Retrieved **10**(07).

Bjorklund, D. F. and K. B. Causey (2017). Children's thinking: Cognitive development and individual differences, Sage Publications.

Block, K., D. Warr, L. Gibbs and E. Riggs (2013). "Addressing ethical and methodological challenges in research with refugee-background young people: Reflections from the field." Journal of Refugee Studies **26**(1): 69-87.

Bond, M., K. Buntins, S. Bedenlier, O. Zawacki-Richter and M. Kerres (2020). "Mapping research in student engagement and educational technology in higher education: A systematic evidence map." International Journal of Educational Technology in Higher Education **17**(1): 2.

Borthick, A. F., D. R. Jones and S. Wakai (2003). "Designing Learning Experiences within Learners' Zones of Proximal Development (ZPDs): Enabling Collaborative Learning On-Site and Online." Journal of Information Systems **17**(1): 107-134.

Boyle, E., T. M. Connolly and T. Hainey (2011). "The role of psychology in understanding the impact of computer games." Entertainment computing **2**(2): 69-74.

Boyle, E. A., T. Hainey, T. M. Connolly, G. Gray, J. Earp, M. Ott, T. Lim, M. Ninaus, C. Ribeiro and J. Pereira (2016). "An update to the systematic literature review of empirical evidence of the impacts and outcomes of computer games and serious games." Computers & Education **94**: 178-192.

Brandt, E. and J. Messeter (2004). Facilitating collaboration through design games. Proceedings of the eighth conference on Participatory design: Artful integration: interweaving media, materials and practices-Volume 1.

Breuer, J. and G. Bente (2010). "Why so serious? On the relation of serious games and learning." Journal for Computer Game Culture **4**: 7-24.

Bruner, J. S. (1966). Toward a theory of instruction, Harvard University Press.

Burkett, I. (2012). "An introduction to co-design." Sydney: Knode.

Calderón, A. and M. Ruiz (2015). "A systematic literature review on serious games evaluation: An application to software project management." Computers & Education **87**: 396-422.

Candappa, M. and M. Ahmad (2007). "Education and schooling for asylum-seeking and refugee students in Scotland: An exploratory study."

Cao, J., J. M. Crews, M. Lin, A. Deokar, J. K. Burgoon and J. F. Nunamaker Jr (2006). "Interactions between system evaluation and theory testing: a demonstration of the power of a mulitfaceted approach to systems research." Journal of Management Information Systems **22**(4): 207-235.

Caracelli, V. J. and J. C. Greene (1997). "Crafting mixed-method evaluation designs." New directions for evaluation **74**: 19-32.

Carrión, M., M. Santórum, J. Benavides, J. Aguilar, Y. Ortiz and M. Pérez (2020). Design and Evaluation of a Virtual Reality Serious Game. International Conference on Applied Human Factors and Ergonomics, Springer.

Carroll, J. M. (2004). "Beyond fun." interactions **11**(5): 38-40.

Chard, S. M. (1999). "Using design science in educational technology research projects." methods **1687**: 1713-1726.

Cheong, C., F. Cheong and J. Filippou (2013). Using Design Science Research to Incorporate Gamification into Learning Activities. PACIS.

Chesbrough, H. W. (2003). Open innovation: The new imperative for creating and profiting from technology, Harvard Business Press.

Chorianopoulos, K. and M. Giannakos (2014). "Design principles for serious video games in mathematics education: from theory to practice."

Christinaki, E., N. Vidakis and G. Triantafyllidis (2014). "A novel educational game for teaching emotion identification skills to preschoolers with autism diagnosis." Computer science and information systems **11**(2): 723-743.

Cleven, A., P. Gubler and K. M. Hüner (2009). Design alternatives for the evaluation of design science research artifacts. Proceedings of the 4th International Conference on Design Science Research in Information Systems and Technology.

Colardyn, D. and J. Bjornavold (2005). The learning continuity: European inventory on validating non-formal and informal learning: National policies and practices in validating non-formal and informal learning, Office for official publications of the European Communities.

Commission, I. O. f. S. I. E. (2001). ISO/IEC 9126–Software Engineering–Product Quality, Geneve.

Connolly, T., M. Stansfield and T. Hainey (2009). Towards the development of a games-based learning evaluation framework. Games-based learning advancements for multi-sensory human computer interfaces: Techniques and effective practices, IGI Global**:** 251-273.

Connolly, T. M., E. A. Boyle, E. MacArthur, T. Hainey and J. M. Boyle (2012). "A systematic literature review of empirical evidence on computer games and serious games." Computers & education **59**(2): 661-686.

Connolly, T. M., M. Stansfield and T. Hainey (2007). "An application of games-based learning within software engineering." British Journal of Educational Technology **38**(3): 416-428.

Corbeil, P. and D. Laveault (2011). "Validity of a Simulation Game as a Method for History Teaching." Simulation & Gaming **42**(4): 462-475.

Cowley, B., D. Charles, M. Black and R. Hickey (2006). User-system-experience model for user centered design in computer games. International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems, Springer.

Creswell, J. W. and V. L. Plano Clark (2011). "Choosing a mixed methods design." Designing and conducting mixed methods research **2**: 53-106.

Csikszentmihalyi, M. (1990). Flow: The psychology of optimal experience, Harper & Row New York.

Csikszentmihalyi, M. (1991). Flow, the psychology of optimal experience, steps towards enchancing the quality of life. harper&row, publishers, Inc.

Csikszentmihalyi, M. (2000). Beyond boredom and anxiety, Jossey-Bass.

Csikszentmihalyi, M. (2014). Toward a psychology of optimal experience. Flow and the foundations of positive psychology, Springer**:** 209-226.

Csikzentmihalyi, M. (1975). Beyond boredom and anxiety, San Francisco: Jossey-Bass.

Culbertson, S. and L. Constant (2015). "Education of Syrian refugee children: Managing the crisis in Turkey, Lebanon, and Jordan."

da Costa, A. C., F. Rebelo and A. Rodrigues (2017). Co-designing a Civic Educational Online Game with Children. International Conference of Design, User Experience, and Usability, Springer.

Daniel, J. (2020). "Education and the COVID-19 pandemic." Prospects **49**(1): 91-96.

de Freitas, S. (2018). "Are games effective learning tools? A review of educational games." Journal of Educational Technology & Society **21**(2): 74-84.

de Freitas, S. and F. Liarokapis (2011). Serious games: a new paradigm for education? Serious games and edutainment applications, Springer**:** 9-23.

de Freitas, S. and M. Oliver (2006). "How can exploratory learning with games and simulations within the curriculum be most effectively evaluated?" Computers & education **46**(3): 249-264.

de Freitas, S. I. (2006). "Using games and simulations for supporting learning." Learning, media and technology **31**(4): 343-358.

de Jans, S., K. Van Geit, V. Cauberghe, L. Hudders and M. De Veirman (2017). "Using games to raise awareness: How to co-design serious mini-games?" Computers & Education **110**(1): 77-87.

de Lima, L. G. R., A. de Lima Salgado and A. P. Freire (2015). Evaluation of the user experience and intrinsic motivation with educational and mainstream digital games. Proceedings of the Latin American Conference on Human Computer Interaction.

de Lope, R. P., N. Medina-Medina, R. M. Soldado, A. M. García and F. L. Gutiérrez-Vela (2017). Designing educational games: Key elements and methodological approach. 2017 9th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games), IEEE.

de Wal Pastoor, L. (2016). "Rethinking refugee education: Principles, policies and practice from a European perspective." Annual review of comparative and international education 2016.

Deane, S. (2016). "Syria's lost generation: Refugee education provision and societal security in an ongoing conflict emergency."

Dedeaux, T. B. (2016). "Not Created Equal: The Effectiveness of Two Types of Educational Computer Games."

Deng, Y., A. N. Antle and C. Neustaedter (2014). Tango cards: a card-based design tool for informing the design of tangible learning games. Proceedings of the 2014 conference on Designing interactive systems.

Derbali, L. and C. Frasson (2010a). Players' motivation and EEG waves patterns in a serious game environment. International Conference on Intelligent Tutoring Systems, Springer.

Derbali, L. and C. Frasson (2010b). Prediction of players motivational states using electrophysiological measures during serious game play. 2010 10th IEEE International Conference on Advanced Learning Technologies, IEEE.

Deterding, S., D. Dixon, R. Khaled and L. Nacke (2011). From game design elements to gamefulness: defining" gamification". Proceedings of the 15th international academic MindTrek conference: Envisioning future media environments.

Diah, N. M., M. Ismail, S. Ahmad and M. K. M. Dahari (2010). Usability testing for educational computer game using observation method. 2010 international conference on information retrieval & knowledge management (CAMP), IEEE.

Díaz, P., P. Paredes, D. Alvarado and E. Giaccardi (2012). Co-designing social games with children to support non formal learning. 2012 IEEE 12th International Conference on Advanced Learning Technologies, IEEE.

Dib, C. Z. (1988). Formal, non-formal and informal education: concepts/applicability. AIP conference proceedings, American Institute of Physics.

Dibitonto, M., F. Tazzi, K. Leszczynska and C. M. Medaglia (2017). The IoT design deck: A tool for the co-design of connected products. International Conference on Applied Human Factors and Ergonomics, Springer.

Dickey, M. D. (2007). "Game design and learning: A conjectural analysis of how massively multiple online role-playing games (MMORPGs) foster intrinsic motivation." Educational Technology Research and Development **55**(3): 253-273.

Djelil, F., E. Sanchez, B. Albouy-Kissi, J.-M. Lavest and A. Albouy-Kissi (2014). Towards a learning game evaluation methodology in a training context: A literature review. European Conference on Games-Based Learning.

Dondi, C. and M. Moretti (2007). "A methodological proposal for learning games selection and quality assessment." British Journal of Educational Technology **38**(3): 502-512.

Drolia, M., E. Sifaki, S. Papadakis and M. Kalogiannakis (2020). "An Overview of Mobile Learning for Refugee Students: Juxtaposing Refugee Needs with Mobile Applications' Characteristics." Challenges **11**(2): 31.

Due, C., D. W. Riggs and M. Augoustinos (2014). "Research with children of migrant and refugee backgrounds: A review of child-centered research methods." Child Indicators Research **7**(1): 209-227.

Dunn, W. E. and J. P. Lantolf (1998). "Vygotsky's zone of proximal development and Krashen's i+ 1: Incommensurable constructs; incommensurable theories." Language learning **48**(3): 411-442.

Dunwell, I., P. Lameras, P. Petridis, S. Arnab and S. de Freitas (2014). Facilitating intuitive-guided learning in a serious game through integration with a learning content management system. 2014 Eighth International Conference on Complex, Intelligent and Software Intensive Systems, IEEE.

Eagle, M. (2009). Level up: a frame work for the design and evaluation of educational games. Proceedings of the 4th International Conference on Foundations of Digital Games.

Ebner, M. and A. Holzinger (2007). "Successful implementation of user-centered game based learning in higher education: An example from civil engineering." Computers & education **49**(3): 873-890.

El-Sattar, A. and H. Hussein (2016). Learning Islamic Principles with Serious Games. Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology, ACM.

Eltahir, M. E., N. R. Alsalhi, S. Al-Qatawneh, H. A. AlQudah and M. Jaradat (2021). "The impact of game-based learning (GBL) on students' motivation, engagement and academic performance on an Arabic language grammar course in higher education." Education and Information Technologies.

Fargas-Malet, M., D. McSherry, E. Larkin and C. Robinson (2010). "Research with children: Methodological issues and innovative techniques." Journal of early childhood research **8**(2): 175-192.

Federoff, M. A. (2002). Heuristics and usability guidelines for the creation and evaluation of fun in video games, Citeseer.

Finneran, C. M. and P. Zhang (2005). "Flow in computer-mediated environments: promises and challenges." Communications of the association for information systems **15**(1): 4.

Flanagan, M., J. Belman, H. Nissenbaum and J. Diamond (2007). A Method For Discovering Values in Digital Games. DiGRA Conference, Citeseer.

Flick, U. (1992). "Triangulation revisited: strategy of validation or alternative?" Journal for the theory of social behaviour.

Fu, F.-L., R.-C. Su and S.-C. Yu (2009). "EGameFlow: A scale to measure learners' enjoyment of e-learning games." Computers & Education **52**(1): 101-112.

Gall, J. P., M. D. Gall and W. R. Borg (1999). Applying educational research: A practical guide, Longman Publishing Group.

Gallacher, J. and M. Feutrie (2003). "Recognising and accrediting informal and non-formal learning in higher education: An analysis of the issues emerging from a study of France and Scotland." European journal of education **38**(1): 71-83.

Garris, R., R. Ahlers and J. E. Driskell (2002). "Games, motivation, and learning: A research and practice model." Simulation & gaming **33**(4): 441-467.

Gasteiger, H., A. Obersteiner and K. Reiss (2015). Formal and informal learning environments: Using games to support early numeracy. Describing and studying domain-specific serious games, Springer**:** 231-250.

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

Gee, J. P. (2003). "What video games have to teach us about learning and literacy." Computers in Entertainment (CIE) **1**(1): 20-20.

Gee, J. P. (2005). "Learning by design: Good video games as learning machines." E-learning and Digital Media **2**(1): 5-16.

Genemo, H., S. J. Miah and A. McAndrew (2016). "A design science research methodology for developing a computer-aided assessment approach using method marking concept." Education and Information Technologies **21**(6): 1769-1784.

Giannakos, M. N. (2013). "Enjoy and learn with educational games: Examining factors affecting learning performance." Computers & Education **68**: 429-439.

Giannakos, M. N., K. Sharma, I. O. Pappas, V. Kostakos and E. Velloso (2019). "Multimodal data as a means to understand the learning experience." International Journal of Information Management **48**: 108-119.

Gibson, B. and T. Bell (2013). Evaluation of games for teaching computer science. Proceedings of the 8th Workshop in Primary and Secondary Computing Education.

Gossen, T., J. Hempel and A. Nürnberger (2013). "Find it if you can: usability case study of search engines for young users." Personal and Ubiquitous Computing **17**(8): 1593-1603.

Graham, C. R., W. Woodfield and J. B. Harrison (2013). "A framework for institutional adoption and implementation of blended learning in higher education." The internet and higher education **18**: 4-14.

Greenberg, B. S., J. Sherry, K. Lachlan, K. Lucas and A. Holmstrom (2010). "Orientations to video games among gender and age groups." Simulation & Gaming **41**(2): 238-259.

Gregor, S. and D. Jones (2007). The anatomy of a design theory, Association for Information Systems.

Gunter, G., R. F. Kenny and E. H. Vick (2006). "A case for a formal design paradigm for serious games." The Journal of the International Digital Media and Arts Association **3**(1): 93-105.

Guo, Z., L. Xiao, C. Van Toorn, Y. Lai and C. Seo (2016). "Promoting online learners' continuance intention: An integrated flow framework." Information & Management **53**(2): 279-295.

Habgood, J. (2005). "Zombie Division: Intrinsic integration in digital learning games." Cognitive Science Research Paper-University of Sussex CSRP **576**: 45.

Hainey, T. (2010). Using games-based learning to teach requirements collection and analysis at tertiary education level, University of the West of Scotland Edinburgh.

Hainey, T., T. M. Connolly, E. A. Boyle, A. Wilson and A. Razak (2016). "A systematic literature review of games-based learning empirical evidence in primary education." Computers & Education **102**: 202-223.

Hamari, J. (2017). "Do badges increase user activity? A field experiment on the effects of gamification." Computers in human behavior **71**: 469-478.

Hamilton, R. and D. Moore (2003). Educational interventions for refugee children: Theoretical perspectives and implementing best practice, Routledge.

Harpstead, E., B. A. Myers and V. Aleven (2013). In search of learning: facilitating data analysis in educational games. Proceedings of the SIGCHI conference on human factors in computing systems.

Harteveld, C. (2011). Triadic game design: Balancing reality, meaning and play, Springer Science & Business Media.

Hays, R. T. (2005). "The effectiveness of instructional games: A literature review and discussion."

Hersh, M. A. and B. Leporini (2013). "An overview of accessibility and usability of educational games." Student usability in educational software and games: Improving experiences: 1-40.

Hevner, A. and S. Chatterjee (2010). Design science research in information systems. Design research in information systems, Springer**:** 9-22.

Hevner, A. R. (2007). "A three cycle view of design science research." Scandinavian journal of information systems **19**(2): 4.

Hevner, A. R., S. T. March, J. Park and S. Ram (2004). "Design science in information systems research." MIS quarterly: 75-105.

Hildén, E., J. Ojala and K. Väänänen (2017). Development of context cards: a bus-specific ideation tool for co-design workshops. Proceedings of the 21st International Academic Mindtrek Conference.

Hill, M. (1997). "Participatory research with children." Child & family social work **2**(3): 171-183.

Hirumi, A. and C. Stapleton (2009). Applying pedagogy during game development to enhance game-based learning. Games: Purpose and potential in education, Springer**:** 127-162.

Hoffman, B. and L. Nadelson (2010). "Motivational engagement and video gaming: A mixed methods study." Educational Technology Research and Development **58**(3): 245-270.

Holstein, K., B. M. McLaren and V. Aleven (2019). "Co-designing a real-time classroom orchestration tool to support teacher–AI complementarity." Journal of Learning Analytics **6**(2).

Hopkins, P. (2008). "Ethical issues in research with unaccompanied asylum-seeking children." Children's Geographies **6**(1): 37-48.

Hornecker, E. (2010). Creative idea exploration within the structure of a guiding framework: the card brainstorming game. Proceedings of the fourth international conference on Tangible, embedded, and embodied interaction.

Hsieh, H.-F. and S. E. Shannon (2005). "Three approaches to qualitative content analysis." Qualitative health research **15**(9): 1277-1288.

Hsu, Y.-C., J.-L. Hung and Y.-H. Ching (2013). "Trends of educational technology research: More than a decade of international research in six SSCI-indexed refereed journals." Educational Technology Research and Development **61**(4): 685-705.

Huang, H., K. H. Ng, B. Bedwell and S. Benford (2020). "A card-based internet of things game ideation tool for museum context." Journal of Ambient Intelligence and Humanized Computing: 1-12.

Huitt, W. (2009). "Humanism and open education." Educational psychology interactive.

Huitt, W. and J. Hummel (2003). "Piaget's theory of cognitive development." Educational psychology interactive **3**(2): 1-5.

Hung, H.-T., J. C. Yang, G.-J. Hwang, H.-C. Chu and C.-C. Wang (2018). "A scoping review of research on digital game-based language learning." Computers & Education **126**: 89-104.

Hussein, M. H., S. H. Ow, L. S. Cheong, M.-K. Thong and N. A. Ebrahim (2019). "Effects of digital game-based learning on elementary science learning: A systematic review." IEEE Access **7**: 62465-62478.

Ibrahim, R. and A. Jaafar (2009). Educational games (EG) design framework: Combination of game design, pedagogy and content modeling. 2009 international conference on electrical engineering and informatics, IEEE.

Ismail, M., N. M. Diah, S. Ahmad, N. A. M. Kamal and M. K. M. Dahari (2011). Measuring usability of educational computer games based on the user success rate. 2011 International Symposium on Humanities, Science and Engineering Research, IEEE.

Ismail, M. A.-A., A. Ahmad, J. A.-M. Mohammad, N. M. R. M. Fakri, M. Z. M. Nor and M. N. M. Pa (2019). "Using Kahoot! as a formative assessment tool in medical education: a phenomenological study." BMC medical education **19**(1): 1-8.

ISO, W. (1998). "9241-11. Ergonomic requirements for office work with visual display terminals (VDTs)." The international organization for standardization **45**(9).

Jagušt, T., I. Botički and H. J. So (2018). "A review of research on bridging the gap between formal and informal learning with technology in primary school contexts." Journal of Computer Assisted Learning **34**(4): 417-428.

Jessen, S., J. Mirkovic and C. M. Ruland (2018). "Creating Gameful design in mHealth: a participatory Co-Design approach." JMIR mHealth and uHealth **6**(12): e11579.

Johnson, R. B. and L. Christensen (2019). Educational research: Quantitative, qualitative, and mixed approaches, Sage publications.

Johnson, R. B. and A. J. Onwuegbuzie (2004). "Mixed methods research: A research paradigm whose time has come." Educational researcher **33**(7): 14-26.

Kadel, R., S. J. Halder, K. Paudel and M. P. Gurung (2018). Analyzing effect of GBL on student engagement and academic performance in computer networking course. Proceedings of the 20th International Conference on Information Integration and Web-based Applications & Services.

Kalyuga, S. and J. L. Plass (2009). Evaluating and managing cognitive load in games. Handbook of research on effective electronic gaming in education, IGI Global**:** 719-737.

Kamii, C. and R. DeVries (1980). Group games in early education: Implications of Piaget's theory, ERIC.

Karlstrøm, M. L. and T. Markussen (2020). Preparing for Colonising Mars through Collaborative and Educational Virtual Reality Games. Master's thesis, Norwegian University of Science and Technolo.

Kato, P. M., S. W. Cole, A. S. Bradlyn and B. H. Pollock (2008). "A video game improves behavioral outcomes in adolescents and young adults with cancer: a randomized trial." Pediatrics **122**(2): e305-e317.

Kaukko, M., K. Dunwoodie and E. Riggs (2017). "Rethinking the ethical and methodological dimensions of research with refugee children." ZEP: Zeitschrift für internationale Bildungsforschung und Entwicklungspädagogik **40**(1): 16-21.

Kebritchi, M., A. Hirumi and H. Bai (2010). "The effects of modern mathematics computer games on mathematics achievement and class motivation." Computers & education **55**(2): 427-443.

Kececi, N. and A. Abran (2001). "Analyzing, measuring & assessing software quality within a logic-based graphical framework." Universite du Quebec a Montreal.

Kelle, S., R. Klemke and M. Specht (2011). "Design patterns for learning games." International Journal of Technology Enhanced Learning **3**(6): 555-569.

Keller, J. M. (1983). "Motivational design of instruction." Instructional design theories and models: An overview of their current status **1**(1983): 383-434.

Keller, J. M. (1987). "Development and use of the ARCS model of instructional design." Journal of instructional development **10**(3): 2-10.

Kenali, H. M. S., N. M. R. N. Yusoff, H. Abdullah and A. M. S. Kenali (2019). "The Effects of Language Games on Smartphones in Developing Arabic Speaking Skills among Non-Native Speakers." Creative Education **10**(05): 972.

Kenny, R. F. and G. A. Gunter (2007). "Endogenous fantasy-based serious games: Intrinsic motivation and learning." International Journal of Social Sciences **2**(1): 8-13.

Khaled, R. and A. Vasalou (2014). "Bridging serious games and participatory design." International Journal of Child-Computer Interaction **2**(2): 93-100.

Khalil, M. K., F. Paas, T. E. Johnson and A. F. Payer (2005). "Interactive and dynamic visualizations in teaching and learning of anatomy: a cognitive load perspective." The Anatomical Record Part B: The New Anatomist: An Official Publication of the American Association of Anatomists **286**(1): 8-14.

Khan, A., F. H. Ahmad and M. M. Malik (2017). "Use of digital game based learning and gamification in secondary school science: The effect on student engagement, learning and gender difference." Education and Information Technologies **22**(6): 2767-2804.

Kiili, K. (2005a). "Digital game-based learning: Towards an experiential gaming model." The Internet and higher education **8**(1): 13-24.

Kiili, K. (2005b). On educational game design: Building blocks of flow experience, Tampere University of Technology.

Kiili, K., S. de Freitas, S. Arnab and T. Lainema (2012). "The design principles for flow experience in educational games." Procedia Computer Science **15**: 78-91.

Kiili, K., T. Lainema, S. de Freitas and S. Arnab (2014). "Flow framework for analyzing the quality of educational games." Entertainment computing **5**(4): 367-377.

Kim, K.-J. and C. J. Bonk (2006). "The future of online teaching and learning in higher education." Educause quarterly **29**(4): 22-30.

Kolb, D. A. (2014). Experiential learning: Experience as the source of learning and development, FT press.

Kolb, D. A., I. M. Rubin and J. M. McIntyre (1971). "Organizational psychology."

Korbach, A., R. Brünken and B. Park (2017). "Measurement of cognitive load in multimedia learning: a comparison of different objective measures." Instructional science **45**(4): 515-536.

Koutromanos, G. and L. Avraamidou (2014). "The use of mobile games in formal and informal learning environments: a review of the literature." Educational Media International **51**(1): 49-65.

Koutromanos, G., A. Sofos and L. Avraamidou (2015). "The use of augmented reality games in education: a review of the literature." Educational Media International **52**(4): 253-271.

Kreimeier, B. (2002). The case for game design patterns.

Krouska, A., C. Troussas and M. Virvou (2019). "A literature review of Social Networking-based Learning Systems using a novel ISO-based framework." Intelligent Decision Technologies **13**(1): 23-39.

Kukulska-Hulme, A. (2007). "Mobile usability in educational contexts: what have we learnt?" The international review of research in open and distributed learning **8**(2).

Kultima, A. and J. Paavilainen (2007). Creativity techniques in game design. Proceedings of the 2007 conference on Future Play.

Kuo, M.-J. (2007). How does an online game based learning environment promote students' intrinsic motivation for learning natural science and how does it affect their learning outcomes? 2007 First IEEE International Workshop on Digital Game and Intelligent Toy Enhanced Learning (DIGITEL'07), IEEE.

Kwiatkowska, J., A. Szóstek and D. Lamas (2014). (Un) structured sources of inspiration: comparing the effects of game-like cards and design cards on creativity in co-design process. Proceedings of the 13th Participatory Design Conference: Research Papers-Volume 1.

Kwok, L. (2012). "Exploratory-triangulation design in mixed methods studies: A case of examining graduating seniors who meet hospitality recruiters' selection criteria." Tourism and Hospitality Research **12**(3): 125-138.

Kwok, S. Y., D. Harrison and A. Malizia (2017). "Designing individualisation of eco information: A conceptual framework and design toolkit." International Journal of Sustainable Engineering **10**(6): 302-312.

Langford, M. and C. Damşa (2020). Online Teaching in the Time of COVID-19.

Laurillard, D. (2013). Rethinking university teaching: A conversational framework for the effective use of learning technologies, Routledge.

Law, E. L.-C., V. Roto, M. Hassenzahl, A. P. Vermeeren and J. Kort (2009). Understanding, scoping and defining user experience: a survey approach. Proceedings of the SIGCHI conference on human factors in computing systems.

Law, E. L.-C. and X. Sun (2012). "Evaluating user experience of adaptive digital educational games with Activity Theory." International journal of human-computer studies **70**(7): 478-497.

Lee, Y. (2008). "Design participation tactics: the challenges and new roles for designers in the co-design process." Co-design **4**(1): 31-50.

Li, C.-S. and B. Irby (2008). "An overview of online education: Attractiveness, benefits, challenges, concerns and recommendations." College Student Journal **42**(2).

Lim, C. P., D. Nonis and J. Hedberg (2006). "Gaming in a 3D multiuser virtual environment: Engaging students in science lessons." British Journal of Educational Technology **37**(2): 211-231.

Linaza, J. (1984). "Piaget's marbles: the study of children's games and their knowledge of rules." Oxford Review of Education **10**(3): 271-274.

Liu, T.-Y. and Y.-L. Chu (2010). "Using ubiquitous games in an English listening and speaking course: Impact on learning outcomes and motivation." Computers & Education **55**(2): 630-643.

Liu, Z.-Y., Z. Shaikh and F. Gazizova (2020). "Using the concept of game-based learning in education." International Journal of Emerging Technologies in Learning (iJET) **15**(14): 53-64.

Lodge, J. M., L. Hansen and D. Cottrell (2016). "Modality preference and learning style theories: rethinking the role of sensory modality in learning." Learning: Research and Practice **2**(1): 4-17.

Loftus, G. R. and E. F. Loftus (1983). Mind at play; The psychology of video games, Basic Books, Inc.

Loh, C. S., Y. Sheng and D. Ifenthaler (2015). Serious games analytics: Theoretical framework. Serious games analytics, Springer**:** 3-29.

Lu, C., M. Chang, E. Huang and C.-W. Chen (2011). "Usability of context-aware mobile educational game." Knowledge Management & E-Learning: An International Journal **3**(3): 448-477.

Lucero, A. and J. Arrasvuori (2010). PLEX Cards: a source of inspiration when designing for playfulness. Proceedings of the 3rd International Conference on Fun and Games.

MacFarland, T. W. and J. M. Yates (2016). Wilcoxon matched-pairs signed-ranks test. Introduction to Nonparametric statistics for the biological sciences using R, Springer**:** 133-175.

MacKenty, B. (2006). "All Play and No Work: Computer Games Are Invading the Classroom--And Not a Moment Too Soon." School Library Journal **52**(9): 46.

Maldonado, H., B. Lee and S. Klemmer (2007). "Patterns of Collaboration in Design Courses: Team dynamics affect technology appropriation, artifact creation, and course performance."

Malone, T. (1981a). What makes computer games fun? Proceedings of the Joint Conference on Easier and More Productive Use of Computer Systems.(Part-II): Human Interface and the User Interface-Volume 1981.

Malone, T. W. (1980). What makes things fun to learn? A study of intrinsically motivating computer games, ProQuest Information & Learning.

Malone, T. W. (1981b). "Toward a theory of intrinsically motivating instruction." Cognitive science **5**(4): 333-369.

March, S. T. and G. F. Smith (1995). "Design and natural science research on information technology." Decision support systems **15**(4): 251-266.

Marciano, J. N., L. C. de Miranda and E. E. C. de Miranda (2014). "Evaluating Multiple Aspects of Educational Computer Games: Literature Review and Case Study." International Journal of Computer Games Technology.

Marconi, A., G. Schiavo, M. Zancanaro, G. Valetto and M. Pistore (2018). Exploring the world through small green steps: improving sustainable school transportation with a game-based learning interface. Proceedings of the 2018 International Conference on Advanced Visual Interfaces.

Markopoulos, P. and M. Bekker (2003a). "Interaction design and children." Interacting with computers **15**(2): 141-149.

Markopoulos, P. and M. Bekker (2003b). "On the assessment of usability testing methods for children." Interacting with computers **15**(2): 227-243.

Mayer, R. E. and R. Moreno (2003). "Nine ways to reduce cognitive load in multimedia learning." Educational psychologist **38**(1): 43-52.

Mayes, J. T. and C. J. Fowler (1999). "Learning technology and usability: a framework for understanding courseware." Interacting with computers **11**(5): 485-497.

McLeod, S. (2007). "Simply psychology." Recuperado de http://www. simplypsychology. org/vygotsky. html.

McMahon, B. and J. P. Portelli (2004). "Engagement for what? Beyond popular discourses of student engagement." Leadership and Policy in Schools **3**(1): 59-76.

Melnic, A.-S. and N. Botez (2014). "Formal, non-formal and informal interdependence in education." Economy Transdisciplinarity Cognition **17**(1): 113-118.

Menashy, F. and Z. Zakharia (2017). "Investing in the crisis: Private participation in the education of Syrian refugees." Education International: 1-40.

Michaud, L. and J. Alvarez (2008). "Serious games." Advergaming, edugaming, training and more, IDATE Consulting & Research.

Mildner, P., N. Stamer and W. Effelsberg (2015). From game characteristics to effective learning games. Joint International Conference on Serious Games, Springer.

Milovanovic, M., M. Minovic, I. Kovacevic, J. Minovic and D. Starcevic (2009). Effectiveness of game-based learning: Influence of cognitive style. World Summit on Knowledge Society, Springer.

Mingers, J. (2001). "Combining IS research methods: towards a pluralist methodology." Information systems research **12**(3): 240-259.

Minovic, M. and M. Milovanovic (2013). Real-time learning analytics in educational games. Proceedings of the first international conference on technological ecosystem for enhancing multiculturality.

Mirnig, A. G., A. Meschtscherjakov, D. Wurhofer, T. Meneweger and M. Tscheligi (2015). A formal analysis of the ISO 9241-210 definition of user experience. Proceedings of the 33rd annual ACM conference extended abstracts on human factors in computing systems.

Mishra, P., M. J. Koehler and K. Kereluik (2009). "Looking back to the future of educational technology." TechTrends **53**(5): 49.

Mitgutsch, K. and N. Alvarado (2012). Purposeful by design? A serious game design assessment framework. Proceedings of the International Conference on the foundations of digital games.

Mohamed, H. and A. Jaafar (2010a). Challenges in the evaluation of educational computer games. 2010 International Symposium on Information Technology, IEEE.

Mohamed, H. and A. Jaafar (2010b). Development and potential analysis of heuristic evaluation for educational computer game (PHEG). 5th International Conference on Computer Sciences and Convergence Information Technology, IEEE.

Mohamed, H., R. Yusoff and A. Jaafar (2012). Quantitive analysis in a heuristic evaluation for usability of educational computer game (UsaECG). 2012 International Conference on Information Retrieval & Knowledge Management, IEEE.

Moore, J. S. and C. Price (2009). Development of an Educational Immersive Environment for Primary School Literacy Education, Academic Conferences Ltd.

Mora, S., F. Gianni and M. Divitini (2017). Tiles: a card-based ideation toolkit for the internet of things. Proceedings of the 2017 conference on designing interactive systems.

Moreno, R. (2006). "Does the modality principle hold for different media? A test of the method-affects-learning hypothesis." Journal of Computer Assisted Learning **22**(3): 149-158.

Mueller, F., M. R. Gibbs, F. Vetere and D. Edge (2014). Supporting the creative game design process with exertion cards. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.

Myers, C., L. Piccolo and T. Collins (2018). "Co-designing cards on social issues for creating educational games."

Nagalingam, V. and R. Ibrahim (2015). "User experience of educational games: a review of the elements." Procedia Computer Science **72**: 423-433.

Nakayama, M., H. Yamamoto and R. Santiago (2007). "The Impact of Learner Characteristics on Learning Performance in Hybrid Courses among Japanese Students." Electronic Journal of E-Learning **5**(3): 195-206.

Narli, N. (2018). "Life, connectivity and integration of Syrian refugees in Turkey: Surviving through a smartphone." Questions de communication(33): 269-286.

Neag, A. (2020). "Unaccompanied Refugee Children and Media Literacy: Doing Media Education Research on the Margins." The Handbook of Media Education Research: 61-74.

Neil, K. (2012). "Game design tools: Time to evaluate." Proceedings of 2012 DiGRA Nordic.

Ni, Q. and Y. Yu (2015). Research on Educational Mobile Games and the effect it has on the Cognitive Development of Preschool Children. 2015 Third International Conference on Digital Information, Networking, and Wireless Communications (DINWC), IEEE.

Nielsen, J. (1994). Usability engineering, Morgan Kaufmann.

Nielsen, J. (2005). Ten usability heuristics, http://www. nngroup. com/articles/ten-usability-heuristics/(acc-essed ….

Nightingale, A. (2009). Methods: Triangulation. International Encyclopedia of Human Geography. R. Kitchen, Thrifts, N, Oxford, Elsevier.

Noel, L.-A. and T. L. Liub (2017). "Using Design Thinking to Create a New Education Paradigm for Elementary Level Children for Higher Student Engagement and Success." Design and Technology Education **22**(1): n1.

Nokelainen, P. (2004). Conceptual definition of the technical and pedagogical usability criteria for digital learning material. EdMedia+ Innovate Learning, Association for the Advancement of Computing in Education (AACE).

Nordhaug, L. M. K. (2016). "Eduapp4Syria [Online]. norad.no: Norwegian Agency for Development Cooperation (Norad)." Retrieved [Accessed April 25th 2019], from Available: http://norad.no/eduapp4syria.

Nousiainen, T. (2009). Children's involvement in the design of game-based learning environments. Design and use of serious games, Springer**:** 49-66.

Nousiainen, T., M. Kangas, J. Rikala and M. Vesisenaho (2018). "Teacher competencies in game-based pedagogy." Teaching and Teacher Education **74**: 85-97.

O'Neil, H. F., R. Wainess and E. L. Baker (2005). "Classification of learning outcomes: Evidence from the computer games literature." The Cirriculum Journal **16**(4): 455-474.

Ojose, B. (2008). "Applying Piaget's theory of cognitive development to mathematics instruction." The mathematics educator **18**(1).

Omar, H. M. and A. Jaafar (2010). "Conceptual framework for a heuristics based methodology for interface evaluation of educational games." Computer and Information Science **3**(2): 211.

Oprins, E., G. Visschedijk, M. B. Roozeboom, M. Dankbaar, W. Trooster and S. C. Schuit (2015). "The game-based learning evaluation model (GEM): measuring the effectiveness of serious games using a standardised method." International journal of technology enhanced learning **7**(4): 326-345.

Orvis, K. A., D. B. Horn and J. Belanich (2006). Videogame-based training success: The impact of trainee characteristics-Year 2, ARMY RESEARCH INST FOR THE BEHAVIORAL AND SOCIAL SCIENCES ALEXANDRIA VA ….

Orvis, K. A., D. B. Horn and J. Belanich (2008). "The roles of task difficulty and prior videogame experience on performance and motivation in instructional videogames." Computers in Human behavior **24**(5): 2415-2433.

Paas, F. and P. Ayres (2014). "Cognitive load theory: A broader view on the role of memory in learning and education." Educational Psychology Review **26**(2): 191-195.

Paas, F., A. Renkl and J. Sweller (2003). "Cognitive load theory and instructional design: Recent developments." Educational psychologist **38**(1): 1-4.

Paas, F., A. Renkl and J. Sweller (2004). "Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture." Instructional science **32**(1/2): 1-8.

Palaiologou, N., G. Fountoulaki and M. Liontou (2019). Refugee education in Greece: Challenges, needs, and priorities in non-formal settings–An intercultural approach. Intercultural and Interfaith Dialogues for Global Peacebuilding and Stability, IGI Global**:** 165-197.

Papastergiou, M. (2009). "Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation." Computers & education **52**(1): 1-12.

Papert, S. (1998). "Does easy do it? Children, games, and learning." Game Developer **5**(6): 88.

Pappa, D. and L. Pannese (2010). Effective design and evaluation of serious games: The case of the e-VITA project. World Summit on Knowledge Society, Springer.

Park, J., S. Kim, A. Kim and Y. Y. Mun (2019). "Learning to be better at the game: Performance vs. completion contingent reward for game-based learning." Computers & Education **139**: 1-15.

Parry, S. B. (1971). "The Name of the Game... Is Simulation." Training and Development Journal **25**(2): 28-32.

Pauline-Graf, D. and S. E. Mandel (2019). "Defining Preliminary Research for Digital Game-Based Learning Evaluation: Best Practices." International Journal of Educational Methodology **5**(4): 623-635.

Pavlas, D. (2010). "A Model Of Flow And Play In Game-based Learning The Impact Of Game Characteristics, Player Traits, And Player States."

Paz, F. and J. A. Pow-Sang (2014). Current trends in usability evaluation methods: a systematic review. 2014 7th International Conference on Advanced Software Engineering and Its Applications, IEEE.

Paz, I. L. and F. Fernandes (2018). "A literature review for game design frameworks towards educational purposes." Lisboa Play2Learn: 321-332.

Pearce, J. M., M. Ainley and S. Howard (2005). "The ebb and flow of online learning." Computers in human behavior **21**(5): 745-771.

Pearce, J. M. and S. Howard (2004). Designing for flow in a complex activity. Asia-Pacific Conference on Computer Human Interaction, Springer.

Peffers, K., T. Tuunanen, M. A. Rothenberger and S. Chatterjee (2007). "A design science research methodology for information systems research." Journal of management information systems **24**(3): 45-77.

Peirce, N., O. Conlan and V. Wade (2008). Adaptive educational games: Providing non-invasive personalised learning experiences. 2008 second IEEE international conference on digital game and intelligent toy enhanced learning, IEEE.

Peña-Lévano, L. (2020). "Personalizing Online Classes: The Use of Evaluation and Participation Tools." Applied Economics Teaching Resources (AETR) **2**(2): TBD-TBD.

Penuel, W. R., J. Roschelle and N. Shechtman (2007). "Designing formative assessment software with teachers: An analysis of the co-design process." Research and practice in technology enhanced learning **2**(01): 51-74.

Pernin, J.-P., C. Mariais, F. Michau, V. Emin-Martinez and N. Mandran (2014). "Using game mechanisms to foster GBL designers’ cooperation and creativity." International Journal of Learning Technology 6 **9**(2): 139-160.

Petri, G. and C. G. von Wangenheim (2016). "How to evaluate educational games: a systematic." Journal of Universal Computer Science **22**(7): 992-1021.

Piaget, J. (1962). "The relation of affectivity to intelligence in the mental development of the child." Bulletin of the Menninger clinic **26**(3): 129.

Piaget, J. (1964). "Part I: Cognitive development in children: Piaget development and learning." Journal of Research in Science Teaching **2**(3): 176-186.

Piaget, J. (1976). Piaget’s theory. Piaget and his school, Springer**:** 11-23.

Pillay, H., J. Brownlee and L. Wilss (1999). "Cognition and recreational computer games: Implications for educational technology." Journal of research on computing in education **32**(1): 203-216.

Pireva, K., R. Tahir, A. S. Imran and N. Chaudhary (2019). Evaluating learners' emotional states by monitoring brain waves for comparing game-based learning approach to pen-and-paper. 2019 IEEE Frontiers in Education Conference (FIE), IEEE.

Pireva, K., R. Tahir, A. I. Wang and A. S. Imran (2021). "Game-based digital quiz as a tool for improving students' engagement and learning in online lectures." Journal of Computer Assisted Learning **under review**.

Pivec, M. (2007). Play and learn: potentials of game-based learning, Wiley Online Library.

Pivec, M., O. Dziabenko and I. Schinnerl (2003). Aspects of game-based learning. 3rd International Conference on Knowledge Management, Graz, Austria.

Plass, J. L., B. D. Homer and C. K. Kinzer (2015). "Foundations of game-based learning." Educational Psychologist **50**(4): 258-283.

Plass, J. L. and S. Pawar (2020). "Toward a taxonomy of adaptivity for learning." Journal of Research on Technology in Education **52**(3): 275-300.

Plass, J. L., K. Perlin and J. Nordlinger (2010). The games for learning institute: Research on design patterns for effective educational games. Game Developers Conference, San Francisco, CA.

Prensky, M. (2001). "Fun, play and games: What makes games engaging." Digital game-based learning **5**(1): 5-31.

Prensky, M. (2002a). "The motivation of gameplay." On the horizon.

Prensky, M. (2002b). "The Motivation of Gameplay or, the REAL 21th century learning revolution." URL http://www. marcprensky. com/writing/Prensky: 2010-2011.

Prensky, M. (2003a). "Digital game-based learning." Computers in Entertainment (CIE) **1**(1): 21-21.

Prensky, M. (2003b). "Escape from planet Jar-Gon: Or, what video games have to teach academics about teaching and writing." On the horizon.

Pries-Heje, J., R. Baskerville and J. R. Venable (2008). "Strategies for design science research evaluation."

Prieto, L. P., M. J. Rodríguez-Triana, R. Martínez-Maldonado, Y. Dimitriadis and D. Gašević (2019). "Orchestrating learning analytics (OrLA): Supporting inter-stakeholder communication about adoption of learning analytics at the classroom level." Australasian Journal of Educational Technology **35**(4).

Pritchard, A. (2017). Ways of learning: Learning theories for the classroom, Routledge.

Privitera, G. J. and L. Delzell (2019). "Quasi-experimental and single-case experimental designs." Research methods for education: 333-370.

Protopsaltis, A., L. Pannese, D. Pappa and S. Hetzner (2011). "Serious games and formal and informal learning." E-Learning Papers: 1887-1542.

Pursi, A. and L. Lipponen (2018). "Constituting play connection with very young children: Adults' active participation in play." Learning, Culture and Social Interaction **17**: 21-37.

Qian, M. and K. R. Clark (2016). "Game-based Learning and 21st century skills: A review of recent research." Computers in human behavior **63**: 50-58.

Quinn, C. N. (1996). "Designing an instructional game: Reflections on 'Quest for Independence'." Education and Information Technologies **1**(3): 251-269.

Radha, R., K. Mahalakshmi, V. S. Kumar and A. Saravanakumar (2020). "E-Learning during lockdown of Covid-19 pandemic: A global perspective." International journal of control and automation **13**(4): 1088-1099.

Ratheeswari, K. (2018). "Information communication technology in education." Journal of Applied and Advanced Research **3**(1): 45-47.

Ritterfeld, U., M. Cody and P. Vorderer (2009). Serious games: Mechanisms and effects, Routledge.

Robson, C. and K. McCartan (2016). Real world research, John Wiley & Sons.

Robson, K., K. Plangger, J. H. Kietzmann, I. McCarthy and L. Pitt (2015). "Is it all a game? Understanding the principles of gamification." Business horizons **58**(4): 411-420.

Rogers, Y., K. Connelly, L. Tedesco, W. Hazlewood, A. Kurtz, R. E. Hall, J. Hursey and T. Toscos (2007). Why it's worth the hassle: The value of in-situ studies when designing ubicomp. International Conference on Ubiquitous Computing, Springer.

Romi, S. and M. Schmida (2009). "Non-formal education: A major educational force in the postmodern era." Cambridge Journal of Education **39**(2): 257-273.

Rooney, P. (2012). "A theoretical framework for serious game design: exploring pedagogy, play and fidelity and their implications for the design process." International Journal of Game-Based Learning (IJGBL) **2**(4): 41-60.

Roser, T. and A. Samson (2009). "Co-creation: New paths to value." London: Promise/LSE Enterprise.

Rushby, N. and D. W. Surry (2016). Mapping the field and terminology. The Wiley handbook of learning technology, Wiley Online Library. **1:** 1-14.

Ryan, R. M. and E. L. Deci (2000). "Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being." American psychologist **55**(1): 68.

Ryan, R. M., C. S. Rigby and A. Przybylski (2006). "The motivational pull of video games: A self-determination theory approach." Motivation and emotion **30**(4): 344-360.

Saariluoma, P. (2005). "Editor's instruction: The challenges and opportunities of Human Technology." Human Technology: An Interdisciplinary Journal on Humans in ICT Environments.

Salah, J., S. Abdennadher, C. Sabty and Y. Abdelrahman (2016). "Super Alpha: Arabic Alphabet Learning Serious Game for Children with Learning Disabilities." Serious Games, Jcsg 2016 **9894**: 104-115.

Salanova, M., R. M. Grau, E. Cifre and S. Llorens (2000). "Computer training, frequency of usage and burnout: the moderating role of computer self-efficacy." Computers in Human Behavior **16**(6): 575-590.

Salas, E. and J. A. Cannon-Bowers (2001). "The science of training: A decade of progress." Annual review of psychology **52**(1): 471-499.

Sampath, K. (1981). Introduction to educational technology, Sterling Publishers Pvt. Ltd.

Sanchez, E. (2019). Game-Based Learning. Encyclopedia of Education and Information Technologies. A. Tatnall. Cham, Springer International Publishing**:** 1-9.

Sanders, E. B.-N. and P. J. Stappers (2008). "Co-creation and the new landscapes of design." Co-design **4**(1): 5-18.

Scaife, M., Y. Rogers, F. Aldrich and M. Davies (1997). Designing for or designing with? Informant design for interactive learning environments. Proceedings of the ACM SIGCHI Conference on Human factors in computing systems.

Schifter, C. C., D. J. Ketelhut and B. C. Nelson (2011). Middle School Children Participation in an Immersive Virtual Game Environment, Presence, and Piaget's Stages of Development. Society for Information Technology & Teacher Education International Conference, Association for the Advancement of Computing in Education (AACE).

Schmidt, M., Y. Earnshaw, A. A. Tawfik and I. Jahnke (2020). "Methods of user centered design and evaluation for learning designers." Learner and User Experience Research.

Schrader, C. and T. Bastiaens (2012). "Learning in educational computer games for novices: The impact of support provision types on virtual presence, cognitive load, and learning outcomes." International Review of Research in Open and Distributed Learning **13**(3): 206-227.

Schuler, D. and A. Namioka (1993). Participatory design: Principles and practices, CRC Press.

Schwier, R. and J. Seaton (2013). "A Comparison of Participation Patterns in Selected Formal, Non-formal, and Informal Online Learning Environments/Comparaison des modes de participation dans des environnements formels, non formels et informels d'apprentissage en ligne." Canadian Journal of Learning and Technology/La revue canadienne de l’apprentissage et de la technologie **39**(1).

Schwier, R. A. (2012). Comparing formal, non-formal, and informal online learning environments. The next generation of distance education, Springer**:** 139-156.

Seeney, M. and H. Routledge (2009). Drawing circles in the sand: Integrating content into serious games. Games-Based Learning Advancements for Multi-Sensory Human Computer Interfaces: Techniques and Effective Practices, IGI Global**:** 84-97.

Semple, A. (2000). "Learning theories and their influence on the development and use of educational technologies." Australian Science Teachers Journal **46**(3): 21.

Shadish, W. R., T. D. Cook and D. T. Campbell (2002). Experimental and quasi-experimental designs for generalized causal inference/William R. Shedish, Thomas D. Cook, Donald T. Campbell, Boston: Houghton Mifflin.

Shaffer, D. W., K. R. Squire, R. Halverson and J. P. Gee (2005). "Video games and the future of learning." Phi delta kappan **87**(2): 105-111.

Shafiq, S. and T. A. Khan (2018). "Role & value of usability in educational learning via game based apps." International Journal of Scientific and Technology Research **7**(11): 70-77.

Sharan, S. and C. Colodner (1976). "Counselor: A simulation game for vocational decision-making." Simulation & Games **7**(2): 193-208.

Sharma, G. (2017). "Pros and cons of different sampling techniques." International journal of applied research **3**(7): 749-752.

Shi, Y.-R. and J.-L. Shih (2015). "Game factors and game-based learning design model." International Journal of Computer Games Technology **2015**.

Shield, L. and A. Kukulska-Hulme (2006). "Are language learning websites special? Towards a research agenda for discipline-specific usability." Journal of Educational Multimedia and Hypermedia **15**(3): 349-369.

Shiratuddin, N. and S. B. Zaibon (2011). Designing user experience for mobile game-based learning. 2011 International Conference on User Science and Engineering (i-USEr), IEEE.

Shute, V. J. and F. Ke (2012). Games, learning, and assessment. Assessment in game-based learning, Springer**:** 43-58.

Shute, V. J., L. Wang, S. Greiff, W. Zhao and G. Moore (2016). "Measuring problem solving skills via stealth assessment in an engaging video game." Computers in Human Behavior **63**: 106-117.

Silva, F. G. (2020). "Practical Methodology for the Design of Educational Serious Games." Information **11**(1): 14.

Sim, G., M. Horton and S. Strong (2004). Interfaces for online assessment: friend or foe. 7th HCI Educators Workshop.

Sim, G., S. MacFarlane and M. Horton (2005). Evaluating usability, fun and learning in educational software for children. EdMedia+ Innovate Learning, Association for the Advancement of Computing in Education (AACE).

Sim, G., S. MacFarlane and J. Read (2006). "All work and no play: Measuring fun, usability, and learning in software for children." Computers & Education **46**(3): 235-248.

Simon, H. A. (1996). "The science of design: creating the artificial."

Simon, H. A. (2019). The sciences of the artificial, MIT press.

Singer, E., M. Nederend, L. Penninx, M. Tajik and J. Boom (2014). "The teacher's role in supporting young children's level of play engagement." Early Child Development and Care **184**(8): 1233-1249.

Sintoris, C., I. Mavrommati, N. Avouris and I. Chatzigiannakis (2018). Out of the box: Using gamification cards to teach ideation to engineering students. European Conference on Ambient Intelligence, Springer.

Siraj-Blatchford, I. (2009). "Conceptualising progression in the pedagogy of play and sustained shared thinking in early childhood education: A Vygotskian perspective."

Skadberg, Y. X. and J. R. Kimmel (2004). "Visitors' flow experience while browsing a Web site: its measurement, contributing factors and consequences." Computers in human behavior **20**(3): 403-422.

Snow, E. L., A. D. Likens, L. K. Allen and D. S. McNamara (2016). "Taking control: Stealth assessment of deterministic behaviors within a game-based system." International Journal of Artificial Intelligence in Education **26**(4): 1011-1032.

Soewardi, H. and M. Perdana (2019). The usability of the educational board game for learning english. IOP Conference Series: Materials Science and Engineering, IOP Publishing.

Soni, V. D. (2020). "Global Impact of E-learning during COVID 19." Available at SSRN 3630073.

Sonnenberg, C. and J. Vom Brocke (2011). Evaluation patterns for design science research artefacts. European Design Science Symposium, Springer.

Spill, T. A. M. and G. Bruinsma (2016). "Designing Serious Games with the Game of Games." Proceedings of the 10th European Conference on Games Based Learning: 634-643.

Squire, K. (2006). "From content to context: Videogames as designed experience." Educational researcher **35**(8): 19-29.

Squire, K. (2013). "Video games and learning: Teaching and participatory culture in the digital age." Alberta Journal of Educational Research **59**(1): 129-132.

Squire, K. D. (2007). "Games, learning, and society: Building a field." Educational Technology: 51-55.

Squires, D. (1999). Usability and educational software design: special issue of interacting with computers, Oxford University Press Oxford, UK.

Standardization, I. O. f. (2008). Ergonomics of Human-system Interaction, ISO.

Steen, M. (2013). "Co-design as a process of joint inquiry and imagination." Design Issues **29**(2): 16-28.

Steen, M., M. Manschot and N. De Koning (2011). "Benefits of co-design in service design projects." International Journal of Design **5**(2).

Stein, A., Y. Yotam, R. Puzis, G. Shani and M. Taieb-Maimon (2018). "EEG-triggered dynamic difficulty adjustment for multiplayer games." Entertainment computing **25**: 14-25.

Steiner, C. M., M. D. Kickmeier-Rus and D. Albert (2015). "Making Sense of Game-Based User Data: Learning Analytics in Applied Games." International Association for Development of the Information Society.

Steinkuehler, C., E. King, C. Martin, Y. Oh, S. Chu, C. Williams, A. Ochsner, S. Harris, V. E. Owen and G. Anton (2011). Mixed methods to study games and learning. Proceedings of the 7th international conference on Games+ Learning+ Society Conference.

Streicher, A. and J. D. Smeddinck (2016). Personalized and adaptive serious games. Entertainment Computing and Serious Games, Springer**:** 332-377.

Strekalova, E. and J. L. Hoot (2008). "What is special about special needs of refugee children?: Guidelines for teachers." Multicultural Education **16**(1): 21-24.

Su, C.-H., K. T.-K. Chen and K.-K. Fan (2013). "Rough set theory based fuzzy TOPSIS on serious game design evaluation framework." Mathematical Problems in Engineering **2013**.

Susi, T., M. Johanesson and P. Backlund (2007). "Serious games-an overview (technical report)." Skövde, Sweden: University of Skövde.

Sweetser, P. and P. Wyeth (2005). "GameFlow: a model for evaluating player enjoyment in games." Computers in Entertainment (CIE) **3**(3): 3-3.

Sweller, J. (1988). "Cognitive load during problem solving: Effects on learning." Cognitive science **12**(2): 257-285.

Sweller, J., J. J. Van Merrienboer and F. G. Paas (1998). "Cognitive architecture and instructional design." Educational psychology review **10**(3): 251-296.

Tahir, R. and A. I. Wang (2017). State of the art in game based learning: Dimensions for evaluating educational games. European Conference on Games Based Learning, Academic Conferences International Limited.

Tahir, R. and A. I. Wang (2018). Insights into design of educational games: Comparative analysis of design models. Proceedings of the Future Technologies Conference, Springer.

Tahir, R. and A. I. Wang (2019a). Exploring methods and guidelines for child-computer interaction research with refugee children. International Conference on Human-Computer Interaction, Springer.

Tahir, R. and A. I. Wang (2019b). "How to Evaluate Educational Games With Refugee Children: Methodological Aspects and Lessons Learned From EduApp4syria."

Tahir, R. and A. I. Wang (2020a). "Codifying Game-Based Learning: Development and Application of LEAGUÊ Framework for Learning Games." Electronic Journal of e-Learning **18**(1): 69-87.

Tahir, R. and A. I. Wang (2020b). "Transforming a Theoretical Framework to Design Cards: LEAGUE Ideation Toolkit for Game-Based Learning Design." Sustainability **12**(Special Issue Design Methodology for Educational Games 20): 8487.

Tahir, R. and A. I. Wang (2021a). Completeness and collaboration in the early design phase of learning games: Do ideation cards provide scaffolding? International Conference on Human-Computer Interaction. Washington, D.C., Springer.

Tahir, R. and A. I. Wang (2021b). "Evaluating the effectiveness of game-based learning for teaching refugee children Arabic using the integrated LEAGUÊ-GQM approach." Computers & Education **Under review**.

Tan, P.-H., S.-W. Ling and C.-Y. Ting (2007). Adaptive digital game-based learning framework. Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts.

Tang, S., M. Hanneghan and A. El Rhalibi (2009). Introduction to games-based learning. Games-based learning advancements for multi-sensory human computer interfaces: Techniques and effective practices, IGI Global**:** 1-17.

Thalmann, M., A. S. Souza and K. Oberauer (2019). "How does chunking help working memory?" Journal of Experimental Psychology: Learning, Memory, and Cognition **45**(1): 37.

Theodosiou, S. and I. Karasavvidis (2015). "Serious games design: A mapping of the problems novice game designers experience in designing games." Journal of e-Learning and Knowledge Society **11**(3).

Thomas, R. L. (2016). "The right to quality education for refugee children through social inclusion." Journal of Human Rights and Social Work **1**(4): 193-201.

Thomas, S., G. Schott and M. Kambouri (2004). "Designing for learning or designing for fun? Setting usability guidelines for mobile educational games." Learning with mobile devices: A book of papers: 173-181.

Tissot, P. (2004). "Terminology of vocational training policy." A multilingual glossary for an enlarged Europe. Luxembourg: Cedefop.

Tobar-Muñoz, H., S. Baldiris and R. Fabregat (2016). Co design of augmented reality game-based learning games with teachers using co-CreaARGBL method. 2016 IEEE 16th International Conference on Advanced Learning Technologies (ICALT), IEEE.

Torrente, J., P. Moreno-Ger, B. Fernández-Manjón and Á. Del Blanco (2009). Game-like simulations for online adaptive learning: A case study. International Conference on Technologies for E-Learning and Digital Entertainment, Springer.

Tran, M. Q. and R. Biddle (2008). Collaboration in serious game development: a case study. Proceedings of the 2008 Conference on Future Play: Research, Play, Share.

Trinidad, S. (2003). Working with technology-rich learning environments: Strategies for success. Technology-rich learning environments: A future perspective, World Scientific**:** 97-113.

Trischler, J., S. J. Pervan, S. J. Kelly and D. R. Scott (2018). "The value of codesign: The effect of customer involvement in service design teams." Journal of Service Research **21**(1): 75-100.

Tseloudi, C. and T. Tsiatsos (2015). Panic in the gallery: An online educational game for art history: Design and evaluation of a matching game. 2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA), IEEE.

Vaajakallio, K. and T. Mattelmäki (2014). "Design games in codesign: as a tool, a mindset and a structure." CoDesign **10**(1): 63-77.

Van Biljon, J., J. Traxler, R. van der Merwe and D. Van Heerden (2015). "Curriculum development for mobile digital literacy skills acquisition using a Design Science Approach." The Journal of Community Informatics **11**(3).

Van Eck, R. (2006). "Digital game-based learning: It's not just the digital natives who are restless." EDUCAUSE review **41**(2): 16.

Van Eck, R. (2007). Building artificially intelligent learning games. Games and simulations in online learning: Research and development frameworks, IGI Global**:** 271-307.

Van Merrienboer, J. J. and P. Ayres (2005). "Research on cognitive load theory and its design implications for e-learning." Educational Technology Research and Development **53**(3): 5-13.

Van Staalduinen, J.-P. and S. de Freitas (2011). "A game-based learning framework: Linking game design and learning." Learning to play: exploring the future of education with video games **53**: 29.

Vandercruysse, S., M. Vandewaetere and G. Clarebout (2012). "Game-based learning: A review on the effectiveness of educational games." Handbook of research on serious games as educational, business and research tools: 628-647.

Vanderhoven, E., B. Willems, S. Van Hove, A. All and T. Schellens (2015). How to evaluate educational games: Lessons learned from the evaluation study of Master FIND. European Conference on Games Based Learning, Academic Conferences International Limited.

Vandewaetere, M., F. Cornillie, G. Clarebout and P. Desmet (2013). "Adaptivity in Educational Games: Including Player and Gameplay Characteristics." International Journal of Higher Education **2**(2): 106-114.

Vann, S. W. and A. A. Tawfik (2020). "Flow Theory and Learning Experience Design in Gamified Learning Environments." Learner and User Experience Research.

Venable, J. (2006). The role of theory and theorising in design science research. Proceedings of the 1st International Conference on Design Science in Information Systems and Technology (DESRIST 2006), Citeseer.

Venable, J., J. Pries-Heje and R. Baskerville (2012). A comprehensive framework for evaluation in design science research. International Conference on Design Science Research in Information Systems, Springer.

Venkatesh, V., S. A. Brown and H. Bala (2013). "Bridging the qualitative-quantitative divide: Guidelines for conducting mixed methods research in information systems." MIS quarterly: 21-54.

Viberg, O., B. Wasson and A. Kukulska-Hulme (2021). "Developing a Framework for Mobile Assisted Language Learning through Learning Analytics for Self-Regulated Learning."

Virvou, M. and G. Katsionis (2008). "On the usability and likeability of virtual reality games for education: The case of VR-ENGAGE." Computers & Education **50**(1): 154-178.

Voulgari, I. and G. N. Yannakakis (2019). Digital games in non-formal and informal learning practices for science learning: A case study. International Conference on Games and Learning Alliance, Springer.

Vygotsky, L. (1978). "Interaction between learning and development." Readings on the development of children **23**(3): 34-41.

Vygotsky, L. S. (1980). Mind in society: The development of higher psychological processes, Harvard university press.

Wallner, G. and S. Kriglstein (2011). Design and evaluation of the educational game DOG eometry: a case study. Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology.

Wang, A. I. (2015). "The wear out effect of a game-based student response system." Computers & Education **82**: 217-227.

Wang, A. I. and R. Tahir (2020). "The effect of using Kahoot! for learning–A literature review." Computers & Education **149**: 103818.

Wang, L., Y. Wang and Y. Li (2015). "Mining experiential patterns from game-logs of board game." International Journal of Computer Games Technology **2015**.

Wang, Y., X. Liu, X. Lin and G. Xiang (2015). An evaluation framework for game-based learning. 15th Global Chinese Conference on Computers in Education.

Ward, N., E. Paul, P. Watson, G. Cooke, C. H. Hillman, N. Cohen, A. Kramer and A. K. Barbey (2017). "Enhanced learning through multimodal training: evidence from a comprehensive cognitive, physical fitness, and neuroscience intervention." Scientific reports **7**(1): 1-8.

Wasson, B. (2007). "Design and use of technology enhanced learning environments." Frontiers In Artificial Intelligence and Applications **162**: 9.

Wasson, B. and K. Morgan (2013). "Information and Communications Technology and Learning. State of the field review."

Weisskirch, R. S. (2003). "Dealing with Piaget: Analyzing Card Games for Understanding Concepts."

Wetzel, R., T. Rodden and S. Benford (2017). "Developing ideation cards for mixed reality game design." Transactions of the Digital Games Research Association **3**(2).

Whitton, N. (2009). Learning with digital games: A practical guide to engaging students in higher education, Routledge.

Whitton, N. (2012). Games-Based Learning. Encyclopedia of the Sciences of Learning. N. M. Seel. Boston, MA, Springer US**:** 1337-1340.

Wilson, K. A., W. L. Bedwell, E. H. Lazzara, E. Salas, C. S. Burke, J. L. Estock, K. L. Orvis and C. Conkey (2009). "Relationships between game attributes and learning outcomes: Review and research proposals." Simulation & gaming **40**(2): 217-266.

Winn, B. and C. Heeter (2006). "Resolving conflicts in educational game design through playtesting." Innovate: Journal of Online Education **3**(2).

Winn, W. (2002). "Research into practice: Current trends in educational technology research: The study of learning environments." Educational psychology review **14**(3): 331-351.

Wu, W.-H., W.-B. Chiou, H.-Y. Kao, C.-H. A. Hu and S.-H. Huang (2012). "Re-exploring game-assisted learning research: The perspective of learning theoretical bases." Computers & Education **59**(4): 1153-1161.

Wu, W. H., H. C. Hsiao, P. L. Wu, C. H. Lin and S. H. Huang (2012). "Investigating the learning-theory foundations of game-based learning: a meta-analysis." Journal of Computer Assisted Learning **28**(3): 265-279.

Yacci, M., H. Anne and E. Rozanski (2004). Operations and strategy learning in edutainment interfaces. EdMedia+ Innovate Learning, Association for the Advancement of Computing in Education (AACE).

Yazgan, P., D. E. Utku and I. Sirkeci (2015). Syrian crisis and migration.

Yelland, N. (2003). Learning in school and out: formal and informal experiences with computer games in mathematical contexts. Learning in School, Home and Community, Springer**:** 3-12.

Yoon, S.-J. and H.-S. Park (2013). "Development of a Quality Evaluation Standard for Educational Serious Games." Journal of information and communication convergence engineering **11**(2): 103-111.

Yu, S.-C., F.-L. Fu and C. H. Su (2009). EGameDesign: guidelines for enjoyment and knowledge enhancement. International Conference on Hybrid Learning and Education, Springer.

Yue, W. S. and N. A. M. Zin (2009). Usability evaluation for history educational games. Proceedings of the 2nd international Conference on interaction Sciences: information Technology, Culture and Human.

Yusoff, A. (2010). A conceptual framework for serious games and its validation, University of Southampton.

Zahedi, M., V. Tessier and D. Hawey (2017). "Understanding collaborative design through activity theory." The Design Journal **20**(sup1): S4611-S4620.

Zaibon, S. B. and N. Shiratuddin (2010). Heuristics evaluation strategy for mobile game-based learning. 2010 6th IEEE International Conference on Wireless, Mobile, and Ubiquitous Technologies in Education, IEEE.

Zeng, J., S. Parks and J. Shang (2020). "To learn scientifically, effectively, and enjoyably: A review of educational games." Human Behavior and Emerging Technologies **2**(2): 186-195.

Zhang, H.-f., X.-y. Fan and H.-f. Xing (2010). Research on the design and evaluation of educational games based on the RETAIN model. 2010 Third International Symposium on Knowledge Acquisition and Modeling, IEEE.

Zusho, A., J. S. Anthony, N. Hashimoto and G. Robertson (2014). "Do video games provide motivation to learn?".

Part II. Research Papers

Research Papers

P1. Tahir, Rabail, and Alf Inge Wang. (2017). "State of the art in game-based learning: Dimensions for evaluating educational games." In Proceedings of the 2017 European Conference on Games Based Learning (ECGBL), Academic Conferences International Limited, pp. 641-650.

P2. Tahir, Rabail, and Alf Inge Wang. (2020). "Codifying game-based learning: Development and application of LEAGUÊ framework for learning games." Electronic Journal of e-Learning 18, no. 1: 69-87.

P3. Pireva, Krenare, Rabail Tahir, Ali Shariq Imran, and Niraj Chaudhary. (2019). "Evaluating learners' emotional states by monitoring brain waves for comparing game-based learning approach to pen-and-paper." In Proceedings of the 2019 IEEE Frontiers in Education Conference (FIE), IEEE, pp. 1-8.

P4. Pireva, Krenare, Tahir, Rabail, Alf Inge Wang, and Ali Shariq Imran. (2021). "Game-based digital quiz as a tool for improving students' engagement and learning in online lectures". Ready for submission.

P5. Tahir, Rabail, and Alf Inge Wang. (2018). "Insights into the design of educational games: Comparative analysis of design models." In Proceedings of the 2018 Future Technologies Conference (FTC), Springer, Cham, pp. 1041-1061.

P6. Tahir, Rabail, and Alf Inge Wang. (2020). "Transforming a theoretical framework to design cards: LEAGUÊ ideation toolkit for game-based learning design." Sustainability, Special Issue Design Methodology for Educational Games 12, no. 20: 8487.

P7. Tahir, Rabail, and Alf Inge Wang. (2021). "Completeness and collaboration in the early design phase of learning games: Do ideation cards provide scaffolding?". Accepted in the 2021 International Conference on Human-Computer Interaction (HCII).

P8. Tahir, Rabail, and Alf Inge Wang. (2019). "Exploring methods and guidelines for child-computer interaction research with refugee children." In Proceedings of the 2019 International Conference on Human-Computer Interaction, Springer, Cham, pp. 70-89.

P9. Tahir, Rabail, and Alf Inge Wang. (2021). "Evaluating the effectiveness of game-based learning for teaching refugee children Arabic using the integrated LEAGUÊ-GQM approach". Ready for submission.

P10. Tahir, Rabail, and Alf Inge Wang. (2019). "How to evaluate educational games with refugee children: Methodological aspects and lessons learned from EduApp4syria." In Proceedings of the 2019 European Conference on Games Based Learning (ECGBL), Academic Conferences International Limited, pp. 722-730.

PAPER 1

P1:
State of the art in game-based learning: Dimensions for evaluating educational games

Rabail Tahir and Alf Inge Wang

In: Proceedings of the 2017 European Conference on Games Based Learning (ECGBL)

State of the art in Game Based Learning: Dimensions for Evaluating Educational Games

Rabail Tahir and Alf Inge Wang
Dept. of Computer Science, Norwegian University of Science and Technology, Norway
rabail.tahir@idi.ntnu.no
alfw@idi.ntnu.no

Abstract An increased use of educational games makes it essential to verify these tools for a sound impact by evaluating them from multiple dimensions. This paper presents a Systematic Literature Review (SLR) on state of the art in Game based Learning (GBL) evaluation. Our research examines the current trends and evaluation practices based on data drawn from search in four open databases along with a manual search of 4 journals proceedings. The paper begins with the context for our study, followed by a depiction of the analysis grid that is used to generate a database of existing literature, and methodology adopted to conduct systematic review of this literature. From initial sample of 1929 articles, a total of 58 relevant articles were identified and further examined for the extent of research carried out in GBL evaluation, highlighting the research topics, type of resources, the highly-cited articles, and the existing evaluation approaches and criteria used for evaluation of GBL. It then analyses the selected studies for outlining the dimensions for evaluating educational games. The findings of this papers provide insights for researchers and evaluators into current trends, evaluation practices and multiple dimensions for which an educational game must be evaluated.

Keywords: game based learning, educational games, evaluation framework, evaluation dimensions

1. Introduction

Computer games are played for many reasons including enjoyment, entertainment as well as educational purposes. Computer games prove to be an effective educational tool as they can provide enjoyment to learners in the learning process, allowing them to engage in education while having fun (Mohamed and Jaafar, 2010). Educators began to acknowledge the power of computer games for educational purposes back in 1980s. Games that embody educational objectives are considered to make learning more enjoyable, interesting, fun and more learner-centered thus making it more effective (Wang et al., 2011). The term GBL refers to the use of computer games or software applications that utilize games, for educational or learning purposes. Now days, mobile phones have also been widely used for GBL under the label “mobile game-based learning”. Although traditional practices for education remain in use, GBL has extensively been implemented in various courses (Alfadhli and Alsumait, 2015).

The rapid increase in the use of educational games makes it essential to verify these tools to provide the learners with a suitable learning environment. This verification is made through the evaluation of these tools from multiple dimensions (De Freitas and Oliver, 2006). However, the study of games has dearth of a consistent research paradigm. To realize the potential of educational games, there is a need to have a scientifically sound approach to evaluate their effectiveness. (Eagle, 2009). Most articles on educational games emphasize on whether to use games for learning or to explore the potential of games in providing effective learning (Wang et al.,2011). The study on educational games evaluation are deficient. What is missing in GBL literature is the dearth of empirical evidence on the validity of the approach (Wang et al.,2011). Generally, the development of any software specifically educational games is a very demanding process involving costly resources and time. Therefore, evaluation is vital to remove imperfections and improve efficiency.

Mohamed and Jaafar (2010) highlight three challenges in the evaluation of educational computer games. These challenges are evaluation criteria, evaluation process and the evaluators. Some researchers (Petri and von Wangenheim, 2016) have examined the evaluation process for educational games. Mostly in the past evaluation of educational games use questionnaires, observations, log files or interviews in an ad-hoc manner without defining any criteria for evaluation. However, there were very few attempt in developing a framework that specify criteria to evaluate educational games and research to explore this area is deficient (Mohamed and Jaafar, 2010). There is a need to define the aspects of educational games (Ak, 2012). The process for identifying criteria is more complex and time-consuming than one would think (Dondi and Moretti, 2007).

This research work try to fill this gap by carrying out a SLR specifically focusing only on the evaluation of educational games and identification of the key dimensions for evaluating GBL. The paper is organized as follows.

Section 2 presents the related work, Section 3 describes the method used for conducting SLR, Section 4 illustrates the search results, Section 5 presents analysis, and lastly Section 6 concludes the paper.

2. Related work

Some literature reviews have been carried out on educational games and GBL. Marciano, Miranda, and Miranda (2014) presented a literature review regarding evaluation of various aspects of software, and describes the evaluation methods and applications. The study intended to understand the context of use of different evaluation technique both general and specific with the aim to be able to select and adapt the method to be used in the specific context of language learning with computer games. Abdul Jabbar and Felicia (2015) reviewed papers to investigate that how the design of game-based activities influence engagement and learning. They developed a set of some general recommendations for the instructional design of GBL. Based on a review of literature, Djelil et al. (2014) proposed to organize an evaluation process in design and experimental phases, and use empirical and analytical evaluation methods to lower the risk of a poor designed learning game. The paper highlighted four criteria classes (ludopedagogical environment, learner affective and cognitive reactions, training context, and learner profile) that effect the evaluation process of learning game. The measurement and analysis criteria are introduced before linking them with the three evaluation dimensions' usability, usefulness and acceptability to evaluate learning games in a training context. However, the research study does not provide a developmental basis or method for the selection of these three dimensions. The review by (Arttu et al., 2017) especially focused on exploring the meaning of flow within the context of serious games in addition investigating the relationship between learning and flow, factors influencing the occurrence of flow and operationalization of flow. The review mainly showed that there are only conceptual considerations and no robust empirical evidence exist about the meaning of flow.

Petri and von Wangenheim (2016) presented an SLR on systematic evaluation of educational games focusing on the evaluation process. The study results are based on 11 relevant articles describing 7 approaches to systematically evaluate educational games. The focus was on how the approaches are defined, operationalize, developed and evaluated. The study confirmed that only a few approaches are available to systematically evaluate educational games. However, the research results are based on only 7 encountered approaches where no clear pattern emerged on which factors are essential to evaluate educational games. This showed that further research is required on educational game evaluations to obtain more valid and uniform results. Another study by Petri and von Wangenheim (2017) presented an SLR that is specific for computing education games and explored how evaluations on computing education games are defined, executed and analyzed. According to the results of this study, most evaluations use a simple research design where the game is used and subsequently a subjective feedback is obtained through questionnaires. Most of the evaluations are carried out without replication, using qualitative methods for data analysis without using a well-defined evaluation framework. Thus, although several reviews exist on educational games, the focus of these studies is either on the design of educational games or on the evaluation process and methodology used. There is not a single review that presented a detailed and complete overview of studies focused on GBL evaluation. Thus, the question of what criteria is important for evaluating educational games remains open.

3. Method

This research is conducted as a Systematic Literature Review (SLR) based on the work by (Kitchenham et al., 2009). The goal of the review is to present state of the art in GBL evaluation and identify dimensions of GBL evaluation. The steps of SLR method are described below.

3.1 Research questions

In accordance with the goal of this research work, we performed an SLR focusing on the following questions.

R1. What are the current trends in GBL and/or educational games evaluation?
R1 further includes: R1.1 How much GBL evaluation research has been carried out? R1.2 Which research topics are being addressed? R1.3 Who is leading GBL evaluation research?
R2**.** What are the evaluation practices in GBL or educational games?
This includes: R2.1 Which evaluation approaches (frameworks/models/guidelines etc.) exist for GBL? R2.2 What criteria has been used for evaluation of educational games?
R3. What are the different dimensions for evaluating GBL and/or educational games?

3.2 Search strategy and process

A systematic literature review was conducted in March and April 2017 from a data pool consisting of four open databases (Google scholar, IEEE Xplore, ACM Digital Library and Directory of Open Access Journals (DOAJ)) and four journal proceedings (International Journal of Game-Based Learning (IJGBL), International Journal of Game Theory and Technology (IJGTT), International Journal of Serious Games (IJSG), and Computers & Education. The journal proceedings were selected based on their relevance in the field of GBL and educational games.

The core concepts include educational games, evaluations, evaluation frameworks and children. Several search strings were constructed using the keywords (including synonyms) based on the research questions. The search strategies were formed and adapted according to the specific syntax of each of the selected data sources however, search terms included the keywords "educational games" or "game based learning" or "serious games" or "educational games for children" with "evaluation", "assessment", "evaluation framework", "evaluation criteria," "assessment criteria," "or "metrics for evaluation". Manual search was also conducted for IJGBL, IJGTT and IJSG. However, the journal proceedings of Computers & Education were searched using search strings due to the extensive set of papers. Initially, we wanted to focus on evaluation of educational games for children, but since there is very little literature on this user group we decided to focus on evaluation of GBL in general.

After the initial search results were obtained the selection of primary studies was conducted as a two-stage process described by (Brereton et al, 2007). In the first stage, the title and abstract (abstract was read in case title did not provide clear idea) of articles were reviewed and all irrelevant papers were rejected and duplications were removed. In the second stage, full copies of all the selected papers were reviewed against inclusion/exclusion criteria to obtain relevant studies for this research.

3.3 Inclusion and exclusion criteria

The inclusion and exclusion criteria were defined according to research objectives and presented in Table 1.

3.4 Quality criteria

To ensure the quality of studies reviewed, search of literature was limited to journal articles, conference proceedings and book chapters. Any unpublished article or grey literature was not included. Only articles written in English were considered. The articles were excluded if full text was not available. Furthermore, after review only articles that provided considerable information on GBL evaluation were considered.

Table 1: Criteria for inclusion/ exclusion

Inclusion criteria	Exclusion criteria
Evaluation approach (frameworks/model etc.) for GBL or educational games Review studies for evaluation of GBL Evaluation/Assessment of one or more dimensions of GBL Guidelines or criteria for evaluating GBL Design guidelines/models for GBL useful for evaluation purpose Case studies, empirical studies etc. of evaluating any educational games (using some evaluation framework or predefined criteria).	Analysis studies of GBL acceptability/applicability in education. Comparison of GBL with traditional learning approaches Effectiveness of games in general for education purpose Evaluation of video/leisure games Evaluation of educational software's (m-learning/e-Learning) i.e. not game. Evaluation of Serious games other than education domain. Evaluation methods/process used to carry out evaluation instead of criteria for evaluation (how to evaluate rather than what to evaluate) Any duplications Different versions of same paper

3.5 Data extraction

The data was systematically extracted for each research question. The selected papers were thoroughly read and data was extracted by the first author and reviewed by the second author. The data extracted for **R1.1** is title, year, resource type and resource name, for **R1.2** research topic and description, for **R1.3** no of citations, country and references of each article, for **R2.1** name of evaluation approach, type of evaluation approach and description. for **R2.2** criteria used for evaluation, and for **R3** dimension(s) for evaluation.

4. Search results

In the initial search, we found a total of 1929 articles (see Table 2). The aim of this research was to include all possible relevant articles on GBL evaluation, and therefore the search queries were not restricted by the year. Almost all search strings retrieved results from between 2000 to 2017. A total of 28 search results were older than 2000; from two data sources; google scholar (3 results) and journal of computer & education (25 results) that ranged from 1940 to 1995. They were specially reviewed and were found not relevant to the topic of GBL and hence excluded.

The first stage of selection resulted in 232 potentially relevant articles which were further reviewed. After the second stage, a total of 162 articles were excluded based on inclusion/ exclusion and 12 based on quality criteria (for 8 papers full text was not available and four papers were not in English). Therefore, resulting in 58 articles that were selected for this research study. All the selected studies are listed in references.

Table 2: Initial search results

Data source	Search results	Year(range)
Computers & Education	712	2003-2017
Google scholar	390	2002-2017
IEEEXplore	311	2003-2017
ACM Digital Library	263	2000-2017
DOAJ	237	2005-2017
IJSG	11	2014-2017
IJGBL	5	2011-2017
IJGTT	0	-
Total	1929	2000-2017

5. Analysis

5.1 R1: What are the current trends in game base learning(GBL) or educational games evaluation?

The analysis of 58 selected studies showed that relevant publications were all from 2004 onwards indicating that the research in GBL evaluation is relatively a new field of study. Results indicated an increasing trend in GBL research with most number of studies in 2009 and 2015. Figure 1 presents how much GBL evaluation research has been carried out per year (R1.1).

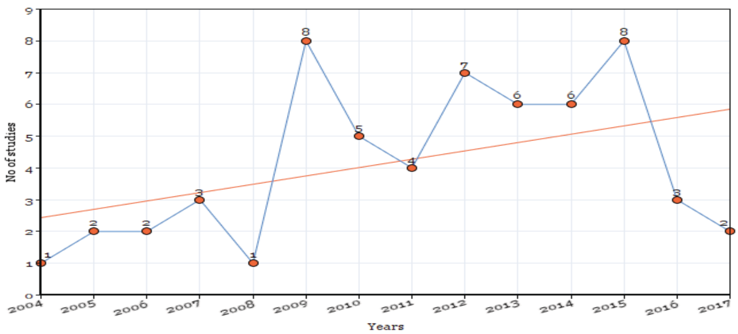


Figure 1: Year wise distribution of GBL evaluation studies (R1.1)

Most the selected studies were journal papers (51.7%) followed by conference papers (44.8%), and then book chapters (3.4%). Most of the research in GBL is published by Elsevier (13 studies) and IEEE (13 studies), followed by ACM (8 studies), Hindawi (5 studies), springer (3 studies), IJSE (3 studies), IGI Global (2 studies) and remaining 11 were from different resources (1 study per resource).

Regarding the research topics being addressed (R1.2), the selected research articles were categorized into five categories: (1) *evaluation approach (32.8%)*: studies presenting some GBL evaluation approach including framework, models, guidelines etc., (2) *development focus approach (5.2%)*: articles presenting an approach with focus on GBL development but can also be used for evaluation of GBL, (3) *design focus approach (25.9%)*: studies presenting any design model, guidelines etc. that can be also used for evaluation, (4) *review Studies*

(17.2%): review articles in GBL evaluation, and (5) *educational game evaluation (18.9%)*: this category includes all the articles that present case studies, empirical evaluation or any type of educational game evaluation studies.

The year-wise distribution of research topics/purpose for the selected relevant studies is presented in Figure 2. According to the results, there is a gradual increase in no of studies for all research topics, with 2009 to 2015 being the peak years of research. Although almost all research topics span over the years except review studies that were not seen in earlier years but have been trending from 2013 onwards with 10 studies in last five years. Moreover, the figure also shows that in the earlier years the design focus approach was the target research topic whereas in the latter years this trend has shifted towards evaluation approaches with 11 studies on this research topic from 2012 to 2016 whereas, only 4 in design focus approach.

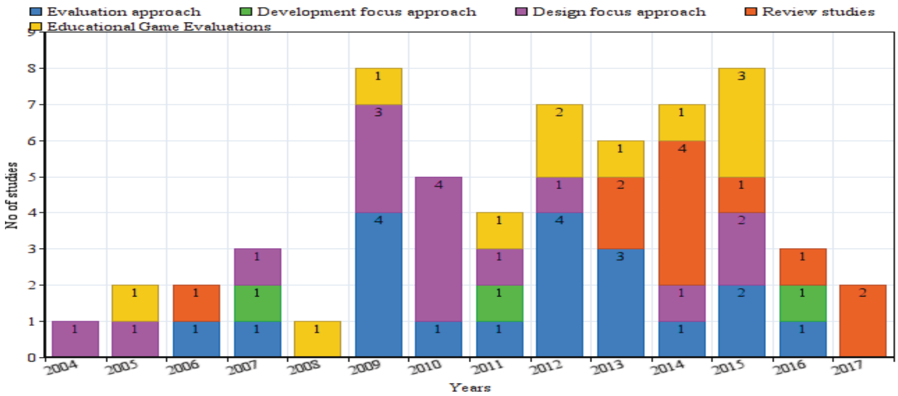


Figure 2: Year wise distribution of research topics

The results on who is leading GBL evaluation research (R1.3) is presented as country-wise distribution of studies, no of citations, and highly cited studies. The country-wise distribution of studies is presented in Figure 3, the results showed that major contribution of GBL evaluation research comes from Malaysia (8 studies).

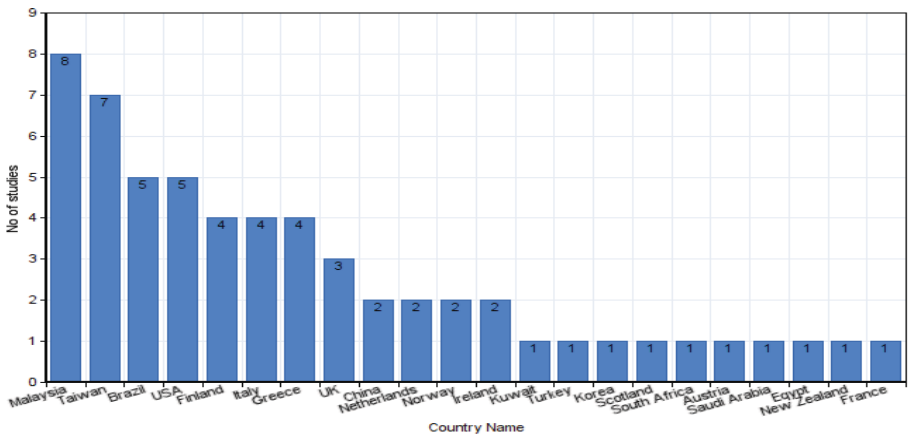


Figure 3: Country wise distribution of studies

Google Scholar was used for the citation counts for article as it indexes and finds more cited references. The results for citation counts are presented in Figure 4. The 10 papers with most citations were further analyzed for research topics, authors names and number of citation, and the results are shown in Table 3.

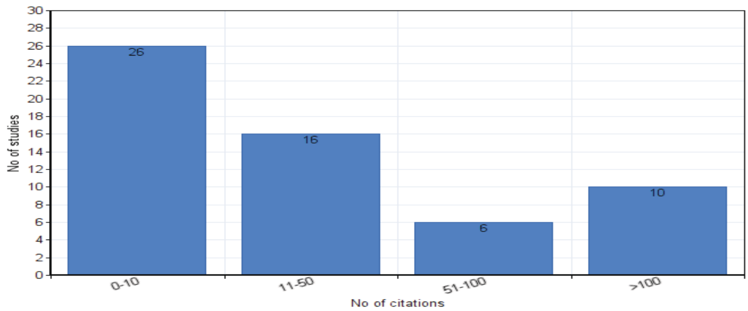


Figure 4: Distribution of studies by no of citation

5.2 R2: What are the evaluation practices in GBL or educational games?

To answer this question, we first wanted to find which evaluation approaches exist for GBL (R2.1). After thorough analysis of the 58 selected studies, 19 evaluation approaches were identified for the evaluation of GBL or educational games

Table 3: Highly cited articles

Author	Research Topic	Country	Citation
(Kiili, K., 2005)	Design Approach	Finland	1070
(Papastergiou, M., 2009)	Educational Game Evaluation	Greece	1043
(De Freitas, S. and Oliver, M., 2006)	Evaluation Approach	UK	623
(Fu, F.L., Su, R.C. and Yu, S.C., 2009.)	Evaluation Approach	Taiwan	376
(Amory, A., 2007)	Development Approach	South Africa	253
(Annetta, L.A., 2010)	Design Approach	USA	238
(Bellotti et al,2013)	Review Study	Itlay	237
(Virvou, M. and Katsionis, G., 2008)	Educational Game Evaluation	Greece	189
(Wouters, P. and Van Oostendorp, H., 2013)	Review Study	Netherland	108
(Mitgutsch, K. and Alvarado, N., 2012)	Evaluation Approach	USA	103

Out of 19 approaches, 8 presents a framework, 4 presents heuristics/guidelines, 2 presents a model, 2 presents a scale, 1 presents a method, 1 presents a standard, and another one presents evaluation constructs. Only 3 approaches (De Freitas and Oliver, 2006), (Su, Chen, and Fan, 2013) and (Mitgutsch and Alvarado, 2012) focus on overall evaluation of GBL. All the other approaches deal with the evaluation of either one or two specific dimensions for evaluating GBL e.g. flow framework for flow dimension, EGameFlow for user enjoyment, framework of UX etc. Table 4 presents the evaluation approaches and their description.

Table 4: Evaluation approaches in GBL

Evaluation Approaches	Description
Four-dimensional framework	Evaluate the potential of using games and simulation based learning in their practice (De Freitas and Oliver, 2006)
Flow Framework	Describes the dimensions of flow experience that can be used to analyze overall quality of playing experience (Kiili et al.,2014)
Playability Heuristic for Educational Games (PHEG)	Heuristic Evaluation for finding usability problems in educational computer games, (Mohamed, Yusoff, and Jaafar, 2012)
Evaluation framework for effective GBL	GBL evaluation with focus on pedagogical perspective. (Connolly, Stansfield and Hainey, 2009)
EGameFlow	Assess user enjoyment of e-learning games (Fu, Su, and Yu, 2009)
Evaluation framework of UX	Evaluate of user experience for adaptive digital educational games (DEGs). (Law and Sun, 2012)
Evaluation Framework for GBL	Guide GBL evaluation from learning perspective. (Wang, Liu, Lin and Xiang, 2011)
Game scale to evaluate Educational games	Evaluate quality in educational computer games in terms of learning and enjoyment characteristics. (Ak, 2012)
Quality Evaluation Standard	Identify quality evaluation elements of educational serious games both technical and non-technical elements. (Yoon and Park, 2013)
Heuristic Evaluation for Educational Games(HEEG)	Heuristic for evaluating educational games in terms of usability and game experience. (Marcelo, Andreza and Igor, 2015)
Heuristics Evaluation Strategy	Evaluate specifically for mGBL. (Zaibon and Shiratuddin, 2010)

Evaluation Approaches	Description
Game-based learning evaluation model(GEM)	Measure the effectiveness of serious games in a practical way. (Oprins et al., 2015)
Guidelines for evaluating games	Identify promising games for teaching computer science based on topics taught, easy to install, engaging, time to use. (Gibson and Bell, 2013)
Evaluation framework for assessing games	Focused on quality aspect in selecting and assessing learning games. (Dondi and Moretti, 2007)
Framework for serious game design evaluation	Evaluate the effectiveness of evaluation model and provide design criteria for multimedia game design educators. (Su, Chen and Fan, 2013)
Quality evaluation model	ISO quality model for mobile games. (Alhuhud and Altamimi, 2016)
Usability evaluation constructs	Present six evaluation constructs for usability evaluation for history educational game design. (Yue and Zin, 2009)
Methodology for interface evaluation	Heuristics based usability evaluation that describe usability factors to evaluate interface of educational games. (Omar and Jaafar, 2009)
serious game design assessment framework	Identified six essential components of the formal conceptual structure underlying a serious game. (Mitgutsch and Alvarado, 2012)

Further, we looked into what criteria has been used for evaluation of educational games (R2.2). The selected papers were classified into three categories for criteria used for evaluation of educational games: (1) *evaluation approach* (framework/model etc.), (2) *predefined criteria* (ad hoc), and (3) *not specified*. Most of the studies did not use any well-defined existing framework or model to conduct the evaluation. From the total of 11 studies on educational game evaluation; 72.7% (8 studies) used some predefined criteria (ad hoc). Most of the studies just outlined the dimensions (goals) of evaluation without explicitly defining the basis for selection or the factors and measures used for evaluation except one study (de Lima, de Lima Salgado and Freire, 2015) that stated the use of game experience questionnaire and intrinsic motivation inventory (IMI) for predefined dimensions of user experience and intrinsic motivation. On the other hand, 27.3% of the studies (3 studies) used some existing evaluation approaches not specific for GBL. Such as Nielsen's heuristics was used by (Mei, Ku and Chen, 2015), and USE scale (Lund, 2001) by (Tseloudi and Tsiatsos, 2015) for evaluating usability. Flow (Csikszentmihalyi, 1992), and a taxonomy of intrinsic motivations for learning (Malone and Lepper, 1987) was used by (Pourabdollahian, Taisch and Kerga, 2012) for measuring engagement. Only one study (Tseloudi and Tsiatsos, 2015) used EGameFlow scale for measuring enjoyment that is a scale developed specifically for educational games.

5.3 R3. What are the different dimensions for evaluating GBL or educational games?

To identify the different dimensions for evaluating educational games, we analyzed the selected studies for the goals of GBL evaluation. The analysis highlighted two critical issues: *first*, there is a wide diversity of elements considered for GBL evaluation and are defined inconsistently across studies; and *second*, the terms such as evaluation dimensions, factors, sub factors and metrics/measures are themselves defined inconsistently across studies and therefore not allowing the proper categorization of these elements and identification of a clear pattern. For example, some studies consider feedback in a broader scope as a dimension (goal) for evaluation whereas other studies use feedback as a factor to achieve a goal (usability). There is no distinction between macro and micro level elements. Therefore, we take the first step towards making this distinction by defining the terms of use. For this research work the term "dimension" is used in a broader scope referring to elements essential for educational game; the main goals/aim of GBL evaluation. "Factors/sub factors" are the elements considered important for achieving a dimension and "metrics/measures" is the gauge to assess that factor. In terms of scope this can be shown as: Dimension > factors> sub factors> metrics/measures. Therefore, the first step is to identify the dimensions. =A total of 37 dimensions were identified in the analysis of 58 studies. Learning is the most widely used dimension (19 studies) followed by usability (12 studies) and game factors (10 studies) including game design, game story and game mechanics. Only one study (Alfadhli and Alsumait, 2015) presents GBL design guidelines that focus on children requirements. The identified dimensions are shown in Table 5 along with the number of studies using them. Dimensions with same frequency (no of studies) are listed in a single row.

Table 5: Dimensions for GBL evaluation

Evaluation dimensions	No
Learning/Pedagogical	28
Usability	12
Game factors (design, story, mechanics)	10

Evaluation dimensions	No
User experience (UX), Motivation	5
Enjoyment, Flow, Engagement	4
Playability	3
Gameplay, Cognitive load, Instructional design, Immersion, Challenges/increased complexity	2
Child requirements, Likeability, feedback, understandability, relevance, interactivity, embedding, transfer, adaptation, naturalization, identity, informed teaching, fidelity, context, learner specification, mode of representation, technical verification, social collaboration, emotional, instructional support, collaborative learning, acceptability, usefulness, Learning Content	1

6. Conclusion

The previous review studies on GBL provided insights on the design of GBL and focused either on exploring either one or two dimensions of GBL or on the evaluation process and methodology (research design, instruments used, data collection and analysis etc.). The existing research fail to provide the state of the art in GBL evaluation. This research fills this gap by exploring the issue from directions such as the trends in GBL evaluation (amount of research in GBL evaluation, research topics, highly cited articles), current practices in GBL evaluation (approaches and criteria), and the dimensions for evaluating GBL.

The main findings of this paper includes: (1) an increasing trend in GBL research within past few years with most studies from 2009 to 2015 (2) Elsevier and IEEE are the two major resources for GBL evaluation research with more journal papers, (3) the research topic/ purpose of most studies focus on an evaluation approach followed by design focused approaches for GBL evaluation, (4) the review studies for GBL evaluation increased over the past few years and there is also a shift in research topics from design to evaluation, (5) most studies focused only on one or two dimensions of GBL and very few focused on overall evaluation specifying all the dimension essential for GBL evaluation, highlighting the need for a comprehensive evaluation framework, (6) current evaluation approaches in GBL does not cater children needs, only one out of all the reviewed studies(design focused approach) considered children requirements, (7) majority of the studies for educational game evaluations do not use existing GBL evaluation frameworks instead they mostly employ pre-defined criteria(ad hoc) for evaluation or few use general guidelines/approaches, and (8) evaluation dimensions, factors/sub factors and metrics are defined inconsistently across the studies and a wide diversity of elements are considered for GBL evaluation, however most extensively used dimension in GBL evaluation are learning, usability and game factors.

For future work, research can be extended to discuss the factors/sub factors that need to be evaluated for each GBL dimensions, why they are important, interrelation of factors and further exploring the metrics for quantifying these factors/sub factors. In sum, this study can help supplement connections with previous studies and forms an important reference base for future research in GBL evaluation.

References

A. Lund, "Measuring usability with the USE questionnaire," STC Usability SIG Newsletter, vol. 8 (2), 2001.

Abdul Jabbar, A.I. and Felicia, P., 2015. Gameplay engagement and learning in game-based learning: A systematic review. Review of Educational Research, 85(4), pp.740-779.

Ak, O., 2012. A game scale to evaluate educational computer games. Procedia-Social and Behavioral Sciences, 46, pp.2477-2481.

Aleven, V., Myers, E., Easterday, M. and Ogan, A., 2010, April. Toward a framework for the analysis and design of educational games. In Digital Game and Intelligent Toy Enhanced Learning (DIGITEL), 2010 Third IEEE International Conference on (pp. 69-76). IEEE

Alfadhli, S. and Alsumait, A., 2015, December. Game-Based Learning Guidelines: Designing for Learning and Fun. In Computational Science and Computational Intelligence (CSCI), 2015 (pp. 595-600). IEEE.

Alhuhud, G. and Altamimi, W., 2016. Quality Evaluation of Mobile Game: Miftah Alfasaha. Mobile Information Systems, 2016.

Amory, A., 2007. Game object model version II: a theoretical framework for educational game development. Educational Technology Research and Development, 55(1), pp.51-77.

Annetta, L.A., 2010. The "I's" have it: A framework for serious educational game design. Review of General Psychology, 14(2), p.105.

Arttu et al, 2017.Flow experience in game based learning – a systematic literature review. International Journal of Serious Games. 2017, 4(1), pp.57-72

Brereton, P., Kitchenham, B.A., Budgen, D., Turner, M. and Khalil, M., 2007. Lessons from applying the systematic literature review process within the software engineering domain. Journal of systems and software, 80(4), pp.571-583.

Catalano, C.E., Luccini, A.M. and Mortara, M., 2014. Best practices for an effective design and evaluation of serious games. International Journal of Serious Games, Volume 1, Issue 1.

Chorianopoulos, K. and Giannakos, M.N., 2014. Design principles for serious video games in mathematics education: from theory to practice. The International Journal of Serious Games, 1(3), pp.51-59.

Connolly, T., Stansfield, M. and Hainey, T., 2009. Towards the development of a games-based learning evaluation framework. Games-based learning advancements for multisensory human computer interfaces: Techniques and effective practices. Hershey PA: IGI Global.

Csikszentmihalyi, M. and Csikszentmihalyi, I.S. eds., 1992. Optimal experience: Psychological studies of flow in consciousness. Cambridge university press.

De Freitas, S. and Oliver, M., 2006. How can exploratory learning with games and simulations within the curriculum be most effectively evaluated? Computers & education, 46(3), pp.249-264.

De Lima, Luis Gustavo Rotoly, André de Lima Salgado, and André Pimenta Freire. "Evaluation of the user experience and intrinsic motivation with educational and mainstream digital games." Proceedings of the Latin American Conference on Human Computer Interaction. ACM, 2015.

Djelil, F., Sanchez, E., Albouy-Kissi, B., Lavest, J.M. and Albouy-Kissi, A., 2014, October. Towards a Learning Game Evaluation Methodology in a Training Context: A Literature Review. In European Conference on Games Based Learning (Vol. 2, p. 676). Academic Conferences International Limited

Dondi, C. and Moretti, M., 2007. A methodological proposal for learning games selection and quality assessment. British Journal of Educational Technology, 38(3), pp.502-512

Eagle, M., 2009, April. Level up: a frame work for the design and evaluation of educational games. In Proceedings of the 4th International Conference on Foundations of Digital Games (pp. 339-341). ACM.

El-Sattar, A. and Hussein, H.K., 2016. Learning Islamic Principles with Serious Games. In Proceedings of 13th International Conference on Advances in Computer Entertainment Technology (p. 30). ACM.

Fu, F.L., Su, R.C. and Yu, S.C., 2009. EGameFlow: A scale to measure learners' enjoyment of e-learning games. Computers & Education, 52(1), pp.101-112.

Giannakos, M.N., 2013. Enjoy and learn with educational games: Examining factors affecting learning performance. Computers & Education, 68, pp.429-439.

Ibrahim, R. and Jaafar, A., 2009, August. Educational games (EG) design framework: combination of game design, pedagogy and content modeling. In Electrical Engineering and Informatics, 2009. ICEEI'09. International Conference on (Vol. 1, pp. 293-298). IEEE.

Kiili, K., 2005. Digital game-based learning: Towards an experiential gaming model. The Internet and higher education, 8(1), pp.13-24

Kiili, K.et al., 2014. Flow framework for analyzing the quality of educational games. Entertainment Computing, 5(4), pp.367-377

Kitchenham et al,2009. Systematic literature reviews in software engineering–a systematic literature review. Information and software technology, 51(1), pp.7-15.

Law, E.L.C. and Sun, X., 2012. Evaluating user experience of adaptive digital educational games with Activity Theory. International Journal of Human-Computer Studies, 70(7), pp.478-497.

Liao, Y.H. and Shen, C.Y., 2012, March. Heuristic evaluation of digital game based learning: a case study. In Digital Game and Intelligent Toy Enhanced Learning (DIGITEL), 2012 IEEE (pp. 192-196). IEEE.

Malone, T.W. and Lepper, M.R., 1987. Making learning fun: A taxonomy of intrinsic motivations for learning. *Aptitude, learning, and instruction*, *3*(1987), pp.223-253.

Marcelo, B., Andreza B. and Igor, de., 2015. HEEG: Heuristic Evaluation for Educational Games. Proceedings of SBGames, 2015, XIV SBGames, pp.224-227.

Mei, S.Y., Ku, O. and Chen, S.Y., 2015, July. Interface Evaluation of a Game-Based Learning System: Experts vs. Novices. In Advanced Applied Informatics (IIAI-AAI), 2015 IIAI (pp. 313-318). IEEE.

Mitgutsch, K. and Alvarado, N., 2012. Purposeful by design? a serious game design assessment framework. In Proceedings of International Conference on the foundations of digital games (pp. 121-128). ACM.

Mohamed, H. and Jaafar, A., 2010, June. Challenges in the evaluation of educational computer games. In Information Technology (ITSim), 2010 International Symposium in (Vol. 1, pp. 1-6). IEEE.

Mohamed, H. and Jaafar, A., 2010, November. Development and potential analysis of heuristic evaluation for educational computer game (PHEG). In Computer Sciences and Convergence Information Technology (ICCIT), 2010 5th International Conference on (pp. 222-227). IEEE.

Mohamed, H., Yusoff, R. and Jaafar, A., 2012, March. Quantitive analysis in a heuristic evaluation for usability of educational computer game (UsaECG). In Information Retrieval & Knowledge Management (CAMP), 2012 International Conference on (pp. 187-192). IEEE.

Omar, H.M. and Jaafar, A., 2009, November. Conceptual framework for a heuristic based methodology for interface evaluation of educational games. In Proceedings of the 2009 International Conference on Computer Technology and Development-Volume 01 (pp. 594-598). IEEE Computer Society.

Oprins, E., Visschedijk, G., Roozeboom, M.B., Dankbaar, M., Trooster, W. and Schuit, S.C., 2015. The game-based learning evaluation model (GEM): measuring the effectiveness of serious games using a standardised method. International journal of technology enhanced learning, 7(4), pp.326-345.
Pappa, D. and Pannese, L., 2010. Effective design and evaluation of serious games: The case of the e-VITA project. Knowledge Management, Information Systems, E-Learning, and Sustainability Research, pp.225-237.
Papastergiou, M., 2009. Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation. Computers & Education, 52(1), pp.1-12.
Petri, G. and von Wangenheim, C.G., 2016. Journal of Universal Computer Science, vol. 22, no. 7, 992-1021.
Petri, G. and von Wangenheim, C.G., 2017. How games for computing education are evaluated? A systematic literature review. Computers & Education.
Ronimus, M., Kujala, J., Tolvanen, A. and Lyytinen, H., 2014. Children's engagement during digital game-based learning of reading: The effects of time, rewards, and challenge. Computers & Education, 71, pp.237-246.
Rooney, P., 2012. A theoretical framework for serious game design: exploring pedagogy, play and fidelity and their implications for the design process. International Journal of Game-Based Learning, 2(4), pp.41-60.
Shi, Y.R. and Shih, J.L., 2015. Game factors and game-based learning design model. International Journal of Computer Games Technology, 2015, p.11.
Shiratuddin, N. and Zaibon, S.B., 2011, November. Designing user experience for mobile game-based learning. In User Science and Engineering (i-USEr), 2011 International Conference on (pp. 89-94). IEEE.
Su, C.H., Chen, K.T.K. and Fan, K.K., 2013. Rough set theory based fuzzy TOPSIS on serious game design evaluation framework. Mathematical Problems in Engineering, 2013.
Tan, P.H., Ling, S.W. and Ting, C.Y., 2007. Adaptive digital game-based learning framework. In Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts (pp. 142-146). ACM.
Thomas, S., Schott, G. and Kambouri, M., 2004. Designing for learning or designing for fun? Setting usability guidelines for mobile educational games. Learning with mobile devices: A book of papers, pp.173-181.
Tseloudi, C. and Tsiatsos, T., 2015, July. Panic in the gallery: An online educational game for art history: Design and evaluation of a matching game. In Information, Intelligence, Systems and Applications (IISA), 2015 6th International Conference on (pp. 1-6). IEEE.
Virvou, M. and Katsionis, G., 2008. On the usability and likeability of virtual reality games for education: The case of VR-ENGAGE. Computers & Education, 50(1), pp.154-178.
Wallner, G. and Kriglstein, S., 2011. Design and evaluation of the educational game DOGeometry: a case study. 8th International Conference on Advances in Computer Entertainment Technology (p. 14). ACM
Wang, Y., Liu, X., Lin, X. and Xiang, G.,2015, An Evaluation Framework for Game-Based Learning. 15th Global Chinese Conference on Computers in Education.
Wouters, P. and Van Oostendorp, H., 2013. A meta-analytic review of the role of instructional support in game-based learning. Computers & Education, 60(1), pp.412-425.
Yoon, S.J. and Park, H.S., 2013. Development of a Quality Evaluation Standard for Educational Serious Games. Journal of information and communication convergence engineering, 11(2), pp.103-111.
Yu, K.C., Hsiao, H.S. and Tsai, F.H., 2005, June. The implementation and evaluation of educational online gaming system. In Information Technology: Research and Education, 2005. ITRE 2005. 3rd International Conference on (pp. 338-342). IEEE.
Yu, S.C., Fu, F.L. and Su, C.H., 2009. EGameDesign: guidelines for enjoyment and knowledge enhancement. In International Conference on Hybrid Learning and Education (pp. 35-44). Springer Berlin Heidelberg.
Yue, W.S. and Zin, N.A.M., 2009. Usability evaluation for history educational games. 2nd international Conference on interaction Sciences: information Technology, Culture and Human (pp. 1019-1025). ACM.
Zhang, H.F., Fan, X.Y. and Xing, H.F., 2010, October. Research on the design and evaluation of educational games based on the RETAIN model. In Knowledge Acquisition and Modeling (KAM), 2010 3rd International Symposium on (pp. 375-378). IEEE.

PAPER 2

P2:
Codifying Game-Based Learning: Development and Application of LEAGUÊ Framework for Learning Games

Rabail Tahir and Alf Inge Wang

Electronic Journal of e-Learning (EJEL)

Codifying Game-Based Learning: Development and Application of LEAGUÊ Framework for Learning Games

Rabail Tahir and Alf Inge Wang
Norwegian University of Science and Technology, Trondheim, Norway
rabail.tahir@ntnu.no
alf.inge.wang@ntnu.no
DOI: 10.34190/EJEL.20.18.1.006

Abstract Educational games are now seen as effective learning tools. However, there is a gap in the literature regarding the core dimensions of Game-based learning (GBL) for comprehensive design, analysis, and evaluation due to inconsistent use of elements. The literature on GBL reports an extensive diversity of elements used for the design and evaluation of GBL without any categorization of micro and macro-level elements. Hardly any studies systematically decompose these aspects to derivate factors/sub-factors, obstructing identification of any clear pattern. The problem is not the scarcity of GBL research but inconsistency in terminology, scope, definition, and usage of elements leading to the absence of a holistic view of GBL for effective design and evaluation. This study bridges the gap by outlining terminology and scope with four conceptual levels and then systematically categorizing GBL elements by scope, definition, and usage. The methodology used is directed content analysis of GBL literature collected through a previous systematic literature review. Dimensionalization of GBL and further decomposition into factor/sub-factors based on theoretical constructs, has resulted in a consistent and clear pattern delineating the structure of the educational game design, analysis, and evaluation. Further codifying metrics and mapping the relationship among GBL dimensions deduce into a conceptual framework (called LEAGUÊ) that facilitates greater insight into the process of learning with educational games, where to focus and what to evaluate. The LEAGUÊ framework can be applied for the analysis, design and evaluation of learning games. The framework is put in practice by utilizing the framework components (dimensions, factors/sub-factors and relations) to develop three items: 1) an analysis instrument, 2)a card-based ideation and design toolkit, and 3) an evaluation guide that can assist educational game designers, researchers, educators and evaluators for the analysis, design and evaluation of learning games respectively. This paper exemplifies the application of framework for the analysis of learning games using the analysis instrument (developed using LEAGUÊ framework) with one case study as an example.

Keywords: game-based learning, educational games, learning games, framework, dimensions, factors, relations, game analysis, evaluation, design

1. Introduction

Game-based learning (GBL) is an innovative educational paradigm that utilizes games as a mode for transferring learning (Tan, Ling et al., 2007). Educational games are considered to have the potential of deeply engaging learners with any topic, allowing active participation in the learning process (Wallner and Kriglstein, 2011). Although much effort is put in game-based learning studies internationally, robust and comprehensive design for effective learning games remains unclear. Game designers create exciting games but neglect the quality of teaching materials in a game. On the other hand, educators focus on educational materials but do not know about how to create exciting games (Shi, Y.-R. and Shih J.-L., 2015). Bellotti, Kapralos et al. (2013) stated that educational games, like any educational tool, must be able to show that necessary learning has occurred. Therefore, it is crucial to systematically evaluate them to affirm their impact (Marciano, Miranda et al., 2014). The diverse characteristics of learning games make their evaluation a difficult task. However, evaluation is the only means to verify that educational goals are achieved and spot any functional vulnerability (Djelil, Sanchez et al., 2014). Dondi and Moretti (2007) highlighted two critical issues related to GBL evaluation.

First, the construction of a general framework is extremely difficult unless it is an abstract one. *Second*, the differences between the processes of analytical evaluation (looking at aspects individually based on identification of single aspects) and global or holistic evaluation (considering the characteristics of GBL product all together). There are some difficulties with both approaches. The analytical approach lacks a theoretical model essential for unifying all the different aspects of GBL evaluation and might cause unnecessary fragmentation. Whereas in holistic approach, the judgement is too comprehensive and there is a high degree of subjectivity which presents a need to use other complementary methods for validation.

Many researchers have attempted to describe what the critical elements are to create a learning game. The review of these theories draws only one conclusion: There is no consensus among researchers about the

Reference this paper: Tahir, R., and Inge Wang, A., 2019. Codifying Game-Based Learning: Development and Application of LEAGUÊ Framework for Learning Games. *The Electronic Journal of e-Learning,* 18(1), pp. 69-87, available online at www.ejel.org

terminology and the comparable importance of GBL elements (Oprins, Visschedijk et al., 2015). Furthermore, it has been a constant challenge to understand the relationships between the different aspects embedded in GBL (Ahmad, Rahim et al. 2015). Our previous review study on GBL highlighted the following problem areas: *1) Most GBL frameworks and studies focus on exploring any single aspect of GBL, making it difficult to identify all core dimensions; 2) Use of a wide diversity of elements for designing and evaluating educational games does not allow the identification of any clear pattern; 3) Very few studies systematically decompose GBL aspects based on their theoretical construct, not allowing the hierarchical decomposition in terms of scope; and 4) The inconsistency in definition, usage, scope and terminology (e.g. dimensions, factors, etc.) of elements in GBL literature*. Therefore, to systematically analyze GBL concepts, there is a need for proper categorization of the wide variety of elements available in the literature (Petri and von Wangenheim, 2017).

This study attempts to overcome the identified problems by performing directed content analysis on the dataset of existing GBL literature collected through a systematic literature review (Tahir and Wang, 2017). The GBL elements extracted from the systematic review are hieratically decomposed (using operations such as coding, categorization, abstraction, comparison and integration) into core dimensions, factors and sub-factors based on scope, frequency of occurrence, relationship between codes, underlying meaning across codes, and mapping to existing theoretical frameworks and constructs defined by researchers in the domain of GBL. The metrics and relations between core dimensions are also detailed (using a similar process) for a complete analysis. The result is a conceptual framework named LEAGUÊ (Learning, Environment, Affective-cognitive reactions, Game factors, Usability, UsÊr) that list the core GBL elements in a hierarchy of scope. The comprehensive hierarchal structure of the framework makes the application of the framework suitable for analysis, design, and evaluation of learning games. Therefore, LEAGUÊ framework is put in practice by utilizing the framework components (dimensions, factors/sub-factors, metrics and relations) to develop three tools: an analysis instrument, a card-based ideation and design toolkit, and an evaluation guide to assist the GBL community. However, this paper will only focus on application of the framework for the analysis of learning games using the developed analysis instrument. The paper is organized as follows: Section 2 discusses related work, Section 3 describes the method for development of the framework, Section 4 presents the results of directed content analysis in the form of LEAGUÊ conceptual framework, Section 5 illustrates the application of the framework and, finally Section 6 concludes the study with discussion and future research.

2. Related work

The multidimensionality of GBL demands to consider several aspects important for its design and evaluation (Furió, D. et al., 2013). However, there is still a debate around which aspects to consider (Oprins, Visschedijk et al., 2015).

2.1 Systematic reviews and evaluation studies on educational games

Many review studies in GBL (Perttula, Kiili et al., 2017; Djelil, Sanchez et al., 2014; Petri and von Wangenheim, 2017; Tahir and Wang, 2017) have reported the use of a wide diversity of evaluation aspects for educational games. These aspects are inconsistently defined, and most studies do not systematically decompose into their constituents (Oprins, Visschedijk et al., 2015; Petri and von Wangenheim, 2016). For example, some studies consider the concept interactivity as one of the main dimensions of GBL (Annetta, 2010), while other studies use interactivity in a narrow scope as a factor to achieve a GBL dimension (Djelil, Sanchez et al., 2014). Moreover, others use it as a sub-factor of a factor interface (Omar and Jaafar, 2010). There is no clear distinction between micro- and macro-level elements.

Our insight into evaluation studies showed a similar trend where most researchers used predefined ad-hoc criteria, selecting different aspects for evaluating educational games. Moreover, existing GBL models/frameworks are less used in empirical research (Tahir and Wang, 2017). Virvou and Katsionis (2008) evaluated usability and likeability in the VR-ENGAGE game for education. Pourabdollahian, Taisch et al. (2012) employed flow dimensions for measuring learner engagement in serious games for manufacturing education.

Papastergiou (2009) focused on evaluating learning effectiveness and motivation of GBL in computer science education. Giannakos (2013) and Yu, Hsiao et al. (2005) used learning performance as a measure in their evaluation studies. The aspect of usability has also been widely used for evaluating educational games, but different studies used different factors to access this aspect (Liao and Shen, 2012; Mei, Ku et al., 2015; Wallner and Kriglstein, 2011). De Lima, de Lima Salgado et al. (2015) evaluated user experience and motivation in educational games.

2.2 Game-based learning concepts in existing frameworks

Although several GBL design and evaluation models/frameworks exist, it is essential to highlight that each of these models/theories focuses on analyzing and understanding educational games using different aspects, where most researchers focus only on one or two specific aspect(s). Thus they could supplement one another, but individually these studies are relatively narrow and may account only for a portion of a complete picture of GBL design and evaluation (Fu, Su et al., 2009; Tan, Ling et al., 2007). Here are some main aspects explored in various GBL frameworks/models:

Learning*:* Most of the researchers in GBL mainly focus on education/learning aspects. Four-dimensional framework by (De Freitas and Oliver, 2006) focuses on learning to help tutors evaluate the potential of employing simulation/GBL in practice. Connolly, Stansfield et al. (2009) describe an evaluation framework that focuses on the pedagogical aspect, introducing attributes to measure the GBL environment with attention on the learner and learning. Another evaluation framework proposed by (Wang, Liu et al., 2015) also emphasizes learning perspective with respect to learning results, learner motivations, and learner experience.

Flow: Conversely, some researchers focus on flow and enjoyment aspects in educational games. Kiili (2005) introduced an experiential gaming model to facilitate flow experience serving as a link between game design and educational theory, but not offering a complete game design. EGameFlow proposed by (Fu, Su et al., 2009) is a scale for assessing the level of enjoyment delivered by e-learning games. Kiili, Lainema et al. (2014) presented a flow framework to analyze overall playing experience of educational games through dimensions of flow.

Game design*:* Serious game design assessment framework implemented by (Mitgutsch and Alvarado, 2012) structures different game design elements to analyze the formal conceptual design of serious games. It recommends on how to shape serious games assessment in terms of design. Chorianopoulos and Giannakos (2014) presented the design principles for serious games in mathematics. Shi and Shih (2015) also focused on game design aspects proposing 11 game factors for GBL design that described a thinking process to design and evaluate educational games using game elements.

Usability: Some researchers focus on usability (Mohamed, Yusoff et al., 2012; Rêgo and de Medeiros, 2015; Omar and Jaafar, 2010) and present heuristics for evaluating the usability of educational games. These researchers incorporated concepts of learning, gameplay, interface, and enjoyment within heuristics for evaluating GBL usability. Yue and Zin (2009) proposed six usability evaluation constructs for the design of history educational games.

Pedagogy and game design: Some researchers have a combined focus on learning and game design as two critical aspects of educational game design and evaluation. Some of the frameworks include educational games design framework by Ibrahim and Jaafar (2009), a framework for the analysis and design of educational games by Aleven, Myers et al. (2010), an adaptive digital GBL framework proposed by Tan P.-H. et al. (2007), a RETAIN model presented by Zhang, Fan et al. (2010), a GBL evaluation model (GEM) by Oprins, Visschedijk et al. (2015), and a Game object model (GOM) proposed by Amory (2007). Rooney (2012) presented a framework consisting of play, fidelity, and pedagogy for serious game design.

2.3 Content analysis

Qualitative content analysis is a data reduction and sense-making effort that requires data samples to comprise of purposively selected texts that can inform research objectives and attempts to identify meanings and core consistencies (Hsieh and Shannon, 2005; Loh, Sheng et al., 2015). Three different approaches exist for content analysis: Conventional, Directed, and Summative. In the *directed approach*, analysis begins with relevant research findings/theory as guidance for the initial codes, and the goal is either to validate or conceptually extend a theory/ framework. Depending on the research question, it has two strategies to begin coding. If the aim is to identify and categorize all possible instances of any specific phenomenon, then it might be useful first to read and highlight the text representing the instances of that phenomenon and then start coding. The second strategy immediately begins coding with predetermined codes (Zhang and Wildemuth, 2005).

3. Development of the framework

This study applied directed content analysis based on the work of (Hsieh and Shannon, 2005). The general objective of this study is both to validate and conceptually extend the existing research on GBL design and evaluation by analyzing, interpreting, and organizing the many aspects to fill the gap in current literature regarding inconsistency in systematic categorization and use of features for GBL design and evaluation.

The content analysis was guided by the following research questions based on problem statements identified in the introduction: *RQ1* What are the core dimensions for the design and evaluation of educational games?; *RQ2* Which factors are important for achieving each of these core dimensions?; *RQ3* What are the sub-factors for assessing these factors based on theoretical constructs (if any)?; *RQ4* What metrics can be used to quantify these factors/ sub-factors for educational game evaluation?; and *RQ5* Are the GBL dimensions interrelated?

3.1 Directed content analysis

The process of content analysis followed the steps defined by (Zhang and Wildemuth, 2005). The steps included: preparing data, defining the unit of analysis, developing coding strategy, testing coding strategy, coding all data, assessing coding consistency, drawing a conclusion from coded data, and reporting method and findings. Content analysis is applied because it plays an important role for theory development. The goal of a directed approach to content analysis was to fill the gap in literature regarding consensus and inconsistency in GBL elements and extend conceptually the existing theory. The existing frameworks reviewed were instrumental in attempting to ensure that there were no omissions and aid in the categorisation process.

The directed content analysis was an iterative process involving progressing through extracted data to further analysis using the following set of operations: coding, categorization, comparison, abstraction, integration, and iteration (Spiggle, 1994; Engl and Nacke, 2013) in such a way that preceding operations shaped the following ones. The analysis was not performed linearly but moving back and forth between stages. The complete process is presented in Figure 1.

3.1.1 Corpus for analysis

The data set for directed content analysis comprised of data extracted from 58 articles on GBL evaluation literature from our previous systematic literature review (Tahir and Wang 2017). The selected articles comprised of GBL frameworks, evaluation studies, and reviews. The corpus completely focused on GBL literature and not on the integration of gaming and learning fields to be in line with (Loh, Sheng et al., 2015).

According to Loh, Sheng et al. (2015) the answer to the question "does learning plus game equals to serious games?" is 20 % Yes and 80 % No because only some measures can be commonly found in all three industries. Other than that, the measures are unlikely to transfer from one industry well into another. Therefore, the core elements that are optimal for use in design and evaluation of game-based learning must be specifically focused on GBL literature to properly assess, measure, and improve educational games. The data items extracted from selected papers include: Dimensions, factors, sub-factors, metrics, interrelated dimensions/factors/sub-factors, relation type and/or description, and definitions of dimensions/factors/sub-factors. All the information was entered into an Excel spreadsheet.

3.1.2 Defining unit of analysis

To remove the inconsistently in the terminology used in varying scope across studies, we introduced and defined four conceptual hierarchical levels concerning scope (dimensions, factors, subfactors, and metrics) for analysis of GBL components. Hierarchy is important when defining attributes for a specific application domain (Kececi and Abran 2001). The scope of terminology is defined as follows: the term "***Dimension***" refers to a broader concept but isolated within its kind and not a composition of different aspects, representing the main goals/objects of GBL. Each dimension represents one specific aspect of GBL. The term "***Factor***" refers to the elements important for achieving a specific dimension, and the term "***Sub-factor***" refers to further categorized elements that constitute that specific factor. The term "***Metrics***" is the gauge to measure a factor/sub-factor either through objective or subjective data. This can be depicted (high to low level) as: Dimension > Factors > Sub-factors > Metrics.

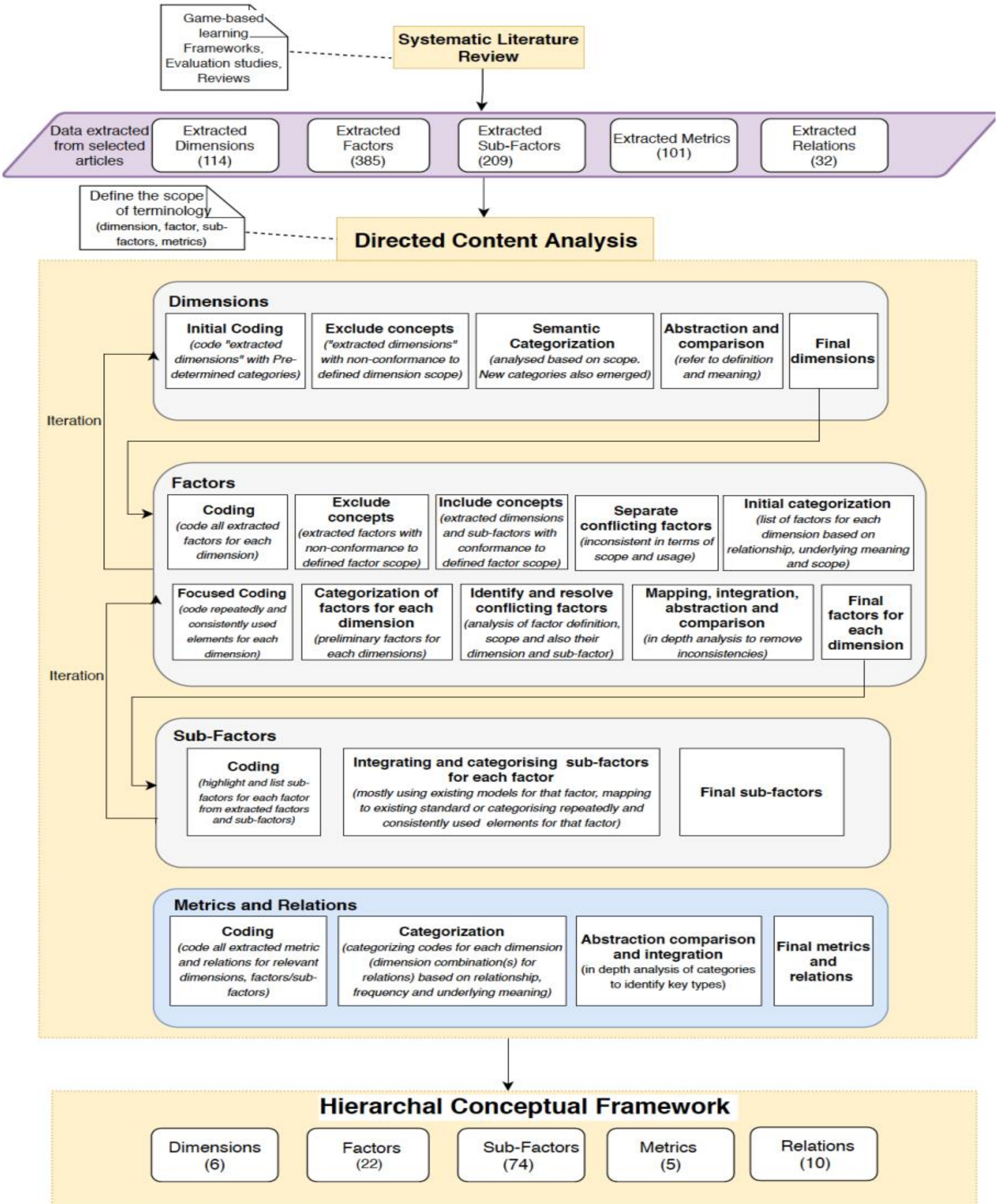


Figure 1: Directed content analysis process

3.1.3 Coding strategy

We adopted the first strategy for coding (see Section 2.3) because the aim is to identify and categorize all possible instances of GBL components in the selected corpus systematically and consistently. Therefore, before starting the analysis, we read the text and extracted the text data for each of the four conceptual levels (dimensions, factors, subfactors and metrics) in the spreadsheet that appeared to represent them on first impression (as used in each study), and then started coding for each level (top-down). The definition of these concepts that appeared in text were also extracted. As we wanted to be sure to capture all possible occurrences of GBL elements therefore first extracting all the identified text without coding might be a good way to increase trustworthiness. Hence, the analysis starts with identifying core dimensions and proceeds with factors & sub-factors (with reference to the level above). The extracted data for dimensions were coded using the predetermined codes that emerged from existing GBL literature during initial review (see section 2.2). The predetermined or initial categories used for coding dimensions were learning, game design, flow, and usability (see Section 2.2). Any dimension that could not be categorized with the initial coding scheme was given a new code. Therefore, new categories appeared during categorization and were further analyzed, using a set of operations mentioned in section 3.1, until the final core dimensions were attained. During analysis and categorization, comparison was performed to explore the similarities and differences across incidents.

Furthermore, several concrete instances were found that shared common features and therefore abstraction was used to group the previously identified categories into more general higher order conceptual classes. Abstraction was also performed on unit of data that was recognised as an empirical indicator of more general construct. After the complete analysis six core dimensions were identified (listed in section 4). The subsequent analysis focused on analyzing the sub-categories, including factors for each dimension, sub-factors for each factor, metrics, and relations using similar operations. The analysis of sub-factors mostly resulted in the integration of constructs where possible by using existing concrete theories/models (e.g. sub-factors of flow were integrated by Csikszentmihalyi's flow model) for aiding the process and enhancing the validity of final GBL components which are theoretically grounded. For metrics and relations, the extracted data was listed for relevant dimension, factor and subfactors. They were than categorized for each dimension (combination of dimension in case of relation e.g. learning and game factors, learning and affective-cognitive reactions etc.) based on frequency, underlying meaning and relationship. The categories formed were further analyzed using operations such as comparison, abstraction and integration to identify the generic metrics types for assessing any factor/sub-factor and key relations between dimensions.

The coding was checked for consistency at each level, where both authors discussed and finalized the categories formed. The analysis and findings resulted in a conceptual framework presented in the next section.

4. The LEAGUÊ conceptual framework

This section presents the results of directed content analysis in the form of a hierarchical integrated conceptual framework called LEAGUÊ (see Figure 2).

4.1 Dimensions

Figure 3 shows the six dimensions in LEAGUÊ identified as key constituents of GBL design and evaluation. The dimensions are presented in the order of letters in LEAGUÊ and not with respect to the importance of one over the other.

4.2 Factors and sub-factors

Each dimension in the framework has a set of factors. Factors are intermediate-level concepts, and the framework entails such 22 factors (see Figure 4). Factors in the framework are further systematically categorized into *sub-factors* based on their theoretical construct, allowing a hierarchical decomposition. The sub-factors are easier to quantify and also serve to reduce the subjectivity often associated with assessing the factors. However, the choice of components for evaluation should depend on the overall evaluation objective and type of data required. Sub-factors are mostly devised by integration and mapping of conceptual elements using well-developed and widely accepted models/theories in areas where researchers had consensus in the literature. Figure 2 presents the complete hierarchy and association, including sub-factors to each factor.

The sub-factors of learning objectives (L1, Figure 2) comes from (Aleven, V., et al., 2010). Learning task/activity (L2/3, Figure 2) is the specific task (designed in line with desired learning objectives and employed learning theory) that outlines the interaction of learners, using specific game characteristics, orientated at specific outcomes (El-Sattar and Hussein, 2016). The sub-factors of learning outcomes (L4, Figure 2) are adapted from the GEM model (Oprins, Visschedijk et al., 2015). The sub-factors of enjoyments (A1, Figure 2) are assimilated from EGameFlow (Fu, F.-L., et al., 2009), which uses flow as a structural foundation, and therefore has some common sub-factors as flow (Rêgo and de Medeiros, 2015;Tseloudi and Tsiatsos, 2015). The sub-factors of engagement (A2, Figure 2) are adopted from the framework by (Pourabdollahian, Taisch et al., 2012). The sub-factors of motivation (A3, Figure 2) are adapted from the well-established ARCS model (Su, Chen et al., 2013).

 ©ACPIL

The sub-factors of flow (A4, Figure 2) are adapted from the flow framework (Kiili, Lainema et al., 2014) and present the original component of flow presented by Csikszentmihalyi (Perttula, Kiili et al., 2017). The sub-factors of the interface (U1, Figure 2) are integrated by mapping the factors found in GBL literature to Nielsen's heuristics (Yue and Zin, 2009). This resulted in 9 sub-factors, one heuristic "help users recognize, diagnose, and recover from errors" could not be mapped to GBL literature. The analysis further clarified that in educational games, the focus is on error prevention and confirmation messages rather than error messages. The review of GBL literature showed the lack of psychosocial indicators used for evaluating educational games. Although the importance of psychological needs and psychosocial stages is highlighted in (Tan, Ling et al., 2007), further details are not provided.

Figure 2: LEAGUÊ hierarchal structure and components

Therefore, the psychosocial well-being indicators (Ê3/2, Figure 2) are obtained from PSWBI (Negovan, 2010). The PSWBI scale is validated with students for psychometric properties, construct validity, reliability, and internal consistency. However, its use for educational games is to be explored.

Learning/Pedagogical	
Characteristics of an educational game that promote and facilitate learning.	(De Freitas and Oliver 2006; Yoon and Park 2013; Rêgo and de Medeiros 2015; Pappa and Pannese 2010; Omar and Jaafar 2010)
Environment	
Conditions for practical use of GBL and accessibility of educational game in terms of its environment	(Zaibon and Shiratuddin 2010; Djelil, Sanchez et al. 2014; Pappa and Pannese 2010)
Affective Cognitive Reactions (ACR)	
Set of emotions, attitudes and feelings triggered by educational game.	(Rêgo and de Medeiros 2015; Zhang, Fan et al. 2010; Rooney 2012; Djelil, Sanchez et al. 2014;Annetta 2010; Pourabdollahian, Taisch et al. 2012)
Game Factors	
Elements and features of a game environment (the game world) of a educational game that stimulate the users.	(Oprins, Visschedijk et al. 2015; Yue and Zin 2009; De Freitas and Oliver 2006; Yoon and Park 2013; Aleven, Myers et al. 2010; Zaibon and Shiratuddin 2010; Tan, Ling et al. 2007; Rooney 2012; Djelil, Sanchez et al. 2014; Liao and Shen 2012)
Usability	
The extent to which an educational game is usable by the learners to accomplish specific goals (including how to learn, understand, control, and user satisfaction).	(Yue and Zin 2009; Rêgo and de Medeiros 2015; Zaibon and Shiratuddin 2010; Djelil, Sanchez et al. 2014)
User	
Characteristics of the target users/user group of an educational game to investigate if the game matches the indented users.	(Tan, Ling et al. 2007; De Freitas and Oliver 2006; Oprins, Visschedijk et al. 2015;Tan, Ling et al. 2007; Djelil, Sanchez et al. 2014)

Figure 3: Dimensions in the LEAGUÊ framework

Learning Objective: Goals and purpose for transfer of knowledge and skills by educational games to users.	(Law and Sun 2012; Alfadhli and Alsumait 2015; Aleven, Myers et al. 2010; Marciano, Miranda et al. 2014)
Learning Strategies: Pedagogical theories, learning models and approaches to achieve learning objectives.	(De Freitas and Oliver 2006; Alfadhli and Alsumait 2015; El-Sattar and Hussein 2016)
Learning Content: Educational material (facts, data and information) provided used for learning in educational games.	(Marciano, Miranda et al. 2014; Alfadhli and Alsumait 2015)
Learning Outcome: Desired learning output (student achievements) from educational games.	(Alfadhli and Alsumait 2015; Wang, Liu et al. 2015; Oprins, Visschedijk et al. 2015)
Technical: The accessibility and easiness of a user to enter the game world with the used technology and specified technical criteria.	(Zaibon and Shiratuddin 2010; Pappa and Pannese 2010)
Context: The particular context where GBL will take place using the educational game.	(De Freitas and Oliver 2006; Perttula, Kiili et al. 2017)
Enjoyment: The extent of playing the educational game is considered pleasurable for an individual user.	(Giannakos 2013)
Engagement: User activity absorption and interest in activity or task, and users' subjective acceptance of games reality combined with degree of focus on this realism.	(Oprins, Visschedijk et al. 2015)
Motivation: Level of user involvement in participating and using an educational game by devoting additional time and effort.	(Djelil, Sanchez et al. 2014; Oprins, Visschedijk et al. 2015)
Flow: State of absolute absorption in an activity which represents an optimal experience.	(Kiili, Lainema et al. 2014; Su, Chen et al. 2013; Fu, Su et al. 2009)
Game Definition: Game goals depending on appropriate set of game rules defining the gaming tasks in games.	(Fu, Su et al. 2009; Oprins, Visschedijk et al. 2015; Djelil, Sanchez et al. 2014)
Game Narrative: Describes what happens in the virtual world including character development and plot.	(Omar and Jaafar 2010; Shi and Shih 2015; Mitgutsch and Alvarado 2012)
Game Mechanics: Defines the functioning, operation and interaction with the game world.	(Yue and Zin 2009; Omar and Jaafar 2010; Mitgutsch and Alvarado 2012; Shi and Shih 2015)
Game Resources: All resources provided to the player (should have both utility and scarcity).	(Alfadhli and Alsumait 2015; Wallner and Kriglstein 2011; Tan, Ling et al. 2007)
Game Aesthetics: Audio-visual language selected, conceptualized and used (images, text, audio, video, animation and multimedia).	(Mitgutsch and Alvarado 2012; Omar and Jaafar 2010)
Game Play: Process by which user reaches the goal which defines the set of challenges of problems the user face to win the game.	(Omar and Jaafar 2010)
Interface: Component through which users interact with a game (elements utilized by user to interact with the educational game that influences its usability).	(Yue and Zin 2009; Omar and Jaafar 2010)
Learnability: Capability of an educational game to enable the user to easily learn to use specific functions.	(Yoon and Park 2013; Tseloudi and Tsiatsos 2015)
Satisfaction: Comfort and acceptability of an educational game to its users.	(Yoon and Park 2013; Tseloudi and Tsiatsos 2015; Ibrahim and Jaafar 2009)
Learner Profile: Attributes of a particular learner/ group.	(De Freitas and Oliver 2006; Djelil, Sanchez et al. 2014)
Cognitive Development: Games should suit learners' cognitive development level.	(Tan, Ling et al. 2007)
Psychological Needs: Games should suit users' psychological needs.	(Tan, Ling et al. 2007)

Figure 4: Factors in the LEAGUÊ framework

The sub-factors of technical (E1, Figure 2) include: *technology type* (technology used for GBL), *technology issues* (e.g., issue of mobility in mobile technology) and meet technical requirements (Zaibon and Shiratuddin, 2010; Pappa and Pannese, 2010). The sub-factors of context (E2, Figure 2) are adopted from the framework by (De Freitas, S. and M. Oliver, 2006). The sub-factors not directly integrated by using existing theories/models were analyzed using a set of operations mentioned in Section 3.1 and selected based on scope, frequency of occurrence, and the relationship between codes and underlying meaning across codes.

4.3 etrics

The metrics represent the lowest level in the hierarchy, which are used to collect evaluation data (Figure 2). The output of a metric interprets the status of sub-factor/factor: the degree to which the educational game possesses a given attribute. The choice of metrics depends on the type of data required, either subjective or objective, qualitative or quantitative. We identified 83 metrics from the corpus of analysis, which were then coded and categorized into five types. The complete exhaustive list of metrics for each factor/subfactor is not provided here. Instead, the aim is to give guidance on the key metrics types used in GBL evaluation that can be utilized and adapted for different evaluation studies depending on the evaluation goals and selected factors/sub-factors. As a result, GBL metrics are coded into five main categories presented in Figure 5. The first three metrics will result in objective data, while the last two will be useful for collecting subjective data. To illustrate, we introduce some typical examples for each category in Figure 6.

Metrics	References
Scores	(Wang, Liu et al. 2015;Petri and von Wangenheim 2017; Papastergiou 2009; Tseloudi and Tsiatsos 2015; Yue and Zin 2009; Giannakos 2013; Yu, Hsiao et al. 2005)
Time	(Connolly, Stansfield et al. 2009; Eagle 2009; Rêgo and de Medeiros 2015; Virvou and Katsionis 2008; Ronimus, Kujala et al. 2014)
Number of occurrences	(Connolly, Stansfield et al. 2009); Mohamed, Yusoff et al. 2012; Eagle 2009; Yoon and Park 2013; Rêgo and de Medeiros 2015; Marciano, Miranda et al. 2014; Virvou and Katsionis 2008; Wallner and Kriglstein 2011)
Rating	(Mohamed, Yusoff et al. 2012; Yu, Hsiao et al. 2005; Fu, Su et al. 2009; Tseloudi and Tsiatsos 2015; Kiili, Lainema et al. 2014; Shiratuddin and Zaibon 2011; Yoon and Park 2013; Zaibon and Shiratuddin 2010; Oprins, Visschedijk et al. 2015; Marciano, Miranda et al. 2014; Shiratuddin and Zaibon 2011; de Lima, de Lima Salgado et al. 2015; Mei, Ku et al. 2015; Pourabdollahian, Taisch et al. 2012; Shi and Shih 2015; Giannakos 2013; Ronimus, Kujala et al. 2014)
Reviews/responses/opinions	(De Freitas and Oliver 2006; Connolly, Stansfield et al. 2009; Wang, Liu et al. 2015; Omar and Jaafar 2010)

Figure 5: Metrics in the LEAGUÊ framework

Metrics	Examples
Scores	Pre/post test score (Petri and von Wangenheim 2017), game performance/score, retention score (Wang, Liu et al. 2015) etc.
Time	Game session time (Rêgo and de Medeiros 2015), learning time, task completion time (Connolly, Stansfield et al. 2009), Time spent on each problem (Eagle 2009) etc.
Number of occurrences	No of usability problems, no of navigation problem (Virvou and Katsionis 2008), no of failures (Yoon and Park 2013), no of levels played(Connolly, Stansfield et al. 2009), no of overall attempts (Wallner and Kriglstein 2011), no of distraction occurrences (Virvou and Katsionis 2008) etc.
Rating	3,4,5,6,7 point Likert scale ((Kiili, Lainema et al. 2014;Shiratuddin and Zaibon 2011;Fu, Su et al. 2009; Zaibon and Shiratuddin 2010; Yu, Hsiao et al. 2005), severity ratings (Mohamed, Yusoff et al. 2012), User task ranking (Connolly, Stansfield et al. 2009)etc.
Reviews/responses /opinions	Expert review (Connolly, Stansfield et al. 2009), self-assessment (Bellotti, Kapralos et al. 2013), Mood and attitude statements (Connolly, Stansfield et al. 2009), game review based on selected (sub)factors (De Freitas and Oliver 2006), comments on positive/negative game aspects(Wang, Liu et al. 2015) etc.

Figure 6: Examples of the metrics

4.4 Relations: High-level abstraction of game-based learning

The highest abstraction of the framework is displayed in Figure 8. We identified ten key relations from directed content analysis presented in Figure 7 (see Figure 8 for the direction of relations).

The structure of GBL is depicted by the hierarchal layout presented in Figure 2. The high-level abstraction of the LEAGUÊ framework describes the internal operation of GBL and classifies the six dimensions into generic and domain-specific. Learning, Game Factors, and Affective-Cognitive Reactions are the core domain-specific dimensions that represent the GBL phenomenon and process. Environment, Usability, and Users are the generic dimensions that influence the core dimensions and are essential for any software application to be effective for its users. An educational game is a game for education purposes that imparts learning by involving learners in the learning process. Game Factors generate Affective-Cognitive Reactions that absorb users in playing the game and positively influence the Learning. The main trick for an effective GBL approach is to keep generic dimensions in line while tweaking the Learning and Game Factors dimensions to integrate, create a balance and work in accordance with each other for enhancing the Affective Cognitive Reactions in order to meet the purpose of the educational game. We have introduced a term T-relation (see Figure 8) for the association between Learning, Game Factors, and Affective-Cognitive Reactions as the core process of GBL, where the integration of game and learning enhance affective reactions (Kiili, K., 2005). The generic dimensions not only influence the GBL phenomenon (domain-specific dimensions) but are also linked with each other. Usability should address the intended users and also cater to the technical and context related specifications of the environment. The technical specification and context (environment) should also map to the learner profile and capabilities (user). The overall process of GBL is a complex phenomenon and requires a multidisciplinary approach.

Dimensions	Relation	References
Learning **&** ***Game Factors***	Integration/ Balance	(De Freitas and Oliver 2006; Aleven, Myers et al. 2010;Zhang, Fan et al. 2010; Wallner and Kriglstein 2011; El-Sattar and Hussein 2016; Pappa and Pannese 2010)
ACR **&** ***Learning***	Positive effect	(Kiili, Lainema et al. 2014; Thomas, Schott et al. 2004; Oprins, Visschedijk et al. 2015);Perttula, Kiili et al. 2017; Bellotti, Kapralos et al. 2013; Yu, Fu et al. 2009; Tseloudi and Tsiatsos 2015; Giannakos 2013)
Game Factors & ACR	Generate	(Rooney 2012;Abdul Jabbar and Felicia 2015; Ronimus, Kujala et al. 2014)
(Integration of Game Factors and Learning) **&** ***ACR***	Enhance/facilitate /increase	(Kiili 2005; Zhang, Fan et al. 2010; Rooney 2012)
Usability **&** ***(Learning, Game factors, ACR)***	Influence	(Gibson and Bell 2013; Shi and Shih 2015)
User **&** ***(Learning, Game factors, ACR)***	Influence	(De Freitas and Oliver 2006; Abdul Jabbar and Felicia 2015;Mei, Ku et al. 2015;Yu, Hsiao et al. 2005;Mitgutsch and Alvarado 2012; Shi and Shih 2015)
Environment **&** ***(Learning, Game factors, ACR)***	Influence	(De Freitas and Oliver 2006; Oprins, Visschedijk et al. 2015; Dondi and Moretti 2007; Virvou and Katsionis 2008)
Usability **&** ***Environment***	Address/Cater	(Virvou and Katsionis 2008)
Usability **&** ***User***	Address/ Cater	(Virvou and Katsionis 2008; Mei, Ku et al. 2015)
User **&** ***Environment***	Map	(De Freitas and Oliver 2006)

Figure 7: Relations in LEAGUÊ

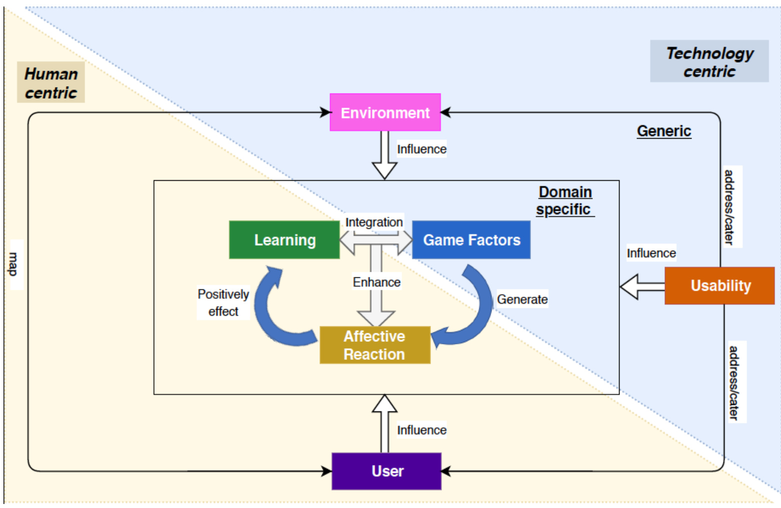


Figure 8: High-level abstraction of LEAGUÊ

There is another viewpoint to the LEAGUÊ framework, which divides it into two views: technology-centric, and human-centric. The technology-centric view includes three dimensions (Game Factors, Usability, and Environment) related to technological aspects of game-based learning and the human-centric view also includes three dimensions (Learning, Affective-Cognitive Reaction, and User) related to human aspects (cognitive, behavioral, identity) of GBL. The idea here is to model the technology-centric dimensions in such a way that they facilitate human-centric dimensions.

5. Application of the LEAGUÊ framework

The framework organizes the GBL dimensions regarded in the literature as significant in producing an effective learning game. For each dimension, a set of configurable factors, sub-factors, and the basic vocabulary is provided to facilitate the application and use of framework components in multiple ways.

The framework can be applied in the process of design, analysis, and evaluation of game-based learning. The dimensions in GBL (depending on the evaluation objective) might be considered in isolation (picking and selecting components) for a specific evaluation study. However, GBL dimensions are linked to each other in terms of cause and effect and can be viewed as a collective whole to understand the process and help in design and analysis. The LEAGUÊ framework is put in practice by utilizing the framework components (dimensions, factors/sub-factors, metrics and relations) to develop three items: LEAGUÊ analysis instrument, LEAGUÊ ideation and design toolkit, and LEAGUÊ evaluation guide for analysis, design and evaluation of learning games respectively. Using each of the three items is divided into a set of stages or activities to scaffold the process. This paper will primarily focus on applying the LEAGUÊ framework for the analysis of learning games using the developed analysis instrument.

5.1 Analysis of learning games using LEAGUÊ analysis instrument

This section presents how the LEAGUÊ framework is applied for the analysis of learning games using the LEAGUÊ analysis instrument. The instrument is aligned with the framework in terms of the elements required to specify a learning game. An example study is also presented which demonstrated its use. It can be useful to understand the potential of educational games in a specified environment by knowing the strengths and areas in which it could improve. The learning game is analyzed based on whether it contains the elements deemed important for game-based learning by splitting the subject matter into its basic and fundamental components (dimensions specified in the LEAGUÊ framework, see figure 3).

The analysis instrument consists of three parts: a primary analysis form, a secondary form to reinforce in-depth analysis, and a reflection form. The primary analysis form (see Figure 9) breaks the complex GBL paradigm into smaller parts based on framework components (factors in the LEAGUÊ framework, see figure 4) to gain a better understanding of how learning takes place in the learning game. Each question in the form regarding all six dimensions is answered for the game being analyzed. The factors laid out in the primary form may help analyze the learning game in terms of individual elements essential for an effective game-based learning approach and highlight any areas that are weak or neglected in the game. At the end of the form there are two rows to list the overall strength and weakness of the game with respect to each dimension by critically analyzing the given answers.

The secondary form (see Figure 10) supports an in-depth analysis of each element of primary form by splitting it further into simpler questions (based on sub-factors in LEAGUÊ framework, see figure 2) to deeply analyze its constituents and facilitate the thinking process to construct the answer for each question in the primary form. The secondary form can be used on its own or can also be used only as a guide to give concrete answers in primary form. Building on this analysis, the relationship between individual components will be considered using the reflection part. The refection questions (see Table 1) determine the effectiveness of the learning game as a collective whole by analyzing the relations between individual components (see figure 7 and 8 for relations in LEAGUÊ framework). This is giving the opportunity to reflect on the design choices made in the game and if they are effective or need improvement. For example, if the game is appropriate for target users in terms of usability, learning content, strategy or game elements used; if the game can be used to support different context and informal or formal learning considerations; and if the game has a balance and harmony in learning and game factors used etc. Similarly, for reflection form each of the seven reflection questions are answered for the game being analyzed and revisions or improvements needed in the game are listed in the last row based on the reflections made. Therefore, the steps of the framework application for analyzing a learning game using LEAGUÊ analysis instrument consists of the following steps in the specified order: 1) description of the different aspects of the game by answering the questions in the primary form (Figure 9), using secondary form (Figure 10) for in-depth analysis and detailed description (if needed), 2) identification of strengths and weaknesses of the game by analyzing the answers, 3) reflection on the relationships between core elements of the game (using Table 1) and 4) critically analyzing the reflections made to highlight the refinements or improvements needed in the game to make it effective.

Learning	Environment	Affective-Cognitive Reactions	Game Factors	Usability	UsÊr
L1. What are the learning objectives of the game?	**E1**. What technical aspects are required for the game to work and best support learning?	**A1**. How can the game provide enjoyment to the users?	**G1**. What are the game objectives to integrate learning?	**U1**. How interface of the game is made easy to use for target users?	**Ê1**. What are the attributes of the target users of the game?
L2. Which learning strategies are being used to enable learning through the game?	**E2**. What is the context for playing the game for learning?	**A2**. How can the game engage the users?	**G2**. What narrative is used to make game compelling and integrate learning?	**U2**. How does the game provide easy learnability to its target users?	**Ê2**. Which cognitive needs (of target users) are considered in the game?
L3. What learning content is being used in the game for target users?		**A3**. How can the game motivate the users?	**G3**. What mechanics are used to make game compelling and support learning?	**U3**. How does the game provide satisfaction to its target users?	**Ê3**. Which Psychological needs (of target users) are considered in the game?
L4. What learning outcome(s) can be acquired from the game?		**A4**. How can the game generate flow?	**G4**. What resources are provided to the users to function effectively that also support learning?		
			G5. What aesthetics are used to make game compelling for target users?		
			G6. What game play is used to make game compelling for target users and support learning?		
Strength of the analyzed game					
Weakness of the analyzed game					

Figure 9: Primary analysis form

The instrument can be used to support an analysis process undertaken by any stakeholder (designers/developers, researchers, and intermediates like teachers/parents etc.) to ensure that they take into account the key issues and essential factors associated with game-based learning to support practice. The analysis of learning games using the three forms provides the analytical as well as holistic picture of whether the learning game is an effective GBL approach and not just the tool itself. This analysis will include all aspects worth considering, from the right content and strategy to appropriate game elements and software for target users in order to apply the game within the specified context. To illustrate the application of the framework for the analysis of learning games, the instrument is used to analyze an empathy game as an example.

5.1.1 Case: Empathy game

The illustrated game is an online empathy game for primary school children (8-14 years old), which can be played in school or at home with friends. The game is about making stories for different characters using personality traits to develop a strategy for achieving a goal. One of the players plays the role of the selected character, and the other two players help the character to complete the tasks assigned according to personality traits and goals, which creates the challenge factor. The game uses this activity to teach empathy and enable players to develop the ability to understand and share the feelings of others. The game is suitable for interactive and collaborative learning.

Learning	Environment	Active-Cognitive Reactions	Game Factors	Usability	UsÊr
(Prior knowledge): What knowledge or skills do players need to have before starting the game? **(Learning and retention)**: What knowledge or skills game is reasonably expected to teach players? **(Potential transfer)**: What knowledge and skills might game teach players that go beyond what they actually encountered in the game? **Learning style:** Which learning styles need to be incorporated in the game? **Learning theory:** Which pedagogic models/learning theory is most effective for attaining learning objectives? **Learning task/activity:** What learning activity/task learners perform in the game to achieve learning objectives? **Learning material:** What learning material need to be incorporated in the game? **Instructional support:** How much instructional support is provided to support learning objectives? **Difficulty levels:** How many levels of difficulty are provided to players to attain learning objectives? **Knowledge/skill/attitude:** What knowledge, skills or attitudes are acquired by the learners by playing this game? **Competencies:** Does learner gain any competencies by playing this game? **Performance:** Does playing this game enhance performance of learners?	**Technology Type:** Which type of technology is used in the game to best support learning? (E.g. mobile, VR, AR, Kinect, etc.) **Technology related issues:** Are there any issues related to the used technology that need special attention or will affect learning? **Technical requirement:** Does the game has any specific technical requirements that must be meet? (E.g. functional req., Portability & conformance to standards, Structure & Organization, Technical quality, Information produced etc.) **Place:** Does the game need to be played on some specific place and how does game facilitate that? E.g. home, school, combination of several. **Setting:** Does the game require any specific settings to be played and does that affect learning? (E.g. access to equipment, technical support, additional resources etc.)	Which elements are used in the game to induce enjoyment? •Immersion •Social interaction •Challenges •Goal clarity •Feedback •Concentration •Control •Knowledge improvement Which elements are used in the game to induce engagement? •Immersion •Challenges •Control •Purpose •Interest Which elements are used in the game to induce motivation? •Attention •Relevance •Confidence •Learner satisfaction Which elements are used in the game to generate flow? •Challenges •Clear clarity •Feedback •Playability •Control •Rewarding experience •Concentration •Loss of self-consciousness •Time distortion	**Game Goals:** What are the goals (that match with the learning objectives) of the game that players need to achieve? **Game rules:** What are the appropriate set of rules/constraints the players must follow to reach the goals of the game? **Game tasks:** What are the tasks (defined by the set of rules and goals) players has to complete in the game? **Player characters:** Does game use characters to tell narrative? **Storyline:** Does game need a storyline to effectively impart learning? **Fantasy:** Does game use fantasy to involve players? **Game Interactions:** How do players interact with the game? **Game Controls:** How do players control the game? **Game tutorial or help:** Does game provide tutorial or help to players if they need support? **Rewards or sources:** Does game provide rewards or other suitable resources to the players (in accordance with game objectives) to function in the game world? **Game customizability:** Does game provide any customizability options important for the players? **Multimedia elements:** What multimedia elements are used to support learning and involve players? E.g. text, audio, images, animations, video, interactive content etc. **Game visualization:** What level of visualization is used to support learning and involve players? E.g. 2D, 3D etc. **Challenges:** Does the game provide suitable challenges to target players? **Strategy:** Does game require players to make a strategy to achieve the game goals? **Pace & adequate levels:** Does game provide different levels with appropriate difficulty and pace? **Game Feedback:** Does game provide appropriate feedback for in-game activity?	What elements are used to make the interface easy to use? •Provide Interface Feedback •Use of metaphors and objects •Provide user control and settings •Provide consistency •Provide error presentation •Provide easy navigation •Provide adaptability and accessibility •Provide attractive screen design •Provide Help and support U2. How does this game provide easy learnability to its target players? U3. How does this game provide satisfaction to its target players?	**Bio-demographics:** What are the bio-demographics of the players that will mostly play this game? (E.g. age, gender, ethnicity, economic background etc. that should be kept in mind.) **Experience:** What level of experience they have with technology and learning domain? **Personality:** What are the learning styles and preferences of target players? **Cognitive Stage:** What is the cognitive stage of players of this game? **Cognitive Load:** How does this game support the level of cognitive load suitable for them? **Psychosocial Stage:** What is the psychosocial stage of players of this game? **Psychosocial Well-being:** How does this game improve their psychosocial Well-being?

Figure 10: Secondary in-depth analysis form

Table 1: Reflection form

No.	Reflection questions
R1.	Are game factors (game objectives, narrative, mechanics etc.) and learning factors (learning objectives, strategy, content etc.) well integrated into the game?
R2.	Are selected game factors (narrative, mechanics, play etc.) effective for generating affective-cognitive reactions (engagement, enjoyment etc.) in target users of this game?
R3.	Does the usability of this game cater to the needs of the target users?
R4.	Does the usability (interface, learnability etc.) of this game cater to any specific needs of the environment (technology, context etc.) in which it will be played?
R5.	Are specific technical and context requirements (if any) for playing this game easy to manage by the target users?
R6.	Are learning factors (objective, content, strategy etc.) appropriate for the target user?
R7.	Are game elements (mechanics, narrative, play etc.) used appropriately for the target user?
Revisions/improvements needed in the analyzed game:	

Completing the analysis instrument highlights the particular challenges, strengths, and weaknesses in terms of essential GBL components needed to embed desired learning through the game into effective practice. The beta version of this game was analyzed using the LEAGUÊ analysis instrument, and the results are presented in Figure 11 for the primary form and Figure 12 for the reflection part. The secondary form was used as a guide to think more in-depth and give a concrete answer for questions in the primary form. The last part of the primary form (the strength and weakness of the analyzed game) and the reflection form (the revisions/improvements needed in the analyzed game) are not shown in the figures but instead described below.

Strength of the analyzed game: In this particular game, a notable strength is a collaborative learning approach that might be used to support the cyclical transition from storytelling towards developing empathy through discussion and reflection on actions. However, learning is not tied to any curriculum content, developing a strategy to achieve a goal using personality traits, and the outcome facilitates reflection and additionally may support creative thinking and imagination. This approach has the potential to additionality teach cooperation, listening, and improve storytelling after time as it uses social interaction between players, as well as this help to engage learners who prefer self-directed and interactive ways of learning. Another strength is the use of characters to tell the story. The game does not require significant technical support and can be easily played with a device with an internet connection. The game also uses bright colors and simple and consistent screens with straightforward controls and navigation.

Weakness of the analyzed game: On the other hand, game aesthetics lack the effective use of multimedia. The game does not provide any help or support for playing the game first time and also lacks the use of rewards or other resources to generate additional purpose to engage in storytelling or facilitate reflection through the use of props. One of the least successful aspects is that the game does not provide any instructional support to facilitate children to feel emotions or differentiate between them to generate empathy unless they self-reflect on their story. The players are not given the control to move back in the game even if they accidentally press a button or miss a task, which is a significant drawback along with no feedback is given for in-game activity. Game is mainly text-based, not much visualization in the game. The tasks are not very clear and challenging to understand at least the first time (reading through text) and could be supported with multimedia usage, e.g., audio, animation, or videos. Also, the game does not provide tasks with an increasing difficulty level.

Reflections made: The reflection tool supports a deeper reflection on the interrelation between different essential elements to apprehend whether the learning game is effective for the purpose. In this case, depending on device availability, learners may want to play the game in school as well as home context supporting formal as well as informal learning processes, reinforcing the learning outcomes. However, the game requires three players to start the game but does not support connecting with players online, so this might be a problem in a context where more children are not present. The player should know how to read as the game is mostly text-based, and also have the vocabulary to create a story as storytelling in the main activity in the game, which makes it more suitable for talkative children who can put their thoughts into words. The challenge for helping a character to achieve the goal give players confidence to verbalize their thoughts into a story without being self-consciousness as they are playing for another character and not themselves. The game uses characters and traits that resemble different types of children in the target age group, and thus provides a safe space for children to talk about situations that may be hard to discuss outside of a game environment. The complete analysis of the reflection part is presented in Figure 12.

Learning	Environment	Active-Cognitive Reactions	Game Factors	Usability	UsÊr
L1. Players by putting thoughts into words can better develop skills on empathy, be able to reflect on everyday situations of self and others, and might also learn cooperation, listening skills, imagination and storytelling by playing game. **L2**. The players learn by telling stories with group discussion and reflection. The game employs behaviorism and social constructivism based on collaborative learning approach. Players develop a story based on personality traits for achieving a life goal by putting oneself in the shoes of the character. **L3**. Prompts (goals) guiding on what to talk about and with what foundation (traits). The game doesn't provide any instructional support and difficulty levels. **L4**. Change in attitude towards others in understanding their situations and not judging based on their few traits. Might support to develop competences in cooperation, listening and improve storytelling after time.	**E1**. A tablet or PC with web browser and internet connection is required. **E2**. The game can be played anywhere (aim is to use in schools) but only collaboratively as it requires three players at a time and the game doesn't facilitate connecting with other players online. One device is required to play the game with three people.	**A1**. Game uses social interaction between players, challenge for helping character to achieve the goal and possibility of immersion in developing a story of the character. The game should also provide feedback and control over task which is major drawback. The sub-goals are not very clear and difficult to understand at least first time. **A2**. Game induces interest through character, provide control over building story and choice of goal. **A3**. Game involve attention by making players listen to each other to develop their part of story; provide relevance of characters to target players' life; gives confidence to verbalize their thoughts into a story. **A4**. Game provide control over the story progression, concentration on listing to peer's part of story, and loss of self-consciousness as by playing for another character.	**G1**. Complete a story by developing a strategy to achieve a goal. The players need to use the traits given for that character for achieving the goal and build on the story of the other players. The players choose a character and goal to work on developing a story both individually and in groups as specified for the task. However, the players always fail to achieve the goal. **G2**. The game uses characters for story telling that is supposed to generate empathy. The game does not use fantasy elements, but players can create a fantasy world around character using their imagination. **G3**. Use simple clicking for selecting a character, goal and traits. Main game mechanic is social interaction that uses high level of interactivity by playing in team. Game is controlled with a continue button, but no back button is provided to go back and change strategy in story. **G4**. There is no game tutorial or help to understand gameplay. No resources (other than the given traits) or rewards are provided. The players can customize their character only by choosing a name. G5. Game is largely text-based, cartoonish images for characters and colorful background. Not much visualization in the game largely roleplay driven rather than immersive 2D/3D interface. **G6**. The only challenge is to build a story around the given traits and goals of selected character. Players need to make a strategy for achieving goal and change it when they fail. Game does provide tasks with increasing difficulty level, it may be easy or hard depending on the matching of the traits and goal. No feedback is given for in-game activity.	**U1**. Provide consistency in UI, easy navigation, adaptability from tablet to computer screen, attractive and playful color scheme and image design, give player the control on when to move forward but they cannot go back if needed. No feedback is provided for interface action accept disabling button. A lot of text on screen. No help/support. **U2**. All screens and activity are similar every time except the content, after first play they are likely to feel comfortable in the next round. **U3**. Use of bright colors and cartoonish characters. Includes variety of characters and traits that resemble different types of children.	**Ê1**. Children aged 8-14, all genders, ethnicities and backgrounds. No experience required as interactions are simple. Player should know to read and have vocabulary to create a story as game is largely text-based and storytelling. Potentially would support all types of personalities but is more fit for talkative children. **Ê2**. The game is made to be rather short, so it wouldn't get boring. **Ê3**. Empathy development is important for target users. Game provides a safe space for children to share stories and interact with each other by having some shared tasks. Gives confidence to talk about situations that may be hard to discuss outside of a game environment.

Figure 11: Analysis instrument for empathy game: primary form

Revisions/improvements needed in the analyzed game: From the analysis, the game designer can identify the need for a better fit between the use of game elements to facilitate the learning outcomes and to emphasize on creating greater challenges for the children and more importantly support increased reflection upon empathy learning through instructional support and feedback and thereby offer improved opportunities to work in a team.

No.	Reflection questions
R1	A good integration is the use of personality traits as a game rule to make a strategy. The connection between the game task and reflection upon learning process was the least successful aspect. No use of instructional support for game task and no use of game elements to facilitate reflection, feel empathy for the character, justify their actions or visualize different personality traits. The game does not provide any feedback to facilitate empathy or trigger feelings, it only depends on the storyteller. Induce the fantasy world created by children's story by making characters react to story and using them in story progression as animation to help children trigger emotions and feel empathy. The players always fail to achieve the goal as a game objective, which initially might help to reflect on traits and feel emotion as sadness for characters' failure but later will confuse them to differentiate between using different traits as they always fail and may end up generating negative emotions only.
R2	Use of characters and social interaction is a major hit otherwise the game is a bit weak in terms of game factors to generate affective reactions. Should give players more control to go back if they misunderstood the task or want to elaborate on their story. Game is monotonous in terms of difficulty level and same result every time. The only challenge is the combination of traits and goal to tell the story that might become boring after players have tried a couple of combinations. No customizability options or use of rewards/other resources to generate interest. No in-game feedback. Multimedia usage is limited, using immersive animation, videos or 3D visualization can increase enjoyment and engagement.
R3	The game is designed to be rather short, so it wouldn't get boring as intension span for children is short. The use of cartoonish characters is in accordance with age group. However, too much text on screen makes screen design less attractive for children. It is difficult to read all text and understand the task and role of each player at least first time. Considering age group, the interface should provide the control to move back in game, if children skip one task because they misunderstood and move forward they are likely to miss the whole game as tasks are interlinked.
R4	The interface of the game is consistent and provides adaptability from tablet to PC. The game needs to be played with three users and therefore provides easy navigation but there should be some feedback and metaphors in screen design (in addition to text) to guide which player is leading.
R5	Game does not require any significant technical support and is therefore appropriate for target age group. The classroom context is more appropriate to play this game as there are other students to play with. In other context the only issue is to have other children to play with as game requires three players. The game should therefore facilitate connecting with friends online who wants to play as the game cannot be played alone.
R6	Learning objective that is empathy development is important for target children. Use of collaborative learning approach provides a safe space for children to share stories and interact with each other by having some shared tasks. It is also appropriate as some children are not talkative and cannot verbalize their thoughts, so teamwork will help and give them confidence to talk about situations that may be hard to discuss outside of a game environment. Overall simple content is used in the game, but some traits are difficult to understand for children age 8-14. Instructional support for generating empathy is missing as children in this age group might need a feedback or some guidance to develop empathy based on their story or to reflect on it but the game does not facilitate children in feeling the emotions or prompts to trigger emotions.
R7	Children can relate to different characters and traits used in the game as they target this age group. Using role playing is appropriate as even with very basic or no interactive image children can pretend to be someone/something else very easily. The use of multimedia is very less for children and no help or support is provided to understand task. Text should be supported with use of animations in the form of tutorial or interactive content. Individual game tasks might be difficult for some children who find it difficult to verbalize their thoughts or are less imaginative and may require assistance.

Figure 12: Analysis instrument for empathy game: reflection form

6. Discussion and conclusion

This study expands on GBL design and evaluation literature to overcome the shortcomings in current research (problem areas highlighted in the introduction) by conducting a directed content analysis. The results of the analysis are translated into a conceptual hierarchal framework LEAGUÊ, which shows that the multidimensionality of GBL requires evaluation of several aspects referred to as core dimensions (RQ1), including Learning, Game Factors, Affective-Cognitive Reactions, Usability, User and Environment. Each dimension focuses on certain factors and sub-factors that constitute that aspect, and metrics are required to assess them. The framework presents 22 factors (addressing RQ2), 74 sub-factors (RQ3), and five metrics categories (RQ4). The dimensions of GBL are related to each other, and it is essential to assess the relations presented as a high abstraction of LEAGUÊ for more significant insights into educational games (RQ5). The framework provides a detailed picture of GBL that will guide not only researchers and evaluators but also designers and developers of educational games. The proposed framework is built on components grounded in theory. Each component has a strong basis for formation that is supported by theoretical constructs in GBL literature and not merely based on suspicion.

Most of the existing GBL frameworks focus on some specific elements, which make them difficult to use in practice when the target genre differs from default game genres used in research (Shi and Shih 2015) or when the objective is design and analysis of complete GBL experience and not just focusing on few individual aspects of it. Thus, the existing models and frameworks could supplement one another, but individually these are relatively narrow and focus on a portion of the complete picture of GBL design and evaluation (such as Learning Mechanics-Game Mechanics (LM-GM) model can help identify the learning and game mechanics to draw the LM-GM map for a game but neglect the other elements that equally account of an effective learning game). The specificities of LEAGUÊ in relation to other frameworks can be highlighted by the comprehensiveness of the framework (detailing the individual parts in order to allow analysis in terms of presence/absence as well as overall picture and interconnection between the core dimensions) that can be equally useful for analytical and holistic evaluation providing a theoretical model essential for unifying all the different aspects of GBL and thereby solving the two critical issues related to GBL design and evaluation highlighted by Dondi and Moretti (2007) (see introduction). Furthermore, dimensions presented in LEAGUÊ are higher-level concepts and not restricted by the game genre.

The LEAGUÊ framework is put into practice by developing three tools (an analysis instrument; an ideation and design toolkit; and an evaluation guide) based on the framework components (dimensions, factors/sub-factors, metrics and relations) to support GBL practitioners and researchers. Therefore, the LEAGUÊ framework can be applied for analysis, design and evaluation of learning games using the three developed items; LEAGUÊ analysis instrument, LEAGUÊ ideation and design toolkit, and LEAGUÊ evaluation guide respectively. This paper described how the framework could be applied for the analysis of learning games using the analysis instrument, exemplified with a case study of an empathy game. The LEAGUÊ analysis instrument can be used by the stakeholders of educational games, including designers, researchers, and intermediates like teachers/parents, to ascertain its effectiveness. Educational game designers/developers can use it to analyze the educational game (both early stage game prototypes or later alpha/beta versions of the game) to identify the loop-holes and make improvements in the design. It is often recommended to carry out analysis early because it is easy to make changes and improvements at an early stage of development as they get more expensive later, and it is useful to carry out analysis before actual evaluation. Educational game researchers are interested in gaining insights regarding the game's effectiveness and suitability in different domains with respect to its designated purpose and application context. Therefore, they can use this instrument to learn more about the different elements used in the games and the relationships between them and gain experience from both successful and failed game concepts in order to improve in designing effective learning games and critically question the effects and consequences games may have on target users, especially in the case of vulnerable groups. Teachers/parents need to be convinced of the positive effect of game-based learning because otherwise, they will choose not to use them. Therefore, they can use the analysis instrument for analyzing the learning game to assess the potential and develop trust and conviction for justification to use the game as an efficient tool or not.

The proposed framework is also employed in workshop sessions for ideation and design of learning games for various learning domains using the LEAGUÊ ideation and design toolkit, and to conduct evaluation studies using the LEAGUÊ evaluation guide. The complete process of using the framework for design and evaluation of learning games along with the results will be presented in another paper, and framework components will be further validated and developed. Future research will focus on automating or partially automating GBL evaluation using the proposed framework and game data logs. The future work will also focus on developing a web-based ideation and evaluation tool that will facilitate the educational game design and evaluation process during different phases of the development lifecycle and help professional and game companies working with game-based learning.

References

Ahmad, M., et al. (2015). Towards an Effective Modelling and Development of Educational Games with Subject-Matter: A Multi-Domain Framework. IT Convergence and Security (ICITCS), 2015 5th International Conference on, IEEE.

Aleven, V., et al. (2010). Toward a framework for the analysis and design of educational games. Digital Game and Intelligent Toy Enhanced Learning (DIGITEL), 2010 Third IEEE International Conference on, IEEE.
Amory, A. (2007). "Game object model version II: a theoretical framework for educational game development." Educational Technology Research and Development 55(1): 51-77.
Annetta, L. A. (2010). "The "I's" have it: A framework for serious educational game design." Review of General Psychology 14(2): 105.
Bellotti, F., et al. (2013). "Assessment in and of serious games: an overview." Advances in Human-Computer Interaction 2013: 1.
Chorianopoulos, K. and M. N. Giannakos (2014). "Design principles for serious video games in mathematics education: from theory to practice." International Journal of Serious Game 1(3): 51-59.
Connolly, T., et al. (2009). "Towards the development of a games-based learning evaluation framework." Games-based learning advancements for multisensory human computer interfaces: Techniques and effective practices. Hershey PA: IGI Global.
De Freitas, S. and M. Oliver (2006). "How can exploratory learning with games and simulations within the curriculum be most effectively evaluated?" Computers & Education 46(3): 249-264.
De Lima, L. G. R., et al. (2015). Evaluation of the user experience and intrinsic motivation with educational and mainstream digital games. Proceedings of the Latin American Conference on Human Computer Interaction, ACM.
Djelil, F., et al. (2014). Towards a learning game evaluation methodology in a training context: A literature review. European Conference on Games Based Learning, Academic Conferences International Limited.
Dondi, C. and M. Moretti (2007). "A methodological proposal for learning games selection and quality assessment." British Journal of Educational Technology 38(3): 502-512.
El-Sattar, A. and H. K. Hussein (2016). Learning Islamic Principles with Serious Games. Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology, ACM.
Engl, S. and L. E. Nacke (2013). "Contextual influences on mobile player experience–A game user experience model." Entertainment Computing 4(1): 83-91.
Fu, F.-L., et al. (2009). "EGameFlow: A scale to measure learners' enjoyment of e-learning games." Computers & Education 52(1): 101-11.
Giannakos, M. N. (2013). "Enjoy and learn with educational games: Examining factors affecting learning performance." Computers & Education 68: 429-439.
Hsieh, H.-F. and S. E. Shannon (2005). "Three approaches to qualitative content analysis." Qualitative health research 15(9): 1277-1288.
Ibrahim, R. and A. Jaafar (2009). Educational games (EG) design framework: combination of game design, pedagogy and content modeling. Electrical Engineering and Informatics, 2009. ICEEI'09. International Conference on, IEEE.
Kececi, N. and A. Abran (2001). "Analyzing, measuring & assessing software quality within a logic-based graphical framework." Universite du Quebec a Montreal.
Kiili, K. (2005). "Digital game-based learning: Towards an experiential gaming model." The Internet and higher education 8(1): 13-24.
Kiili, K., et al. (2014). "Flow framework for analyzing the quality of educational games." Entertainment Computing 5(4): 367-377.
Liao, Y. H. and C.-Y. Shen (2012). Heuristic evaluation of digital game based learning: a case study. Digital Game and Intelligent Toy Enhanced Learning (DIGITEL), 2012 IEEE Fourth International Conference on, IEEE.
Loh, C. S., et al. (2015). Serious games analytics: Theoretical framework. Serious games analytics, Springer: 3-29.
Marciano, J. N., et al. (2014). "Evaluating multiple aspects of educational computer games: literature review and case study." International Journal of Computer Games Technology 2014: 14.
Mei, S.-Y., et al. (2015). Interface Evaluation of a Game-Based Learning System: Experts vs. Novices. Advanced Applied Informatics (IIAI-AAI), 2015 IIAI 4th International Congress on, IEEE.
Mitgutsch, K. and N. Alvarado (2012). Purposeful by design?: a serious game design assessment framework. Proceedings of the International Conference on the foundations of digital games, ACM.
Mohamed, H., et al. (2012). Quantitive analysis in a heuristic evaluation for usability of educational computer game (UsaECG). Information Retrieval & Knowledge Management (CAMP), 2012 International Conference on, IEEE.
Negovan, V. (2010). "Dimensions of students' psychosocial well-being and their measurement: Validation of a students' Psychosocial Well Being Inventory." Europe's Journal of Psychology 6(2): 85-104.
Omar, H. M. and A. Jaafar (2010). "Conceptual framework for a heuristics based methodology for interface evaluation of educational games." Computer and Information Science 3(2): 211.
Oprins, E., et al. (2015). "The game-based learning evaluation model (GEM): measuring the effectiveness of serious games using a standardised method." International Journal of Technology Enhanced Learning 7(4): 326-345.
Papastergiou, M. (2009). "Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation." Computers & Education 52(1): 1-12.
Pappa, D. and L. Pannese (2010). "Effective design and evaluation of serious games: The case of the e-VITA project." Knowledge Management, Information Systems, E-Learning, and Sustainability Research: 225-237.
Perttula, A., et al. (2017). "Flow experience in game based learning–a systematic literature review." International Journal of Serious Games 4(1).

Petri, G. and C. G. von Wangenheim (2016). "How to Evaluate Educational Games: a Systematic." Journal of Universal Computer Science 22(7): 992-1021.
Petri, G. and C. G. von Wangenheim (2017). "How games for computing education are evaluated? A systematic literature review." Computers & Education 107: 68-90.
Pourabdollahian, B., et al. (2012). "Serious games in manufacturing education: Evaluation of learners' engagement." Procedia Computer Science 15: 256-265.
Rêgo, M. B. B. A. B. and I. de Medeiros (2015). HEEG: Heuristic Evaluation for Educational Games. Proceedings of SBGames 2015.
Rooney, P. (2012). "A theoretical framework for serious game design: exploring pedagogy, play and fidelity and their implications for the design process." International Journal of Game-Based Learning (IJGBL), 2(4), 41-60.
Shi, Y.-R. and J.-L. Shih (2015). "Game factors and game-based learning design model." International Journal of Computer Games Technology 2015: 11.
Spiggle, S. (1994). "Analysis and interpretation of qualitative data in consumer research." Journal of consumer research 21(3): 491-503.
Su, C.-H., et al. (2013). "Rough set theory based fuzzy TOPSIS on serious game design evaluation framework." Mathematical Problems in Engineering 2013.
Tahir, R. and A. I. Wang (2017). State of the art in Game Based Learning: Dimensions for Evaluating Educational Games. European Conference on Games Based Learning (pp. 641-650). Academic Conferences International Limited.
Tan, P.-H., et al. (2007). Adaptive digital game-based learning framework. Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts, ACM.
Tseloudi, C. and T. Tsiatsos (2015). Panic in the gallery: An online educational game for art history: Design and evaluation of a matching game. Information, Intelligence, Systems and Applications (IISA), 2015 6th International Conference on, IEEE.
Virvou, M. and G. Katsionis (2008). "On the usability and likeability of virtual reality games for education: The case of VR-ENGAGE." Computers & Education 50(1): 154-178.
Wallner, G. and S. Kriglstein (2011). Design and evaluation of the educational game DO Geometry: a case study. Proceedings of the 8th International Conference on Advances in Computer Entertainment Technology, ACM.
Wang, Y., et al. (2015). An Evaluation Framework for Game-Based Learning. In 15th Global Chinese Conference on Computers in Education.
Yu, K.-C., et al. (2005). The implementation and evaluation of educational online gaming system. Information Technology: Research and Education, 2005. ITRE 2005. 3rd International Conference on, IEEE.
Yue, W. S. and N. A. M. Zin (2009). Usability evaluation for history educational games. Proceedings of the 2nd international Conference on interaction Sciences: information Technology, Culture and Human, ACM.
Zaibon, S. B. and N. Shiratuddin (2010). Heuristics evaluation strategy for mobile game-based learning. Wireless, Mobile and Ubiquitous Technologies in Education (WMUTE), 2010 6th IEEE International Conference on, IEEE.
Zhang, H.-f., et al. (2010). Research on the design and evaluation of educational games based on the RETAIN model. Knowledge Acquisition and Modeling (KAM), 2010 3rd International Symposium on, IEEE.
Zhang, Y. and B. M. Wildemuth (2009). "Qualitative analysis of content." Applications of social research methods to questions in information and library science, 308, 319.

PAPER 3

P3:
Evaluating Learners' Emotional States by Monitoring Brain Waves for Comparing Game-Based Learning Approach to Pen-and-Paper

Krenare Pireva, Rabail Tahir, Ali Shariq Imran, and Niraj Chaudhary

In: Proceedings of the 2019 IEEE Frontiers in Education Conference (FIE)

This paper is not included due to copyright
available at https:/doi.org/10.1109/FIE43999.2019.9097262

PAPER 4

P4:
Game-Based Digital Quiz as a Tool for Improving Students' Engagement and Learning in Online Lectures

Krenare Pireva, Rabail Tahir, Alf Inge Wang, and Ali Shariq Imran

Ready for Submission

This paper is awaiting publication and is not included in NTNU Open

PAPER 5

P5:
Insights into the Design of Educational Games: Comparative Analysis of Design Models

Rabail Tahir and Alf Inge Wang

In: Proceedings of the 2018 Future Technologies Conference (FTC)

Insights into Design of Educational Games: Comparative Analysis of Design Models

Rabail Tahir(✉) and Alf Inge Wang

Norwegian University of Science and Technology, Trondheim, Norway
rabail.tahir@ntnu.no

Abstract. The study reports on an ongoing research that intends to identify and validate the core dimensions for Game-Based-Learning (GBL) and further explore the shift in dimensional focus between different phases of educational game development life cycle: pre-production (design), production (development) and post-production (testing and maintenance). Hence, this paper presents the initial work focusing on design phase by presenting a comparative analysis of educational game design models using GBL attributes, validity and framework attributes as analytical lens. The main objective is to analyze the fundamental GBL attributes in existing design models to identify the common attributes which demonstrate their importance for design phase and highlight any need for further research in terms of attribute validation and framework improvement. This study also highlights the strengths and weakness of existing design frameworks. The results of analysis underline learning/pedagogical aspects and game factors as the most essential attributes for design phase of educational games. Comparative analysis also guides researchers/practitioners to better understand GBL through various properties of different existing design models and highlights the open problems such as lack of tool support, empirical validation, independent evaluations, adaptability and absence of concrete guidance for application to make more informed judgments.

Keywords: Educational games · Game-Based learning · Serious games
Design models · Frameworks · Comparative analysis · Design attributes

1 Introduction

Over the past decade, educational games or game-based learning systems have greatly impacted the learning industry. However, it has been a constant challenge for educational game designers to understand the different aspects embedded in game-based learning [1]. Lately, several researchers have proposed design frameworks/models/guidelines to guide educational game design [2–16]. According to Neil [17] usually all proposed design models tend to communicate some core foundational elements, yet they differ in their approach and results. As there is a lack of dialogue between researcher and practitioners and also among researchers themselves. Therefore, also at completely theoretical level, there is a lack of work providing comprehensive comparative analysis in the field [17]. To the best of our knowledge, we found only two

© Springer Nature Switzerland AG 2019
K. Arai et al. (Eds.): FTC 2018, AISC 880, pp. 1041–1061, 2019.
https://doi.org/10.1007/978-3-030-02686-8_78

such attempts of comparison studies for learning game design frameworks. Dos Santos et al. [18] presented a comparison of 5 digital learning game design methodological frameworks and highlighted their differences and similarities to identify selection criteria for guiding framework choice and promote methodological frameworks as a way to encourage principled educational games design. However, the framework selection is not explicitly stated. Likewise, Ahmad et al. [19] presented a survey of different educational design frameworks; against criteria such as well-designed games, effective video games, four learning theories and key elements of a games and analyzed them from software engineering perspective for the development of effective educational games. However, the keywords are not specifically focused on educational games. Malliarakis et al. [20], however, did not present a comparative analysis but studied existing frameworks for educational game design to document the features supported by current educational games to teach computer programming in order to establish a framework for the design of their computer programming specific educational game.

Often the underlying purpose of comparison entails valuing one model over another. However, this is not the sole focus of this study. Rather, the approach here is to analyze the existing design models/frameworks against core GBL dimensions to pinpoint elements specifically focused for the design phase based on similarities in analyzed frameworks. The GBL dimensions selected as analytical lens comes from our previous research results [33]. Although all core dimensions are considered important for an effective educational game product but dividing them in different phases might help education game designer and developers to emphasize the focus in that phase and ease the process. Further, the design frameworks are also compared in terms of validation of used dimensions and exploring framework attributes to highlight strengths and weaknesses which would aid researchers and designer in better understanding the issues in educational game design. The objectives of this study are the following:

1. Exploring game-based learning attributes used in existing design models.
2. Validation of game-based learning attributes by existing models and frameworks: Support for being theoretically grounded and empirically sound.
3. Comparison of existing GBL frameworks using analytical lens to identify open issues and highlight their strengths and weaknesses.

The paper is organized into following sections. Section 2 describes background by presenting an overview of educational game design frameworks/models, Sect. 3 describes the method, Sect. 4 illustrates the comparative analysis, Sect. 5 presents discussion and finally, Sect. 6 concludes the study with conclusion and future work.

2 An Overview of Educational Game Design Frameworks/Models

Our previous research study examined the state of the art in game-based learning by conducting a systematic literature review. The work reported in [21] highlighted the existing design focused approaches for educational games and these frameworks/models were selected for the comparative analysis described in this paper. In this section, the existing educational game design models/frameworks are presented, and their objectives are briefly described.

2.1 Level Up

The goal of Level Up [6], is to build new modes to design and evaluate the future game-based learning systems. The author hypothesized that the framework will increase the production speed of educational games, increase the quality and offer scientific evaluation of educational content of the games. According to the author Level Up framework will make use of a collection of empirical experiments as well as log-data driven analyses using empirical learning curves for understanding learning in educational games. The aim is to model learning of students and identify gaps to improve game development by using educational data mining on game-log data of students. The learning models could be dual fold: assessing the quality of learning in educational game and identifying the exact spots for applying in-game feedback (e.g. hints on more difficult problems). The author makes use of game-log data for evaluating learning in an educational game. The evaluations and logging system together are considered to provide foundation for developing design principles for an effective educational game.

2.2 Experiential Gaming Model

The experiential gaming model [8] is developed based on the idea of integrating experiential learning theory, flow theory and game design. Experiential gaming model emphasizes the importance of clear goals, providing immediate feedback, and matching challenges to skill level of players. The model comprises of an experience loop, ideation loop, and a challenge depository. The model uses the operational principle of human blood-vascular system as metaphor. The heart of the model is formed by challenges based on educational objectives. The flow theory is applied and factors contributing to flow experience are discussed in the model to enhance positive user experience and maximize educational game impact.

2.3 Framework for the Analysis and Design of Educational Games

This framework for design of educational game [2] is developed based on existing components including a method for specifying the educational objectives, principles for instructional design supported by empirical research in learning sciences and a framework for linking game dynamics, mechanics and aesthetics. The framework directs the levels which are essentials for an educational game to be effective. The framework discusses the three components: Learning objectives, MDA and Instructional principles highlighting the support they can provide to game designer by the analytical angle. The author highlights that success of educational game is more prospective when learning objectives of educational game are clearly established early in development process and if designers carefully think about linking the desired game aesthetic in game mechanics, via proper game dynamics observing the proven instructional design principles.

2.4 RETAIN Model

Zhang et al. [16] presented the RETAIN model consisting of six elements (relevance, embedding, transfer, adaptation, immersion and naturalization). The model is constructed on instructional design principles and describes the notorious concepts between instructional design and game, providing a common framework for educators and game designers by comprehending the effective integration of game and learning content to even them out.

2.5 Adaptive Digital Game-Based Learning Framework

The author [13] has identified essential components and features of best practice to be considered for the design of games-based learning environments based on existing models and frameworks. The author discusses four frameworks/models in this paper: The Design Framework for Edutainment Environment, Adopted Interaction Cycle for Games, The Engaging Multimedia Design Model for Children and Game Object Model. Based on analysis the developed framework focuses on the learners and the game design. The framework also highlights some important features such as challenge, goals, story and objectives not included as part of the framework.

2.6 A Theoretical Framework for Serious Game Design

Rooney [10] investigated a triadic theoretical framework consisting of the elements of pedagogy, play and fidelity for the design of serious games. The author points out that the inherent inconsistencies between pedagogy, game design and fidelity make it difficult to balance these elements during serious game design process and integrating them in one coherent framework. Another challenge is the multidisciplinary nature of serious game that require collaboration between members from different disciplines bringing in the conflicting interests, priorities and from diverse backgrounds can complicating the process of "balancing".

2.7 The "I's" Have It (A Framework for Serious Educational Game Design)

The framework "I's have it" for the design of serious educational games is a nested model of six elements: identity, immersion, interaction, increased complexity, informed teaching and instructional [4]. The elements of the framework are derived from studies on design and development of games from Grade 5 to graduate level. The elements are grounded in theory and research within education, instructional technology, psychology, and learning sciences. According to the framework educational games contain these six elements that come into view in the order of magnitude staring from the element identity and ending at instructional. According to the author the backbone of his work is based on the research in constructivist viewpoint which shows that people learn based on discovering prior schema and eventually building the new knowledge by connecting their new experience with prior ones.

2.8 e-VITA Framework for SGs

The framework for serious games developed as a part of e-VITA project [9] focuses on three key dimensions including technical verification, user experience and pedagogical aspects (learning outcome). The project highlights serious games as a game, an IT product, and a learning instrument. It argues that with respect to development and evaluation, an educational game should have three critical dimensions to be effective (1) it should be easy-to-use and technically sound; (2) it should be engaging and fun game; and (3) it should be an effective learning instrument providing desired learning outcomes. To improve motivation and learning, all the three dimensions should be targeted, the failure to meet any one dimension could compromise the effectiveness of serious games.

2.9 Educational Games (EG) Design Framework

The focus of Ibrahim et al. [7] was to develop an educational game design framework for higher education. This author compared few available frameworks and recommend the required criteria based on his analysis both from pedagogy and game design viewpoint. The idea behind this framework is to combine three factors that include pedagogy, game design and learning content modelling into the educational game design. The focus of game design is on multimodality and usability. As usability studies in educational games are not much focused by researchers. Similarly, the focus of pedagogical factor is learning outcomes and motivation theory. The factors of fun, problem solving, and syllabus matching are also highlighted.

2.10 Game Factors and Game-Based Learning Design Model

Shi et al. [11] underlined the fact that prior models are designed based on specific game genres making them difficult to use when target game genre is different from default game genres applied in research. Therefore, the author presents macro level design concepts comprising of 11 key factors for game-design. The factors include game goals, game fantasy, game mechanism, game value, narrative, interaction, challenges, freedom, sociality, sensation, and mystery. The author verifies the usability of the model and performance of identified factors for designing educational games by analyzing two applications.

3 Method

The methodology used in this paper is the comparative analysis of educational game design models/frameworks using appropriate analytical tools. The Quasi-formal comparison technique proposed by [22] and used by many researches [23–25] for comparative reviews is employed in this study.

The comparison of existing frameworks and models with one another is useful to get an insight into a specific area and identify the gaps for future research. Although, it is a very difficult task, but the result is often considered to have some sort of researcher bias as it is based upon the subjective judgment of the researcher. Two alternative approaches have been proposed for comparative analysis, informal and quasi-formal comparison. However, informal comparison lacks a systematic framework to direct the analysis and therefore is more likely to have a subjective bias. Quasi-formal comparison on the other hand attempts to subdue the subjective limitations by presenting a strategy and creating a baseline for comparison in the form of an analytical tool. Quasi-formal comparisons can be conducted using different techniques. One technique is to select a set of critical perspectives or attributes and then compare the objects against them and this is considered closer to a traditional scientific method [22]. This approach is adopted for conducting the quasi-formal comparison in this study. For this purpose, appropriate analytic tools are needed to make analysis and comparison. Although many researchers have proposed and used analytical tools for comparative analysis [26–29] but not all fit for the purpose and specific area of this research. The analytical lenses seen as appropriate for the research objective of this study are classified as: GBL/educational game attributes; validity and framework attributes. The GBL attributes were selected based on our earlier research study which categorized game-based learning into six fundamental dimensions using directed content analysis [33] of GBL literature selected through a systematic literature review [21]. The analytical lenses of validity and framework attributes are taken from [23, 26, 27]. These analytical lenses are described along with the references in Table 1. The research study outlines three research questions, which are as follows:

RQ1. Which GBL attributes are essential for design phase of educational game development life cycle. (comparison of attributes covered in each model/framework).

Table 1. Analytical lens for comparative analysis of existing educational game design models/frameworks

Analytical lens	Description	Reference
GBL Attributes	How many and which GBL attributes are covered by the educational game design model/framework?	[21, 33]
Learning/pedagogical	Does the model/framework consider learning/pedagogical attribute, or any elements related to it?	
Game factor	Does the model/framework consider game factor attribute, or any elements related to it?	
Affective Reactions	Does the model/framework consider affective reaction attribute, or any elements related to it?	
Usability	Does the model/framework consider usability attribute, or any elements related to it?	
User	Does the model/framework consider user attribute, or any elements related to it?	
Environment	Does the model/framework consider environment attribute, or any elements related to it?	
Validity	Does the model/framework have support for its claims?	[18, 23, 26]
Theoretical evidence (Development basis)	Is the model/framework grounded in appropriate theory? (author provide development basis for the model/framework)	

(*continued*)

Table 1. (*continued*)

Analytical lens	Description	Reference
Empirical evidence (Validation/application)	Does the model/framework have empirical support for its claims? (details of application/validation of framework/model: game name, sample size, validated elements)	
Framework attributes	What type of attributes are provided by the model/framework?	[18, 23, 27, 28]
Tool/instrument Support	Does the model/framework offer tool/instrument support for its artefacts?	
Assessment and stakeholders	What types of assessment approaches are used for the model/framework? Which groups of stakeholders are required to participate in assessment?	
Applicable Stage	What is the most appropriate educational game development lifecycle phase(s) to apply the model/framework?	
Application domain	In which application domain(s) the model is mostly applied?	
Guidance for application (abstract principles vs concrete guidance)	Does the model/framework rely only on abstract principles or it provides concrete guidance? (offer guidelines on how to practically use it for educational game design)	
Target/adaptability	Is the model/framework fit for all educational games (universal/generic) or is it situation appropriate (specific)? Does it offer adaptability in actual use?	
Strength/weakness	What are the strengths and weaknesses of the model/framework?	

RQ2. To what extent are the attributes being used in existing models validated? Are they theoretically grounded? Is empirical evidence available?
RQ3. What type of characteristics are provided by existing design models to operationalize and use them and their strengths and weaknesses?

4 Comparative Analysis

The frameworks described above aimed at establishing guidelines and patterns for designing effective educational games. A comparison of these models, highlights not only the fundamental common characteristics to be considered during GBL design phase but also highlights the distinct aspects and approaches of each framework plus bringing forward the open issues that still needs to be addressed in GBL design research. In this section, 15 existing educational design models/guidelines (including 10 models/frameworks and 5 design guidelines/principals) are compared and analyzed using the three categories of analytical lenses (GBL attributes, validity and framework attributes) described in Table 1.

4.1 Key GBL Attributes

Among the most significant comparison features is the number of key attributes a model/framework deal with [26]. Six fundamental GBL elements were selected for comparative analysis of design frameworks (see Table 2). These include learning/pedagogy, game factors, affective reactions, usability, user and environment. The reason for selecting specifically these six attributes as analytical lens is because they are identified as core dimensions of GBL in our earlier research study [33]. Therefore, the aim here is to identify if any of these six attributes should be more focused or particularly essential for the design phase of effective educational games.

Table 2. Comparative analysis of educational game design models/frameworks based on key GBL attributes

Design-focus frameworks	Learning/pedagogy	Game factor	Affective reactions	Usability	Users	Environment	Total
Game-Based Learning Guidelines [3]	**X** (learning objectives)	**X** (game req.)		X (User Interface)	X (child req.)		4
Level Up [6]	**X** (learning)						1
Experiential gaming model [8]	**X** (experiential learning)	**X** (Game design)	X (flow)				3
Usability guidelines for mobile educational games [14]				**X** (Usability)		X (Context)	2
Framework for analysis and design of educational games [2]	**X** (learning objectives, Instructional design)	**X** (MDA					2
RETAIN Model [16]	X (Relevance, Embedding, Transfer, Adaptation, Naturalization		X (immersion)				2
Adaptive Digital Game-Based Learning Framework [13]		X (Game design)			**X** (Learner)		2
A Theoretical Framework for Serious Game Design [10]	**X** (pedagogy)	**X** (fidelity)	**X** (play)				3
"I's" have it [4]	**X** (instructional)	X (identity)	X (immersion)				3

(*continued*)

Table 2. (*continued*)

Design-focus frameworks	Learning/pedagogy	Game factor	Affective reactions	Usability	Users	Environment	Total
User Experience for Mobile Game-Based Learning [12]	X (learning content)	X (game play)		X (usability)		X (mobility)	4
EGameDesign [15]	X (Knowledge enhancement)		X (Enjoyment)				2
e-VITA framework for SGs [9]	X (Pedagogical aspects)		**X** (affective aspects)	X (usability)		X (Technical verification)	4
Educational Games (EG) Design Framework [7]	**X** (pedagogy, learning content)	**X** (Game design)					2
Design principals for serious game [5]		**X** (design principal)					1
Game Factors and Game-Based Learning Design Model [11]		**X** (Game Factors)					1
Total models:15	**11**	**10**	**6**	**4**	**2**	**3**	

*Bold **X** is used when all factors of that attribute are covered by a framework and X when only some are covered.*

Learning/pedagogical entails the elements related to pedagogy and learning such as learning objective, strategy, content and outcome. Game Factors include the features of a game world that encompass every perspective of game environment (game definition, mechanics, narrative, aesthetics, resources). Affective Reactions depict the emotions and feelings stimulated during interaction with educational game such as (flow, engagement, motivation, enjoyment). Usability signifies how usable is the educational game by its users in achieving its goals (learnability, satisfaction, interface). User is the learner/player playing the educational game and their characteristics such as profile, cognitive and psychological needs. Lastly, environment describes the technical and context-related aspects of educational game. Table 2 presents the comparative analysis based on these GBL attributes.

4.2 Validity: Theoretical and Empirical Evidence

This section analyzes the design frameworks in terms of their validity, examining the theoretical and empirical support available for each framework. The theoretical validity is examined to explore the development basis and foundations of these design frameworks/models. Empirical support is required to see if the existing design models are grounded in empirical evidence or applied to any educational game. It is important to see if the existing educational game design models have strong practical orientation in real life educational game design and development using empirical studies or just present in research work. Table 3 details the models/frameworks with their development basis, empirical validation or application, educational games on which the model is applied, sample size of empirical study and the elements of model/framework validated in the study.

4.3 Framework Attributes

The existing design frameworks are also analyzed with analytical lens of framework attributes mentioned in Table 1. The comparative analysis of educational game design frameworks in terms of tool support, assessment and stakeholders, application stage, domain, guidance for application and target/adaptability is presented in Table 4. Table 5 highlights the strengths and weaknesses of each mentioned framework. For this part of analysis, we have only included the design frameworks/models and not design principals/guidelines. Therefore, a total of 10 frameworks are compared here.

The framework attributes are briefly described here: a tool support facilitates to capture the design artefacts together with evaluation outcomes, decision rationales and measurements that are invaluable assets [23]. A stakeholder is any representative or person having interest in the system [23]. A perspective of abstract versus concrete guidance allows to assess guidance for application, whether the frameworks offer any concrete guidance for their application in designing educational games or just rely on abstract rules e.g. to illustrate this “respect people” without providing any guidelines on how to perform it is an abstract principle [23]. The target of analyzed design models can be categorized as general or specific based on whether model can be used for the design of any kind of educational game and for any target audience or they focus on any specific platform, audience or game genre, providing specific guidelines for their target. Design models are used for the design process of educational games therefore, the application stage is the design phase. However, some of these models claim to be equally applicable to other stages of development lifecycle.

Table 3. Comparative analysis of educational game design models/frameworks based on validity

Model Ref	Theoretically grounded (Development basis)	Empirical validation/application	Educational game(s)	Sample size	Validated elements
[3]	Reviewed literature* (not specified)	No validation			
[6]	Intelligent tutoring system literature	Yes (empirical study)	Wu's Castle video game	61	Learning curve
[8]	Experiential learning theory, flow theory and game design	Yes [31, 32]	IT-Emperor game, Day Off	221	Flow antecedents
[14]	Interviews with educational game developers, game design theory, and game analyses	No validation			
[2]	Existing components: method for specifying educational objectives, framework for relating game's mechanics, dynamics, and aesthetics, and principles for instructional design	Yes*(case study), applied framework to analyze the game	Zombie Division	NI	
[16]	Game and instructional design principals (Keller's ARCS Model, Gagne's events principles of Bloom's scaffolding)	Yes*(case study), applied for evaluation of educational game	Knowledge Discovery	NI	Relevance, Embedding, Transfer, Adaptation, Immersion, Naturalization

(*continued*)

Table 3. (*continued*)

Model Ref	Theoretically grounded (Development basis)	Empirical validation/application	Educational game(s)	Sample size	Validated elements
[13]	Four models: Design Framework for Edutainment Environment, Adopted Interaction Cycle for Games, Engaging Multimedia Design Model for Children and GOM	No validation			
[10]	NI	No validation			
[4]	Experience of developing and testing educational games and using research from commercial video games	No validation (example only)	The Great Entomologist Escape	NI	
[12]	NI	Case study	1Malaysia	64	
[15]	Four-dimensional game-design evaluation framework and Bloom six levels of knowledge	Yes* (case study), applied to design a learning game	VIEW		
[9]	NI	Yes* (preliminary validation of game (results not provided)	e-VITA-European life experience	NI	
[7]	Compares a few frameworks: Adaptive Digital Game-Based Learning Framework, Three Layered Thinking Model, Experiential Gaming Model and Model for Educational Game Design	No validation			

(*continued*)

Table 3. (*continued*)

Model Ref	Theoretically grounded (Development basis)	Empirical validation/application	Educational game(s)	Sample size	Validated elements
[5]	Literature review of related work* (not specifically stated)	Yes *(case study), applied in 2 Math video games but no evaluation performed	Gem Game, Grandma's Garden Game	NI	
[11]	Literature search of studies whose primary concerns were game factors	Yes	Slice it, Xiao-Mao	31	All 11 factors

*NI = Not identified. * is used when it is stated but not explained, not empirical validation or when results are not provided*

5 Discussion

A comparison among existing models/frameworks clarifies the underlying common features and distinctive aspects. Mainly such comparison provides two benefits: first to help educational game designer/researchers understand and contrast the alternative approaches available for selecting an appropriate one, and second to highlight the open problems for future research. However, this study has a third key benefit of guiding educational game designers in design phase by highlighting the essential attributes for design of educational games. This study performs the comparative analysis of educational game design models/frameworks through the perspective of important GBL features that in our viewpoint could be considered as the core dimensions and are fundamental for an effective GBL product. Although all of these attributes are important for educational game development life cycle, but the view or focus may change in different phases of design, development and evaluation; leading to some attributes more important in one phase than the other. Therefore, the idea is to explore this shift and focus.

RQ1: The comparison among existing models/frameworks in terms of GBL attributes clarifies the underlying common features for design phase. 11 design models included learning attribute mostly focusing on learning objectives, learning content, instructional design, knowledge enhancement/transfer and pedagogical aspects. 10 frameworks focused on game factors with emphases on game design including factors such as goals, mechanics, dynamics, aesthetics, narrative and fidelity. However only 6 design frameworks focused on affective reactions such as experiential gaming model emphasized on flow experience, RETAIN and I's focused on immersions, EGameDesign focused on enjoyment. Although it is a common feature of digital games and considered equally important in educational games as well, but in design models it comes after learning and game factors. Usability is approached by 4 frameworks/guidelines including e-VITA, experience for mobile game-based learning, usability guidelines for mobile educational games and game-based learning guidelines.

Table 4. Comparative analysis of educational game design models/frameworks based on framework attributes

Model Ref	Tool support	Assessment/stakeholder	Assessment method	Guidance for application	Target/adaptability	Applicable stage	Domain
[6]	NO	Mixed (user &model)/ students, user	Qualitative	Partial guidance	Specific/NI	Design and evaluation	Computer science
[8]	NO	NI	NI	Abstract	General/NI	Design and analysis	IT
[2]	NO	Expert assessment/ designer	Qualitative	Concrete/application and use of components	General/NI	Design	Math
[16]	Yes/Specified design and evaluation criteria	Expert assessment/Teachers and instructional designers	Quantitative	Concrete/criteria and case study to apply it	General/NI	Analysis, design, development and evaluation	Chinese, math, foreign languages
[13]	NO	NI	NI	Abstract	General/NI	Design	NI
[10]	NO	NI	NI	Abstract	General/NI	Design	NI
[4]	NO	NI	NI	Abstract	General/NI	Design	NI
[9]	NO	Mixed approach/expert and users	Quantitative and qualitative	Abstract	Specific/Yes (used based on game scope &characteristics)	Evaluation and design	Intergenerational and intercultural learning
[7]	NO	NI	NI	Abstract	Specific/NI	Design	Higher education
[11]	Yes	User-based/player	Qualitative	Concrete	General/Yes (macro elements for different genre)	Design	Geometry/history, geography, culture

Table 5. Strength and weakness of existing educational game design models/frameworks

Model Ref	Strength	Weakness
[6]	Uses data-driven analysis of learning experiences through visualizations, educational data mining, and statistical techniques applied to game logs. Game-log data are used to model learning and identify places of improvements	The steps in the process of designing educational games are not clearly defined
[8]	Model links gameplay with experiential learning to facilitate the flow experience	It only provides a link between game design and educational theory not guiding the whole game design project. Several issues such as engaging storyline, appropriate graphics and sounds, and game balance are not included. Only good gameplay cannot save learning game
[2]	Useful analytical tool and also assist to improve the creativity of educational game designer by guiding the brainstorming of game ideas from both game design and educational angles. Encourage thinking across components rather than individual approach	The framework is descriptive and difficult to apply. It does not offer any tool or instrument support as well
[16]	Offers a common framework for educators and game designers by comprehending the effective integration of curriculum and game. The model also aids in evaluating the effectiveness of games used in educational settings as well as to select valuable games for use in classrooms	The model provides guidance to assess already developed games for classroom use. However, does not provide practical guidelines to structure the design process for educational game development. The criterion for design and evaluation should be refined further to be perfect for educational game design in practice
[13]	Emphasize the pedagogical aspects in designing educational games	key features presented for designing educational games are based on four frameworks and not all are specific for educational games. No guidance is provided on practical application of framework
[10]	The triadic theoretical framework provides a rich theoretical basis and present serious game design elements by outlining underpinning theories and associated challenges	Does not provide any concrete guidance on steps to integrate them in design process or how to operationalize them in serious game design

(*continued*)

Table 5. (*continued*)

Model Ref	Strength	Weakness
[4]	Provides a hierarchy with identity as core foundational element. Includes informed learning concept as an important element in hierarchy. It exhibits a game concept to demonstrate learning game design process	Model does not provide design steps and practical application of these concepts in design process with reference to their magnitude
[9]	Framework emphasize the threefold nature of educational game and include technical verification and user experience along with pedagogical dimension, highlighting critical aspects of each	The framework does not focus on game specific dimensions and doesn't provide practical guidelines to educational game design
[7]	The model emphasizes on higher education with game design, pedagogy and learning content modelling as main factors and is designed specifically for student self-learning with incorporated self-assessment modules	The model does not provide concrete guidance for application. Although model focuses on higher education, but the compared frameworks used as development basis are not specific for higher education
[11]	Presented macro game design concepts that can be adapted to different game genre. To build a GBL design model it defines all factors and also analyze the relationships among them	GBL combines game and education but the model only discussed the game factors

Environment is covered by three frameworks [9, 12, 14] focusing on context, mobility and technical verification. User attribute is only focused by adaptive digital game-based learning framework and game-based learning guidelines that included learner and children requirements respectively. Majority of analyzed frameworks focus on two attributes (learning and game design) highlighting their importance in design phase. None of the design frameworks or even guidelines covered all six attributes.

RQ2: The analytical lens of validity highlighted that all analyzed frameworks to some extent cited some theory or literature to justify their development. The selection of a theoretical basis for development of framework is based on the specific objectives and approach of each framework towards game-based learning. The knowledge of underlying developmental base is also important for educational game designer to select the framework appropriate to their objectives. Most frameworks are theoretically grounded in literature for a pedagogical base and game design principals. Some of the pedagogical theories used include Blooms taxonomy, Piaget's schemes and Gagne's events of instruction, Vygotsky zones of proximal development, experiential learning theory and instructional design principals [4, 15, 16, 31].

Some frameworks (Adaptive digital game-based learning framework and Educational Games (EG) design framework) compared existing models as developmental base of their framework. Moreover, "I's" combined the practical experience from field with research from commercial games as the development base. When it comes to empirical validation or application of design frameworks, only two frameworks level up and experiential gaming model had empirical evidence of their validity with sample size of 61 and 221 respectively. Learning curve, flow antecedents and game factors in [11] were the only elements validated by empirical study. However, the frameworks are validated by the authors who proposed them, and no other educational game so far reported to use these frameworks in its design. All the other mentioned frameworks were not empirically validated, only mentioning it as a future work. However, four frameworks: Framework for analysis and design of educational games, RETAIN, EGameDesign and design principals for serious game illustrated the application of framework on educational game as a case study without actual implementation.

RQ3: The comparison on the basis of framework attributes highlighted some open problems. Surprisingly, no tool support is available by existing educational game design frameworks except Game Factors and Game-Based Learning Design Model that provided an instrument called "Game factor questionnaire" and RETAIN model which provided design and evaluation criteria in terms of level points, higher the points, better is the designed educational game. The studied models also differ in terms of assessment and stakeholders involved. Framework for analysis and design of educational games and RETAIN model focused on expert-based assessment with teachers and designers as stakeholders, e-VITA framework for SGs focused on mixed approach of both expert and user assessment and Game Factors and Game-Based Learning Design Model emphasized on user-based assessment. While the authors of remaining frameworks and models did not provide any information.

Based on comparative analysis, six frameworks (Experiential gaming model, Adaptive digital dame-based learning framework, A theoretical framework for serious game design, "I's", e-VITA framework for SGs and Educational Games (EG) design framework) emphasized on abstract principles rather than concrete guidance and are limited to high-level concepts without providing any procedural guidance to structure the design process of educational games. The other three frameworks provided some form of concrete guidance to support educational game design. Framework for analysis and design of educational games provided guidance on each of the three components by illustrating their application on a zombie game and also guided how to think across component during brainstorming. RETAIN provided a criterion with level points to assess already developed educational game and a case study to illustrate it. However, it did not provide guidance for designing a new educational game. Game Factors and Game-Based Learning Design Model suggested macro elements and represented a thinking process with a model to help educational game designers incorporate it in their game along with an instrument (game factor questionnaire) for assessment.

The comparative analysis also illustrated that most of the models are general for any educational game design and audience. However, there were three specific models, two of these focused on a specific domain (computer science games in level up, intergenerational in e-VITA framework) and one focused on specific audience (higher education students in Educational Games (EG) design framework). The framework attribute of “adaptability in use” is addressed by only two models: e-VITA framework which emphasized that framework should be employed depending on the characteristics and scope of game and Game Factors and Game-Based Learning Design Model that not only emphasized but also provided the opportunity for adaptation by offering macro elements that can be adapted for different genre. According to the comparative analysis, most of the analyzed frameworks focused only on design stage but three models (Level up, Experiential gaming model and e-Vita) can be used for evaluation or analysis as well along with design stage. Moreover, RETAIN model claims to be applicable for all stage (Analysis, design, development and evaluation) of educational games development life cycle. However, no practical usage is available. The educational game design models are applied in various educational domains such as computer science, math’s, geography, culture, language and history are particularly mentioned among the compared models.

6 Conclusion and Future Work

This paper particularly focuses on design of educational games and reports on the comparative analysis of design models/frameworks for game-based learning. The study analyzes the use of GBL dimensions and validation in existing frameworks to identify essential elements for design stage. Secondly it also highlights the differences and similarities between different GBL design frameworks/models by exploring framework attributes to guide educational game designer/researchers in making more informed decisions and also to underline the open research issues in this area. The results of comparative analysis conclude that: Learning/pedagogy (Learning objective, instructional design, learning content and knowledge enhancement/outcome) and game factors (mechanics, dynamics, narrative, aesthetics, goals) are the most essential attributes for the design of educational games. The attributes of affective reactions (flow, enjoyment, immersion) comes after learning and game factors. Whereas, usability (user interface), user (learner requirements) and environment (including technical and context related aspects) are less emphasized by the analyzed educational game design models. Therefore, the design phase of educational game should emphasize more on linking learning objective with game objective in an efficient way to facilitate the affective reactions such as flow in order to engage and immerse the player [8, 10]. The importance of these three attributes in the design of educational game is also evident from the developmental basis of these models, most of which are theoretically grounded in learning and game design theories with focus on ARCS models and flow theory. However, there is a scarcity of evidence for empirical validity and practical application of educational game design models for educational game development. A few empirical studies and developed educational games that exist for framework validation are conducted by the same researchers who developed the framework in order to validate it and few elements such as learning curve, flow antecedents and some game design factors are empirically validated. A bigger community of educational game designers and researchers is needed who are willing to apply these models for designing educational games to bring useful insights from industry and go beyond the researchers who developed these frameworks.

Therefore, the analysis brings forward two extremely important issues which are in line with the results of [18]; lack of independent evaluation and absence of practical application of these design models in educational game industry for designing effective educational games. This lack of usage and assessment can also be seen as a result of absence of tool support, lack of adaptability and concrete guidance for practical application of framework concepts in the design process of educational game development. However, one aspect could also be that most of industry work is not published in research community and a collaboration between industry and research is important for thorough insights. Also, most of the frameworks do not provide any information on assessment approach, method or stakeholder(s) that are required to participate in assessment.

For overcoming these issues, future research should focus on providing concrete guidelines and steps to use the framework's principals for educational game design in practice for example if a framework focuses on linking gameplay and learning so researcher should provide practical insights about how certain learning objective such as problem solving can be seamlessly embedded in game mechanics or if focus is challenges then how to increase learning complexity along with increasing game challenges and mapping learning content to game tasks and narrative. The future research should also guide the game designers on assessment of the design principals (that the models provides) embedded in their educational game as part of design phase. Finally, there is an extreme lack of tool support for available educational game design models which need to be addressed to make ways for framework-based educational game design by providing tool support for practical application. The future work will focus on the development and evaluation models for educational games to investigate and compare the shift in dimensional focus between different stages of educational game development lifecycle.

References

1. Ahmad, M., Rahim, L.A., Arshad, N.I.: Towards an Effective Modelling and Development of Educational Games with Subject-Matter: A Multi-Domain Framework. IEEE (2015)
2. Aleven, V., Myers, E., Easterday, M., Ogan, A.: Toward a Framework for the Analysis and Design of Educational Games. IEEE (2010)
3. Alfadhli, S., Alsumait, A.: Game-Based Learning Guidelines: Designing for Learning and Fun. IEEE (2015)
4. Annetta, L.A.: The, "I's" have it: a framework for serious educational game design. Rev. General Psychol. **14**(2), 105 (2010)
5. Chorianopoulos, K., Giannakos, M.N.: Design principles for serious video games in mathematics education: from theory to practice. Int. J. Serious Game **1**(3), 51–59 (2014)
6. Eagle, M.: Level Up: A Frame Work for the Design and Evaluation of Educational Games. ACM (2009)
7. Ibrahim, R., Jaafar, A.: Educational Games (EG) Design Framework: Combination of Game Design, Pedagogy and Content Modeling. IEEE (2009)

8. Kiili, K.: Digital game-based learning: Towards an experiential gaming model. Int. Higher Educ. **8**(1), 13–24 (2005)
9. Pappa, D., Pannese, L.: Effective design and evaluation of serious games: the case of the e-VITA project. In: Knowledge Management, Information Systems, e-Learning, and Sustainability Research, pp. 225–237 (2010)
10. Rooney, P.: A theoretical framework for serious game design: exploring pedagogy, play and fidelity and their implications for the design process (2012)
11. Shi, Y.-R., Shih, J.-L.: Game factors and game-based learning design model. Int. J. Comput. Games Technol. **2015**, 11 (2015)
12. Shiratuddin, N., Zaibon, S.B.: Designing User Experience for Mobile Game-Based Learning. IEEE (2011)
13. Tan, P.-H., Ling, S.-W., Ting, C.-Y.: Adaptive Digital Game-Based Learning Framework. ACM (2007)
14. Thomas, S., Schott, G., Kambouri, M.: Designing for learning or designing for fun? Setting Usability Guidelines for Mobile Educational Games. Learning with Mobile Devices: A Book of Papers, pp. 173–181 (2004)
15. Yu, S.-C., Fu, F.-L., Su, C.: EGameDesign: guidelines for enjoyment and knowledge enhancement. In: Hybrid Learning and Education, pp. 35–44 (2009)
16. Zhang, H., Fan, X., Xing, H.: Research on the Design and Evaluation of Educational Games Based on the RETAIN Model. IEEE (2010)
17. Neil, K.: Game Design Tools: Time to Evaluate (2012)
18. dos Santos, A.D., Fraternali, P.: A Comparison of Methodological Frameworks for Digital Learning Game Design. Springer, Heidelberg (2015)
19. Ahmad, M., Rahim, L.A., Arshad, N.I.: An analysis of educational game design frameworks from software engineering perspective. J. Inf. Commun. Technol. **14** (2015)
20. Malliarakis, C., Satratzemi, M., Xinogalos, S.: Designing Educational Games for Computer Programming: A Holistic Framework. Electr. J. e-Learning **12**(3), 281–298 (2014)
21. Tahir, R., Wang, A.I.: State of the art in game based learning: dimensions for evaluating educational games. In: Academic Conferences International Limited (2017)
22. Song, X., Osterweil, L.J.: Toward objective, systematic design-method comparisons. IEEE Softw. **9**(3), 43–53 (1992)
23. Abrahamsson, P., Oza, N., Siponen, M.T.: Agile Software Development Methods: A Comparative Review1. Springer, Heidelberg (2010)
24. Chowdhury, A.F., Huda, M.N.: Comparison between Adaptive Software Development and Feature Driven Development. IEEE (2011)
25. Katayama, E.T., Goldman, A.: From manufacture to software development: a comparative review. In: Agile Processes in Software Engineering and Extreme Programming, pp. 88–101 (2011)
26. Tripathi, P., Kumar, M., Shrivastava, N.: Theoretical Validation of Quality Metrics of Indian e-Commerce Domain. IEEE (2009)
27. Babar, M.A., Gorton, I.: Comparison of Scenario-Based Software Architecture Evaluation Methods. IEEE (2004)
28. Yusof, N., Rias, R.M.: Serious games in psychotherapy: a comparative analysis of game design models (2006)
29. Abrahamsson, P., Warsta, J., Siponen, M.T., Ronkainen, J.: New Directions on Agile Methods: A Comparative Analysis. IEEE (2003)

30. Ahmed, F.F.: Comparative Analysis for Cloud Based e-learning. Procedia Comput. Sci. **65**, 368–376 (2015)
31. Kiili, K.: Evaluations of an experiential gaming model. Hum. Technol. Interdisc. J. Hum. ICT Environ. (2006)
32. Kiili, K.: Content creation challenges and flow experience in educational games: the IT-Emperor case. Int. Higher Educ. **8**(3), 183–198 (2005)
33. Tahir, R., Wang, A.I.: Codifying game-based learning: LEAGUE for evaluation. In: Proceedings of the 12th European Conference on Games Based Learning (2018)

PAPER 6

P6:
Transforming a Theoretical Framework to Design Cards: LEAGUÊ Ideation Toolkit for Game-Based Learning Design

Rabail Tahir and Alf Inge Wang

Sustainability, Special Issue Design Methodology for Educational Games

sustainability

MDPI

Article

Transforming a Theoretical Framework to Design Cards: LEAGUE Ideation Toolkit for Game-Based Learning Design

Rabail Tahir * and Alf Inge Wang

Faculty of Information Technology and Electrical Engineering, Norwegian University of Science and Technology, 7491 Trondheim, Norway; alf.inge.wang@ntnu.no

* Correspondence: rabail.tahir@ntnu.no

Received: 27 July 2020; Accepted: 7 October 2020; Published: 14 October 2020

Abstract: Educational game design is a complex process demanding multi-dimensional focus in a heterogeneous team to balance multiple aspects. The existing Game-based learning (GBL) frameworks detail the required knowledge but are hard to use in design practice. Conversely, card-based design tools are a lightweight approach used to assist the early design phase. While several game design cards exist, none is specific for informing GBL knowledge. There is a lack of operationalizable approaches for designing learning games that integrate research based GBL knowledge into the actual ideation process. This paper presents a card-based GBL ideation toolkit to reduce the complexity of framework application and introduction of key GBL concepts in the design process as a tangible reference point to facilitate multi-dimensional focus, supporting idea generation, critical reflection, and creation of a shared understanding in the collaborative design process. The paper describes a ten-step process of transforming the LEAGUE framework into the LEAGUE toolkit (GBL ideation cards), an evaluation of the toolkit with design workshop participants, and design lessons detailing strengths and limitations to support GBL design practices.

Keywords: game-based learning; educational games; ideation tool; card-based; learning game design; collaborative design

1. Introduction

Game-based learning (GBL) offers a rich design space encompassing multiple dimensions (such as learning, game factors, usability, affective reactions, user, and environment), which are interrelated, and GBL design teams need to consider these as they influence the design of learning games [1–5]. The interdisciplinary teams for designing educational games consists of experts in different domains (e.g., game design, learning domain, technology, and human factors), and it is essential to involve them in the design process to explore the design from different perspectives to create appealing and successful solutions [6,7].

Several GBL frameworks structure the design concepts of learning games and can be used to justify and reason the design decisions. However, most of the GBL frameworks are not used in learning game design practice because it is hard to apply these theoretical principles [4,8–11], and they lack tool-support and guidance for the practical application [8,10,12,13]. This points to the lack of an operationalizable approach that is more accessible and lightweight to integrate theoretical GBL design knowledge into actual ideation and hands-on practice of learning game design to improve the collaborative process of designing a learning game and guide the GBL design team [4,6,11,12,14–16]. To bridge the gap between theory and practice, Hornecker [9,15] proposed transforming a framework into a design tool by converting the theoretical concepts into easy questions that designers can relate to, which can be introduced in the design process.

Researchers have found many useful characteristics of cards: they provide a common vocabulary to designers for use in design, enable the transfer of knowledge from academia to design practice, help kick off the design discussion, assist in different viewpoints aiding multi-dimensional focus, help refine ideas, structure design discussions to guide the design process, facilitate shared understanding and communication, and offer a playful approach to involve in design tasks [4,17,18]. These characteristics can address GBL design practice challenges, making them a viable and affordable tool (instead of other approaches such as design patterns or guidelines) to provide the intermediate-level knowledge [4,5]. They offer an approachable way to introduce information as part of the collaborative design process, and their abstraction level has enabled researchers to successfully use them in a wide variety of fields [4,17,19,20]. However, the existing game design cards are not specific for GBL knowledge and cannot be used to ideate learning game design to reinforce GBL key concepts and facilitate the required multi-dimensional focus. We found only two card-based tools for ideation and design of learning games: Tango cards [4] (for tangible learning games) and Exertion cards [15] (for exergames). However, both these tools cannot be generalized to other design situations because of their specific nature and focus on these particular areas. One limitation of existing card-based design tools is that researchers have not always articulated the design knowledge embedded in their cards [4]. Therefore, in order to make the GBL design knowledge easily accessible in the early design process (addressing the gap between theory and practice), we transformed the game-based learning framework (LEAGUE framework [2]) into a card-based GBL design tool (LEAGUE ideation toolkit) to examine if this approach is useful and valuable for educational game design practice. The LEAGUE framework [2] is selected because it was developed to support game-based learning's multi-dimensional nature. It emphasizes the key GBL components (with detailed hierarchy) and their interrelations, informing the design knowledge for learning games.

An educational game is different from an entertainment game as it requires a multi-dimensional focus (involving different aspects) in a collaborative design process with different stakeholders working together [1,2,6,14,16]. Therefore, retaining and balancing the different aspects of GBL is challenging for team members, as they often have limited knowledge beyond their expertise and, consequently, not the same interpretation of the overall design space [17,21]. Therefore, our LEAGUE ideation cards' objective was to make theoretical knowledge about designing learning games easily accessible to team members involved in the GBL design process by informing them during their work and providing inspiration. Thus, the LEAGUE toolkit's target audience would mostly be the GBL design teams, including researchers, students, and practitioners in the industry from the GBL community. We intended for a GBL specific yet a generic tool that can be used to design educational games for diverse learning domains and game genres, supporting GBL designers in the initial ideation phase. The developed toolkit consists of four types of card decks (primary cards, trigger cards, reflection cards, and custom cards), scaffolding GBL design, and five design activities (idea generation, idea development, idea refinement, idea illustration, and idea documentation), scaffolding collaborative ideation process, carried out in a workshop technique for learning game design ideation. To investigate the toolkit's effectiveness in supporting the GBL design ideation process, the toolkit was employed and evaluated in three design workshop sessions with 34 participants. For this study, we focused on participants' experience (perceived usefulness, understandability, level of fun, and satisfaction) using the toolkit and not examining produced design artifacts and team dynamics. This paper contributes by describing a 10-step process for transforming a framework into an ideation toolkit, providing tool support for applying theoretical GBL framework knowledge in design practice (support for GBL design team), and discussing design lessons by highlighting the strengths and limitations of the developed toolkit that can serve as guidelines for other researchers intending to do a similar task.

2. Background

This section presents challenges in the GBL design practice identified by relevant research studies, introduces the LEAGUE framework (used for transformation into card-based GBL toolkit), highlights the general characteristics of cards found effective for the design practice and the use of card-based tools in different domains that inspired the use of cards as an operationalizable approach for our work. Further, it underlines the limitations of existing design cards for GBL design practice.

2.1. Challenges in the GBL Design Practice

GBL is a complex multi-dimensional phenomenon, and several key factors influence the design of learning games such as learning, game factors, usability, affective reactions, user, and environment [2,3]. Designing learning games is complicated because it includes embedding learning content into gameplay, selecting game features that motivate the learner to repeat learning cycles within the game context, considering user characteristics to ensure proper usability, context requirements, and technical conditions for selecting appropriate technology [2,17,22,23]. Many researchers have explored the essential elements of educational games, including theoretical concepts and design knowledge of GBL. However, different researchers focus on different GBL design and evaluation elements producing a scattered picture [1]. The research work by Tahir and Wang [1] identified a lack of a holistic view and identification of the core dimensions of GBL. They addressed this problem by introducing the LEAGUE framework [2]. Although several frameworks can provide techniques to structure the design concepts of learning games and justify their design decisions, there is still a lack of research on improving the process of designing a learning game [6,14,16]. Most design frameworks/models are not used in the learning game industry because they lack tool-support and guidance for practical applications [8,10,12,13]. Furthermore, learning games require to be developed in interdisciplinary teams involving experts in different areas (technology, game design, pedagogy, and usability). These stakeholders sometimes have limited knowledge (of other areas) beyond their expertise and often do not have the same interpretation of the design space [17]. However, it is crucial to involve a variety of stakeholders in the design process [6]. Therefore, there is a need to translate the intricate multidisciplinary theoretical knowledge of GBL into some easily accessible design practice that can support the hands-on practice and guide the designers to develop effective learning games in a playful collaborative manner [4].

2.2. The LEAGUE Framework

The LEAGUE framework [2] provides a holistic view of GBL design outlining the core components in a hierarchy by defining four conceptual levels (dimensions, factors, sub-factors, and metrics) for comprehensive GBL design. The dimension level is at the highest abstraction and metric the lowest. The framework focuses on six dimensions, and each has factors (22 in total) and sub-factors (total 74). The GBL dimensions in the framework are related to each other in terms of cause and effect. Table 1 presents the dimensions, factors, and interrelations (between dimensions).

Table 1. LEAGUE framework components.

Conceptual Level	**Elements**
Dimensions	Learning, Environment, Affective-Cognitive Reactions (ACR), Game Factors, Usability, and User
Factors	Learning objectives, learning strategy, learning content, learning outcome, technical aspects, context, enjoyment, engagement, motivation, flow, game definition, game narrative, game mechanics, game resources, game aesthetics, gameplay, interface, learnability, satisfaction; learner profile, cognitive needs, and psychological needs
Interrelations	Learning (*integrate*) Game Factors; Game Factors (*generate*) ACR; Usability (*address/cater*) User; Usability (*address/cater*) Environment; Environment (*map*) Use; User (*influence*) Learning & Game factors, ACR; User (*influence*) Learning, Game factors, ACR

Such detailed frameworks are complex to use in design practice, as reported by other researchers [10,12,15,24,25]. Additionally, there is uncertainty applying such frameworks for educational game design in practice [8–11,14]. Currently, the LEAGUE framework [2] does not provide any step-by-step guidance for a GBL design process. Therefore, a more accessible and lightweight approach than the existing framework is needed to effectively transfer knowledge between theory and design practice [4], and tool support is needed.

2.3. Cards as Design Tools: Useful Characteristics for Design Practice

Design researchers have found many important characteristics of cards that make them effective for design practice [4,26]. According to Lucero et al. [19], cards are great for collaborative design owing to their general characteristics (i.e., triggers of combinatorial creativity, tangible idea containers, and collaboration enablers). Cards facilitate the design process by functioning as conversation-starters, orienting devices, and pacemakers by structuring the creative sessions [15,17]. Cards facilitate the ideation process [15,26] as they are used to bookmark ideas [4], rate, or evaluate those ideas, thus enabling critical reflection [26]. Furthermore, they make design practice more playful and engaging by providing a source of inspiration through provocative questions or triggers and preventing the discussion from becoming unproductive [26]. Researchers in different areas [4,18,26,27] have created card-based design tools to make domain concepts and knowledge easily accessible to designers in their design practice as they provide a tangible representation of abstract concepts that are easy to use. Design cards can support focus shifts, as evident from the work of [15,26], which is essential for multidisciplinary domains like GBL that require multi-dimensional focus. Cards act as physical props during design discussions that help articulate thoughts and make arguments tangible, thus aiding shared understanding and communication [4,15,26].

These cards' characteristics provide a low-tech and approachable way to communicate the LEAGUE framework categories in learning game design practice, leading us to develop a card-based toolkit to make GBL design concepts easily accessible to the design team, facilitate collaboration, guide the ideation process and stimulate design ideas.

2.4. Card-Based Tools in Various Domains

Card-based tools have previously been used by researchers to put together knowledge from diverse areas (such as tangible interfaces, IoT, playfulness, and eco information) into an easily accessible form to stimulate design thinking and aid in design practice [7,24,26,28]. Many researchers focused on game design or gamification. Relevant examples include the Verbs, Nouns, and Adjectives (VNA) cards [29], card-based toolset for gamification design [18], three brainstorming games for game designers [30], ideation cards for mixed-reality game design [17], Playful Experiences framework (PLEX) cards [25], and a deck of lenses [31]. We found only two design card toolkits that focused on educational games, but these are limited to the specific domain's knowledge. Deng et al. [4] developed "Tango cards" for designing tangible learning games. Tango cards summarize the design knowledge in two areas: "tangibles" and "games". The focus is more on tangible and games rather than the GBL approach itself. Mueller et al. [15] developed a design tool specific to sports (exertion games), focusing perspectives on the body, and designing exertion experiences. Therefore, these cards do not inform complete GBL design knowledge.

For this study, we developed a card-based tool for GBL design to facilitate the ideation of learning game design in practice. The development of our toolkit focused on customizable and context-specific design patterns [32]. Our LEAGUE toolkit shares some core aspects with the previously discussed design cards. They communicate domain knowledge using provocative questions and tasks similar to [4,15,26,31], utilizing the characteristic of cards as tangible containers; use different card decks, each serving a specific purpose similar to [17,24,26,28]; we also structure the ideation process and organize participation using playful design activities similar to [17,26,33]; provide inspiration using examples (as triggers) from ad hoc external sources as means of supplementing and developing design concepts similar to [26,28]; facilitate critical thinking using reflection criteria similar to [17,26] and, supporting multi-dimensional focus using categories similar to [7]. However, our toolkit differs from existing card-based tools by extending the ideation process to include illustration and documentation (creating a small version Game design document) to support shared understanding and in-depth discussion and a log sheet along with ideation sheets for recording and tracking design decisions, thus supporting awareness and traceability on the design process [34]. The next section will focus on the toolkit development.

3. Toolkit Development Process: Turning the Framework into Ideation Cards

This section presents the process of turning the framework concepts into ideation cards. As discussed earlier, such transformation is valuable for bridging the gap between theory and practice [9,15] and several researchers followed this approach, e.g., PLEX cards based on the PLEX framework [25], Eco information individualization design toolkit based on the conceptual framework of Eco Information Individualization [24], exertion cards based on exertion framework [15]. However, except Mueller [15], none of the others explicitly detail the steps of transforming the framework into a design tool, and not many well-defined processes exist. The five-stage process described by Mueller [15] is not validated beyond their work. Therefore, we took their work as a starting point and further adapted and extended the process based on our experience to validate and extend the prior work that might be useful for other researchers for converting framework into design cards.

Our extended process of transforming the framework into ideation cards consisted of the following ten steps:

1. Define goals/objectives: The following objectives were defined for the toolkit: (1) Summarize and communicate GBL design knowledge (LEAGUE framework [2] categories): making GBL concepts easily accessible to learning game designers in practice; (2) Support collaborative design process: fostering multidisciplinary focus shift by focusing on different dimensions; (3) Inspire designers: supporting the initial generation of ideas (brainstorming) by providing triggers to facilitate the creative thinking; (4) Support in-depth reflection of ideas: providing criteria to enable critical thinking and a trade-off between different aspects; and (5) Structure and guide the ideation process: orienting the ideation process from start to end with structured design activities.
2. Establish target boundaries: We decided to aim for a relatively large number of cards (ultimately 176) to provide a comprehensive tool but targeted to keep the main cards (GBL concepts) to a limited number (28 in total) in order to minimize the chances of designers feeling overwhelmed.
3. Scrutinize framework to extract concepts: The LEAGUE framework [2] provides the GBL design space. As described by [35], the design space is the set of decisions and choices that need to be made about the designed product, and it captures the essential elements that the design product must-have. We looked at the components of the LEAGUE framework and picked 6 key dimensions, 22 factors, and relations (see Table 1) for converting to ideation cards, as these components can fully communicate the GBL concepts required by designers to make design decisions in the learning game design process without overwhelming them with detailed sub-factors and metrics.
4. Decide the type of cards: The extracted dimensions, factors, and relations were translated into a set of ideation cards. The main traits we wanted in LEAGUE cards are (i) informative and collaborative: to define and inform GBL design concepts and support multi-dimensional focus, (ii) inspirational: to support brainstorming, (iii) reflective: to support the refinement of ideas, and (iv) customizable: to facilitate the creative thinking. Therefore, we decided on four different decks of cards (primary, trigger, reflection, and custom) to focus on a particular task. In addition to the four card types, primary and trigger cards also belong to a sub-type. The two sub-types are dimensions and factors.
5. Formulate the content: For primary cards, we focused on extracted dimensions and factors (see Table 1) from the LEAGUE framework [2]. The goal here was to translate the framework components into directive yet colloquial questions/tasks. For trigger cards, the goal was to provide some example answers/ideas to exemplify the possible design choices to stimulate brainstorming. Triggers were collected from ad-hoc external sources, existing educational games, and GBL

literature [2,36]. For reflection cards, we focused on extracted interrelations (see Table 1) from the LEAGUE framework and translated them into critical thinking questions. The goal was to emphasize the trade-offs that need to be negotiated. Custom cards were blank cards to leave room for custom choices and support creativity. Tables 2 and 3 illustrate the translation of framework concepts into primary and reflection cards questions.

Table 2. Translation of framework concepts into primary cards questions.

Framework Elements	Conceptual Level	Definition of GBL Element	Primary Card ID	Translated Primary Card Task/Question	Related Trigger Cards
Learning (domain)	Dimension	The learning area(s) focus in an educational game to promote and facilitate learning.	DL	Decide the learning domain for the game.	Math; Climate change; Smart city; Dance
Learning strategy	Factor	Pedagogical theories or approaches used to achieve learning objectives.	FL2	What strategy should be used to enable learning through the game?	Drill and Practice; Organize; Compare/contrast; Judge

Table 3. Translation of framework interrelations into reflection cards questions.

Interrelated Dimensions in Framework	Identified Relation	Translated Question for Reflection Cards
Learning & Game Factors	Integration/Balance	Are game elements (game objectives, narrative, etc.) and learning elements (learning objectives, strategy, content, etc.) well integrated into this game?
Game Factors & ACR	Generate	Are selected game elements (narrative, mechanics, play, etc.) effective in generating user reactions (engagement, enjoyment, etc.) in this game?

6. Reduce items: The translation of framework dimensions and factors resulted in 28 primary cards (one question for each GBL element), and the translation of framework relations resulted in 7 reflection cards (focusing on questions that could challenge designers to reflect); thus, 35 question cards (primary and reflection cards) in total. To reach our target boundary, we limited the number of triggers (possible choices/examples) for each GBL element. This resulted in 113 examples called trigger cards.
7. Define rules/process: The LEAGUE toolkit uses structured design activities to guide the ideation process (one of the defined objectives). We defined five design activities. Each design activity had a required output and used a different set of cards and ideation sheets. We also imposed time limitations for each activity to make participants active and prevent them from being unproductive.
8. Visualize: All cards have a standard "playing card" size approximately 2.5 × 3.5 inches (64 × 89 mm). All cards are color-coded by deck (type) and category (six dimensions) to be distinct. Figures 1 and 2 shows an example of developed cards. Each of the six categories has a different color (taken from the LEAGUE framework [2]). For Trigger cards, the categories are defined by the border color of each card. All cards have a consistent graphical layout and information architecture. We made sure to keep the card design minimal and easy to follow, not overcrowded with too much text, and balance text and images [4]. The text on trigger cards (presenting example answers or triggers) is limited to only a few words, as they are intended for inspirational use and should only provide a hint and not a concrete design [4,26]. The card's backside consists of four elements (see Figure 2): type of card deck, card title, an image icon to visualize card type, and a short description of the role of the card or the definition of the GBL-concept. The card's frontside consists of five main elements: a unique ID, card name,

the sub-type, the main question/concept/ idea, and graphics (icon or image) illustrating the question/concept/idea. However, the custom cards are blank. We also developed a board with a playbook to make the process easy to understand and structured.

9. Gather feedback: The feedback from the co-author was incorporated iteratively at each stage of the process of developing the LEAGUE toolkit. After the completion, the toolkit was discussed in detail with fellow researchers to verify that cards were understood without much explanation. They mainly provided feedback on improving the wording and presentation of cards and playboard. Afterward, the toolkit was employed in three design workshop sessions to explore the toolkit's potential through feedback from participants and inspect the workshop session, design outcomes, and team dynamics. We used a questionnaire, focus group, observation, and video recording for the data collection.
10. Refine and improve: The toolkit was iteratively refined with feedback from fellow researchers and design workshops. In the first iteration, definitions and questions on the cards were rephrased for clarity, preciseness in meaning, and their presentation based on fellow researchers' feedback. In the second iteration, in addition to these changes, the design activities were adjusted and re-organized by changing the allocated time and rearranging debriefing sessions based on feedback from the first workshop session. In the third iteration, we plan to improve the cards' searchability using accessories and precisely define the criteria for reflection cards to facilitate critical thinking based on collective results from three design workshop sessions. The toolkit is not ultimate and will still be improved based on future studies.

4. The LEAGUE Ideation Toolkit: Developed Card-based Tool for GBL Design

This section presents the developed LEAGUE ideation Toolkit. The toolkit consists of four decks of cards, five design activities (each with an idea sheet collectively called ideation sheets), a board with a playbook, and a log sheet. They are all used together in a workshop format to ideate learning game design (see Figure 1). The LEAGUE toolkit uses ideation cards with structured design activities to make theoretical and conceptual knowledge of GBL design (from LEAGUE framework) accessible to the GBL design team and guide the ideation process.

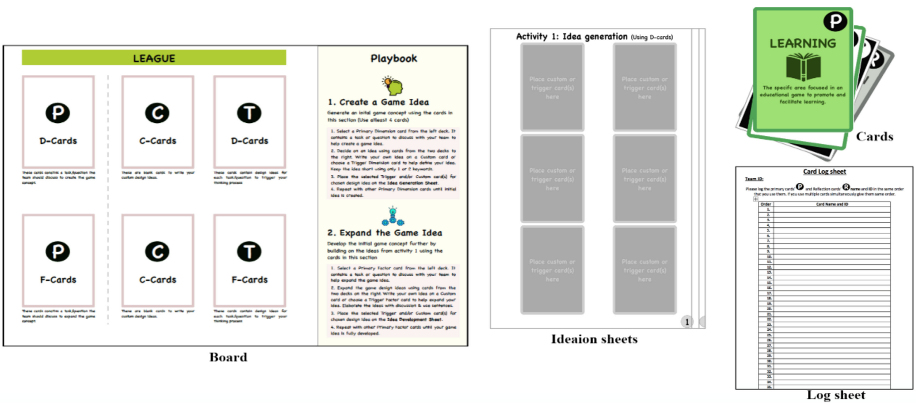


Figure 1. LEAGUE toolkit items.

The toolkit is intended to facilitate and guide the collaborative ideation of learning game design and thus is designed to be played in a group (with a recommended group size of four to six players). The target audience of the LEAGUE toolkit is both academia and industry. Currently, the toolkit has only been used with researchers/students.

4.1. LEAGUE Cards

This section describes the LEAGUE cards and the rationale behind the different decks of cards. The toolkit consists of four distinct card types (Decks): Primary, Trigger, Custom, and Reflection cards. Figure 2 shows the different card types. The complete box of LEAGUE cards includes 28 primary cards, 113 trigger cards, 28 custom cards, and 7 reflection cards (176 cards in total). However, the main question cards are only 35 (primary and reflection). Each deck has a specific purpose in the overall ideation and design process and is introduced in a design activity to serve the goal set for that specific activity.

Figure 2. Example of four card decks.

The primary and trigger cards are grouped into six main categories (based on six dimensions in LEAGUE framework): Learning, Game, Reaction, Usability, User, and Environment. Each of the two decks (Primary and Trigger) consists of two sub-decks (sub-types): dimensions and factors (see Table 1), used in different design activities (explained in Section 4.2).

The four card types are described as follows:

- Primary Cards (Present GBL design concepts): The Primary cards are the main deck of cards that are the building blocks for GBL design. Each primary card presents one particular GBL concept. The card poses a question, or a task related to that concept, which should be discussed in a team to develop a design idea (using either custom, trigger, or any combination of these cards). The team successively answers these tasks/questions to gradually build the learning game design idea through collaborative team discussion. There are 28 primary cards in total posing 28 different tasks/questions, out of which six are primary-dimension cards (focused on framework dimensions), and twenty-two primary-factor cards (focused on framework factors).
- Trigger Cards (Support for brainstorming): Trigger cards are examples of possible design ideas or hints for primary cards' tasks or questions. These cards trigger the thinking process by giving a direction to think. Each primary card has multiple trigger cards (with the same name as the primary cards). For example, for the primary-dimension card "reaction", there are three different trigger-dimension cards (emotional, behavioral, and cognitive) with the same name

"reaction", as shown in Figure 3. There are 113 trigger cards in total, out of which twenty-two are trigger-dimension cards (for primary-dimension cards), and ninety-one are trigger-factor cards (for primary-factor cards). We do not claim that the trigger cards are absolute and complete. However, we believe that they cover a range of different domains and areas of the GBL design space, which are enough to trigger the brainstorming and ideation.

- Custom Cards (Allow out-of-the-box thinking): This deck consists of blank cards used by the participants to write their creative design ideas. This provides an opportunity for out-of-the-box thinking and provides room for the creative impulses of participants.
- Reflection Cards (Aid refinement of generated ideas): Reflection cards present seven evaluation criteria to reflect on the generated ideas and design choices to refine them. Each reflection card contains a question pointing to a critical relation between different GBL dimensions that can negatively impact learning games' effectiveness if not considered. It encourages the team to critically think about the trade-off and look for design iterations if problems exist.

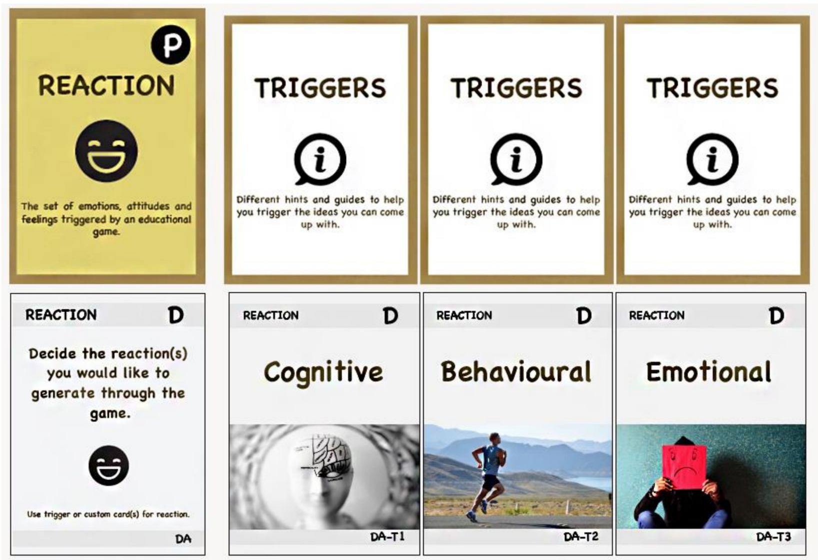


Figure 3. Trigger cards for the primary card "reaction".

Depending on the deck they belong to, the cards are either informative, inspirational, or reflective. Primary cards are informative, presenting GBL design elements and used as building blocks for the learning game design. The Trigger cards have an inspirational role and help trigger brainstorming by providing many provocative ideas as creative triggers. The Reflection cards provide a critical lens to validate or improve the developed design ideas by reflecting on them.

4.2. Design Activities

The LEAGUE toolkit play procedure is divided into five ideation design activities (see Figure 4). The five design activities are (1) Idea generation, (2) Idea development, (3) Idea refinement, (4) Idea illustration, and (5) Idea documentation.

Each design activity has a different goal and uses different toolkit items (cards and/or idea sheets). Each of the five design activities involves a separate idea sheet(s), which the team uses to produce the intended design outcome for that activity. A log sheet is used for logging the order of use of cards in the first three activities. The idea sheet has the same name as the design activity, e.g., the idea sheet for idea generation activity is called idea generation sheet. Collectively idea sheets of all five activities will be referred as "ideation sheets". The idea sheet(s) for the first three design activities provides a layout for the placement of cards used as design ideas in that activity. They provide a space for annotating how the cards have been used to support discussion and also record the team's decisions. The idea sheet for the fourth activity is a blank sheet to draw and visualize the design. Finally, the idea sheet for the fifth design activity provides a template for documenting the learning game design. Each used idea sheet is the design artifact/outcome of that specific design activity and provides the visual display of generated ideas, which helps the team summarize each design activity's outcome. They are also crucial for data collection and recording not only the ideas but the complete ideation process.

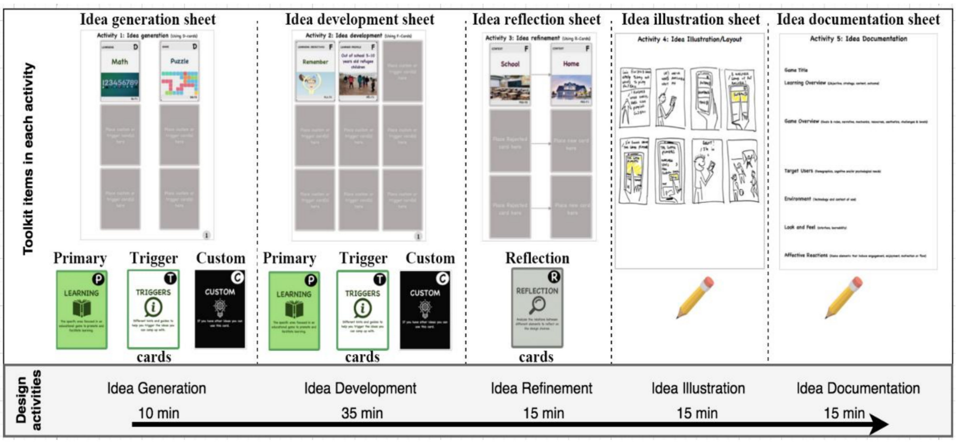


Figure 4. Ideation process with five design activities.

The five design activities are as follows:

1. Idea generation (Coming up with initial ideas): This activity aims to generate an initial concept for a learning game design. For this activity, the team uses sub-deck dimensions (see Section 4.1) and has six primary-dimension cards (6 dimensions), to solve using 22 trigger-dimension cards and 6 custom cards. Solving different primary cards (using trigger or custom cards) gradually generates an initial game idea. There is no right or wrong order of using the cards. Participants can shuffle through cards and pick one. The id of used primary cards is logged in the log sheet (in the order of use). The idea generation sheet is used to stick the trigger and custom cards to compose the initial idea.
2. Idea development (Expanding the idea): The goal here is to expand and further develop the initial ideas from the first activity into more detailed and concrete ones. For this activity, the team uses sub-deck factors (see Section 4.1) and has 22 primary-factor cards (22 factors), to solve using 91 trigger-factor cards and 22 custom cards. The team can select and use the cards in any order. The idea development sheet is used to stick the trigger and custom cards to develop the design idea. The id of used primary cards is recorded in the log sheet in the order of use.
3. Idea refinement (Reflecting on the idea): The goal is to improve or refine the developed ideas by reflecting on the design choices made using the reflection cards to identify the limitations and uncover questionable decisions. A team has seven reflection cards for this activity, and similar to the first two activities, they can shuffle through the cards and select in any order. The idea refinement sheet is used to add or replace the trigger and custom cards used to refine the developed idea. The idea refinement sheet has two sections for the placement of trigger/custom cards: one for rejected/replaced cards and one for new/added cards. In this activity, the team can use the ideation sheets from activity one and two to get an overview of design choices and

stimulate reflection on what needs improvement. Both used and unused trigger and custom cards from the previous two activities can improve the design choices by discarding previously used cards and adding new adds. A log sheet is used for logging the order of the use of reflection cards.

4. Idea illustration (Visualizing the game idea): This activity aims to plan the overall flow of the game in terms of how users will play the game from launching the game to quitting it. The idea illustration sheet is used to sketch the flow, and the team can choose from different ways (such as flow diagram, user scenarios, or screen prototypes) to illustrate the overall picture of a refined design idea. This activity allows for sketching the user experience and enables a transition from a static representation of ideas to a more dynamic view of how game players will play or interact with the learning game.
5. Idea documentation (Archiving the final idea): This last activity aims to document the final state of the learning game design idea, producing a short version of a game design document (GDD). The idea documentation sheet is used that provides a format to fill in details of the final idea.

4.3. Board and Playbook

The board is provided as scaffolding and comprises two main parts: a layout structure (card deck placeholders for design activities) and a playbook (describing how to play along with required toolkit artifacts (cards and idea sheets), intended goal, and outcome). Figure 1 illustrates a portion of the board. It provides visual affordance and describes the play sequence to guide the ideation process, reducing the need for supervision. The playbook explains the card decks required for each design activity, and the layout structure provides the space for placing these card decks.

4.4. Workshop Technique

The workshop format (inspired by [17,26]) provides support for a collaborative design process. A workshop session is approximately 2 h, where the participants work in teams (four to six participants) to generate and develop learning game design ideas, reflect on them, and finally illustrate and document their design ideas using the LEAGUE toolkit. One or two facilitators organize the design workshop to lead the team(s) through the ideation session by sequentially presenting the design activities. Each workshop starts with a short introduction (10 min) of GBL concepts and the LEAGUE toolkit description. Afterward, all the teams are provided with the LEAGUE toolkit and are asked to start the five-step ideation session. First, one of the organizers presents each activity individually, and the other organizer simultaneously provides each team with the toolkit artifacts (cards and idea sheets) required for that activity (see Section 4.2). Subsequently, the teams start working on that specific activity. Each design activity is time-bound (activity 1 is of 10 min, activity 2 of 30 min, activity 3 of 10 min, activity 4 and 5 of 15 min each) and must be completed following certain rules specifying the use of particular cards in each design activity and required output (see Section 4.2). A time constraint is added to avoid getting stuck or reaching the game idea too early before exploring the different cards. After activities 2 and 3, teams very briefly present their ideas in 5 min. One team member takes the role of a logger and records the order of use of cards in activity 1–3 using a log sheet. The log sheet is useful for both data collection and for the team to reflect on their design strategy to make improvements in the future. In the end, teams summarize their learning game designs with group presentations. The workshop ends by collecting participants' feedback.

5. Toolkit Evaluation: The User Study

We conducted design workshops with 34 participants (21 males and 13 females) in three different sessions to understand the value and utility of the LEAGUE toolkit in informing and guiding the ideation of learning games design in practice, which would, in turn, strengthen the argument for the transformation of the framework and also enabled the refinement of the design toolkit and the process. The participants were a convenient sample of university students and researchers (25 to 45 years old) at the faculty of computer science and engineering. The sample included 16 master students and 18 researchers (Ph.D./postdoc) that formed seven teams (comprising 4 to 6 members). Two teams had four participants each, four teams had five participants each, and one team had six participants. All participants' primary subject of study was computer science except one researcher from electrical engineering. It should be mentioned that 24 participants had no background in learning game design, 3 had minimal experience, and 7 had some experience. We selected participants with no to less experience to fully explore the toolkit's support in informing GBL design and not coming from their previous experience and knowledge to ensure the validity of data. None of the participants had previous experience with the LEAGUE toolkit. Each session had different participants. All participants were asked to sign a consent form and were informed that their participation was voluntary. Each session was approximately two hours long and was supervised by two organizers. The workshops were organized as described in Section 4.4. At the end of the workshop, data regarding participants' experience with the LEAGUE toolkit was collected using a questionnaire (with a 3-point rating scale) and a short focus group session to get feedback on their collaborative design process using the toolkit and suggestions for improvement. The questionnaire consisted of 23 questions (inspired by [15,24,26]) related to understandability, satisfaction, fun, and usefulness. Data was also collected through observation (researchers taking notes during workshop sessions) and video recording. The focus group session was recorded, and data analysis was guided by the Corbin and Strauss process [37].

We focused on three evaluation goals:

1. Participants experience using the toolkit: How did participants experience learning game design using the toolkit in terms of fun, satisfaction, understandability, and usefulness?
2. Roles (defined objectives) of the toolkit in the GBL design practice: Were the five defined objectives (i) inform GBL design knowledge, (ii) support collaborative design process, (iii) brainstorming, (iv) reflection, and (v) guidance for GBL ideation for the toolkit achieved?
3. Refinement of the toolkit: How to further improve the toolkit?

6. Results and Analysis

This section presents the design workshops' results using the toolkit through five stages of ideating an educational game design. First, we very briefly describe a few of the educational game design ideas participants came up with during the ideation sessions to exemplify the different types (range of learning domain) of ideas participants could achieve in ~70-min ideation session using the toolkit. Some of the ideas developed include: Team 1) "My swinging 20's (or Die)": An augmented reality dance class where the elderly with mobility issues learn to dance with their famous idol by following the indicated move patterns shown by colored areas using a dance pad otherwise, they will die. Team 2) "Smart city simulator (SCS)": A 3D simulation VR game for young adults to understand smart cities by developing and organizing a smart city to increase the inhabitants' happiness level. Team 3) "University runners!!!": A campus-based location-enabled game (using sensors installed at a campus that are linked to GPS location used in the web game) for all students at the university with assignments to learn to work as a team (teamwork skills) to achieve a common objective that is deadline extension by running away. The design outputs of these teams for each activity are shown in Appendix A.

Next, we will elaborate on the results from the questionnaire, focus group, and observation, which will be presented in three categories focusing on evaluation goals. An overview of the participants' responses to statements on fun, satisfaction, understandability, and usefulness is presented in Table 4 (rating scale is, 1 = Agree, 2 = Neutral, 3 = Disagree).

Table 4. Participants' feedback from the questionnaire.

Aspects	Key Concepts of the Questions	Agree	Neutral	Disagree
Fun	Interacting with cards was fun	74%	21%	6%
	Fun to do different activities	88%	9%	3%
	First activity (idea generation)	76%	18%	6%
	Second activity (idea development)	85%	9%	6%
	Third activity (idea refinement)	62%	32%	6%
	Fourth activity (idea illustration)	62%	29%	9%
	Fifth activity (idea documentation)	47%	35%	18%
Satisfaction	Visual design of cards	85%	12%	3%
	Time given for each activity	41%	32%	26%
	Sequence of use-primary cards	71%	26%	3%
Understandability	Cards	79%	12%	9%
	First activity (idea generation)	68%	12%	21%
	Second activity (idea development)	88%	9%	3%
	Third activity (idea refinement)	76%	24%	0%
	Fourth activity (idea illustration)	82%	18%	0%
	Fifth activity (idea documentation)	88%	12%	0%
Usefulness	Informing GBL design concepts (Primary Cards)	74%	18%	9%
	Supporting brainstorming (Trigger Cards)	76%	24%	0%
	Reflecting on ideas (reflection cards)	50%	41%	9%
	Information on card	74%	26%	0%
	Easy to ideate educational game design	62%	29%	9%
	Process provided guidance for GBL design	85%	12%	3%
	Considered elements I would not have without cards.	71%	26%	3%

6.1. Participants Experience Using the Toolkit

The aspects fun, satisfaction, understandability, and usefulness (shown in Figure 5) presents the participants' responses about their experience using the toolkit. The results reveal that overall, 71% agreed (only 7% disagree) that using the toolkit was fun, 66% agreed (11% disagree) with overall satisfaction, 80% agreed (only 5% disagree) that overall toolkit was easy to understand and 70% agreed (only 5% disagree) that toolkit was useful for the defined roles. The results for individual questions of these aspects are shown in Table 4. The questions for "usefulness" are linked to the toolkit's roles (defined objectives) and are discussed in the next section.

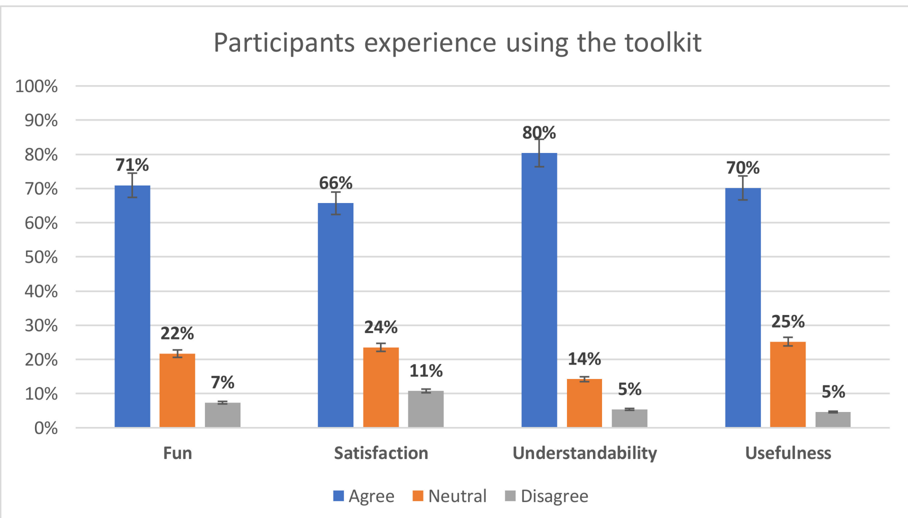


Figure 5. Overall rating on fun, satisfaction, understandability, and usefulness.

The results for specific questions related to “fun” show that most participants had fun using cards and activities (between 74 and 88% agreeing to the statements). The responses to how fun each design activity was show that most workshop participants agree that the first four design activities were fun to do (only 6 to 9% disagreed). However, only approximately half of the participants agreed that the fifth design activity (idea documentation) was fun. This can also be justified by the nature of the documentation task being time-consuming and tedious. Despite that, only 18% disagreed, which means the toolkit made it more accessible and fun to some extent, at least. Idea development was regarded as the most fun, followed by idea generation, which was the most creative thinking tasks. The results for specific questions related to “satisfaction” show that participants were satisfied with the visual design of cards and the sequence of use of primary cards (only 3% disagreed). The latter was asked to know if additional guidance was required for the sequence of use of cards to further scaffold ideation. Fewer participants agreed that sufficient time was given for each activity (41% agreed). This implies that some refinements are required in the workshop technique to readjust time distribution for design activities. The specific questions related to "understandability" show the participants’ response concerning how easy it was to understand cards and each of the five design activities. In general, most respondents (68% to 88% agreed) thought that cards and design activities were easy to understand. Nobody thought the three activities idea refinement, idea visualization, and idea documentation were hard to understand. However, 21% of the respondents disagree that idea generation was easy to understand and 3% for idea development. One reason for idea generation being slightly tricky to understand compared to others is because it was the first activity right after the introduction and also time-bound (only 10 min), participants felt rushed as they needed some time to understand the complete picture together as a team before they get started with the activity. From the results, we can conclude that the second activity (idea development) was the most fun and easy to understand. On the contrary, the fifth activity was also the easiest to understand. However, it was considered the least fun. Therefore, the nature of the task also affects the experience (fun). The feedback from focus group sessions also supported this. The following comments were received when asked about the most and least fun activity: “Second activity was the most fun to develop the idea more”, “Nobody likes documentation”.

Observations and feedback from focus group sessions support the results from the questionnaires. We received positive feedback from the participants highlighting that the toolkit was fun to use. “The sense of time diminished, time passed quicker than it actually does. It didn’t feel like a 2-h workshop”, “All group members were engaged”, “fun to play and engaging”. Overall, there were very few questions and misunderstandings during workshop sessions, suggesting that it was not difficult to understand the toolkit artifacts (cards and activities). However, in terms of time provided for each activity, participants sometimes felt rushed, although it kept them motivated and engaged. The participants’ feedback from the focus group include: “More time please! sometimes we felt a bit rushed”, “Since it is a group work, I think even if you give more time everyone is going to use it anyway, so it is good that it was time-restricted”. It was observed that the first activity was most difficult to understand. Some teams needed additional time to explore and read the playbook to understand the process before starting the activity, while others jumped too soon on the details about the game idea in activity 1.

6.2. Roles (Defined Objectives) of the Toolkit in the GBL Design Practice

This section focuses on investigating the five objectives (see Section 3, step 1) defined for the LEAGUE toolkit. The questions for usefulness in Table 4 are related to the roles (defined objectives) of the toolkit in GBL design ideation. Overall, 70% of the participants agree (only 5% disagree) that the toolkit was useful for the roles presented in questions. We will discuss the results for each role (objective) of the toolkit below.

Role 1 (Inform GBL design concepts): Only 9% disagree that the toolkit was useful in informing GBL design concepts through primary cards (achieving objective one). Figure 6 shows the percentage of teams' usage for individual primary cards. Interestingly, all primary cards were used at least by one team, indicating that not a single primary card can be considered irrelevant. The feedback from focus group sessions includes: "Good discussions about the game design and what is important to make a good educational game".

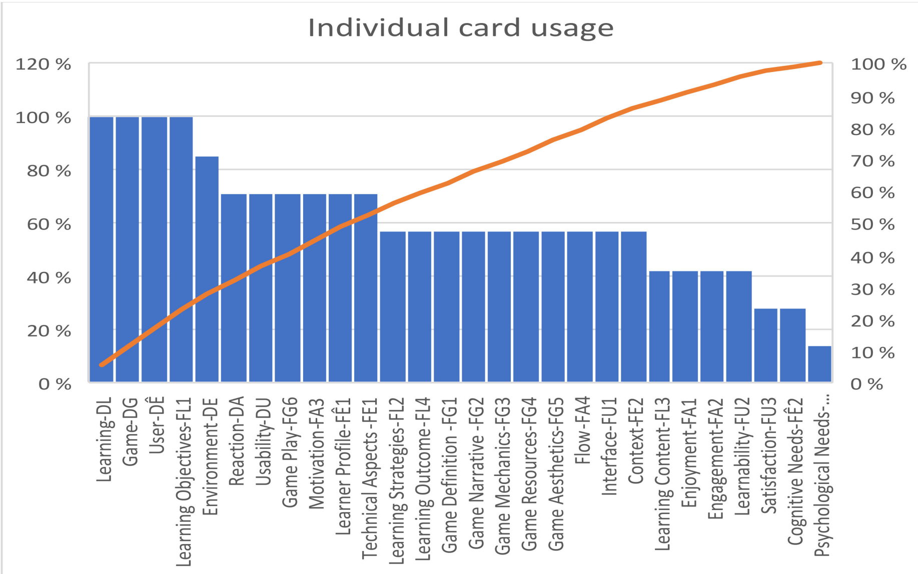


Figure 6. Primary cards usage by seven teams.

Role 2 (Support collaborative design process): The observation and focus group session provided some useful insights regarding play strategy emphasizing the toolkit's role to support a collaborative design process (achieving objective two). Most teams played collaboratively, selected one card with consensus, and then discussed as a group. One team used a turn-taking strategy and divided the cards to have a better flow. In both cases, the cards were selected through voting and debating. Most teams did not define any roles based on the area of expertise (although all teams overall focused all six GBL dimensions in the ideation process) but developed roles for practical work such as logging, drawing, documenting, taping cards, or to spare time such as finding triggers. The feedback from focus group participants includes: "It was very collaborative, and it is fun to discuss ideas in a team and build on them", "All group members were engaged", "It was a good approach for initiating team discussion".

Role 3 (Support brainstorming/idea generation): None of the participants disagreed that trigger cards supported brainstorming and information on cards was useful (achieving objective three). Based on observations, the teams used trigger cards in three different ways: (1) Use a trigger card as a design idea, (2) Use trigger cards to extend their ideas by combining different trigger card(s) and/or a custom card, and (3) Use trigger card as an example or inspiration to understand the concept and come up with their own ideas. However, all the teams almost always browse through the trigger cards either before initiating or finalizing their ideas based on their approach. Combining different cards (mixing trigger and custom cards) sparked the potential to generate creative design ideas. The feedback from focus group participants includes: "Trigger cards help to come up with ideas", "Trigger cards also work to confine the idea", "As a first-time user they worked really well but I wonder if it could be a bit restricting when you use the toolkit multiple times, but it is good the you are able to write your own ideas as well using custom cards".

Role 4 (Support reflection): The toolkit's role in providing support for reflection was agreed by only 50% of participants. However, only 9% disagreed with the statement (achieving objective four at least partially). Based on observations, most teams made only one or two changes to their designs using the reflection cards which is in line with the questionnaire results. The feedback from focus group participants include: "We refined based on reflection cards, but the questions in reflection cards were overlapping so we found only one problem", "It set our purpose for the whole design, but they were not imposing any new idea changes", "We did not change anything using reflection cards but we refined the idea further", "Maybe reflection should go after idea illustration because if you haven't visualized, you cannot change anything".

Role 5 (Guidance for GBL ideation): The majority of the participants agreed that the toolkit provided guidance for educational game design (only 3% disagree), the cards prompted to consider new elements (only 3% disagree), and it was relatively easy to ideate and design an educational game using the toolkit (only 9% disagree) thus achieving objective five. The feedback from focus group sessions includes: "For a novice in learning game design like me, it was very helpful because it was not difficult, and the cards were guiding me on what to do. Otherwise, I do not know what to do in game design and how a game is designed", "It is 'meta-game', a game to design a game", "Ideation sheets are very useful to visualize because you don't remember everything". When asked if more guidance was required for primary card selection, most teams were satisfied with their selected order of use of primary cards and thought that open choice is better as the order may vary in different games and the team should decide what is important for their game idea, guiding order would constraint the process. However, few participants thought it would be helpful to guide the sequence. The feedback from focus group participants includes: "Since we already defined the purpose of our game in activity one it was easier for us to follow that path and select the cards that satisfy our purpose", "We had many options in the second activity, so we were picking the concepts we thought were more important for our game idea", "We can just browse through the cards to select the ones that base on our initial idea to further develop it".

6.3. Refinement of the Toolkit

The analysis of the questionnaire, observation, and focus group data revealed that some elements played a role in hindering the ideation process of learning game design. This section highlights the issues in the developed toolkit (reflecting on the features that limited its use) and discuss the refinements that would improve its effectiveness. We identified four challenges that are presented along with the recommended refinements in the next two subsections.

6.3.1. Challenges in the Workshop Format

Challenge 1 (Introduction phase of the workshop): The observations during design workshop sessions and participants' feedback from focus groups revealed that the workshop's introduction phase could impact the toolkit's understandability. For example, few teams (in the first session of the design workshop) felt that activity 1 was a bit difficult to understand due to which they focused too soon on the details of the initial game concept in design activity 1 and were also unclear that trigger cards are for inspiration and not the definite answer which made them rely more on trigger cards (instead of using custom cards) for generating their initial game idea. This is also evident from the questionnaire results where the participants least agreed (68%) with the understandability of the first activity compared to the other four activities where they agreed between 76 and 88%. The introduction phase was short (only 10 min) and used PowerPoint slides to introduce GBL concepts and the LEAGUE toolkit, after which the teams immediately started working on activity 1 without giving any time for free exploration of the toolkit.

Recommended refinement 1 (Use of video/demo and free play): Therefore, to address this issue, the introduction must be more focused and include a demo/video explaining and visualizing the process instead of just slides. One participant suggested similar improvements: "Show the cards when the slides presentation is happening so that there is an easier translation of knowledge from talk to gameplay". It is also critical to give some time for free play after introduction, so team members can familiarize themselves with the toolkit components and understand the desired goal and outcome for each activity using the board and playbook before starting the ideation process.

Challenge 2 (Time-bounded design activities): We added a time constraint to the process similar to [26] in an attempt to avoid participants' converging on an idea too early (without exploring different cards) or getting stuck. However, the study results revealed that time distribution for design activities is not an easy undertaking. Participants acknowledged the benefit of time-bounded activities. However, it was vital for them that it did not get in the way of the creative process. For example, some participants thought they could use more time in idea development (second activity) to further elaborate their ideas and make it more concrete. Similarly, some teams required more time for activity one as participants needed time to explore the play process as a team before getting started. For activity 3, a difference in opinion was observed, where some teams needed more time, and the others needed less, depending on their reflective thinking and improvements. This is also evident from the questionnaire results, where only 41% of the participants were satisfied with the time given for each activity.

Recommended refinement 2 (Flexible format to introduce relaxation in the time constraint): Therefore, to address this issue, it helps to make the workshop format more flexible and accommodating by adding room for relaxation in time constraint if and when needed. This can be done by adding rapid group debriefing slots after each activity, which can also be used for extending the time of the activity if needed. Another solution could be to run two parallel activities, for example, running the activities 4 and 5 in parallel (where team members can divide the task and simultaneously work on illustration and documentation), leaving more time for the first three activities. Based on our experience, it is also necessary to plan some extra time for setup and practical arrangements in case of minor setbacks.

6.3.2. Challenges in Working with the Cards

Challenge 3 (High number of trigger cards): Some participants thought that there were too many cards in the second activity, as evident from the focus group feedback, which includes: "They are many", "yeah! Quite a lot". Although, this was not a problem since the second activity was considered the most fun (the majority of the participants (85%) enjoyed) and easy to understand (88% participants agreed) according to the questionnaire results. However, due to the high number of cards provided, it is vital to make the cards highly searchable.

Recommended refinement 3 (Make cards highly searchable using accessories): Some teams were observed splitting the card decks among all the participants who browsed through the cards and then collectively selected the relevant cards to discuss (also evident from Role 2 in Section 6.2). This was a way to speed up and simplify the card-selection process by dividing tasks and can be further facilitated by introducing mechanics such as turn-taking, defending and attacking, etc. Previous card research has recommended applying visual design (such as color-coding) to make cards highly searchable [4,26]. The LEAGUE cards are also color-coded by the six categories (six GBL dimensions). The use of the same card-name and initial letters of card-id for linked cards (primary and trigger cards) worked effectively as an identifier for the cards. However, some participants suggested that working with cards should be simplified further by having accessories such as a card division box (for dividing each category or using alphabetical order) that would improve the searchability of cards during design activities. Deng et al. [4] proposed a similar approach to introduce accessories (such as clothes pegs) for designers to mark important cards. Another solution could be to reduce the number of cards.

Challenge 4 (Facilitate critical reflection): A reflection card in the LEAGUE toolkit presents a question concerning interrelation between GBL dimensions to uncover the questionable decisions in the game design idea. Therefore, each reflection card focused on two dimensions to encourage the team to critically think about the trade-off by urging and attesting the generated ideas against these cards' criteria. However, this did not work very well, as some participants thought that questions in reflection cards were overlapping. Therefore, although they set the rationale for the learning game design but were not inflicting any new idea changes. This was also in line with the questionnaire results, where only 50% of the participants agreed with the usefulness of reflection cards in refining ideas.

Recommended refinement 4 (Precisely define criteria with examples imposing design change): Therefore, to address this issue, reflection cards need to be more directive to impose new design changes and challenge designers to reflect on developed design ideas. This can be done by offering clear guidance about what is required for critical reflection, for example, using additional directive questions guiding how to judge (similar to [26]), or providing some examples of possible refinements concerning the interrelation. Another approach can be to focus on one specific factor of each dimension (rather than a high-level concept) to make the question more specific, imposing a design change.

7. Discussion and Conclusions

In summary, the toolkit was found useful for the GBL design practice and contributed to informing and introducing the GBL concepts to the participants during the ideation process. In a short period of time, teams could ideate, develop, refine, illustrate, and document their educational game design ideas using the toolkit artifacts.

This section discusses some of our approach's strengths and limitations by reflecting on the results from user studies and identifying which elements facilitated the ideation process of learning game design using the toolkit. These reflections can serve as useful design lessons and guidelines for designing similar GBL ideation tools, and finally, we conclude the paper with directions for future work.

7.1. Strengths of the Design Toolkit

There are five main successful aspects of transforming the framework into such a design tool that facilitated the collaborative design process of learning games.

1. Easy to use in practice: The structured design activities systematically break down the creative process into individual steps that are easier to understand and operate. The cards, on the other hand, supported users to carry out the individual tasks. This is consistent with results from previous card-based tools, e.g., [15,26]. The cards helped the participants recognize that several elements combine to make an effective learning game and further helped them identify these essential elements. The team can shuffle through cards (owing to their tangible form [4,19]) to select them to cover the important aspects until they feel satisfied with their idea.
2. Stimulate brainstorming and creative thinking: All participants found trigger cards useful (none disagreed) for stimulating creative thinking and as a kick start for brainstorming. They not only provided the existing ideas but also helped generate new ones. These results are in line with previous research on design cards [19,26]. Some teams would select a trigger card to elaborate on the idea with team discussion and end up combining the trigger card with a custom card to generate a new idea.
3. Creative elements in the toolkit generate fun: The majority of participants considered that trigger and primary cards were more useful than the reflection cards, which can also be explained with the results for reflection activity that was considered comparatively less fun than idea generation and development (see Table 4). The fun element was led by the creativity involved in the design activities. The design activities which required more creativity were considered more fun (even if they were lengthy or less easy to understand at first) as compared to activities like reflection and illustration, which were comparatively less creative and required more critical and analytical thinking, were comparatively less fun (although they were fun for more than 60%). Lastly, the documentation activity was the least fun part, although it was the easiest to understand.
4. Guide the design process in a playful manner: Cards and design activities together provided a structured path that offered guidance on how to proceed with the design process. They give a clear direction and order by providing guidelines to follow five steps (design activities) and building blocks to use (different card decks). The use of different types of cards was successful for individually supporting each design activity, introducing new elements specific to that step not only guided that activity but also added newness and individuality avoiding them to become

boring. Participants were engaged in exploring new cards to achieve a new goal. Each card type was useful for their specific design task, and the card content was useful and easy to understand. Therefore, the results confirmed that the cards were useful for idea generation, development, and refinement, which is in line with the previous finding [17,26]. The majority of the participants enjoyed using the cards (74%) and thought that design activities were useful and fun (85–88%).

5. Inform and encapsulate theoretical concepts: The primary cards were useful for informing and encapsulating theoretical GBL design concepts (only 9% participants disagreed). Such an assessment is similar to previous findings by [24]. The majority of the respondents (71–74%) thought that they considered elements they might have overlooked otherwise, and the information on the cards was useful. The cards' information acts as a quick reminder for designers to the related knowledge/experience, which helps them focus on "all GBL aspects" during idea generation, development, and refinement resulting in a more concrete design. Using all six primary cards in the first activity resulted in a strong foundation, as the initial design idea comprised all six GBL aspects to expand on in the next activity. One of the participants praised the potential of the toolkit for academia: "This can be used by the teachers in the learning game design course since it explains all the important dimensions".

7.2. Some Design Decisions that Proved Helpful

Unlike other design tools, we combined the playbook within the board to simplify its use and provide a structured step by step guidance in combination with the board layout. The board size was kept moderate. These design alternations proved helpful as it reduced the effort of handling two artifacts and managing large space and provided one point of reference for both layout and play procedure. The moderate board size encouraged participants to keep cards in place, which limited other projects [26]. Further, we used ideation sheets for each activity that provided the layout for used/selected cards and kept them organized in one place. The use of ideation sheets and a log sheet is a novel feature of this toolkit, which is not present in previous work and was found very useful for tracking each team's design process and capturing the design decision rationales. These artifacts can be used to facilitate awareness and traceability of the design process which is vital for the design practice [34]. It is helpful to use all idea sheets from previous activities to form a comprehensive and meaningful description of the discussed ideas. These idea sheets serve as a useful visual reminder and help form a story around the overall game idea that ensures that all team members share the same understanding of the game idea. Each design activity was ordered and time-bound, so a team cannot skip to the next activity without completing the prior one, and also, the order was well though the following activity required outcomes from the preceding activity to work on. This created not only the sequence but also motivation for the next step. It was also useful to have a debrief session between different design activities. It made it possible to follow the game idea's progress and change, and motivated teams to do better.

7.3. Limitations of the Study

There are a few limitations to this work. The toolkit has only been evaluated with researchers and students and not with design practitioners and learning game experts. Therefore, results are not a representation of the overall design community. There were some issues with time management and workshop organization that affected the understandability of a few activities and working with the cards. It is evident from the results in Table 4 that the first activity was most difficult to understand. The observed issue can be mitigated by simple modification in the workshop technique, such as letting participants explore the toolkit for five minutes instead of directly jumping into the design process and giving breaks between [4]. The first author has predominantly led the workshop. Therefore, we cannot conclude about the level of supervision needed when other researchers use the toolkit. However, the second facilitator was different in the three conducted sessions, and they presented different design activities indicating that the knowledge is in the toolkit and not the person introducing them, which suggests that the toolkit can also be used in settings with someone other than the lead author as the main facilitator. Moreover, although workshop sessions were recorded, this paper only focused on the participants' perception of using the toolkit collected from a questionnaire, focus group, and an observation. We triangulated the questionnaire data by confirming statements from focus group feedback and/or observations to minimize the limitations of questionnaires. However, the quality of the generated ideas was not ranked or evaluated. Hence, no conclusions can be drawn in terms of their novelty. The further work will examine video recording and toolkit artifacts focusing on generated ideas, team dynamics, and multi-dimensional focus to explore the full potential of the toolkit.

7.4. Conclusion and Future Work

This paper points to a lack of operationalizable approaches for designing learning games that integrate the research based conceptual GBL design knowledge in educational games' practical design process [38]. GBL design frameworks provide theoretical design knowledge but are challenging to use in practice without tool support [12]. Design cards are a well-accepted form of intermediate-level knowledge facilitating effective transfer from theory to practice [4,5]. However, none of the existing design cards entails complete GBL design knowledge. Therefore, to bridge this gap, we transformed the LEAGUE framework [2] into a card based GBL ideation toolkit to support the learning game design team in early design practice. The toolkit contains a set of four card deck types (Primary, Trigger, Custom, and Reflection cards) containing GBL design concepts and a workshop technique with five structured design activities that provide step-by-step guidance for the ideation process enabling team members to design learning games in a collaborative and playful manner. The cards are grouped into six key categories, each focusing on one dimension of GBL design. The results from three design workshops illustrated the toolkit's value and utility in informing and guiding educational game design in practice. The toolkit can function as both a practitioner tool and a research instrument to further the domain of GBL design. Researchers in other domains can also learn from transforming the theoretical knowledge of the framework into a lightweight card-based tool.

Future work focuses on multiple directions. We would like to revise the toolkit and workshop format based on the findings from the user study. Future work will also focus on exploring the toolkit's in-depth potential, examining how the toolkit supports multi-dimensional focus and collaboration by exploring design outcomes (toolkit artifacts and generated ideas) and video recordings (team dynamics) of ideation sessions. Future studies will involve industry practitioners to evaluate toolkit in industry-based projects with real-life constraints and more extended periods, investigate team dynamics, and track design decisions to identify design patterns leading to effective educational games. Furthermore, we also plan to use the toolkit with different workshop techniques, mechanics, and game rules to explore if it further facilitates the ideation process. We are also interested in investigating the feasibility of the developed game ideas by complementing ideation with prototyping to develop digital prototypes of the learning game design ideas generated in the workshops.

Author Contributions: Conceptualization, R.T.; methodology, R.T.; validation, A.I.W.; formal analysis, R.T.; investigation, R.T.; data curation, R.T.; writing—original draft preparation, R.T.; writing—review and editing, R.T. and A.I.W.; visualization, R.T. All authors have read and agreed to the published version of the manuscript.

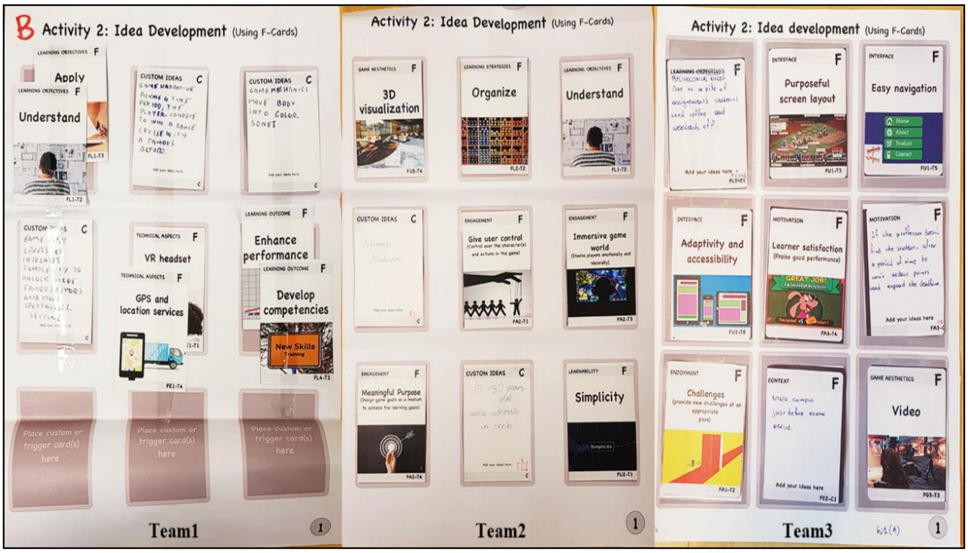
Funding: This research was funded by the Norwegian Agency for Development Cooperation (NORAD) through the EduApp4Syria project (norad.no/eduapp4syria).

Acknowledgments: Special thanks to Serena Lee-Cultura (Norwegian University of Science and Technology), who reviewed the toolkit artifacts and helped in planning, organizing, and conducting the first two design workshops. We are also thankful to fellow researchers and design workshop participants who provided valuable feedback for refinement of the toolkit.

Conflicts of Interest: The authors declare no conflict of interest.

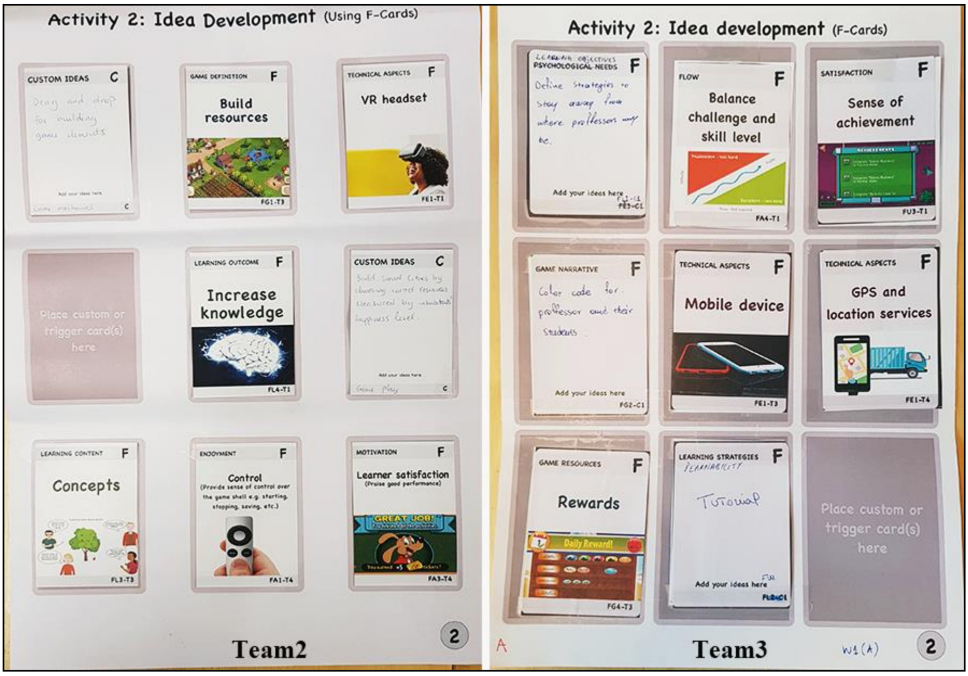
Appendix A

Figure A1. Output of the three teams for design activity one (idea generation).



(**a**)

Figure A2. *Cont.*



(**b**)

Figure A2. (**a**) Output of the three teams for design activity two (idea development). (**b**) Output of the teams (2 and 3) for design activity two (idea development) cont.

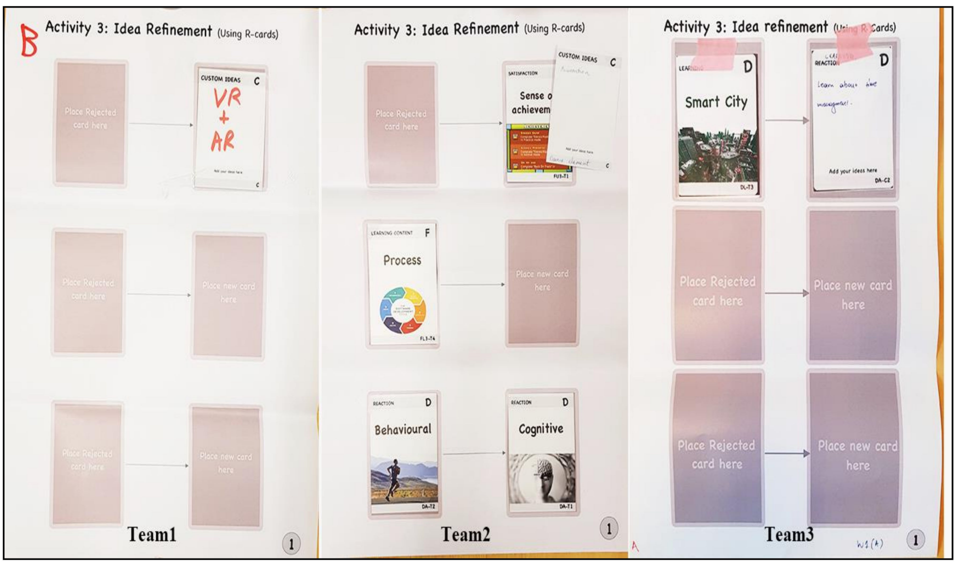


Figure A3. Output of the three teams for design activity three (idea reflection).

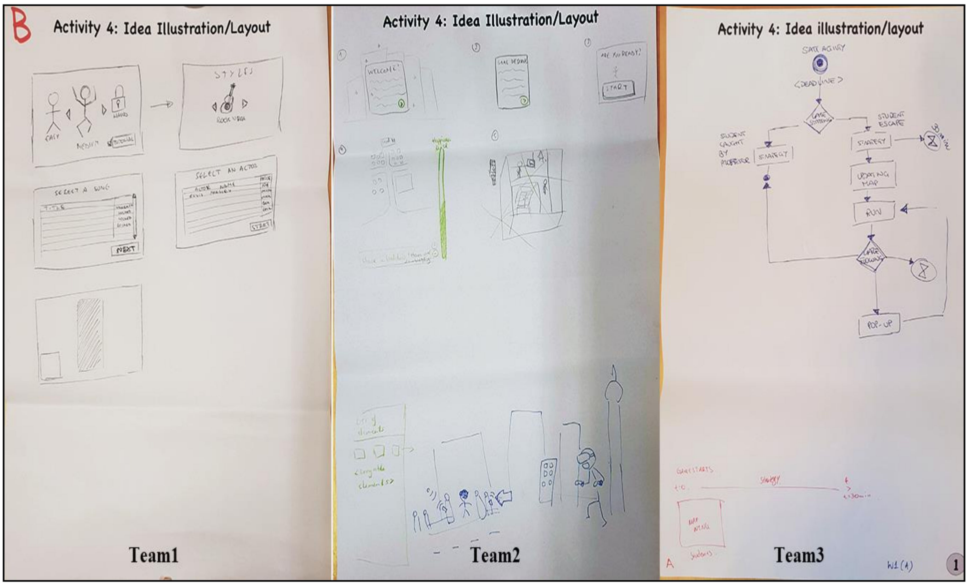


Figure A4. Output of the three teams for design activity four (idea illustration).

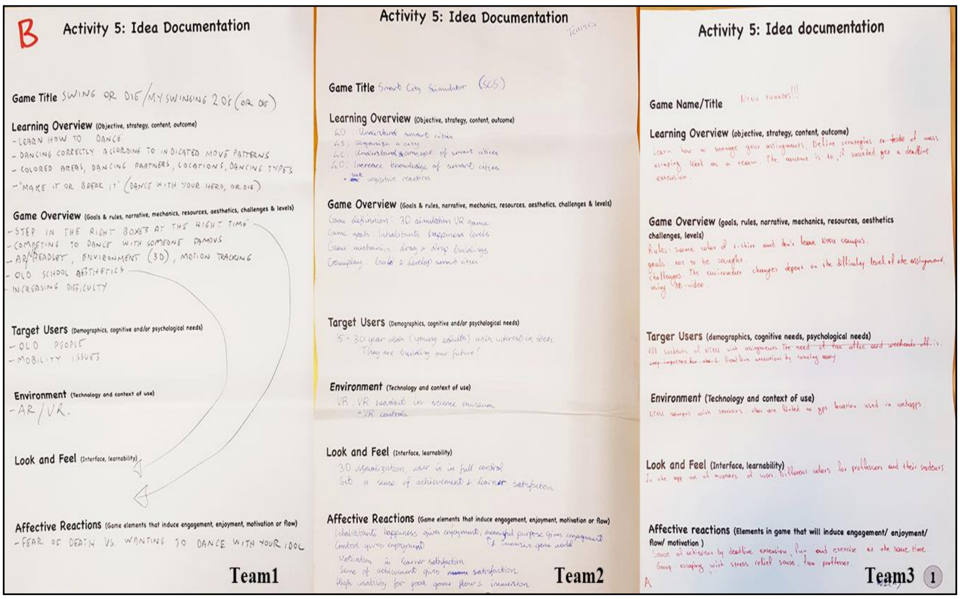


Figure A5. Output of the three teams for design activity five (idea documentation).

References

1. Tahir, R.; Wang, A.I. State of the art in Game Based Learning: Dimensions for Evaluating Educational Games. In Proceedings of the 11th European Conference on Games Based Learning (ECGBL 2017), Graz, Austria, 24–27 August 2015; pp. 641–650.
2. Tahir, R.; Wang, A.I. Codifying Game-Based Learning: The LEAGUE framework for Evaluation. In Proceedings of the 12th European Conference on Game Based Learning (ECGBL 2018), Sophia Antipolis, France, 4–5 October 2018; pp. 677–686.

3. Ávila-Pesántez, D.; Rivera, L.A.; Alban, M.S. Approaches for serious game design: A systematic literature review. *ASEE Comput. Educ. (CoED) J.* **2017**, *8*, 1–11.
4. Deng, Y.; Antle, A.N.; Neustaedter, C. Tango cards: A card-based design tool for informing the design of tangible learning games. In Proceedings of the DIS '14: Designing Interactive Systems Conference 2014, Vancouver, BC, Canada, 21–25 June 2014; pp. 695–704.
5. Höök, K.; Löwgren, J. Strong concepts: Intermediate-level knowledge in interaction design research. *ACM Trans. Comput. Hum. Interact. (TOCHI)* **2012**, *19*, 1–18. [CrossRef]
6. De Freitas, S.; Liarokapis, F. Serious games: A new paradigm for education? In *Serious Games and Edutainment Applications*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 9–23.
7. Bekker, T.; De Valk, L.; Eggen, B. A toolkit for designing playful interactions: The four lenses of play. *J. Ambient Intell. Smart Environ.* **2014**, *6*, 263–276. [CrossRef]
8. dos Santos, A.D.; Fraternali, P. A comparison of methodological frameworks for digital learning game design. In Proceedings of the 4th International Conference on Games and Learning Alliance (GALA 2015), Rome, Italy, 9–11 December 2015; pp. 111–120.
9. Hornecker, E. Creative idea exploration within the structure of a guiding framework: The card brainstorming game. In Proceedings of the TEI '10: Fourth International Conference on Tangible, Embedded, and Embodied Interaction, Cambridge, MA, USA, 25–27 January 2010; pp. 101–108.
10. Ahmad, M.; Rahim, L.A.; Arshad, N.I. A review of educational games design frameworks: An analysis from software engineering. In Proceedings of the 2014 International Conference on Computer and Information Sciences (ICCOINS), IEEE, Kuala Lumpur, Malaysia, 3–5 June 2014; pp. 1–6.
11. Theodosiou, S.; Karasavvidis, I. Serious games design: A mapping of the problems novice game designers experience in designing games. *J. e-Learn. Knowl. Soc.* **2015**, *11*, 133–148.
12. Tahir, R.; Wang, A.I. Insights into Design of Educational Games: Comparative Analysis of Design Models. In Proceedings of the Future Technologies Conference (FTC) 2018, Vancouver, BC, Canada, 15–16 November 2018; pp. 1041–1061.
13. Ahmad, M.; Rahim, L.A.; Arshad, N.I. An analysis of educational games design frameworks from software engineering perspective. *J. Inf. Commun. Technol.* **2015**, *14*, 123–151.
14. Paz, I.L.; Fernandes, F. A literature review for game design frameworks towards educational purposes. *Lisboa Play2Learn* **2018**, 321–332.
15. Mueller, F.; Gibbs, M.R.; Vetere, F.; Edge, D. Supporting the creative game design process with exertion cards. In Proceedings of the CHI '14: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, ON, Canada, 26 April–1 May 2014; pp. 2211–2220.
16. Fernández-Manjón, B.; Moreno Ger, P.; Martinez-Ortiz, I.; Freire, M. Challenges of serious games. *EAI Endorsed Trans. Game-Based Learn.* **2015**, *2*, 150611. [CrossRef]
17. Wetzel, R.; Rodden, T.; Benford, S. Developing ideation cards for mixed reality game design. *Trans. Digit. Games Res. Assoc.* **2017**, *3*, 175–211. [CrossRef]
18. Raftopoulos, M. Playful card-based tools for gamification design. In Proceedings of the OzCHI '15: The Annual Meeting of the Australian Special Interest Group for Computer Human Interaction, Parkville, VIC, Australia, 7–10 December 2015; pp. 109–113.
19. Lucero, A.; Dalsgaard, P.; Halskov, K.; Buur, J. Designing with cards. In *Collaboration in Creative Design*; Springer: Cham, Switzerland, 2016; pp. 75–95.
20. Ren, X.; Lu, Y.; Oinas-Kukkonen, H.; Brombacher, A. Perswedo: Introducing persuasive principles into the creative design process through a design card-set. In Proceedings of the IFIP conference on human-computer interaction, Mumbai, India, 25–29 September 2017; pp. 453–462.
21. Ahmad, M.; Rahim, L.A.; Arshad, N.I. Towards an Effective Modelling and Development of Educational Games with Subject-Matter: A Multi-Domain Framework. In Proceedings of the 5th International Conference on IT Convergence and Security (ICITCS), Kuala Lumpur, Malaysia, 24–27 August 2015; pp. 1–5.
22. Pivec, M.; Dziabenko, O.; Schinnerl, I. Aspects of game-based learning. In Proceedings of the 3rd International Conference on Knowledge Management (I-KNOW '03), Graz, Austria, 2–4 July 2003; pp. 216–225.
23. Olsen, T.; Procci, K.; Bowers, C. Serious games usability testing: How to ensure proper usability, playability, and effectiveness. In Proceedings of the International Conference of Design, User Experience, and Usability, Orlando, FL, USA, 9–14 July 2011; pp. 625–634.

24. Kwok, S.Y.; Harrison, D.; Malizia, A. Designing individualisation of eco information: A conceptual framework and design toolkit. *Int. J. Sustain. Eng.* **2017**, *10*, 302–312. [CrossRef]
25. Lucero, A.; Arrasvuori, J. PLEX Cards: A source of inspiration when designing for playfulness. In Proceedings of the 3rd International Conference on Fun and Games, Leuven, Belgium, 15–17 September 2010; pp. 28–37.
26. Mora, S.; Gianni, F.; Divitini, M. Tiles: A card-based ideation toolkit for the internet of things. In Proceedings of the DIS '17: Designing Interactive Systems Conference 2017, Edinburgh, UK, 10–14 June 2017; pp. 587–598.
27. Bekker, T.; Antle, A.N. Developmentally situated design (DSD) making theoretical knowledge accessible to designers of children's technology. In Proceedings of the CHI '11: CHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 7–12 May 2011; pp. 2531–2540.
28. Halskov, K.; Dalsgård, P. Inspiration card workshops. In Proceedings of the DIS06: Designing Interactive Systems 2006, University Park, PA, USA, 26–28 June 2006; pp. 2–11.
29. Kultima, A.; Alha, K. Using the VNA Ideation Game at Global Game Jam. In Proceedings of the 2011 DiGRA International Conference: Think Design Play (DiGRA '11), Hilversum, The Netherlands, 14–17 September 2011.
30. Kultima, A.; Niemelä, J.; Paavilainen, J.; Saarenpää, H. Designing game idea generation games. In Proceedings of the FuturePlay08: FuturePlay 2008 Academic Games Conference, Toronto, ON, Canada, 3–5 November 2008; pp. 137–144.
31. Schell, J. *The Art of Game Design: A Deck of Lenses*; Schell Games: Burlington, MA, USA, 2008.
32. Wölfel, C.; Merritt, T. Method card design dimensions: A survey of card-based design tools. In Proceedings of the 14th IFIP TC 13 International Conference, Cape Town, South Africa, 2–6 September 2013; pp. 479–486.
33. Brandt, E. Designing exploratory design games: A framework for participation in Participatory Design? In Proceedings of the PDC'06: Expanding Boundaries in Design, Trento, Italy, 31 July–5 August 2006; pp. 57–66.
34. Lopez, M.G.; Rovelo, G.; Haesen, M.; Luyten, K.; Coninx, K. Capturing design decision rationale with decision cards. In Proceedings of the IFIP Conference on Human-Computer Interaction, Mumbai, India, 25–29 September 2017; pp. 463–482.
35. Shaw, M. The role of design spaces. *IEEE Softw.* **2011**, *29*, 46–50. [CrossRef]
36. Adams, N.E. Bloom's taxonomy of cognitive learning objectives. *J. Med. Libr. Assoc. JMLA* **2015**, *103*, 152–153. [CrossRef]
37. Corbin, J.; Strauss, A. *Basics of Qualitative Research: Techniques and Procedures for Developing Grounded Theory*; Sage publications: Los Angeles, CA, USA, 2014; p. 358.
38. Moreno-Ger, P.; Burgos, D.; Martínez-Ortiz, I.; Sierra, J.L.; Fernández-Manjón, B. Educational game design for online education. *Comput. Hum. Behav.* **2008**, *24*, 2530–2540. [CrossRef]

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

PAPER 7

P7:

Completeness and Collaboration in the Early Design Phase of Learning Games: Do Ideation Cards Provide Scaffolding?

Rabail Tahir and Alf Inge Wang

In: Proceedings of the 2021 International Conference on Human-Computer Interaction (HCII)

Completeness and collaboration in the early design phase of learning games: Do ideation cards provide scaffolding?

Rabail Tahir [1] and Alf Inge Wang [1]

[1] Norwegian University of Science and Technology (NTNU), Trondheim, Norway
rabail.tahir,alf.inge.wang@ntnu.no

Abstract. Game-based learning (GBL) has proliferated rapidly in recent years, with both industry and academic research communities calling for collaborative work practices in the educational game design process that need to address all the key GBL aspects and create a shared understanding among team members. Design cards have the potential to improve idea generation and communication between stakeholders. However, potential scaffolding for completeness (focusing on all key GBL dimensions) and collaboration (working together to produce something) in learning game design are not explored. Therefore, in this paper, we investigate how this design approach can scaffold for collaboration and completeness in the early phase of the learning game design process using a card-based GBL ideation toolkit in design workshops. Seven teams were analyzed using design artifacts and video recordings of the workshop session. The results are encouraging in terms of the applicability of ideation cards in the GBL design process to scaffold completeness and collaboration.

Keywords: Game-Based Learning Design, Learning Games, Collaboration, Completeness, Early Design Process, Ideation.

1 Introduction

Game-based learning (GBL) is a multidimensional phenomenon depending on several aspects (e.g., users, learning, game factors, usability, environment, and affective reactions) for it to be effective [1, 2]. There is no single way to design a learning game. Not many specific methods scaffold for incorporating all the vital elements of an educational game considering different experts' involvement in its development, making it a different task [1, 3]. Researchers have pointed out that complex design products need to be understood from multiple aspects [4]. Therefore, an essential requirement for a learning game design process is to focus on all the critical dimensions of GBL [5]. We refer to this as "completeness" in GBL design. Additionally, there is a need to achieve an adequate balance between these key elements (i.e., different aspects such as learning factors, game factors, technical factors, and user experience related factors) in the learning game to keep learning integral but still providing an enjoyable user experience for learner engagement [5, 6].

The design process of learning games is complex involving several professionals from different domains working together for a common end product [7]. Therefore, team collaboration is a critical factor in determining and maintaining effectiveness in design [8]. Researchers have highlighted that team members (i.e., experts in particular domains, e.g., designers and educators) often face difficulties in sharing knowledge in a multidisciplinary setting. Each has a different area of expertise, ways to communicate, operating procedures, and use different idea representation approaches [4]. Therefore, communication between them is not very simple to manage [1]. Researchers argue that it is important for a design team to communicate and negotiate with each other to make decisions by entering compromises [9]. Industry and research communities both require collaborative work practices in the design process [9]. "Collaboration" stresses knowledge co-creation through a common design process, and peer collaboration stimulates cognitive engagement and motivation [10].

Researchers have already used design games to understand design as a social activity or for staging collaborative design efforts involving many stakeholders [9]. Playful tools and design games have been used to structure the design dialogues between stakeholders and are suggested to support and enhance collaborative ideation and concept design. The focus on play downplays the power relations and factors hindering idea generation [11]. Some researchers have used ideation cards for designing exertion games [12] and tangible games [13]. They found them useful for idea generation, articulation, offering guidance, expanding participants' horizons, focusing on the aim, formative evaluation, and providing common vocabulary. However, how completeness and collaborative process for GBL design are facilitated through ideation cards is not explored.

According to Markopoulos et al. [14], the use of novel methods in the early design phase can help adopt a broader perspective, and Lucero et al.[15] advocates that the general characteristics of design cards make them an effective tool for collaborative design practice. In this paper, we hypothesize a card-based tool as a scaffolding for collaboration and completeness in the ideation process of learning game design. We chose to focus on these elements for two reasons. *First*, considering the nature of learning game design, these are vital for GBL design practice. *Second*, they can also be used as means for learning about the GBL design process as a collaborative design activity engaging various stakeholders. Our research objective is to investigate ideation cards as scaffolding for completeness and collaboration in the early phase of the learning game design process. For this purpose, the LEAGUÊ ideation toolkit (see Section 2.4) was used as the intervention in this study. Our analysis describes how collaboration and completeness are facilitated by using the card-based tool in the ideation process of learning game design. The contribution of the paper is twofold: 1) it demonstrates the usefulness of ideation cards in the GBL design process (specifically in terms of completeness and collaboration), and 2) it reflects on factors and design decisions in the employed card deck/activities that advance the key outcomes: completeness and collaboration.

2 Related Work

This section presents relevant research studies that explored or demonstrated the importance of collaboration and completeness in the GBL design process, the use of innovative approaches to aid the design process, and card-based methods in various domains. Moreover, we underline efforts in the GBL domain to acknowledge areas where future research may take shape.

2.1 Collaboration and Completeness in the Design Process

Several researchers focus on "completeness" in GBL design, i.e., addressing all the essential elements of a learning game in the design process [1, 5, 16]. De Lope et al. proposed a five-stage methodology (in which the story plays a key role) suitable for designing learning games focusing on five key elements [5]. The study focused on the design phase, which structures the game with these five essential elements and proposes modeling tasks resulting in design artifacts such as diagrams or descriptive documents that can facilitate communication between design team members. Similarly, Silva [1] also presented a methodology divided into steps to support educational games' design process to be more all-encompassing. It identifies the steps required to define the learning mechanisms in an educational game starting from the topic choice and ending with the user experience. Another study by Kellner et al. [16] presented guidelines for developing adventure learning games (based on existing guidelines and frameworks) that help evoke all key aspects in the design. However, these studies are limited in scope, focusing on specific game types or lack thorough validation to provide evidence to support completeness in the design process based on generated game designs. Flexibility and the ability to work in a broader perspective are recognized as key skills required for the 4th generation industrial revolution. They should also be addressed in the field of educational game design [17]. The collaborative design emphasizes that all people are creative, and if provided with appropriate tools and settings, can effectively contribute to the design [18]. Da Costa et al. [19] described a co-design process based on a user-centered design approach in defining the concepts of a civic educational game. They relied on including the institution and users in the initial phase of the design process to provide an effective learning game. However, the results are limited in scope and showed that experience with only 4 or 5 children was productive. Tran and Biddle [7] presented an ethnographic study focusing on the studio team's day-to-day collaboration for development practices in a small company working in the domain of serious game development. Their finding emphasizes that social and technical factors influence collaboration in the development process of serious games. They found that co-location and a positive social environment facilitate the participation of different professionals in game development. The study reports on collaboration occurring within the game development team (consisting of six members) in a real context and not using any method or tool for scaffolding the team collaboration efforts. The team members had experience working together for at least six months to two years, which might have influenced collaboration. Marne et al. [20] aimed to create a language with a design pattern library based on their six facet approach that should enable the team of designers and teachers to brainstorm and communicate their ideas and work together for holistic coherence. This study's results are limited in scope to indicate support for collaboration, as initial results were with single designers (either teacher or game designer working alone). Some researchers [18] followed an event-driven design process for co-design. In this process, the collaboration with team members is enabled through co-design events consisting of a predesigned structure, tasks, and facilitation resulting in a co-constructed understanding concerning potential designs, experiences, and context.

2.2 Use of Innovative Approaches to Aid the Design Process

Hannula and Irrmann [11] studied a design game to plan a service co-design project using video recording of interaction between an inter-organizational group of participants playing the game. The case selected for the study consisted of six players. Four out of six were from the platform provider organization, while two other players had no prior experience of the case before. The results highlight the ability of design games to scaffold for co-creation and interaction in the early phase of service co-design projects. Kayali et al. [21] used a mixed-method approach to develop informatics and society learning games with the collaboration of high school students, university students, and researchers. They employed playing research and game analysis (which require students to learn about games by playing them reflectively) to prepare students for educational game design tasks. In these tasks, they use explorative design and design thinking methods to create the game. The research advocates the possible success of playful participation (without explicitly stating the encouraging aspects) for GBL design. However, complete results are not presented, and the project was still at the early stage. Schmoelz [22] investigated playful activities in the classroom for enabling co-creativity. The classroom activity design involved students playing the C2L storytelling card game called 4Scribes to explore different ways to deal with problems and find solutions. He used qualitative data collection methods for analysis, including narrative-Socratic dialogues, gameplay videography, and field notes. The results support the use of playful classroom activities to facilitate co-creative reframing, co-creating a shared story, expressing emotions, and engaging in dialogue.

2.3 Use of Card-Based Tools in Various Domains

Card-based tools have been used in various domains to facilitate user participation and creativity [23]. According to [24], the process most supported by creativity support tools (such as design cards) is ideation or idea generation. Roy and Warren [23] analyzed 155 card-based tools, with most aiming to aid human-centered design, creative thinking, or domain specific-methods. According to the review, some scientific trials indicated the usefulness of these tools to help designers generate innovative ideas. Feedback showed that cards could provide relevant information in handy form and support the design process. However, more testing and independent trials are required to confirm their effectiveness. Bekker et al. [25] presented a card-based design tool that describes the five perspectives on play. Only two of the five lenses were evaluated, which showed promising results such as applicability for a variety of users. The cards proved inspirational for the design process (such as brainstorming and other design activities) and useful in analyzing the initial concepts, structuring information, and reflecting on design decisions. Similarly, Chasanidou [26] also presented a design tool named DEMO to design for motivation and found the use of artifacts such as cards and the structured processes as effective practices for the early phase of the design stage. Sintoris et al. [27] used a card-based gamification approach in two engineering courses to teach ideation. They examined the produced design ideas and students' opinions regarding the tool and the design process. The students showed a positive response. However, there was a contradiction between students' responses and results of the workshops, as not many innovative ideas were produced, and there were issues with the feasibility of some cards.

Pernin et al. [28] employed the tangible version of the ScenLRPG method (built on visual formalism) based on a board game to design GBL systems specific for vocational training context. They investigated the use of game mechanisms to promote GBL designers' creativity and cooperation and the effectiveness of board game-based design tool. Some researchers, such as Mueller et al. [12] and Deng et al. [13], used card-based tools to support the design process of creating exertion games and tangible learning games, respectively. They got positive results from the participants' survey. However, they have not investigated if these tools facilitate completeness and collaboration in the design process from the generated game ideas and team interactions.

From the previous work, we find evidence for the importance of completeness and collaboration in the design process of learning games and the use of card-based tools to aid the design process by supporting initial idea generation, structuring information, reflecting on design decisions, offering guidance, introducing different perspectives, help in focus shift, and evaluation. However, not much work has been done exploring using a card-based approach to scaffold for collaboration and completeness, particularly in the GBL design process and investigating the contributing factors. Most of the existing card-based design tools were specific for a game genre or type, e.g., [12, 13, 28]. Therefore, they could not be used for our study as they did not incorporate the key GBL concepts, which are essential to investigate support for completeness. However, a particular tool focusing on GBL design is the LEAGUÊ ideation toolkit [29] used in this study. The motivation for using this card-based tool has been the particular focus of the toolkit on key GBL concepts.

3 Material and Methods

This section presents the research questions and approach, the LEAGUÊ card-based ideation toolkit for GBL design, research context, participants and procedure, and data collection and analysis methods.

3.1 Research Questions and Research Approach

The research goal of the user study presented in this article was to investigate how ideation cards facilitate completeness and collaboration in the learning game design process. We organized workshops as the research approach for the user study [30]. We conducted three design workshops using the LEAGUÊ ideation toolkit as the intervention to investigate card-based ideation tools as scaffolding for completeness and collaboration in the ideation phase of the learning game design process. The produced design artifacts and video recordings from the design workshop sessions were the primary data sources [31, 32]. We achieved our objective by focusing on the following two research questions:

- RQ1: Does the card-based toolkit support teams address all GBL key dimensions when ideating learning game design?
- RQ2: Which factors contribute to collaboration among team members when using the card-based tool to ideate learning game design?

3.2 LEAGUÊ Ideation Toolkit

The LEAGUÊ toolkit [29], containing four card decks, a board with a playbook, five design activities with ideation sheets, and a log sheet, is used for this study to ideate learning game design in a workshop format (see Fig. 1). It focuses on the multidimensionality of GBL design and offers cards concentrating on six key GBL dimensions (each detailing specific concepts). These six dimensions are *learning*, *game factors*, *affective reactions*, *usability*, *user*, and *environment* that need to be considered in any learning game design to be effective.

The toolkit contains the following different cards: *Primary* (28 cards), presenting 28 GBL design concepts (categorized in six GBL dimensions) in the form of a question or task, *trigger* (113 cards), providing hints and example ideas for GBL design concepts, *custom* (28 cards), blank cards to come up with own design ideas or custom solutions and lastly, seven *reflection* cards providing critical lenses or evaluation criteria to reflect on generated design ideas and further refine them.

Primary cards are the main deck of cards presenting 28 GBL concepts (the building block of learning game design) grouped in categories emphasizing the six key GBL dimensions (using color-coding). Out of these 28 primary cards (GBL concepts), there are five cards for the dimension "learning", three cards focusing the dimension "environment", five cards of "affective reactions", seven cards for "game factors", four cards for "usability" and, four cards for the dimension "user". The playing team successively selects the primary cards through collaborative discussion to ideate their learning game design using trigger or custom cards. Therefore, these cards are useful for investigating scaffolding for completeness (achieving multidimensional focus) in the ideation phase of learning game design.

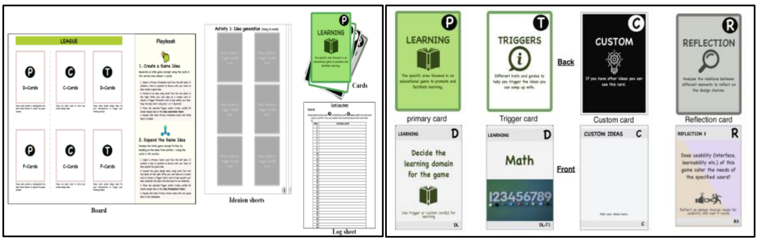


Fig. 1. The LEAGUÊ toolkit items (on the left) and four card types (on the right)

The *playbook* of the LEAGUÊ toolkit introduces five design activities for ideating learning game design in a team of four to six participants using cards, ideation sheets, and a log sheet. Each design activity has a separate ideation sheet used to produce the required design outcome of that activity. The design activities are played in sequence and are as follows: 1) *Idea generation*: coming up with an initial concept of a learning game using provided primary, trigger, and custom cards. 2) *Idea development*: expanding the initial idea from the first activity and developing it further into a more detailed and concrete one using provided primary, trigger, and custom cards. 3) *Idea refinement*: improving or refining the developed ideas by reflecting on the design choices and identifying the limitations and questionable decisions using the reflection cards to think about the trade-offs between different GBL aspects that can negatively affect the design of the learning game. 4) *Idea illustration*: planning the game's overall flow (illustrating how a user will play the game from start to exit) using a flow diagram, screen prototypes, or user scenarios. 5) *Idea documentation*: recording the final state of learning game design idea by producing a short version of a game design document (a format is provided to fill in the final idea details).

3.3 Research Context, Participants, and Procedure

The learning game design workshops (using LEAGUÊ toolkit) were organized in three different contexts: as a research study, in a doctoral summer school, and a graduate "Game development" course. In total, 34 people (ages 25-40) including, 16 master students and 18 researchers (Ph.D./postdoc), participated in the workshops that formed seven teams (each with 4-6 members). Two teams had 4 participants each, four teams had 5 participants, and one team had 6 participants. There were 13 females and 21 males. The primary subject of study was computer science for all participants except two researchers from electrical engineering. Most participants (24 out of 34) had no background in learning game design, 3 had little experience, and 7 had moderate experience. The participants with no to moderate experience were selected to fully explore the support for completeness provided by the card-based toolkit and not influenced by their experience and knowledge, ensuring data validity. The participants were selected through opportunity sampling, and none of the participants had previously used the LEAGUÊ toolkit. The participants were informed about the study's research objective, asked to sign a consent form, and were informed that their participation was voluntary.

The duration of design workshops was approximately two hours, and two organizers facilitated them. At the beginning of the workshop, the participants were given a 10-minute introduction to the LEAGUÊ ideation toolkit and key concepts of GBL. Subsequently, participants in teams were asked to start the ideation session for learning game design with five design activities. Each design activity was first individually presented by one of the organizers, followed by the teams working on that activity. One team member acted as a logger and recorded the sequence of primary and reflection cards used by the team in a log sheet during the first three activities. All activities were time-bound and organized in sequence. The first design activity (idea generation) was 10 minutes duration, and teams had six primary cards (focusing six GBL dimensions) to solve using trigger or custom cards. After that, all teams summarized their initial ideas in a minute. The second activity (idea development) was 30 minutes in which teams had 22 primary cards (categorized in six GBL dimensions) to solve using trigger or custom cards, followed by teams presenting their developed ideas in a minute. The third design activity (idea refinement) was 10 minutes, and teams had seven reflection cards to refine their ideas. Subsequently, each team in a minute reported the refinements they made in their design idea. The fourth (idea illustration) and fifth activity (idea documentation) were run in parallel (20 minutes duration in total). Finally, after completing all design activities, there were group presentations in which each team summarized the idea of their learning game design. The ideation and log sheets of teams were collected, and the play sessions of teams were also video recorded. Fig. 2. presents one of the teams using the toolkit during the workshop and their ideation sheets and log sheet.

3.4 Data Collection and Analysis

Previous work shows that participants' subjective opinion is not enough to evaluate design cards [27]. Therefore, for this study, we used the *toolkit artifacts* (ideation sheets and log sheets), along with *video recording* for collecting data to investigate the ideation process with the card-based toolkit (see Fig. 2.). For data analysis, we used descriptive statistics and the grounded theory approach [33]. The study focused on two main aspects: *completeness* and *collaboration* in the ideation process. Below we detail the data collection and analysis process for these aspects.

Fig. 2. GBL design workshop session (left side); teams' ideation and log sheets (right side)

Completeness (focus on all key GBL dimensions to ideate learning game design): means that a team must focus on and incorporate at least one or more concepts for each of the six GBL dimensions (categories) in their learning game design during the ideation process. For this study, the "completeness" is examined by investigating six key GBL dimensions (learning, game factors, affective reactions, usability, user, and environment) in the learning game design ideas produced by the teams. Although it is not essential to use all the 28 GBL concepts for ideating a learning game to achieve completeness as different concepts might be more or less relevant for different types of learning games, this thinking is in line with [34]. Nonetheless, it is crucial to cater to all high-level dimensions (looking at the game from multiple angles achieving multidimensional focus) in every learning game design to be effective, focusing on the concepts/factors deemed important for that specific game.

The toolkit artifacts were used to capture teams' design decisions to investigate "completeness" in their learning game design ideas. The team's log sheet details the order of primary cards (GBL concepts) they used in the ideation process. Each team's ideation sheets provide insights into the total GBL dimensions (out of six) covered in each activity. We used descriptive statistics to analyze the data for completeness.

Collaboration is explored by recording and analyzing the instances of interaction, discussion, and communication between team members facilitated by the card-based tool. The video recording of the play sessions (using a single fixed-point video camera next to the table, as shown in Fig. 2.) provided the data for team dynamics during the learning game design's ideation process. Here, we were interested in investigating the ability of the card-based toolkit to scaffold collaboration. We used video-based micro ethnography [11], a qualitative research method, to gather information and understand how collaboration occurred in the teams using the toolkit and the main contributing factors that initiated it. Many researchers have applied ethnography to study speech and moment-to-moment gestures in contexts such as workplaces, virtual environments, or classrooms [35, 36]. Our analysis focused on investigating the design dialogues between team members to ideate learning game design throughout the video data. One case (team) was selected for video analysis in the context of this study to focus on detailed analysis and moment-to-moment interaction. The selection was based on random sampling. We used video analysis software

V-Note Pro for analyzing the data for collaboration. The complete video recording for the selected case was 1 hour and 44 minutes in length and included the team's ideation session (consisting of design activities, debriefing, and the focus group after the workshop closing). The video analysis was guided by the process presented by Heath et al. [37]. The analysis consisted of three rounds: in the *first round*, we watched the whole video and created a content log; in the *second round*, we identified the events of interest in the data corpus; and in the *third round* of analysis, we selected the segments for detailed speech act level analysis. We selected five segments related to the five design activities for detailed analysis because these segments were most active concerning collaboration and relevant to illustrate how the tool affected the team's collaborative design process. Next, a grounded theory approach by Gioia et al. [33] is followed to model, analyze and interpret the qualitative data collected through video analysis and present it as a data structure. We coded the events and actions using the V-Note Pro tool. Events are the episodes in the video recording referring to different activities. The selected segments were coded using data-driven categories that resulted in actions. The actions are the collaborative acts (instances of collaboration) undertaken by the players within the activities. The result of the analysis is presented in the next section.

4 Results

This section presents the design workshops' results regarding the card-based ideation tool's effectiveness to scaffold for completeness and collaboration in the early phase of the learning games design process. The LEAGUÊ ideation toolkit was used as the intervention in this study to analyze the scaffolding provided by the ideation cards. The results are compiled from the ideation session of seven teams using the toolkit through five design activities to ideate the educational game design.

4.1 Research Question 1: Completeness (GBL Dimensions Covered)

This section reports the use of primary cards (28 GBL concepts grouped in six categories) and total categories (six key GBL dimensions) covered by each team in different design activities and, overall, in produced game ideas. The used primary cards detail the GBL concepts focused on by each team.

Most to Least Used GBL Concepts. Fig. 3. shows the classification of primary cards with regard to team usage. The figure highlights three categories: most used cards (that were used by more than 70% of teams), moderately used cards (used by nearly half of the teams), and less used cards (used by less than 30% of teams).

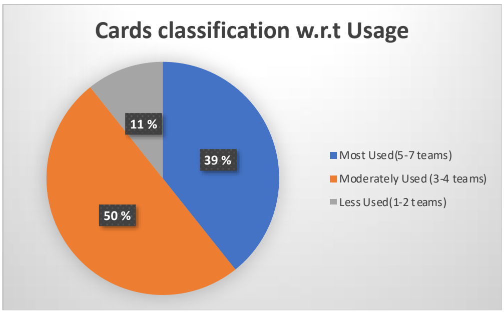


Fig. 3. Classification of primary cards (GBL concepts) according to team usage

The results from Table 1 show that learning domain, game genre, user, and learning objectives are the GBL aspects that were addressed by all seven teams, i.e., 100% usage. Following these aspects were environment, reaction, usability, gameplay, motivation, learner profile, and technical aspect considered by more than 60% of teams. The least important aspects, according to usage, were satisfaction, cognitive needs, and psychological needs, which were used by only 1 or 2 teams. It is also evident from these aspects' nature as they require much deeper focus and analysis, which is hard to realize within 30 minutes duration of the activity. Interestingly, all primary cards were used at least by one team, indicating that all GBL concepts were useful for ideation. However, it is also important to note that not even a single team used all primary cards. It does not necessarily mean that all GBL concepts are not important or required but more possibly that different concepts are more important for different types of games. Also, the time restrictions explain why not all cards were used.

Table 1. Team usage for individual GBL concepts

Primary Cards (GBL concepts)	No of Teams	Team Usage (%)
Learning domain-DL	7	100%
Game genre-DG	7	100%
Reaction-DA	5	71%
Usability-DU	5	71%
User-DÊ	7	100%
Environment-DE	6	86%
Learning Objectives-FL1	7	100%
Learning Strategies-FL2	4	57%
Learning Content-FL3	3	43%
Learning Outcome-FL4	4	57%
Game Definition -FG1	4	57%
Game Narrative -FG2	4	57%
Game Mechanics-FG3	4	57%
Game Resources-FG4	4	57%
Game Aesthetics-FG5	4	57%
Game Play-FG6	5	71%
Enjoyment-FA1	3	43%
Engagement-FA2	3	43%

Motivation-FA3	5	71%
Flow-FA4	4	57%
Interface-FU1	4	57%
Learnability-FU2	3	43%
Satisfaction-FU3	2	29%
Learner Profile-FÊ1	5	71%
Cognitive Needs-FÊ2	2	29%
Psychological Needs-FÊ3	1	14%
Technical Aspects -FE1	5	71%
Context-FE2	4	57%

Total GBL Concepts Used. Fig. 4. presents the percentage of primary cards used by each team. Primary cards are 28 different GBL concepts (grouped in six categories focusing on GBL dimensions). There are six primary cards in the first activity and twenty-two in the second, with six categories (dimensions) offered in both activities. Four teams (57%) used all six primary cards in the first activity, meaning they focused on all six categories (key GBL dimensions).

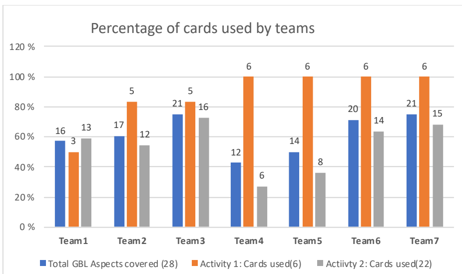


Fig. 4. Percentage of primary cards used by teams

The minimum number of cards used in activity 1 to generate a learning game idea was three (out of six) cards used by “team 1”. However, this did not affect the overall "completeness" of this team, as shown in Fig. 5. In the second activity, none of the teams used all 22 cards. The maximum number of cards was used by “team 3” (16 cards), meaning they addressed 16 GBL concepts (out of 22) in the second activity. The minimum number of cards was used by “team 4” (only six cards). Similarly, the teams' total cards also vary, “team 3” and “team 7” used 21 cards (maximum in total), and “team 4” used only 12 cards (minimum in total).

GBL Dimensions Covered. Fig. 5. (left) shows the percentage of key GBL dimensions (out of six) addressed by teams in each design activity and overall, in their produced game idea. It is interesting to note that overall, the teams addressed all six GBL dimensions in their produced game idea, which shows that the employed toolkit was useful in scaffolding for "completeness" in the GBL design process. However, it is important to note that they were not fully covered (teams used not all primary cards/GBL concepts within a category/dimension). We further analyzed the percentage of each category/GBL dimension covered by the seven teams, presented in Fig. 5. (right).

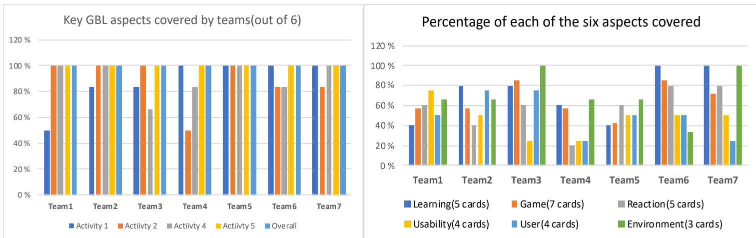


Fig. 5. GBL dimensions covered (left); percentage of each GBL dimension covered (right)

We also investigated for any common pattern (order of primary card usage) that most teams followed in developing their learning game design idea. These patterns could be useful to guide the process to other novice GBL designers. As primary cards are different GBL concepts used for building learning game design, a successful pattern could guide the GBL community regarding the best practice to tackle multiple GBL aspects for efficiency and effectiveness in the learning game design practice.

As explained before, the 28 primary cards are divided into the first two design activities: six in the first activity and twenty-two in the second activity. Therefore, to identify any pattern, we calculated the mode of "order of use of each card" for the seven teams for activities 1 and 2. Fig. 6. highlights the pattern (concerning the order of primary card usage) in activity 1 (left) and activity 2 (right).

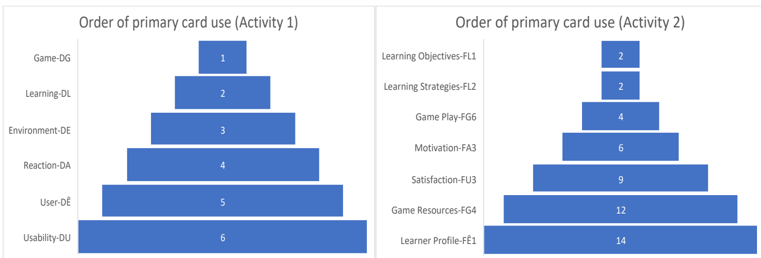


Fig. 6. Pattern in activity 1 (left); pattern in activity 2 (right)

The typical pattern identified in activity 1, "idea generation", as shown in Fig. 6. (left), was: First, the primary card "game" was used, then "learning", followed by the "environment" of the game. Further, the teams typically used the "reaction" card to address the affective-cognitive reaction the learning game intends to generate, followed by target "users", and lastly, the "usability" aspect. For the second activity, "idea development", the set of data values for only seven primary cards (out of 22) had a mode. These cards' order is presented in Fig. 6. (right), where the remaining cards had no mode value. One reason for this is the small sample size (only seven teams), and the percentage of usage for these cards was low; therefore, no frequent number was identified. Thus, we assume that it is possible to identify a clear pattern if the study is repeated with more teams (large sample size). From the identified order of use, we can see that "learning objectives" and "learning strategies" are mostly addressed at the second number (out of 22). It also means that once the vital GBL dimensions are addressed in the first activity in a specified order, most teams focus on addressing concepts related to the "learning" dimension followed by "game". Therefore, the teams' initial focus in the second activity is also on factors related to "learning" and "game", the same as the first activity. After that, "motivation" is addressed mostly in sixth place, followed by "satisfaction" in ninth place. Finally, "game resources" and "learner profile" was mostly focused on number 12 and 14, respectively.

The Produced Learning Game Design Ideas. The learning game designs ideated by the seven teams are presented below to exemplify the multidimensional focus in each idea.

Team 1 (NTNU runners!): All students at NTNU university with assignments learn to work as a team to achieve a common objective that is deadline extension by running away. The NTNU campus has sensors that are linked to GPS location used in the web game. The students work as a team and define mass escaping strategies, and the goal is not to get caught by the professors. The players use different avatars and colors for professors and students. Each student team has to use the same color of t-shirt and cannot leave the NTNU campus. The environment changes depending on the difficulty level of the assignment using 3D-videos. If the team succeeds in escaping from the professor, they get a deadline extension for that specific assignment, which gives them a sense of achievement and stress relief.

Team 2 (Math-ur-mind!): A puzzle-based mobile game for kids aged 8-12 years to understand math concepts through drill and practice using great graphics. The game can be played anywhere to develop math competencies and improve processing speed by solving exciting tasks in an interactive and fun way providing immediate feedback on actions for satisfaction.

Team 3 (Save the planet!): An outdoor tablet game for the elderly to change attitude and behavior regarding global warming and shopping behavior because they are not well informed about climate change. The game has vibration keys with easy navigation and audio features. Different interesting tasks (e.g., earn points by picking up the trash to clean the planet, shoot the plastic bags to free the planet) with constructive feedback allow the elderly to learn about recycling and mass production. The game provides fun facts on how to recycle and avoid global warming.

Team 4 (Swing or Die/ My swinging 20's (or Die)): An augmented reality game for the elderly with mobility issues to learn how to dance. The players get to dance with their idols. They have to learn to dance correctly according to the indicated move patterns shown by colored areas. The game uses an AR headset and motion tracking to indicate player to step in the right boxes at the right time to compete in dancing with some famous idols. The game uses a 3D environment with old-school aesthetics. Players can choose between different levels with various dancing patterns, locations, and dance types, along with increasing difficulty. If the player loses the competition, they will die.

Team 5 (PROGBOT): A cross-platform game for school children grades 5-7 to learn programming and related concepts. The player guides the robot through the levels by using simple symbolic programming as the main mechanic. The game has different levels on a world map, and players complete each level to conquer the area and defeat enemies. They can upgrade the robot with coins from completing levels. The player controls the robot by programming it. The game has a purposeful and consistent interface, and gameplay provides clear feedback when running the "program/solution" and encourages confidence by allowing for small growth steps.

Team 6 (Code and Conquer): A mobile game for children in primary school interested in technology to understand, apply, and develop competencies in programming skills through drill and practice and scenarios. The goal is to eliminate all the opponents. The game uses animation and tutorials and provides feedback and hints to develop competencies.

Team 7 (Smart city simulator (SCS)): A 3D simulation VR game for young adults (15-30 years old) to understand smart cities. A player uses a VR headset and VR controls to organize a smart city to increase knowledge of smart cities' concepts and seek cognitive reaction. Players use drag and drop to build buildings using 3D visualization. The game goal is to increase the inhabitants' happiness levels by developing smart cities. Inhabitants' happiness gives enjoyment and meaningful purpose, and an immersive game world gives engagement.

4.2 Research Question 2: Collaboration (Main Contributing Factors)

We followed the grounded theory approach by Gioia et al. [33] in conducting and presenting the analysis. The analysis started with finding recurring actions where collaboration occurred, forming first-order concepts (denoted as actions) from the data. We recorded the occurrences of these actions in V-Note, each with a start/end time. Hence, it was also possible to count the number of occurrences of each action during the design process, making it easier to investigate the frequency of different actions in events, in specific time intervals, or over the whole ideation session. The next step was to extract the themes guiding these actions of collaboration. Based on first-level codes (actions), we started seeking similarities and differences in the codes and grouped them to generate second-order themes (theoretical concepts from the data) explaining how codes relate to each other. The second-order themes represent the main factors contributing to collaboration among team members using a card-based toolkit. Finally, the second-order themes were compared against each other to distill them into "aggregate dimensions" that explain how card-based toolkit scaffold collaboration in the early phase of the GBL design process. The resulting data structure for collaboration among team members in ideating learning games using a card-based toolkit is shown in Fig. 7.

Our analysis resulted in the following six themes that characterize interaction in the GBL ideation process when using a card-based tool:

- *Interacting with the material*: In this theme, the contributing factor was the toolkit material that mediated the player's interactions. These instances of collaboration revolve around actions such as presenting cards to other team members, discussing different cards, working on the ideation sheets, or pointing to previous idea sheets. Interactions also included players together arranging, decluttering, or looking through cards for either initiating a discussion or further elaborating on it.
- *Focusing on Play*: The acts of collaboration in this theme were focused on play-related interactions. The players were engaged in discussing the plan, making play decisions, e.g., which aspect to take first, postponing something for later, asking questions about play rules, or explaining play rules to other players. Team members would also update each other on play status, e.g., what has been already done and what is still left.

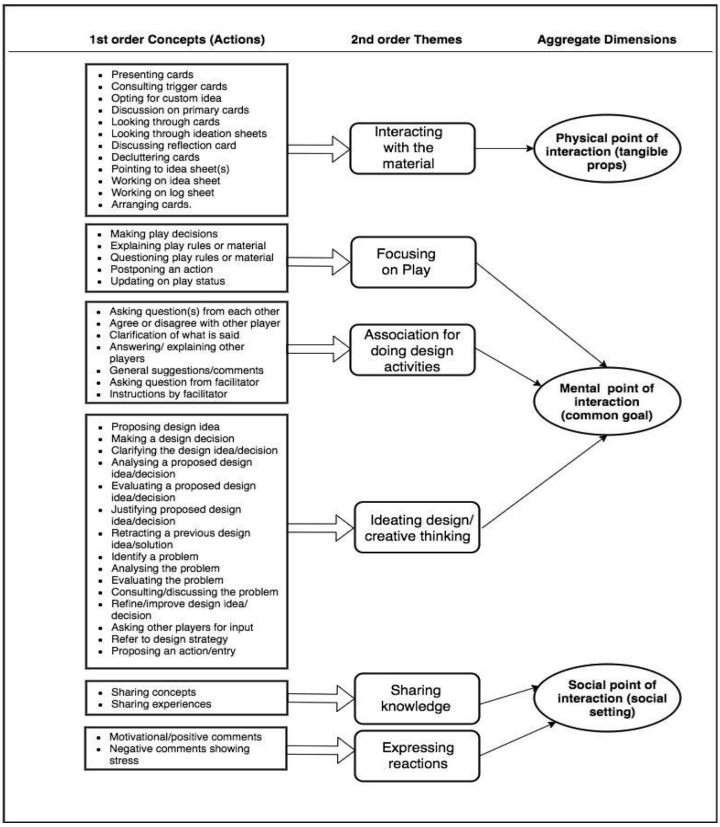


Fig. 7. The data structure for collaboration with a card-based toolkit

- *Association for doing design activities*: Here, the contributing factor for collaboration was the association needed for collectively performing the design activities. It consisted of players asking questions and giving answers to each other (e.g., Player1: Who wants to write?; Player 5: I can write); asking questions from the facilitator (e.g., Player 1: Do we build on the previous activity?); agree or disagree with other players (e.g., Player 3: You look like you like drawing!; Player 2: No no! who said.); clarifying their point in a discussion (e.g., Player 2: No, I mean this is something that is already available), or giving general suggestions or comments (e.g., Player 2: It is better to stick them on the sheets at the end; Player 4: Let's move on!). The facilitator also enabled these interactions by often giving some instructions or presenting new information or choices (e.g., you can use more sheets; you can look through other sheets for getting an overview) to the team.
- *Ideating design/creative thinking*: In this theme, all collaborative interactions were instigated by co-creating the design where the players proposed design ideas, made collective design decisions by asking other players for input or analyzing, clarifying,

evaluating a design idea. Players developed their design strategy as a team, justified proposed design ideas, identified problems, analyzed and evaluated them through discussions, or referred to previous design solutions to improve or refine the design.

- *Sharing knowledge*: The contributing factor that mediated interactions in this theme was sharing knowledge and information. Players referred to their past experiences related to topics under discussion or generally built a rapport with others. Players also explained concepts to each other they knew it would be useful in creating shared understanding and awareness.
- *Expressing reactions*: These instances of collaborations were triggered by the feelings that players experienced within the playful and collaborative setup. The team members expressed their positive and negative thoughts and reactions at different points (sometimes within an activity and sometimes at the beginning or end) that indicated their motivation or stress.

The six themes aggregated to identify the three central features of a card-based toolkit that scaffolds collaboration. A card-based toolkit provides three points of interaction that effectively instigate and foster collaboration among team members: *Physical point of interaction*, *mental point of interaction*, and *social point of interaction*. The physical point of interaction is created by tangible props that serve as director and structure the activity forming building blocks of play. The mental point of interaction is created by the common goals that serve as actors that lead to performing the stated activity, i.e., ideation of learning game design. The social point of interaction is formed by the social setting that serves as a supporter that encourages participation by providing a friendly environment.

These three points have mutually beneficial relationships that together support the collaborative design process. The tangible props provide a steppingstone and generate a physical point of interaction that supports both attaining the common goal (by posing questions and providing hints) and encouraging social interaction (by providing initial grounds for initiating interaction). The social setting provides a friendly environment making it easier to share knowledge and information, leading to improved ideation and confidence for creative thinking. On the other hand, the common goal is the driver that motivates to make an effort to strengthen all types of collaborative interactions for achieving the objective.

5 Discussion

The results discussed the two research questions concerning scaffolding provided by the card-based ideation tool for completeness and collaboration in the GBL design process.

From the analysis of generated ideas in different design activities and observation of workshop sessions, we have established that the employed card-based toolkit performed well in scaffolding for completeness. It facilitated the teams to address all six key GBL dimensions (categories) in the produced game design ideas. However, not all GBL concepts (primary cards) related to each of the six categories (dimensions) were considered. One reason is the nature of a time-bound design workshop, which restricted the freedom to complete all cards. In a real-world setup, this can be controlled by changing activity rules from time-bound to finishing cards. It is not compulsory to cover all GBL concepts within a dimension as different games might give more weight to different concepts [38, 39]. However, considering the multidisciplinary nature of GBL [40, 41], it is essential to focus on six key GBL dimensions (targeting relevant concepts within each dimension) for effective GBL design, which was successfully achieved using the card-based toolkit. All primary cards were used at least by one team, indicating that every card (GBL concept) was relevant. However, some GBL concepts were more focused than others. The learning domain, game genre, target users, and learning objectives of the game are the concepts that were addressed by all seven teams. Whereas, satisfaction, cognitive needs, and psychological needs were least focused by the teams, perhaps because these concepts require more in-depth focus and analysis, and thus more time was required.

The different categories (color-coded for easy searchability) of primary cards supported achieving multidimensionality. Since cards act as tangible idea containers, by converting the six key GBL dimensions into different card categories, the primary cards acted as design building blocks that team members used to develop and complete their design ideas from multiple angles (achieving multidimensionality in design). The final activity format required documenting all the key GBL dimensions, which served as a reminder for the team to revisit the design decisions and ideation sheets and improve their idea by working on the missing aspects. It directed the teams to focus on all six GBL dimensions in the last activity. However, it is not the only feature that led to the completeness; almost all teams focused on all six GBL dimensions in at least one other design activity in addition to activity 5 (see Fig. 5). Therefore, breaking the ideation task into different activities supports completeness. Each new activity puts things into perspective, providing an opportunity to revisit the design decisions and further add or modify them if needed.

Our study also highlighted the potential of ideation cards to facilitate collaboration among team members in the early phase of the GBL design process. We identified three aggregate dimensions from six contributing factors that facilitate collaboration in the specific context of using a card-based toolkit (see Fig. 7). The toolkit scaffolds for collaboration by providing three points of interaction in the design process: Physical point of interaction (tangible props); Social point of interaction (social setting), and Mental point of interaction (common goal/task). Individual card items' physicality makes them different from other approaches such as design model/framework or checklist by affording actions such as grabbing, pointing, and sorting or grouping [13]. Team members focused on individual items deemed important for their learning game idea, area of expertise, or previous experience to start a discussion or bookmark their ideas. The cards help participants externalize the design rationale, making the ideas concrete and more accessible to themselves and other team members [13]. Also, as each card focused on one specific GBL element, it provided a comprehensive enough description of that element (using definition, examples, or images), making it easier for all stakeholders (from different areas) to understand the concept. Moreover, it also made it easy for team members to use that tangible information to further extend and explain their ideas to other team members. This type of card-based tool also has a strong potential of being a framework for analyzing the GBL's collaborative ideation process of multidisciplinary teams.

5.1 Limitations of the Study

One of the limitations of this paper is that there was no control group to compare the results and assess the intervention's effect. We could use a control group employing some other approach (such as a checklist or framework). However, we wanted to demonstrate its effectiveness in designers' practice where no such approach is typically used. We conducted design workshops for this study instead of using the toolkit in the designer's day-to-day practice in a game studio with professionals as it was practically difficult to achieve.

However, the previous work [12] suggests that design workshops are a way to approximate design practice as it offers a similar environment with team-based design exercises and time-constrained format, similar to the environment to which designers are exposed. Another limitation of this study is that the LEAGUÊ toolkit is not representative of all GBL ideation cards. Therefore, the results are only generalizable to ideation cards presenting similiter features to LEGAUÊ or providing enough knowledge of GBL concepts. One could also argue that completeness was evaluated empirically using toolkit artifacts (counting the number of GBL concepts covered by each team). In contrast, expert evaluation could provide useful insights into the quality of generated ideas. However, for this paper's context, we were merely interested in understanding the scaffolding provided by the toolkit for achieving multidimensional focus in generated ideas (considering the learning game idea from multiple angles). The quality or effectiveness of generated ideas is important but was not the main focus of this study. Lastly, the workshop participants had no or little experience of GBL design; this was useful to explore the support for completeness provided by the card-based toolkit and not influenced by their experience and knowledge, ensuring the validity of data. It also allowed us to examine the cards' usefulness for early-career GBL designers but not senior designers.

6 Conclusion

Collaboration and completeness (considering the game from multiple angles) are vital in the GBL design process [5-7, 9] but are difficult to manage in practice [1, 3]. This paper attempts to solve this problem by postulating ideation cards as scaffolding for collaboration and completeness in the early design phase of learning games, advancing the state-of-the-art. The paper investigates the ideation process of learning games when using a card-based ideation toolkit, focusing on contributing factors and design recommendations for improvements. The data collected from the design workshops highlighted the usefulness of a card-based tool for scaffolding completeness and collaboration. All teams focused on GBL's six key dimensions in ideating their learning game design using the toolkit. The toolkit features that most contributed to scaffold completeness were different card categories and different tasks (design activities) in addition to the general characteristics of cards as tangible idea containers. The toolkit features that most contributed to collaboration were tangible props, common goals, and social setting.

Future work will focus on identifying GBL design patterns in the ideation process that can result in effective and efficient learning game designs to further help GBL designers in learning game design practice. A larger sample size is needed for this purpose. Moreover, we intend to use the toolkit with professional GBL design teams consisting of multidisciplinary experts. Future work should also focus on considering other existing card-based tools for GBL ideation and design to act as ready-made scaffolds for completeness and collaboration to validate this approach's effectiveness in the GBL design process. We will also extend the study dimensions to include creative thinking (creativity), which is also essential for the early design phase of learning games.

References

1. Silva, F. G. Practical Methodology for the Design of Educational Serious Games. *Information*, 11, 1 (2020), 14.
2. Tahir, R., Wang, A. I., Mjølsnes, S. F., Amundsen, V. E., Sivertsen, E., Time, B., Kvande, T., Sinding, M. H. S., Gopalakrishnan, S. and Vieira, F. G. *Codifying Game-Based Learning: The LEAGUE framework for Evaluation*. Academic Conferences International Limited, City, 2018.
3. Harteveld, C. *Triadic game design: Balancing reality, meaning and play*. Springer Science & Business Media, 2011.
4. Zahedi, M., Tessier, V. and Hawey, D. Understanding collaborative design through activity theory. *The Design Journal*, 20, sup1 (2017), S4611-S4620.
5. De Lope, R. P., Medina-Medina, N., Soldado, R. M., García, A. M. and Gutiérrez-Vela, F. L. *Designing educational games: Key elements and methodological approach*. IEEE, City, 2017.
6. Zemliansky, P. *Design and Implementation of Educational Games: Theoretical and Practical Perspectives: Theoretical and Practical Perspectives*. IGI Global, 2010.
7. Tran, M. Q. and Biddle, R. *Collaboration in serious game development: a case study*. City, 2008.
8. Maldonado, H., Lee, B., Klemmer, S. R. and Pea, R. D. *Patterns of collaboration in design courses: team dynamics affect technology appropriation, artifact creation, and course performance*. City, 2007.
9. Brandt, E. and Messeter, J. *Facilitating collaboration through design games*. City, 2004.
10. Kangas, M. Creative and playful learning: Learning through game co-creation and games in a playful learning environment. *Thinking skills and Creativity*, 5, 1 (2010), 1-15.
11. Hannula, O. and Irrmann, O. Played into collaborating: design games as scaffolding for service co-design project planning. *Simulation & Gaming*, 47, 5 (2016), 599-627.
12. Mueller, F., Gibbs, M. R., Vetere, F. and Edge, D. *Supporting the creative game design process with exertion cards*. City, 2014.
13. Deng, Y., Antle, A. N. and Neustaedter, C. *Tango cards: a card-based design tool for informing the design of tangible learning games*. City, 2014.
14. Markopoulos, P., Martens, J.-B., Malins, J., Coninx, K. and Liapis, A. *Creativity and collaboration in early design*. Springer, City, 2016.
15. Lucero, A., Dalsgaard, P., Halskov, K. and Buur, J. *Designing with cards*. Springer, City, 2016.
16. Kellner, G., Sommeregger, P. and Berthold, M. *Towards Guidelines for Educational Adventure Games Creation (EAGC)*. Springer, City, 2012.
17. Avouris, N., Mavrommati, I. and Sintoris, C. *Designing through ideation cards for internet of things: can cards help engineers out of the box*. City, 2018.
18. Vaajakallio, K. and Mattelmäki, T. Design games in codesign: as a tool, a mindset and a structure. *CoDesign*, 10, 1 (2014), 63-77.
19. da Costa, A. C., Rebelo, F. and Rodrigues, A. *Co-designing a Civic Educational Online Game with Children*. Springer, City, 2017.
20. Marne, B., Wisdom, J., Huynh-Kim-Bang, B. and Labat, J.-M. *The six facets of serious game design: a methodology enhanced by our design pattern library*. Springer, City, 2012.
21. Kayali, F., Schwarz, V., Götzenbrucker, G. and Purgathofer, P. Learning, Gaming, Designing: Using Playful Participation to Create Learning Games together with High School Students. *Conjunctions. Transdisciplinary Journal of Cultural Participation*, 3, 1 (2016), 1-10.
22. Schmoelz, A. On co-creativity in playful classroom activities. *Creativity. Theories–Research-Applications*, 4, 1 (2017), 25-64.
23. Roy, R. and Warren, J. P. Card-based design tools: a review and analysis of 155 card decks for designers and designing. *Design Studies*, 63 (2019), 125-154.
24. Frich, J., MacDonald Vermeulen, L., Remy, C., Biskjaer, M. M. and Dalsgaard, P. *Mapping the landscape of creativity support tools in HCI*. City, 2019.

25. Bekker, T., De Valk, L., Rijnbout, P., de Graaf, M., Schouten, B. and Eggen, B. *Investigating perspectives on play: the lenses of play tool*. City, 2015.
26. Chasanidou, D. Design for motivation: evaluation of a design tool. *Multimodal Technologies and Interaction*, 2, 1 (2018), 6.
27. Sintoris, C., Mavrommati, I., Avouris, N. and Chatzigiannakis, I. *Out of the box: Using gamification cards to teach ideation to engineering students*. Springer, City, 2018.
28. Pernin, J.-P., Mariais, C., Michau, F., Emin-Martinez, V. and Mandran, N. Using game mechanisms to foster GBL designers' cooperation and creativity. *International Journal of Learning Technology 6*, 9, 2 (2014), 139-160.
29. Tahir, R. and Wang, A. I. Transforming a Theoretical Framework to Design Cards: LEAGUE Ideation Toolkit for Game-Based Learning Design. *Sustainability*, 12, 20 (2020), 8487.
30. Ørngreen, R. and Levinsen, K. Workshops as a Research Methodology. *Electronic Journal of E-learning*, 15, 1 (2017), 70-81.
31. Savenye, W. C. and Robinson, R. S. Qualitative research issues and methods: An introduction for educational technologists (
32. Given, L. M. *The Sage encyclopedia of qualitative research methods*. Sage publications, 2008.
33. Gioia, D. A., Corley, K. G. and Hamilton, A. L. Seeking qualitative rigor in inductive research: Notes on the Gioia methodology. *Organizational research methods*, 16, 1 (2013), 15-31.
34. Carson, R. S., Aslaksen, E., Caple, G., Davies, P., Gonzales, R., Kohl, R. and Sahraoui, A. E. K. *5.1. 3 Requirements Completeness: Prepared by the Requirements Working Group of the International Council on Systems Engineering For information only; not approved by INCOSE Technical Board Not an official position of INCOSE*. Wiley Online Library, City, 2004.
35. Streeck, J. and Mehus, S. Microethnography: The study of practices. *Handbook of language and social interaction* (2005), 381-404.
36. Kohonen-Aho, L. and Alin, P. Introducing a video-based strategy for theorizing social presence emergence in 3D virtual environments. *Presence: Teleoperators and Virtual Environments*, 24, 2 (2015), 113-131.
37. Heath, C., Hindmarsh, J. and Luff, P. *Video in qualitative research*. Sage Publications, 2010.
38. Maruyama, Y., Masoodian, M. and Rogers, B. *A survey of Japanese gamers' ratings of experience elements for different game genres*. City, 2011.
39. Rapeepisarn, K., Wong, K. W., Fung, C. C. and Khine, M. S. *The relationship between game genres, learning techniques and learning styles in educational computer games*. Springer, City, 2008.
40. De Freitas, S. Are games effective learning tools? A review of educational games. *Journal of Educational Technology & Society*, 21, 2 (2018), 74-84.
41. Winn, B. and Heeter, C. Resolving conflicts in educational game design through playtesting. *Innovate: Journal of Online Education*, 3, 2 (2006).

PAPER 8

P8:
Exploring Methods and Guidelines for Child-Computer Interaction Research with Refugee Children

Rabail Tahir and Alf Inge Wang

In: Proceedings of the 2019 International Conference on Human-Computer Interaction (HCII)

Exploring Methods and Guidelines for Child-Computer Interaction Research with Refugee Children

Rabail Tahir(✉) and Alf Inge Wang

Norwegian University of Science and Technology, Trondheim, Norway
{rabail.tahir,alf.inge.wang}@ntnu.no

Abstract. There exist many guidelines and methods on how to do Child-Computer Interaction (CCI) research, but very few focusing specifically on refugee children with a challenging background. The complex situations and multiple changes refugee children undergo, including community, culture, schooling, friendships, language, war, displacement, physical violence and even identity, makes them different from children who are not refugees. They suffer learning disabilities, mental health issues, poor physical health, trust issues and overall developmental disabilities. As there are a large number of refugee children in the world, who are displaced and out of school, it is important to help these children using available technology and assess the effectiveness of the use of technology. This paper presents a literature study on available research guidelines and methods for CCI. The literature has been reviewed for guidelines and evaluation methods, starting from more general research with children, moving to more specific research with refugee children, and finally to identify gaps, present common grounds and directions for research with this specific population. The results from 55 articles reveal that although guidelines and methods for research with children can be used for refugee children, special attention and additional guidelines are needed to address specific needs of this group. Further, the review reveals a lack of CCI research and research methods for refugee children and most adapted/new children-friendly research methods are not fully employed in research with refugee children. The results of this review could serve as a starting point for researchers entering the CCI field to work with refugee children.

Keywords: Research methods · Research guidelines · Evaluation · Refugee children · Child-Computer Interaction

1 Introduction

With the emergence of Child-Computer Interaction (CCI) initiative, researchers have highly acknowledged the importance of children's viewpoint in research. Evaluation of children-friendly products also requires adapted research methods and guidelines due to the difference in children's skills, nature and complexities [1]. United Nations Convention on the Rights of the Child (UNCRC) states: "All Children and Young People who can form their own views, have a right to express those views freely in all

© Springer Nature Switzerland AG 2019
M. Kurosu (Ed.): HCII 2019, LNCS 11568, pp. 70–89, 2019.
https://doi.org/10.1007/978-3-030-22636-7_5

matters affecting them, with the views of the child being given due weight in accordance with their age and maturity" [2]. In psychology, research with children is considered more complex compared to adults, since researchers must carefully plan the data collection process to avoid additional stress, time and effort [3]. Many researchers see the need for distinguishing between research with adults and research with children which introduce additional issues [48]. Further, this research study investigates how research with refugee children distinguish itself in characteristics and context from research with children in general. More specifically, this study investigates if there are special areas you have to take into account when conducting Child-Computer Interaction research with refugee children. Our research goal is to investigate whether research guidelines and methods for refugee children must be different considering the extraordinary circumstances of this vulnerable population. The increasing number of refugees has intensified the interest of research within this population, and a need for new knowledge and understanding of this particular group [6]. This extension of research involves uncovering unique requirements relevant to the design of research protocols and ethics. Therefore, there must be particular attention on methodological and ethical dimensions in research with refugee children [7]. Some researchers have reported that refugee children suffer from high rates of mental health issues such as psychological disturbance, stress, anxiety, and learning difficulties [49–51]. Furthermore, the barriers they encounter, such as diverse traumatic experiences, different languages, parent separation, socio-economic issues, identity issues, and cultural shock, add to the special needs making them different from children without the same experiences [52]. The question here is whether these barriers and special issues infuse the need for additional guidelines and research methods for refugee children. This paper aim to address this failing by exploring guidelines and methods for CCI research, and examining, in a structured process, how it differs from research with refugee children, and by highlighting areas where future work might be required.

The literature study presented in this paper emphasized on how CCI research is carried out focusing on methods and guidelines, and we were especially interested in research where refugee children were involved. Owing to the fact that CCI began with work driven from interest in childrens' technology use within education, further extending to involvement in design and evaluation process [70] and also for this specific group (refugees) there has been a great focus on educational technology which can help these children where many do not have access to school or at least do not have an opportunity to learn to read and write their own mother tongue [21, 36, 39]. This meant that in addition to searching for literature on Human-Computer Interaction (HCI) and Child-Computer Interaction (CCI), the study also included research on educational technology including educational games. Moreover, as there is limited work on evaluation of CCI involving refugee children, this study also include literature from social science research and evaluation studies with refugee children to compile a list of guidelines and methods used with this population. The results of this review could serve as a starting point for many novice researchers in CCI community to conduct research with refugee children. The remaining paper is structured as follows: Sect. 2 describes the background, Sect. 3 explains the methodology used for the review, Sect. 4 illustrate the results with respect to research methods and guidelines, Sect. 5 presents discussion and limitations, and finally Sect. 6 concludes the paper.

2 Background

An increasing interest for children as users of technology has led to efforts to understand these users' impact on the methodology and how this influence evaluation (in terms of guidelines) where children participate [9]. This section introduces a background on research with children, specifically refugee children.

2.1 Research with Children

Samantha [71] investigated seven methodological issues to explain problems in research with children and claim that it is different because children are inherently different from adults. Other researchers highlighted the issues of verbalization and gender differences in children [1, 29]. Research with children is considered more complex as compared to adults owing to the strict requirements regarding ethical principles and preparation of environment etc. Although involvement of children in the design and evaluation process of a product is highly encouraged [4], the opinion of young children is difficult to collect and different methods have been explored for this challenging task and many new/adapted methods are devised [3, 5, 25, 26, 29].

Many researchers address research involving children with specific focus on guidelines and methods [22–28]. According to Read and Mathilde [70], CCI is a research area within HCI that grew from work mainly driven from interest in the use of educational technology with children and involving them in design and evaluation process. Druin proposed a framework for understanding the children's role in the design and evaluation process of learning technologies [10]. Jenkinson presented the shortcoming of traditional methods to measure the effectiveness of educational technology, identifying a need for more fine-grained research studies taking a flexible approach [18]. Appropriate evaluation methods are required to conduct evaluation with children [22]. Sim and Zaman proposed a method impact assessment framework that can be used by the CCI community as a critical lens for assessing evaluation methods with children [24]. Several researchers highlight methods and guidelines for usability research with children [9, 11–14]. However, research on educational game evaluation goes beyond just usability and includes constructs such as learning, flow and game factors [15]. Playing games is one of the most natural forms of learning. Children learn to talk by playing with sounds, and even learn strategic and collaborative thinking by playing games [20]. Prensky revealed that combining games with educational goals could not only trigger learning motivation but also offer interactive learning opportunities [19], which makes them relevant and important in CCI research.

2.2 Research with Refugee Children

According to the 2016 UNHCR report, the estimated number of refugees is 21 million, and half of them are less than 18 years old [7]. In recent years, refugee children who have faced experiences of war and violence have been the subject of a number of research studies [16]. The special circumstances of this group demand extra emphasis on research ethics and more careful selection of research methods [7].

What Makes Refugee Children Different? The definition of a refugee is: "A person who has been forced to leave his or her country to escape war, persecution, and natural disaster" [40]. As refugees end up in another country than their own, they face cultural challenges in addition to other problems [40]. Research shows long-lasting effects of pre- and post-displacement risk-factors on refugee children and their caregivers [7]. A number of challenges are associated with the displacement of refugee children such as experiences of trauma in the past, several overlapping transitions, and unfamiliar social setup [7, 40]. Most refugee children have interrupted education, and during their displacement they experience multiple language transitions which affect their learning, their wellbeing, and overall development. Further, many refugee children have experienced psychological and physical violence, threats of harm, separation or disappearance of family members, and have been under combat fire. Moreover, settlement and relocation produce additional stress in their lives, when these families have to compromise their needs in new environments with minimal social support facing experiences such as poverty, food insecurity, accusation, stress and discrimination [7]. These complex situations and multiple changes refugee children undergo, including community, culture, schooling, friendships, language and even identity, makes them different from children who are not refugees [7, 40].

The Role of Human-Computer Interaction (HCI) in Refugee Context? The HCI community has started to give attention to the refugee crisis leading to several initiatives developing technologies to aid refugee and assist them in their camps, and in their new relocated countries and communities [36]. Some of these contributions include: Deana and Rebecca's work to aid refugee resettlement processes by utilizing asynchronous interactive voice response and setting a translator as a mediator sharing same culture and language as the refugee [37]. Jennifer and her colleagues used field communication tags to help guide refugees through the city by providing information in their preferred language [38]. Some studies highlight that the use of smart phones is common among refugees [36]. A few technology applications have been developed to help refugees, such as "Refugee Info" to help refugees overcome the language difficulties; "Refugees Welcome" which connects refugees looking for accommodation to landlords, and "Hababy" which helps refugees find health services in Europe. However, there is very limited number of HCI studies focusing on research methods and guidelines for the context of refugee children. Reem and her colleagues identified some key deficiencies regarding the role of the HCI community in refugee context and emphasized the need to adapt HCI research methods and guidelines [36]. Most studies within HCI focusing where refugee children are involved are within educational technology and game-based learning and are described in the following section.

Educational Technology and Evaluation with Refugee Children. Some educational technology research projects have been launched for refugee children displaced by conflict, but most of these projects are in initial stages or under development, and little research has yet been published [8]. Two projects with some initial evaluation results include "Learning Sudan" - a computer game that is custom-built and offers supplementary mathematics learning opportunities to out-of-school children in Sudan [21, 36], and "EduApp4Syria" that introduces innovative smartphone educational games to improve Arabic literacy skills for Syrian children [39]. Despite the evident motivational appeal of learning technology and its effectiveness, little evaluation research has been conducted regarding the use of educational games with refugee children [8, 21]. George and his collogues developed and evaluated a reusable process for the design and evaluation of educational technology for war-affected displaced children [73]. However, most of the evaluation research conducted with refugee children comes from social science researchers exploring the complex humanitarian and political aspects in which these children live, exploring areas to improve their wellbeing, research on education of refugee children and their social and cognitive development [7]. Although it is highly emphasized that methodological dimensions and ethical engagement is crucial in research with refugee children and is identified as a challenging process [7], it has not been sufficiently addressed so far in the CCI community. To the best of our knowledge, no comprehensive research guidelines and methods have been proposed for refugee children by researchers in this field.

3 Methodology

In this study, we performed a systematic review initially with the aim of identifying and compiling research methods and guidelines for educational games evaluation with refugee children. As little CCI research is available for this specific population within the area of interest and also otherwise, we approached this research objective by investigating the extent to which research with refugee children can be regarded as similar, or different from research with children who are not refugees in terms of research methods and guidelines. The research questions include: RQ1 What evaluation methods are used for conducting research with children in CCI and how do they compare to research methods used with refugee children?; RQ2 What guidelines are used for conducting research with children in CCI and how do they compare to guidelines for research with refugee children?; and RQ3 Are there specific guidelines and methods for the refugee context in addition to those generally used with children in CCI?

The methodological approach followed the steps mentioned in [53]. The literature search was performed in five digital databases (Google scholar, ACM Digital Library, Science Direct, IEEE Xplore, and Springer Link) for conference papers, journal papers and published reports in the period from December 2017 to January 2018. The search strings used for the literature search included the keywords: “research guidelines”, “children”, “child computer interaction”, “human computer interaction”, “refugee children”, “evaluation”, “research methods”, “evaluation methods”, “educational”, and “games”. The keywords educational and games were included as we knew there were relevant CCI studies that focused specifically on these areas. Search strings were constructed using the keywords (including synonyms) based on the following criteria: (1) Methods for research with children in CCI or educational game evaluation, (2) Guidelines for research with children in CCI or educational game evaluation, (3) Methods for research with refugee children in CCI or game evaluation, (4) Guidelines for research with refugee children in CCI or educational game evaluation, (5) Methods for research with refugee children in general, and (6) Guidelines for research with children or refugee children in general. Search strings were modified and adapted for the specific syntax of each selected data source.

The article selection process included three cycles: First, an initial search using search strings to examine titles and keywords. Second, the abstracts of the papers were read for relevance, all irrelevant papers were rejected, and duplications were removed, which resulted in 129 articles. Third, the articles were filtered using inclusion/exclusion criteria resulting in 52 articles selected for this review. For an article to be included, it had to focus on one of the six criteria described above and written in the English language. The articles were also excluded if full text was not available. Since almost a year was passed until publication, the search was performed again in same five digital databases following same procedure in December 2018 to add any new relevant articles published during this year. After completing the cycles of selection process, 3 new articles were added, resulting in 55 primary studies for this review.

To ensure the quality of reviewed studies, only the articles providing sufficient information on guidelines and methods were considered. After assessing the quality of the relevant papers, data was extracted from each article and organized using a spreadsheet. The information included methods and guidelines for children/refugee children concerning RQ1 and RQ2. For RQ3, data from first two questions was further analyzed for differences to highlight specific methods/guidelines for refugee children.

4 Results

This section presents the results from reviewing 55 articles. 36 papers focused on children, and 19 papers focused on refugee children. The selected articles are listed in Table 1. We focused on the approach of investigating the extent to which research with refugee children can be regarded as similar, or different from research with children who are not refugees. After extracting data for methods (RQ1) and guidelines (RQ2) for children and refugee children separately from selected articles (see Table 1), the data was initially grouped into two main categories to initiate comparison: similarity in research methods/guidelines (methods/guidelines that were found common or similar in both corpus of literatures on research with children and refugee children) and difference in research methods/guidelines (methods/guidelines that were found uncommon or different for each corpus of literature on children vs refugee children). The main findings for each research question are summarized in the following subsections.

Table 1. Selected articles

Category	Research papers
Children methods and guidelines	[1–5, 9–13, 17, 22–35, 47, 54–58, 67–69, 74, 75]
Refugee children methods and guidelines	[7, 16, 21, 40–46, 59–66, 73]

4.1 RQ1: Research Methods with Children vs. Refugee Children

This section highlights the methods used in research with children in general as well as research methods used with refugee children. Table 2 provides a summary of methods and recommendation for use. According to the results of this literature review, three categories emerged from the content of data collected for RQ1 using inductive approach during analysis. The categories are: Preferred methods (explicitly mentioned as preferred for each target group), General methods (normally used with any user group regardless of differences), and Specific methods (used or adapted with focus on each target group). Preferred and general methods used with children with and without refugee background were mostly same and come under the category of similarity in research methods, whereas specific methods are different for children and refugee children and come under the category of difference in research methods. Furthermore, recommendations for use of each method with children or refugee children were categorized into 4 categories based on type of results provided by the selected articles regarding method usage. These categories are listed under Table 2.

Similarity in Research Methods Used with Children and Refugee Children in Reviewed Literature. First, in the category "preferred methods" for both children with or without refugee background; the methods found were the mixed method approach, the participatory method and the observation method using an observation form/checklist. However, our study found that details regarding how the methods are used with refugee children slightly differ on areas such as flexibility and the special needs of refugee group (for details see Sects. 4.2 and 4.3). Furthermore, visual methods are specifically preferred for research with refugee children, as their refugee experiences can make them silent and less expressive, and these techniques help them to speak [60]. Second, there are some "general methods" which are reportedly used with any user group including children with or without refugee background. Further, there are some recommendation found in literature for their use with children. E.g. although questionnaires are used with children, research has found that this method is not recommended as an effective child-friendly method. Quasi-experimental methods are mostly used with children for educational game evaluation employing a mixed methods approach [30, 32, 33]. However, for refugee children specifically, there is a lack in research focusing on applicability or effectiveness of employing these research methods.

Difference in Research Methods Used with Children and Refugee Children in Reviewed Literature. Third, the review results also highlighted some "specific methods" in research with children both with and without refugee background. For children these include think-aloud protocol, co-discovery, active intervention and most of the specific methods for children (see Table 2) are new/adapted methods for research with children: for example, adapted survey techniques (fun sorter, smileyometer, again-again, tangible interface), interview techniques such as contextual laddering (adapted from laddering technique), and techniques such as constructive interaction, peer tutoring and video diary. The specific methods found in literature with refugee children mostly include: clinical evaluations, case study, individual in-depth interviews and self-reports, which typically come from the social science research where focus was more on the social aspects and behaviors rather than the effectiveness of the methods used.

There is a lack of research in CCI community for this specific area. Also, there are very few new/adapted research methods for this specific group of refugee children. The review highlighted only three methods: communicative focus groups, social network mapping with group debriefing and self-report with pictorial questionnaire, which were adapted specifically for solving issues concerning research with refugee children [46].

Table 2. Research methods with children and refugee children

Children			Refugee children		
Research methods	Used w/children	Ref.	Research methods	Used w/refugee children	Ref.
Similarity in research methods used with children and refugee children in reviewed literature					
Preferred methods with children in CCI			*Preferred methods with refugee children*		
Mixed method/multi-methods	Yes	[4, 5, 13, 22, 23, 54, 55, 74]	Mixed method	Yes	[16, 21, 42, 45, 46, 61]
Participatory techniques	Yes	[26, 28, 34, 57, 58]	Participatory method	Yes	[7, 43, 46, 60, 63, 64]
Observation using checklist/observation form	Yes	[22, 55, 74]	Observation with observation form	Yes	[45]
			Visual methods	*Yes*	[46, 60, 63, 64, 73]
General methods with children in CCI			*General methods with refugee children*		
Interview (*structured*)	Yes	[12, 27, 29]	Interview (*general/semi structured*)	Yes	[16, 41, 42, 44, 45]
Experiment/quasi-experimental methods: pre-test and post/test with/without experimental and control groups	Yes	[13, 30–33]	Quasi-experimental methods: pre-post-test with/without experimental and control groups	Yes	[16, 21, 42, 66]
Observation	Yes	[4, 13, 29, 33]	Observation	Yes!	[21, 44, 46]
Questionnaire	No	[2, 13, 25, 27, 74]	Questionnaire	Yes!	[16, 42, 45]
User field test	Yes	[23, 27, 56]	User field test	Yes!	[21, 45]
Data log	Yes	[5]	Logged data	Yes!	[21]
Difference in research methods used with children and refugee children in reviewed literature					
Specific methods with children in CCI			*Specific methods with refugee children*		
Think-aloud method	Yes*	[1, 9, 12, 13, 22, 23, 29, 55]	Communicative focus groups	Yes	[46]
Video recording	Yes	[5, 13, 22, 23, 27, 74]	Social network mapping with group debriefing	Yes	[46]

(*continued*)

Table 2. (*continued*)

Children			Refugee children		
Research methods	Used w/children	Ref.	Research methods	Used w/refugee children	Ref.
Smileyometer	Yes*	[25, 27, 29, 33, 55]	Self-report with pictorial questionnaire	Yes	[42]
Drawings	Yes*	[17, 23, 28, 29, 74]	Sticky note activity	Yes	[73]
Again - Again	Yes	[25, 29, 55]	Case reports	Yes!	[16, 44]
User laboratory test	Yes	[13, 27, 56]	Wellbeing survey/computerized surveys	Yes	[42, 46]
*Photographs**	Yes	[47, 67]	Clinical evaluations	Yes!	[16, 44]
Peer tutoring	Yes	[23, 69]	Oral test	Yes!	[21]
Contextual laddering	Yes	[4, 22]	Individual in-depth interviews	Yes!	[46]
Fun sorter	Yes	[25, 29]	Self-reports	Yes!	[16, 44]
Active intervention	Yes*	[1, 22]			
Constructive interaction	Yes*	[13, 23]			
Tangible survey/tangible interface	Yes	[5, 23, 75]			
Video diary	Yes*	[57]			
Picture cards method	Yes	[68]			
Structured/unstructured checklist	Yes	[2]			

** is used with methods that fall under the subcategory (preferred, specific, general) but does not comply with the main category (similarity, difference).*
Yes**: used & recommended for children, **Yes: used with children but doubt/disagreement among researchers if recommended or not, **Yes!**: used with children but article does not mention whether it was effective or not. **No**: used with children but ineffective and thus not recommended.*

4.2 RQ2: Guidelines for Research with Children vs. Refugee Children

To a lesser or greater extent, participation in the research does influence the participants. Likewise, the research methods and the research process itself has the potential to influence the phenomenon being studied [46]. This section presents the guidelines for conducting research with children in general and specifically for refugee children. Also, for RQ2, three categories emerged from the content of data extracted for guidelines, using inductive approach during analysis. These categories are: ethical, practical and methodological. Table 3 provides a summary of these guidelines. Ethical category comprises of guidelines that focus on “ethical complexities linked with research while protecting research participants and reducing potential harms"; Practical category encompass guidelines focusing on "developing the research processes that maximize the benefits"; and methodological category contain guidelines which focuses on "adapting research methods to enhance their relevance to the specific circumstances of participants' and heighten their engagement in research."

Similarity in Research Guidelines Used with Children and Refugee Children in Reviewed Literature. The results show that some guidelines appear in both for research with children in general and in refugee context and can be considered as general guidelines for conducting research (in children context). However, deeper analysis reveals that the specific refugee context makes the application of these general guidelines different for this specific group. To illustrate this, consider the issue of obtaining consent from parents which becomes more difficult for refugee children; where the extraordinary circumstances such as separation from parents and their unaccompanied status can make parental consent impossible and further raises issues of obtaining consent from caretakers or social workers responsible, depending on local laws [65]. Similarly, for ensuring confidentiality of data collected from research participants in the case of refugee children, special attention must be paid to the ethnic culture and context, as things considered confidential in the west are public knowledge in many tight-knit communities and cultures and vice versa which might confuse the participants rather than comforting them. For example, in refugee context where many participants are not familiar with the research protocol, sometimes research respondents spontaneously reveal the adverse incidents, such as exploitation, self-harm and abuse which are normal experiences for refugees, in these cases researcher must make clear the limits of confidentiality, especially when researchers have a duty to report based on disciplinary norms [72]. Another example is of collecting video recording, where some conservative refugee societies have reservations and therefore should be further ensured of the opportunity to request destruction of videos in which they appeared [72]. In the same way, obtaining a written signed confidentiality agreement which is normal in western culture might be different in refugee context as in some cultures signing a document is considered dangerous matter and should be avoided [40]. Although the general guidelines look the same, refugee context induce additional details to implementation.

Difference in Research Guidelines Used with Children and Refugee Children in Reviewed Literature. The results of our literature review also brought forth specific guidelines for research with children and refugee children (see Table 3). Difference in specific guidelines for research with children with and without refugee background highlight that needs of refugee children are different from children with normal background. For example, in refugee context wellbeing, trust and respect becomes more of a concern than just emphasizing on fun or creativity. Instead of just focusing on simple language and limited writing you must focus on additional issues of language barriers, low literacy rates and gaining access. Furthermore, the review also highlighted that specific guidelines for research with refugee children are more focused on ethical category, which is also reflected in practical guidelines being more directed on translating the ethical reflections into practice in the research process. In contrast, specific guidelines with children in CCI have strong emphasis on methodological category in addition to ethical and practical. Whereas, no specific methodological guidelines are found in literature reviewed of refugee children that underline the lack of methodology guidelines for research with refugee children which is in accordance with the results of Sect. 4.1 (subheading difference in research methods) emphasizing the need for adapted methods for this specific group (refugee children).

4.3 RQ3: Specific Methods or Guidelines for Refugee Children

According to the review, although participatory, mixed method and observation with checklist are preferred methods generally with children with or without refugee background. However, details on using these methods with refugee children differ with focus on guidelines. Participatory and visual methods are particularly focused by many researchers as useful for refugee context in addressing the issues of power, vulnerability, ethics and language by following guidelines (Table 3) in research process [60]. The visual methods found useful for refugee children included photovoice, fotonovela, digital storytelling and quilting [60, 64]. The specific methods for refugee children were mostly found to be the general methods used in social science research with any user group such as case reports, laboratory evaluations and in-depth interviews. Most articles did not provide any details on usefulness of the employed method, which illustrate the lack of research on effectiveness of methods for research with refugee children. Unfortunately, review results did not highlight many new/adapted methods developed for refugee children, which emphasizes the need of methodology research for this specific user group. However, *communicative focus groups*, *social network mapping with group debriefing* and *self-report with pictorial questionnaire* are three specific methods found in the reviewed literature adapted specifically for the context of research with refugee children [46]. The fact that despite there are not many adapted/new methods for this specific group, the methods developed/adapted for children in general are also not yet fully employed for research with refugee children. Future research is required to explore their effectiveness for this specific group. The review highlighted only two methods: sticky note activity that used smiley faces and visual methods including photographs that were employed for refugee children considering their effectiveness as the children friendly methods.

The results highlight that there are some differences in research guidelines for children with and without refugee background (see “specific guidelines” in Table 3). The majority of the differences comes from specific ethical and practical guidelines pertaining to refugee paradigm. For refugee children there is a need for additional guidelines that take into account issues such as language barriers, culture, diverse background (illiteracy or mental health issues), refugee status (more vulnerable due to separation from family), relocation, and gaining access and reaching out to refugee communities. This review did not highlight any specific methodological guidelines for refugee children, which is in line with the results from Sect. 4.1. However, the reason for this as deduced from current review, is more inclined towards the scarcity of research in this area than concluding that no additional methodological guidelines or adapted/new methods are needed for refugee children. Most of the studies conducted with refugee children focused on the intervention results sidelining the effectiveness or outcome of methods used for research, and to a greater extent using general research methods without much discussion about method selection or their perceived impact.

Table 3. Guidelines for research with children and refugee children

Guidelines with children		Ref.	Guidelines with refugee children		Ref.
Similarity in research guidelines used with children and refugee children in reviewed literature					
General guidelines					
Ethical	Obtain consent from children and parents	[2, 26, 34, 35, 48, 57, 58]		Provide complete explanation and obtain informed consent from both children and parents or caretaker	[40, 46, 65, 73]
	Confidentiality	[2, 35, 57, 58]		Confidentiality (with respect to ethnic culture)	[40, 46, 62, 65]
	Impact of research on child/protection from harm	[2, 26, 35, 48, 58]		Protection from harm and distress	[40, 46, 62, 65]
	Build rapport	[26, 28, 48, 57]		Build trust: show interest, empathy and care	[7, 40, 46, 60, 64, 73]
Practical	Present and discuss results with children/not inflicting researchers' own perceptions	[2, 26, 28, 48]		Involve children to help researchers to interpret the findings	[64, 65]
				Feedback the research results	[62, 65]
Methodological	Conduct a pilot study	[21, 24]		Conduct a pilot study	[46, 60]
	Use appropriate methods and tools (age, language, content, gender, capability etc.)	[2, 25, 28, 48]		Use/modify methods and tools appropriate for them instead of universal standard: using standardized research instruments may be invalid when applied to different cultural groups	[40, 46, 62, 65]
	Use participatory approach	[34, 48]		Use collaborative and participatory research approaches	[46, 59, 62, 65]
	Use more than one evaluation methods	[23, 48]		Use mixed methods to engage young people with refugee background	[46, 61]
Difference in research guidelines used with children and refugee children in reviewed literature					
Specific guidelines					
Ethical	Cater children interest and allow them to be creative	[2, 26, 28, 34]	*Ethical*	Contribute to their wellbeing: research should add value to the lives of refugee children	[7, 46, 62]
	Payment or gift/reward*	[26, 57, 58]		Don't misinform them or make promises that cannot be kept	[7]
Practical	Provide assistance	[2, 25]		Work with them, not on them: treat them with respect and not just as a source of data	[7, 46]
	Make it Fun	[25, 57]		Recognize, learn and accept their diverse backgrounds (culture, religion, education, experiences etc.)	[40, 59, 60, 62, 64, 73]

(*continued*)

Table 3. (*continued*)

Guidelines with children		Ref.	Guidelines with refugee children		Ref.
	Be nice	[2, 25]		Use oral consent if written form is difficult to obtain (considering certain reservations (distrust) or illiteracy)	[40, 46, 60, 62, 64]
	Limit the writing	[2, 25]		Get approval of all procedures by ethics committee to ensure sensitivity	[46, 62, 65]
	Create an open and informal atmosphere	[2, 25]	*Practical*	Consider context and surrounding conditions of refugees	[59, 60, 73]
Methodological	Keep it short	[2, 25, 26]		Debriefing session with children and as well as caretakers after research	[62]
	Use simple language	[2, 25, 28, 57]		Thinking carefully about overall design of the research process for it to be ethical and sensitive to refugee context (research material, approach, schedule, children involvement etc.)	[64, 65, 73]
	Research context and setting (open, stress-free, child friendly environment)	[2, 28, 48, 57]		Recognize language barriers and need for a translators/interpreter	[40, 46, 62, 64, 65]
	Work in small groups	[26, 34, 57]		Flexible rather than tightly defined approach: expand the concept of 'ethical research' by applying both the relational and procedural ethical frameworks. For example, oral consent if written is not possible	[7, 59, 65]
				Ways of gaining access to refugee communities and children (collaborating with trusted members and leaders of host community)	[59, 73]

5 Discussion and Limitations

According to the review, there are several issues highlighted in CCI that demanded for new/adapted methods for research with children such as verbalization, skills, nature, gender differences, attention span, cognitive load etc. These issues were the driving force for methodology research which not only justified the need for new/adapted methods with children but also made sense to prioritize certain methods over the others. For example, researchers found that think aloud method worked only with children who can verbalize making it a difficult method to apply with children as not many children are naturally talkative [23]. Therefore, many researchers focused on active intervention method and found it effective to elicit verbal comments from children and consequently decided to combine the think-aloud method and the active intervention method which solved this issue to some extend [1]. However, another issue with children is that they are more inclined to answer what they feel adults like to hear in order to please them. This explained the reason for preferring a multi-method/mixed method approach by some researchers when working with children [54], e.g. using observation or recording children' s facial expressions and behaviors in addition to other methods used. Often nonverbal communication reveals more information than the verbal communication [1]. While some other researcher advocated the use of participatory or collaborative methods to solve this issue [23] e.g. using drawing intervention method which is considered to elicit extra information as children are involved in doing an activity that they were familiar with and in a large group, so they are more relaxed and feel less conscious when talking; or using Peer Tutoring method which require little input from the researcher and children are engaged in teaching their friends or helping them to carry out the tasks and therefore less conscious about their answers. The same rationale is true regarding the need for changes/adaptations in research guidelines in conducting research with children. For example, the issue of short attention span for children demanded for the short sessions [2, 25, 26] and the issue that children have not attained the legal right to consent required adaption in research guidelines and justified the need for the new guideline of obtaining consent from parents which has now become a standard in research with children [2, 26, 34, 35, 48, 57, 58].

Similarly, concluding from the above discussion where issues were seen as the driving force for changes and adaptations in research with children. The issues in research with refugee children as describe in this paper (Sect. 2.2) goes far beyond the general issues in research with children as a user group [52]. They have faced experiences of war and violence, dislocation, poverty, stress, discrimination, language barrier, loss of family members, difference in culture etc. These special circumstances result in learning disability, mental health issues, insecurity, distrust, physical health issues, access issues etc. Therefore, these children must be represented as special target group as compared to the general user group of children because it is impossible to ignore these specific issues and unavoidable to control their impact on conducting research with refugee children. Consequently, the above discussion implies that this group demand additional emphasis on research guidelines and ethics, and more careful selection of research methods. The prior is also depicted in the results of this review (see Table 3). For example, the issues of low literacy, distrust and dislocation in refugee context demanded adaption in research guidelines which require more flexible approach of obtaining oral consent [41], approaching them through trusted member of their community to build trust and in case of unaccompanied or separated children, it is required to gain access to local authority social worker or other officials responsible for the child in accordance with the law [65]. However, regarding research methods little has been contributed by researchers in reviewed literature but the need for such effort is highlighted by many [46, 60] which shows a lack of research in this area and a potential direction for future work for CCI community. Some researchers have highlighted the importance of visual and participatory methods in research with refugee children which to some extent solve the issues of trust, language, power and vulnerability [60]. Also, it is argued that in the context of refugee children most of the methodological challenges can be resolved by ethical reflexivity that further supports the results of this review where more focus is on ethical guidelines in research with refugee children [46]. To illustrate this, we mention the example of an adapted research method for refugee children where ethical reflexivity led the adaptation. For example, inclusion of group debriefing with hypothetical example of a social network circle with some gaps (that depicts the case of most participants) in social network mapping method solved the issues of trust and normalizing refugee experiences (missing parents or family members). However, further research is required by focusing on the effectiveness of different research methods when used with ethical reflexivity in refugee context to validate this argument. Conversely, sometimes you cannot solely rely on ethical reflexivity to guide adaptation because methodological approach is essential to solve a particular issue. To illustrate this, we give an example of another adapted research method for refugee children known as communicative focus group. Here focus group method (which resulted in simplistic responses) is adapted to solve the issues of eliciting complex experiences of refugees and addressing ethical risk of inflicting harm (through symbolic violence) by incorporating methodological approach of critical communicative methodology (CCM) and using visual prompts to stimulate discussion on issues of interest [46]. Therefore, we need further research and innovative methods in CCI to conduct research with this specific population of refugee children.

The review also highlighted that research with children focused mostly on design and evaluation of products such as educational games, prototypes, educational toys or children experiences and the constructs/aspects used for research were fun, ease to use, usability, likability, experience, attractive to use. Whereas for refugee children, research focused more on evaluation and effects of interventions, creative programs, psychosocial treatments and just recently on educational games. The constructs/aspects mostly used in research with refugee children included emotional distress, behavioral problems, learning, knowledge acquisition, wellbeing, settlement experience, perceived difficulty, cooperation, psychosocial wellbeing, mental health care, enjoyment and motivation. This difference in research focus and constructs/aspects is also depicted in the specific methods used with children and refugee children. Where most of the specific methods used with refugee children came from social science.

One of the limitations of this study could be the choice of databases and search strings used for selecting articles. Although we included articles from social science research on refugee children, we might have missed some important work and including other databases and different keywords might result in additional papers.

6 Conclusion

This paper has addressed challenges related to research methods and guidelines for CCI research with children with or without refugee background. Our literature study resulted in three identified categories of research methods: Preferred, General and Specific methods. To a large extent the methods used in research with children with and without refugee background are similar for preferred and general methods, with more variation found for specific methods (RQ1). For research guidelines we found two categories general (similar) and specific (different) guidelines. Our review also showed that even for general guidelines there are some differences in details for research with refugee children that must take additional issues into account (RQ2). Further, guidelines were introduced in the three groups ethical, practical, and methodological. Our study revealed the need to adapt guidelines for research with specific emphasis on the context of refugee children (RQ3). This need comes from specific issues such as language barrier, culture, war traumas, mental health issues, separation, and socio-economic conditions due to relocation of this population. Thus, there is a need to take into account additional ethical, practical and methodological parameters when conducting CCI research with refugee children to make sure the results of introducing technology includes a good understanding of its users. Unfortunately, only three new or adapted research methods were found in review specifically for refugee children, but there are some preferred and specific methods used with this population which we have highlighted and can guide researchers.

The review also highlighted some gaps in current literature: Firstly, there is a lack of research on new/adapted research methods for refugee children and/or effectiveness of general research methods when used in this context. Secondly, most children-friendly research methods are not fully employed in research with refugee children, and existing evaluation methods that work well with children might need to be adopted or tailored before they can be used with refugee children. Thirdly, there is a gap in literature regarding focus on methodological guidelines for the specific group of refugee children which is in line with the scarcity of research on effectiveness of methods for research with this user group. However, this study presents a starting-point to guide researchers and evaluators in the CCI community in conducting research with the specific population of refugee children and, methods and guidelines identified in this review for working with refugee children might be helpful to guide the adaption of the research process.

References

1. Zaman, B., Geerts, D.: Gender differences in children's creative game play. Young People and New Technologies, Northampton (2005)
2. Shaw, C., Brady, L.-M., Davey, C.: Guidelines for research with children and young people. National Children's Bureau Research Centre, London (2011)
3. Lanna, L.C.: How to collect audiovisual data in a research with very young children. Digit. Educ. Rev. **24**, 43–52 (2013)

4. Zaman, B.: Introducing contextual laddering to evaluate the likeability of games with children. Cogn. Technol. Work **10**(2), 107–117 (2008)
5. Gallacher, S., Golsteijn, C., Rogers, Y., Capra, L., Eustace, S.: SmallTalk: using tangible interactions to gather feedback from children. ACM (2016)
6. UNHCR: Missing out. Refugee education in crisis (2016)
7. Kaukko, M., Dunwoodie, K., Riggs, E.: Rethinking the ethical and methodological dimensions of research with refugee children. ZEP: Zeitschrift für Internationale Bildungsforschung und Entwicklungspädagogik **40**(1), 16 (2017)
8. Lewis, K., Thacker, S.: ICT and the education of refugees: a stocktaking of innovative approaches in the MENA region (English). World Bank Education, Technology and Innovation: Systems Approach for Better Education Results (SABER)-ICT Technical Paper Series; no. 17. World Bank Group, Washington, D.C. (2016)
9. Markopoulos, P., Bekker, M.: How to Compare Usability Testing Methods with Children Participants. Shaker Publisher, Herzogenrath (2002)
10. Druin, A.: The role of children in the design of new technology. Behav. Inf. Technol. **21**(1), 1–25 (2002)
11. Hanna, L., Risden, K., Alexander, K.: Guidelines for usability testing with children. Interactions **4**(5), 9–14 (1997)
12. Vermeeren, A.P.O.S., Bekker, M.M., Van Kesteren, I.E.H., de Ridder, H.: Experiences with structured interviewing of children during usability tests. British Computer Society (2007)
13. Als, B.S., Jensen, J.J., Skov, M.B.: Exploring verbalization and collaboration of constructive interaction with children. In: Costabile, M.F., Paternò, F. (eds.) INTERACT 2005. LNCS, vol. 3585, pp. 443–456. Springer, Heidelberg (2005). https://doi.org/10.1007/11555261_37
14. Tahir, R., Arif, F.: A measurement model based on usability metrics for mobile learning user interface for children. Int. J. E-Learn. Educ. Technol. Digit. Media **1**(1), 16–31 (2015)
15. Tahir, R., Inge Wang, A.: State of the art in game based learning: dimensions for evaluating educational games. In: European Conference on Games Based Learning, pp. 641–650. Academic Conferences International Limited (2017)
16. Möhlen, H., Parzer, P., Resch, F., Brunner, R.: Psychosocial support for war-traumatized child and adolescent refugees: evaluation of a short-term treatment program. Aust. N. Z. J. Psychiatry **39**(1–2), 81–87 (2005)
17. Light, R.: Accessing the inner world of children: the use of student drawings in research on children's experiences of game sense. In: Proceedings for the Asia Pacific Conference on Teaching Sport and Physical Education for Understanding, pp. 72–83. University of Sydney, Sydney (2006)
18. Jenkinson, J.: Measuring the effectiveness of educational technology: what are we attempting to measure? Electron. J. E-Learn. **7**(3), 273–280 (2009)
19. Prensky, M.: Digital game-based learning. Comput. Entertain. (CIE) **1**(1), 21 (2003)
20. Sung, H.-Y., Hwang, G.-J.: A collaborative game-based learning approach to improving students' learning performance in science courses. Comput. Educ. **63**, 43–51 (2013)
21. Stubbé, H., Badri, A., Telford, R., van der Hulst, A., van Joolingen, W.: E-Learning Sudan, formal learning for out-of-school children. Electron. J. E-Learn. **14**(2), 136–149 (2016)
22. Zaman, B.: Evaluating games with children. In: Proceedings of Interact 2005 Workshop on Child Computer Interaction: Methodological Research (2005)
23. Xu, D.: Design and evaluation of tangible interfaces for primary school children. In: Proceedings of the 6th International Conference on Interaction Design and Children, pp. 209–212. ACM (2007)
24. Sim, G., Zaman, B., Horton, M.: A method impact assessment framework for user experience evaluations with children. In: Proceedings of the 31st British Computer Society Human Computer Interaction Conference, p. 82. BCS Learning & Development Ltd. (2017)

25. Read, J.C., MacFarlane, S.: Using the fun toolkit and other survey methods to gather opinions in child computer interaction. In: Proceedings of the 2006 Conference on Interaction Design and Children, pp. 81–88. ACM (2006)
26. Gielen, M.A.: Exploring the child's mind-contextmapping research with children. Digit. Creat. **19**(3), 174–184 (2008)
27. Razak, F.H.A., Hafit, H., Sedi, N., Zubaidi, N.A., Haron, H.: Usability testing with children: laboratory vs field studies. In: 2010 International Conference on User Science and Engineering (i-USEr), pp. 104–109. IEEE (2010)
28. Barendregt, W., Bekker, T.: Exploring the potential of the drawing intervention method for design and evaluation by young children. In: CHI 2013 Extended Abstracts on Human Factors in Computing Systems. ACM (2013)
29. Yusoff, M.Y., Landoni, M., Ruthven, I.: Assessing fun: young children as evaluators of interactive systems. In: Proceedings of the Workshop on Accessible Search Systems Held at the 33rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval, New York, USA (2010)
30. Hung, C.-M., Huang, I., Hwang, G.-J.: Effects of digital game-based learning on students' self-efficacy, motivation, anxiety, and achievements in learning mathematics. J. Comput. Educ. **1**(2–3), 151–166 (2014)
31. Herodotou, C.: Young children and tablets: a systematic review of effects on learning and development. J. Comput. Assist. Learn. **34**(1), 1–9 (2018)
32. Bragg, L.A.: Testing the effectiveness of mathematical games as a pedagogical tool for children's learning. Int. J. Sci. Math. Educ. **10**(6), 1445–1467 (2012)
33. Sim, G., MacFarlane, S., Horton, M.: Evaluating usability, fun and learning in educational software for children. In: EdMedia: World Conference on Educational Media and Technology, pp. 1180–1187. Association for the Advancement of Computing in Education (AACE) (2005)
34. Thomas, N., O'kane, C.: The ethics of participatory research with children. Child. Soc. **12** (5), 336–348 (1998)
35. Neill, S.J.: Research with children: a critical review of the guidelines. J. Child Health Care **9** (1), 46–58 (2005)
36. Talhouk, R., Ahmed, S.I., Wulf, V., Crivellaro, C., Vlachokyriakos, V., Olivier, P.: Refugees and HCI SIG: the role of HCI in responding to the refugee crisis. In: Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 1073–1076. ACM (2016)
37. Brown, D., Grinter, R.E.: Designing for transient use: a human-in-the-loop translation platform for refugees. In: Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, pp. 321–330. ACM (2016)
38. Baranoff, J.R., Gonzales, I., Liu, J., Yang, H., Zheng, J.: Lantern: empowering refugees through community-generated guidance using near field communication. In: Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pp. 7–12. ACM (2015)
39. Norad Innovation Competition: EduApp4Syria (2017). https://www.norad.no/eduapp4syria. Accessed 2 Dec 2017
40. Kabranian-Melkonian, S.: Ethical concerns with refugee research. J. Hum. Behav. Soc. Environ. **25**(7), 714–722 (2015)
41. Candappa, M., Ahmad, M.: Education and Schooling for Asylum-Seeking and Refugee Students in Scotland: An Exploratory Study. The Scottish Government, Edinburgh (2007)
42. Rousseau, C., Drapeau, A., Lacroix, L., Bagilishya, D., Heusch, N.: Evaluation of a classroom program of creative expression workshops for refugee and immigrant children. J. Child Psychol. Psychiatry **46**(2), 180–185 (2005)

43. Fisher, K.E., Yefimova, K., Yafi, E.: Future's butterflies: co-designing ICT wayfaring technology with refugee Syrian youth. In: Proceedings of the 15th International Conference on Interaction Design and Children, pp. 25–36. ACM (2016)
44. Jordans, M.J.D., Tol, W.A., Komproe, I.H., De Jong, J.V.T.M.: Systematic review of evidence and treatment approaches: psychosocial and mental health care for children in war. Child Adolesc. Mental Health **14**(1), 2–14 (2009)
45. Stubbé, H., Badri, A., Telford, R., Oosterbeek, S., van der Hulst, A.: Formative evaluation of a mathematics game for out-of-school children in Sudan. In: Cai, Y., Goei, S.L., Trooster, W. (eds.) Simulation and Serious Games for Education. GMSE, pp. 61–79. Springer, Singapore (2017). https://doi.org/10.1007/978-981-10-0861-0_5
46. Block, K., Warr, D., Gibbs, L., Riggs, E.: Addressing ethical and methodological challenges in research with refugee-background young people: reflections from the field. J. Refug. Stud. **26**(1), 69–87 (2012)
47. Einarsdottir, J.: Playschool in pictures: children's photographs as a research method. Early Child Dev. Care **175**(6), 523–541 (2005)
48. Morrow, V., Richards, M.: The ethics of social research with children: an overview. Child. Soc. **10**(2), 90–105 (1996)
49. Fazel, M., Stein, A.: The mental health of refugee children. Arch. Dis. Child. **5**, 366–370 (2002)
50. Rousseau, C.: The mental health of refugee children. Transcult. Psychiatr. Res. Rev. **32**(3), 299–331 (1995)
51. Fazel, M., Stein, A.: Mental health of refugee children: comparative study. BMJ **327**(7407), 134 (2003)
52. Strekalova, E., Hoot, J.L.: What is special about special needs of refugee children? Guidelines for teachers. Multicult. Educ. **16**(1), 21–24 (2008)
53. Kitchenham, B., Pearl Brereton, O., Budgen, D., Turner, M., Bailey, J., Linkman, S.: Systematic literature reviews in software engineering – a systematic literature review. Inf. Softw. Technol. **51**(1), 7–15 (2009)
54. Obrist, M., Förster, F., Wurhofer, D., Tscheligi, M., Hofstätter, J.: Evaluating first experiences with an educational computer game: a multi-method approach. IxD&A **11**, 26–36 (2011)
55. Sim, G., Read, J.C., Horton, M.: Practical and ethical concerns in usability testing with children, games user research: a case study approach, p. 1 (2016)
56. Dimitracopoulou, A.: Learning environments and usability: appropriateness and complementarity of evaluation methods. In: Proceedings of 8th Conference on Informatics, Towards the Information Society, pp. 545–554 (2001)
57. Iivari, N., Kinnula, M., Kuure, L., Molin-Juustila, T.: Video diary as a means for data gathering with children-encountering identities in the making. Int. J. Hum.-Comput. Stud. **72**(5), 507–521 (2014)
58. Powell, M.A., Fitzgerald, R.M., Taylor, N., Graham, A.: International literature review: ethical issues in undertaking research with children and young people, for the Childwatch International Research Network. Southern Cross University, Centre for Children and Young People, Lismore/University of Otago, Centre for Research on Children and Families, Dunedin (2012)
59. White, A., Tyrrell, N.: Research with children seeking asylum in Ireland: reflecting on silences and hushed voices. In: Skelton, T., Evans, R., Holt, L. (eds.) Methodological Approaches, pp. 179–197. Springer, Singapore (2017). https://doi.org/10.1007/978-981-287-020-9_5

60. Due, C., Riggs, D.W., Augoustinos, M.: Research with children of migrant and refugee backgrounds: a review of child-centered research methods. Child Indic. Res. **7**(1), 209–227 (2014)
61. Boeije, H., Slagt, M., van Wesel, F.: The contribution of mixed methods research to the field of childhood trauma: a narrative review focused on data integration. J. Mix. Methods Res. **7** (4), 347–369 (2013)
62. Thomas, S., Byford, S.: Research with unaccompanied children seeking asylum. BMJ Br. Med. J. **327**(7428), 1400 (2003)
63. Su-Ann, O.: Photofriend: creating visual ethnography with refugee children. Area **44**(3), 282–288 (2012)
64. Vecchio, L., Dhillon, K.K., Ulmer, J.B.: Visual methodologies for research with refugee youth. Intercult. Educ. **28**(2), 131–142 (2017)
65. Hopkins, P.: Ethical issues in research with unaccompanied asylum-seeking children. Child. Geogr. **6**(1), 37–48 (2008)
66. Sirin, S., Plass, J.L., Homer, B.D., Vatanartiran, S., Tsai, T.: Digital game-based education for Syrian refugee children: project hope. Vulnerable Child. Youth Stud. **13**(1), 7–18 (2018)
67. Jones, S., Hall, L., Hall, M.: The photograph as a cultural arbitrator in the design of virtual learning environments for children and young people. In: Proceedings of the 2011 International Conference on Electronic Visualisation and the Arts, pp. 258–265. BCS Learning & Development Ltd. (2011)
68. Barendregt, W., Bekker, M.M.: Development and evaluation of the picture cards method. In: Proceedings of Interact 2005, Tenth International Conference on Human Computer Interaction (2005)
69. Höysniemi, J., Hämäläinen, P., Turkki, L.: Using peer tutoring in evaluating the usability of a physically interactive computer game with children. Interact. Comput. **15**(2), 203–225 (2003)
70. Read, J.C., Bekker, M.M.: The nature of child computer interaction. In: Proceedings of the 25th BCS Conference on Human-Computer Interaction, pp. 163–170. British Computer Society (2011)
71. Punch, S.: Research with children: the same or different from research with adults? Childhood **9**(3), 321–341 (2002)
72. Clark-Kazak, C.: Ethical considerations: research with people in situations of forced migration. Refuge: Canada's J. Refugees/Refuge: revue canadienne sur les réfugiés **33**(2), 11–17 (2017)
73. Alain, G., Coughlan, T., Adams, A., Yanacopulos, H.: A process for co-designing educational technology systems for refugee children. In: Proceedings of the 32nd International BCS Human Computer Interaction Conference (HCI 2018), Belfast, UK, 4–6 July 2018
74. Fleck, S., Baraudon, C., Frey, J., Lainé, T., Hachet, M.: Teegi's so cute! Assessing the pedagogical potential of an interactive tangible interface for schoolchildren. In: IDC 2018-17th Interaction Design and Children Conference (2018)
75. Veytizou, J., Bertolo, D., Baraudon, C., Olry, A., Fleck, S.: Could a tangible interface help a child to weigh his/her opinion on usability? In: 30eme conférence francophone sur l'interaction homme-machine, p. 8 (2018)

PAPER 9

P9:
Evaluating the Effectiveness of Game-Based Learning for Teaching Refugee Children Arabic using the Integrated LEAGUÊ-GQM Approach

Rabail Tahir and Alf Inge Wang

Ready for Submission

This paper is awaiting publication and is not included in NTNU Open

PAPER 10

P10:
How to Evaluate Educational Games with Refugee Children: Methodological Aspects and Lessons Learned from EduApp4syria

Rabail Tahir and Alf Inge Wang

In: Proceedings of the 2019 European Conference on Games Based Learning (ECGBL)

How to Evaluate Educational Games With Refugee Children: Methodological Aspects and Lessons Learned From EduApp4syria

Rabail Tahir and Alf Inge Wang
Dept. of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway
rabail.tahir@ntnu.no
alf.inge.wang@ntnu.no
DOI: 10.34190/GBL.19.136

Abstract: Educational game evaluation is a multidimensional and complex phenomenon. The growing interest in game-based learning (GBL) results in an increasing need to evaluate the effects of this approach, which requires appropriate methods, techniques, and principles that can be applied by the GBL community. This paper reflects on the methodological aspects of evaluating educational games with refugee children drawing on practical experience and evaluation studies conducted in the EduApp4Syria project. The paper gives an overview of the project and presents three field studies conducted, including the GBL evaluation methods used in the context of refugee children. The methods used included quasi-experimental design, mixed-method approach, observation with/without checklist, questionnaires, interviews, pre/post-test (using EGRA), screen recording, game-logs, and expert evaluation. The evaluations illustrate the application and assessment of these methods. This paper presents the findings and pitfalls related to the applicability of evaluation methods in various phases of the game development life cycle and methodological and practical challenges in conducting research and eliciting data in the context of evaluating educational games with refugee children. This article provides an up-to-date examination of both methodological challenges common to GBL evaluation and those unique to the user group of refugee children, culminating in guidance for researchers on methods and critical issues that need to be considered when designing research studies involving educational games and children. The paper assists researchers to critically reflect on these methodological issues and methods they use as they will have implications on the data obtained.

Keywords: game-based learning, language-learning games, evaluation methods, methodological aspects, children, refugee children

1. Introduction

Digital games have become the native language of children growing up in this technological era (Prensky and Berry, 2001). They learn to talk by playing with the sounds, they even learn strategic thinking and collaboration by playing games, making it one of the most natural forms of learning (Sung and Hwang, 2013). Several researchers have identified the potential of games for generating possible positive impact for learning on the digital generation (Connolly et al., 2012, Kinzie and Joseph, 2008, Prensky and Berry, 2001). Prensky, (2003) emphasized that integrating games with educational goals not only trigger motivation but also provide new interactive learning opportunities for children. However, the use of games to teach educational content brings into question their compatibility with learning which has prompted many researchers to investigate the actual benefits of games for learning (Erhel and Jamet, 2013). To assess the effectiveness of games for learning, evaluations are necessary, and the GBL approach requires a multidimensional evaluation that requires appropriate methods, techniques, and principles that can be applied by the GBL community (Tahir and Wang, 2017, Connolly et al., 2012).

The studies on the effectiveness of educational games encountered a series of methodological issues and awareness of these challenges is essential to bring research to a higher level (Vanderhoven et al., 2015). Furthermore, research with refugee children raises several additional practical, ethical, and methodological issues such as consent, access, privacy, and confidentiality. Although these are not unique to refugee children, they do present researchers with specific dilemmas to cater differences in research with refugee children that must take additional issues into account such as language barrier, culture, war traumas, mental health issues, separation, and socio-economic conditions due to relocation of this population considering their extraordinary experiences (Tahir, 2019). The increased number of refugee populations has led to the interest of research and a need for understanding and new knowledge of this particular group (Out, 2016). A child-centered approach to data collection views children as subjects rather than objects of research in order to address this difficulty (Mauthner, 1997). There exist guidelines and methods on child-computer interaction (CCI) research, but very little focus has been given to refugee children with a challenging background. The review (Tahir, 2019) highlighted the gap in the literature regarding methodological guidelines uncovering issues in research with refugee children and the effectiveness of different research methods for researching with this user group.

Hill Malcolm (Hill, 1997) argues that it is crucial that research-based publications should provide details of the employed methods and also give assessments and feedback regarding how satisfactory particular techniques are (Fargas-Malet et al., 2010). This research paper addresses this argument by presenting details of methodological underlining for evaluating learning games with refugee children based on experiences from the EduApp4Syria research project. Several evaluation studies were carried out throughout this project, and this paper points out certain pitfalls and reflects on methodological aspects along with any practical and ethical considerations of evaluating educational games with refugee children. Based on the lessons learned, several factors need to be taken into account when devising an appropriate methodology for researching with refugee children to ensure that a good understanding of users is included in the resultant introduced technology.

2. Background

This section presents the background and main goals of the project EduApp4Syria, evaluation conducted in each phase of this project and the overview of the field studies describing the main focus, participants, setup and methods used.

2.1 Research project (EduApp4Syria)

The EduApp4Syria project is an international innovation competition where the aim is to help Syrian children learn how to read Arabic and improve their psychosocial wellbeing through game-based learning apps on smartphones (Nordhaug, 2016). The motivation for the project is that 2.25 million Syrian children are out of school both within Syria and in other countries because of the conflict. Many Syrian children who attend school face difficulties in learning, because they have endured long-term stress or because they are being taught in a language they do not master. There is a risk that a whole generation grows up that cannot read or write in their mother tongue. The EduApp4Syria competition is funded by the Norwegian government and coordinated by the Norwegian Agency for Development Cooperation (Norad) in cooperation with Norwegian University of Science and Technology, All Children Reading, USAID, World Vision, the Australian Department of Foreign Affairs and Trade, Orange, INEE and UNICEF Ventures.

2.2 Evaluation in each phase of the project

The EduApp4Syria project is organized as a multi-stage innovation competition over three phases, as shown in Figure 1. Before the competition launched on January 1st, 2016, a field study was conducted among Syrian refugees in Istanbul and Gaziantep in Turkey. The goal of this field study was to check the status quo on game-based learning apps for Syrian children, to elicit user requirements for the apps and to understand the situation and context of Syrian refugees. Before the launch of the competition, a jury of experts with experts within literacy, psychosociology, game-based learning, e-learning, intellectual property (IP), Arabic language, and Syrian culture was established. The jury was selected based on their credentials and their experience with similar projects. To establish a common understanding of the project and its aims, the jury was introduced to the problem and theories on GBL. A provided checklist and evaluation form, based on theories of intrinsic motivation, cultural and language suitability, and technical requirements, were used by the jury to select games for the next phases. After the three phases of submission and evaluations, two winning games were selected: Feed The Monster, which is a game where the player feeds monsters letters, words and sentences, and Antura and the Letters, which is a collection of mini-games playing with alive and animated letters and the shepherd dog Antura.

2.3 Overview of the three field studies in the project

This section provides an overview of the three field studies carried out in the EduApp4Syria project conducted at different phases of the project, where the main focus was to investigate the effectiveness of GBL approaches for refugee children based on usability, learning, engagement, game elements and technical and user requirements.

2.3.1 Field study one: User test in Trondheim, Norway, August 9th and 10th, 2016

The goal of this user test was to evaluate the enjoyment, appearance, audio and perceived learning for five learning games from phase one (see Figure 1) and pick the three best ones to be funded for the next phase. The user test was organized through the qualification program for immigrants in Trondheim municipality, and 50 Syrian children were recruited for the test. The user tests were organized over two days in five sessions, where ten children participated in each session. In each session, a pair of children would play through five games. They would play one game for 12 minutes; then answer a questionnaire and be asked about the game for 3 minutes before they would continue with the next game. After playing through all five games, the pair of children would be asked to rate all five games and be interviewed about their experiences with the games. For each session, five pairs of children were organized at five tables where the observer and if required, an interpreter would sit. Many of the children were fluent in Norwegian and did not need an interpreter. The evaluation team consisted of the head of the jury (GBL and technical expert), the Syrian language and cultural expert of the jury, and three Norad-employees. The head of the jury introduced the team to the evaluation and explained the evaluation process and tools. The observer's responsibility was to aid the children in starting the game, observe when they were playing and interview the children. Each group in a session started with a different game, to account for the perception of games due to the order they were played. Each table had one smartphone running the games, which meant that one child would play while the other watched. The observer asked the children to change on who was playing. The questionnaires used for each game had four questions. Parts of the forms are shown in Figure 2**.** The observers were also asked to observe the children, and some children were video recorded. The jury ended up picking the two games the children ranked as best, and one of two games the children ranked as number three and four with similar scores.

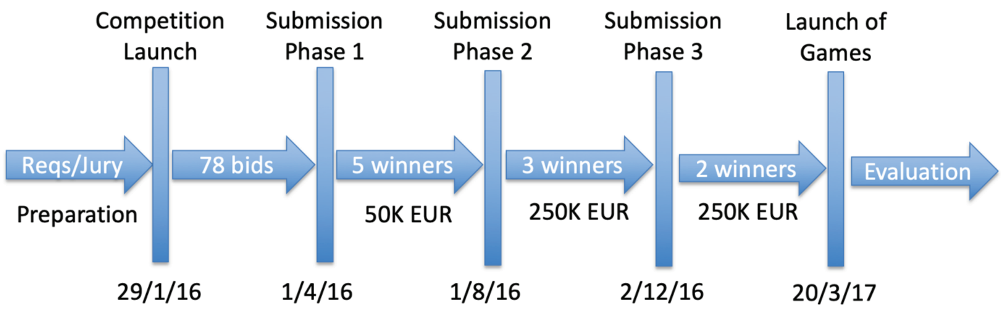


Figure 1: Overview of the EduApp4Syria multi-stage innovation competition

Figure 2*:* Left: Part of the Per-game evaluation form. Right: Part of the final evaluation form

2.3.2 Field study two: User test in Amman, Jordan, December 8th, 2016

In the second user test, three learning games were evaluated in one day. Twenty-eight children and parents were invited to four sessions, where seven children and their parents participated in each session. Each session had seven tables where one child could play all three games. Every child was asked to play each game for 25 minutes, then have a 5-minute break, before playing the next game until all three games had been played. For this user test, the sequence of which games to be played changed for each session. No questionnaires or forms were used for the children for this test, only observation. The observers were allowed to help the children after some time if the children got stuck or if the app crashed. The focus of the user test was to assess the usability and engagement of the games, and more specifically to investigate whether the games were appropriate for the target audience, if the game interfaces were fun and engaging, if the games were easy enough to use and understand, if the games held the user's concentration over a period of time, if the game matched the users' literacy skills, if the game rewarded the user appropriately, if the user had a sense of control of the game, if the user understood the goal of the game, if the feedback was appropriate, and if the user got emotionally involved and immersed. The observers were asked to look for positive and negative responses in forms of facial expressions, body language, laughter, celebrations, shout-outs, focus, play without help, concentration, engagement, frustration, anger, boredom, lost focus or interest, need for help, giving up, and stop playing the game. The observers were given an observation form, guidelines for filling out the observation forms, and an example observation form. Between games, the children were interviewed about their experiences. The parents of the children were also asked to play the games and were afterward interviewed about whether they thought the games were appropriate and useful for the children, and if they would let their children continue to play such games.

2.3.3 Field study three: User test in Trondheim, Norway, April 7th to May 6th, 2018

In the third user study, the final released version of one of the two winner games (Feed the Monster) was evaluated with migrant refugee children for one month. The focus of this user test was on learning gain, game factors that generate affective cognitive reactions, engagement, enjoyment, usability, and user characteristics that influence these factors. A quasi-experiment design was used for this study with a sample size of 30 children between 5 to 10 years old who could speak but could not read or write Arabic. Data collection enabled appropriate methods to be tested and combined; which included a demographic questionnaire, a VARK questionnaire, a pre/post-test using EGRA, an observation using a checklist, game logs, screen recordings, usability tasks, follow-up interviews, and a follow-up questionnaire. The user test was organized through the weekend Arabic class program for children in the Muslim Society in Trondheim (MST) which is a non-profit, religious and cultural organization that aims to serve the interests of the Muslim community in Norway. The children and their parents were contacted with the help of two teachers of the weekend class program, and a total of 30 Syrian children were recruited for the user study. The user tests were organized over one month in nine sessions, where 3-5 children participated in each session. The sessions took place twice a week, on Saturday and Sunday. The experiment was designed to be one week long and consisted of two parts: a playtest session and 1-week play at home. The playtest session started with a pre-test (based on EGRA), followed by the game play session and a short follow-up interview at the end. After this session, parents were handed mobile devices with the game installed to let children play the game at home daily for at least 20 minutes for one week. Many of the parents were not fluent in English, and an interpreter was required who could speak both Arabic and English. A translator and three to four observers/evaluators (2 experts in GBL and 2 novice) participated in each session. The observers were given an observation checklist, guidelines for filling out the observation forms, and an example observation form. The session had several tables where one child would sit with one observer to play the game. Every child was asked to play the game for 20-25 minutes, then have a 5-10-minute short follow-up interview.

3. Lessons learned

This section presents the results in terms of the lessons learned and the experience of research with refugee children obtained from the project EduApp4syira. The findings are categorized according to two research questions:

- *RQ1*. What is the applicability of evaluation methods in different phases of educational game design and development for refugee children?
- *RQ2*. What are the issues in methodology and practical and ethical challenges in conducting research and eliciting data when evaluating educational games with refugee children?

3.1 Applicability of methods in various phases of the game development life cycle (RQ1)

During the various phases of educational game development, there is a set of evaluation methods that can be applied. Table 1 presents an overview of different evaluation methods and their applicability in various phases in educational games design and development for refugee children based on the experience from the EduApp4Syria project (also applicable for other user groups). The *Concept phase* of the project focused on constructs such as project plan, learning components embedded in the game, expected effect on psychosocial wellbeing, and user requirements. The *Pre-Production* focused on constructs such as enjoyment, appearance, audio, perceived learning, *Production* focused on usability and engagement, and *Post-Production* focused on learning gain, game factors that generate affective cognitive reactions, engagement, enjoyment, usability and user characteristics that influence these factors.

Expert evaluations are more feasible in the early phases of development as educational games require a multi-disciplinary approach from the early concept to ripe all the benefits. Whereas, methods such as observation, experiments, game logs, pre/post-test, screen recordings are particularly useful in later phases, providing an opportunity for impact evaluation with user data. The game logs can be useful for monitoring long term engagement and learning, especially for playing in a real context. Methods such as interviews, questionnaires, and user testing can be employed in multiple phases with slight variations in details. Observations without a checklist are beneficial in the earlier phases of development when the objective is more exploratory, and it helps uncover problems and provide insights into the children's experiences. In later phases, it is more effective to adopt observation with a checklist to be more focused and concrete. Interviews with children should be kept short, it is useful also to conduct interviews with parents to get useful insights about childrens' game usage and this also increases parents' confidence in learning games. Questionnaires are not very effective concerning data obtained from children. However, it is useful to obtain children demographics and learning preferences data from parents. User testing is particularly useful to discover bugs in the production phase and also later to test usability and engagement. Based on our experience from this project, it is recommended to use a multi-method approach when researching with children. Different methods provide a different level of details and insights, and it is often useful to combine qualitative and quantitative data as children of this age often say things to please adults, others are shy and there is a considerable variation in what games the children think are most fun and which games they actually want to play more.

Table 1: Educational game development phases and applicability of methods

Evaluation Methods	Concept (Phase 1)	Pre-Production (Phase 2)	Production (Phase 3)	Post-Production (Launch of game)
Expert evaluation using checklist criteria	X	X	X	
User Testing		X	X	X
Interview		X	X	X
Observation without checklist		X		
Observation with checklist			X	X
Questionnaire		X		X
Pre/Post Test				X
Screen/Video Recording		X	X	X
Game logs				X
Quasi-experiment				X

3.2 Methodological, practical and ethical issues (RQ2)

This section focuses upon methodological challenges and practical and ethical considerations identified in research with refugee children.

3.2.1 Consent, gaining access and privacy

When researching with refugee children, consent for accessing children needs additional details as children and parents are commonly reached through trusted NGOs, religious societies, or qualification programs for immigrants. Researchers need to provide a thorough explanation of the research study to their collaborative partners to gain their trust before obtaining informed consent from children and parents or caretakers. In EduApp4Syria, children were accessed through Trondheim municipality, the Al Arj association, and the Muslim Society in Trondheim (MST) respectively in the three studies. It is also essential to ask for children's consent to participate (verbally) after getting a consent form signed from parents. In the third user test, some children were not willing to participate in the study, although their parents had signed the consent form.

Another constraint is to find suitable timeslots for participants. In the third user test, the study had to be conducted on weekends during the time for classes in the mosque and had to be adjusted according to the participants' particular needs. Furthermore, it was also essential to find a suitable room that would not be in conflict with other activities and provide sufficient privacy without interruptions. This can be a sensitive issue as initially, the members and leaders of the host community do not have trust, and they want to observe your study to understand your objective and research. In the third user test, during the first session, the administrator of weekend classes monitored the study. However, once she was familiar with the process, they facilitated privacy.

3.2.2 Language barrier and technology experiences

The language and technology experiences of the refugee children participating in the study can have a significant effect on the results. Most of the children in the first user test knew how to read in Norwegian and were experienced at playing games on smartphones, which made it easier for them to play the games in the test. However, in the second user test in Amman, most of the children could not read in any language before the test and had little or no experience using a smartphone. This caused some initial problems as the children did not know how to do gestures such as drag-and-drop and did not know typical user conventions used in most smartphone games. However, the limited exposure to smartphones and smartphone games was found to be very useful for testing the usability of the games. In games with lousy user design, limited user feedback, unclear goals, and progression, many users got stuck and needed help. The test showed very clearly the games which had good user design and who did not. In the third user test, most of the children knew how to read in Norwegian and were experienced at playing games on smartphones. Most of the children could speak two languages, either Arabic and Norwegian or Arabic and English with a few exceptions who could only speak in Arabic. However, it was found that some children knew a different dialect of Arabic and were unable to understand the game audio. Most of the children had access to video games, mobile technology, and digital gadgets at home. It was observed that children who had multiple options and with greater exposure to the digital world were more challenging to engage with the game for a longer period.

3.2.3 Learner-related

The target population may vary concerning demographics, cognitive and psychological factors (such as age, gender, verbalization, previous knowledge, attention span, learning disabilities etc.) which should be taken into account when devising methodology based on the type of data required. The user characteristics might influence the produced results, as children have different experiences of learning and playing with educational games. The first user test showed some significant differences in the children demeanor, where some were open and vocal, while others were quieter and shyer. This could have been caused by the variation of how long the refugees had been in Norway. Some families had lived in Norway for over two years, where one family only one week. The choice of research method should take this into consideration.

During the second user test, it was found that younger children and those who did not know any Arabic letters preferred Feed the Monster, while older children with some familiarly to the Arabic letters preferred Antura and the Letters. The first game was easier to play but could be a bit repetitive, especially with previous reading experience. The second game was a bit harder to play but provided more variation. A similar trend was noticed in the third user test, where younger children were more engaged in Feed the Monster than older ones.

Children are also different in terms of their personalities and preferences. Some were more into physical activities, and it was difficult for them to sit for 20 minutes. They started jumping and playing around as soon as the session was over. Game session time should be shorter than 20 minutes for more physically active children. Another critical factor is the attention span of children in general and children with learning disabilities. Some refugee children had learning difficulties, and although their learning gain was not satisfactory; they spent more time playing the game even after the game session was over as compared to others.

Gender differences also affect the engagement of children with educational games to some extent, where girls and boys had different preferences of game genre and characters. However, the two winner games (Antura and Feed The Monster) engage both genders.

3.2.4 Environment and setup

The context and setup of field studies play an important role in evaluation and research. The setup can sometimes be challenging to manage based on factors such as gaining access and privacy discussed earlier. In the first user test, all the children played in the same room. This gave some challenges as these games require the players to hear the audio properly to be able to play the game (e.g., what letter or word being said). Even though the children were spread out in the room, it was sometimes hard to hear the sounds from their own game. One solution would be to provide headsets for the children, but this was impractical as they were playing in pairs. Also, we wanted to test the games in a realistic environment, and most children play smartphone games without headphones. Using headphones would also make it harder to study the dynamics between the pair of children and would have made the observations harder to do (see Figure 3 from the first user test).

The third user test took place within the mosque area, and we were given two rooms for most of the sessions. Only two or sometimes three children would have parallel sessions in one room, and the rest of the children who were not playing were in a separate room. There were not many issues regarding hearing audio this time, as the room was big, the children were spread out, some children were pre-tested while others played the game. However, there were still some challenges. Children who were playing the game were distracted when other children whose session had not started yet began to play with each other and jumping around in the other room, and vice versa children outside wanted to start playing the game soon. Also, children would get distracted and started looking around when it was noisy, mostly when people coming and leaving the room. The organizers from the mosque were frequently coming in during the first days to monitor research.

Figure 3: Pictures from the user test in Trondheim, Norway

3.2.5 Parents involvement

The presence of parents provides positive effects when researching with refugee children as they become more comfortable and responsive. We noticed that for younger children (mostly five), it was essential to have their mother around and help communicate with them. Some children would not play or respond when asked by the observer, but when their mother asked the same thing, they would reply to her in detail. One girl stopped playing in between and ran looking for her mother, but when she came and sat with her, the girl started playing again. For user study in Amman, most of the children came with their mothers or their grandmothers, which gave a very positive side-effect. As the mothers and grandmothers were invited to play the games, it was revealed that several of them did not know how to read Arabic either. Some of them learned their first letters from the game, and became very enthusiastic, and wanted to learn by playing the games together with their children.

3.2.6 Cultural issues

Some cultural issues were also noticed during the field studies. During the third user test, one boy did not want to talk to the female evaluator and only responded when a male evaluator came. Then he agreed to play the game and was comfortable. The family had recently come to Norway. One interesting observation from study one was that almost all the children came with their father and not their mother. However, in user test in Amman most came with their mothers. Based on experience during the three evaluations, the immigrant children who are born and raised in Norway were different concerning cultural and economic aspects than refugee children in countries like Jordan. Children are influenced by the country they have immigrated to. Children in Amman, Jordan were more obedient and interested in technology as they were less exposed to it. Whereas children in Norway were less engaged by technology because of their previous exposure and they also had a strong opinion when they wanted to stop, not play or answer a question. However, in Amman, the children to a larger degree completed the session.

3.2.7 Translator

It is essential to recognize the need for a translator when researching with refugee children. In field study three, most of the parents did not know Norwegian or English, and the researcher running the experiment could only communicate in English. Therefore, a translator was needed for communication. There were three or four evaluators during each session who could speak one or two of the three languages (Arabic, English, Norwegian), and for a few children, a translator was required. Parents required the help of the translator to understand the process and documents they had to fill. The documents were translated into three different languages. However, it is sometimes difficult to work with a translator and other evaluators if they do not focus on the instruments or do not understand the research objective completely. In our case, some evaluators helped the children with the language pre-test by giving hints, which made the results invalid. It is vital to give the evaluators specific instructions for conducting the task following the research objectives.

3.2.8 Child-Friendly interactions

The interaction between the researcher or evaluator and children is a delicate process. It is crucial to establish a tone of informality with children in order to make them comfortable and create space for them to express themselves clearly. Is it essential to maintain a balance between asking too few and too many questions. During the third field study, it was noticed that when evaluators were not talking at all with children and only concentrating on observing them, children were a bit nervous during the session. This might be because they gave a pre-test before, so they felt they were also being tested on the game as well. Children in the second session were less nervous because evaluators talked and interacted with them. Asking questions within the gameplay (e.g., do you know what happened when the monster threw the letter) proved very helpful and gave additional insights to children's understanding rather than just relying on observations. However, it is also crucial not to break the tempo of children playing the game. Asking questions when they are less engaged makes them active again. To conclude, it is essential to talk to children and create a friendly atmosphere.

3.2.9 Effect of Information provided

Care needs to be taken to present the research purpose and objective to the parents and children. The difficulty of describing the purpose of research to children remains an important issue that needs special attention. If a researcher or translator convey wrong information or incomplete information, it negatively affects the results and research process. To avoid this problem, it is essential that translators or evaluators (other than the researcher) involved in a research study have a clear understanding of the research objective. In the third user test, some children knew they were getting the mobile for home as the translator gave this information to the parents. Therefore, they were more excited about taking the mobile and playing the game at home rather than during the game session. During one of the sessions in the third user test, the pre-test and gameplay were started before the translator had entirely explained the research purpose to the parents and before they started filling the forms. As a result, parents were unaware of the research agenda and were trying to help children to get a good score. Maybe they had the impression that the children will get the mobile and gift prize if they do good in the pre-test. However, after the translator explained the research, parents had a clear understanding of everything and did not try to help their child even if they were asked by the researcher to assist in conducting pre-test because the child needed their presence. Therefore, it is imperative that the researcher or translator clearly explain the research purpose to parents or caretakers before the session start and take them to a separated place if their presence is not required. It is also crucial that the researcher and translator have the same understanding of the research study so that the translator conveys the exact meaning to the participants.

4. Conclusion

Research with refugee children in child-computer interaction remains at an exploratory stage. This paper reflects on the methodological and practical aspects of evaluating educational games with refugee children based on pragmatic experience and reflections made during evaluation studies conducted in the EduApp4Syria project. We put forth the findings and a few pitfalls that need to be taken into account when evaluating educational games with refugee children. However, most of these findings can be useful for general research with children as well.

From our experience, a successful evaluation of learning game starts with the selection of a set of factors or dimensions based on the evaluation objective. There is not just one construct, but the complete game-based learning experience is made up of several dimensions that influence each other. Therefore, different methods might be beneficial for different objectives and at different stages of the development lifecycle. The paper identifies the applicability of various methods for different phases of development as well as highlighting essential constructs. According to experiences from the EduApp4Syria project, when devising an appropriate methodology for evaluation, several factors must be taken into consideration: such as language barrier, learner related issues (verbalization, previous knowledge, personality, technology experiences, attention span, learning disabilities, age and gender of the children etc.), parents involvement, cultural issues, need for translator, environment and setup of research, child-friendly interactions and the effect of information provided on research participants. In this paper we present the evaluation methods and approaches used in the project (EduApp4Syria) for research with young refugee children and highlight the issues, so new researchers in this field can learn from the experience and find the most appropriate way to apply them, so as to diminish their drawbacks as far as possible and also maximize their benefits. This paper aims to encourage researchers to critically reflect on the methodological and practical issues and the methods they choose to employ since they will have implications for the data produced.

References

Connolly, T. M., Boyle, E. A., Macarthur, E., Hainey, T. & Boyle, J. M. (2012). A systematic literature review of empirical evidence on computer games and serious games. *Computers & Education*, 59, 661-686.

Dubeck, M. M. & Gove, A. (2015). The early grade reading assessment (EGRA): Its theoretical foundation, purpose, and limitations. *International Journal of Educational Development*, 40, 315-322.

Erhel, S. & Jamet, E. (2013). Digital game-based learning: Impact of instructions and feedback on motivation and learning effectiveness. *Computers & Education*, 67, 156-167.

Fargas-Malet, M., Mcsherry, D., Larkin, E. & RobinSson, C. (2010). Research with children: Methodological issues and innovative techniques. *Journal of Early Childhood Research*, 8, 175-192.

Fleming, N. D. (2001). Teaching and learning styles: VARK strategies, IGI global.

Hill, M. (1997). Participatory research with children. *Child & Family Social Work*, 2, 171-183.

Kinzie, M. B. & Joseph, D. R. (2008). Gender differences in game activity preferences of middle school children: implications for educational game design. *Educational Technology Research and Development*, 56, 643-663.

Nordhaug, L. M. K. (2016). Eduapp4Syria [Online]. norad.no: Norwegian Agency for Development Cooperation (Norad). [Online]. Available: Available: http://norad.no/eduapp4syria [Accessed April 25th 2019].

Out, M. (2016). *Refugee education in crisis*. United Nations High Commissioner for Refugees, Geneva.

Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment* (CIE), 1, 21-21.

Prensky, M. & Berry, B. D. (2001). Do they really think differently. *On the Horizon*, 9, 1-9.

Sung, H. Y. & Hwang, G. J. (2013). A collaborative game-based learning approach to improving students' learning performance in science courses. *Computers & Education*, 63, 43-51.

Tahir, R. & Wang, A. I. (2017). State of the art in game-based learning: Dimensions for evaluating educational games. European Conference on Games Based Learning, Academic Conferences International Limited, 641-650.

Tahir, R. & Wang, A. I. (2019). Exploring methods and guidelines for child-computer interaction research with refugee children. 21st International Conference on Human-Computer Interaction, Florida, USA: Springer.

Vanderhoven, E., Willems, B., Van Hove, S., All, A. & Schellens, T. (2015). How to evaluate educational games: Lessons learned from the evaluation study of Master FIND. European Conference on Games Based Learning, Academic Conferences International Limited, 548.

ISBN 978-82-326-6522-8 (printed ver.)
ISBN 978-82-326-5979-1 (electronic ver.)
ISSN 1503-8181 (printed ver.)
ISSN 2703-8084 (online ver.)

NTNU

Norwegian University of
Science and Technology