

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

Rabail Tahir & Alf Inge Wang

To cite this article: Rabail Tahir & Alf Inge Wang (2024) Evaluating the effectiveness of game-based learning for teaching refugee children Arabic using the integrated LEAGUE-GQM approach, Behaviour & Information Technology, 43:1, 110-138, DOI: [10.1080/0144929X.2022.2156386](https://doi.org/10.1080/0144929X.2022.2156386)

To link to this article: <https://doi.org/10.1080/0144929X.2022.2156386>



© 2022 The Author(s). Published by Informa UK Limited, trading as Taylor & Francis Group



Published online: 19 Dec 2022.



[Submit your article to this journal](#)



Article views: 5035



[View related articles](#)



[View Crossmark data](#)



Citing articles: 2 [View citing articles](#)

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

Rabail Tahir and Alf Inge Wang

Norwegian University of Science and Technology, Trondheim, Norway

ABSTRACT

Game-based learning (GBL) is widely utilised in various domains and continues to receive interest and attention from researchers and practitioners alike. However, there is still a lack of empirical evidence concerning its effectiveness, making GBL evaluation a critical undertaking. This paper proposes an integrated 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 focuses on how children's age group, learning modality preference, and prior mobile experience affect their learning, usability, and gameplay performance. A quasi-experiment with a one-group pretest-posttest design was conducted with 30 children (5–10 years old) from migrant refugee backgrounds. The results show a 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. Moreover, 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 more inclusive design focusing on user characteristics.

ARTICLE HISTORY

Received 12 January 2022
Accepted 4 December 2022

KEYWORDS

Game-based learning;
evaluation approach;
learning games; refugee
children; Arabic reading skills

1. Introduction

In recent years, educational games have become more prevalent, and game-based learning (GBL) is now a well-established and growing research area that receives substantial attention from researchers and practitioners (Salah et al. 2016; Backlund and Hendrix 2013; Ariffin and Sulaiman 2014). Although GBL is considered an alternative learning tool for education and training and is widely utilised in various settings and domains (Prensky 2003), there is still a need for more empirical studies to prove its effectiveness (Ariffin and Sulaiman 2014; Boyle et al. 2016). Thus, it is essential to evaluate educational games' effectiveness before being used in a real context (Becker 2013).

With the increase in complexity and cost of learning game design, evaluating them is essential (Becker 2013; Wouters, Van der Spek, and Van Oostendorp 2009). Although plenty of resources are available for the design of educational applications, approaches to guide the evaluation are not (Becker 2013; Schleyer and Johnson 2003; Kafai, Franke, and Battey 2002). Mohamed and Jaafar (2010) identified three challenges in educational

game evaluation: evaluation criteria, evaluators, and the evaluation process. The evaluation criteria are important as they address the essential elements that need to be evaluated in the educational game to fulfil the evaluation goal. According to the literature on GBL evaluation (Calderón and Ruiz 2015; Tahir and Wang 2017), educational games are evaluated at different development stages and depending on the evaluation goal, different characteristics are assessed by selecting different criteria. The GBL evaluation literature also shows that different methods, techniques, and procedures have been used to evaluate learning games. However, when assessing the impact of learning games, the main interest mostly is to determine the educational effectiveness (learning outcomes), usability, and the user experience of the game. Most authors prioritise the evaluation criteria with respect to evaluation goals and verifying that the game has satisfied its specified objectives (Calderón and Ruiz 2015). Patton (2008) and other researchers pointed out that there is no single universal approach for designing an evaluation study (Diamond, Horn, and Uttal 2016). Therefore, we

can say that each evaluation study is distinct and unique. When developing the evaluation plan, it is important to select the criteria and measures specific to the evaluation goal to guide the learning game's assessment. Calderón and Ruiz (2015) proposed a taxonomy of models useful for evaluating different quality characteristics. Hence, a guiding approach for developing an evaluation plan would be useful for planning and designing an evaluation study, from articulating the purpose/goal to selecting evaluation characteristics/criteria and establishing evaluation questions to specify metrics and analysis methods to guide the assessment of our learning game (Diamond, Horn, and Uttal 2016).

Several researchers have highlighted the importance of GBL in language learning to improve students' performance and make learning more active, mainly focusing on classroom practice (Godwin-Jones 2014; Ju and Adam 2018; Sahrir and Yusri 2012). Language learning becomes even more important for migrants, as it is crucial to learn the language to integrate into a new society (Lou and Noels 2020). The Syrian crisis (Yazgan, Utku, and Sirkeci 2015) deprived over 2.25 million Syrian children of school education both within Syrian and other countries. Refugee children have to cope with high levels of stress and traumas affecting their learning ability, and little means of education were available for these children. It is important to teach these children basic literacy skills in Arabic (Wofford and Tibi 2018) for further integration into schools, or we might end up with a whole generation that cannot read or write in their mother tongue. Since smartphones are commonly used among migrants to stay connected (Gordano Peile and Ros Hajar 2016; Ros 2010), this can be a means for language learning purposes (Gaved and Peasgood 2017; Castaño-Muñoz, Colucci, and Smidt 2018). In their research, Bradley et al. (2020) focused on mobile literacy of Arabic-speaking migrants and the use of mobile technology to support migrants' language learning process and integration in Sweden. Researchers (Sahrir and Yusri 2012) have highlighted a lack of research regarding instructional and learning support in Arabic. Many have made an effort to improve this situation by focusing on Arabic language learning games and indicated positive results concerning GBL effectiveness for students' acquisition of Arabic language skills. However, here also, the focus has been mostly on classroom teaching (Sahrir and Yusri 2012; Eltahir et al. 2021; Sahrir and Alias 2012). According to Boyle et al. (2016), investigating informal learning in games can provide important insights into game mechanisms that can improve learning game design.

Moreover, limited educational gaming research has focused on user characteristics and their influence on

performance outcomes and game experience in GBL environments (Orvis, Horn, and Belanich 2009). Learner characteristics affect online learning (Lim and Kim 2003), and the growing use of learning technologies demands a sound understanding of learner characteristics that affect learning with technology (Nakayama, Yamamoto, and Santiago 2007). However, this relation is less explored in learning games since the greater focus has been on learning effectiveness and usability (Calderón and Ruiz 2015; Tahir and Wang 2017).

This paper outlines the simple and effective integrated LEAGUE-GQM approach to create an evaluation plan for assessing learning games' effectiveness by establishing the goals, defining questions, and identifying measures for the evaluation process with the LEAGUE evaluation guide. The main novelty of the presented approach is its simplicity and the GBL-specific guidance. The proposed approach is applied in practice for planning and conducting an evaluation study on the Arabic language learning game Feed the Monster (FTM). With the backdrop of the Syrian crisis and the 'EduAppSyria' initiative project focusing on Arabic language learning games for Syrian refugee children (Nordhaug 2019), this user study's main objective was to evaluate the effectiveness of the GBL approach for migrant refugee children to learn Arabic reading skills in an informal learning environment. The study employed a quasi-experimental design with 30 children (5–10 years old) and collected quantitative and qualitative data to answer five research questions. The research questions focused on the learning gain using GBL intervention; age-group differences between younger and older children; the correlation between learning modality preferences of children and their learning gain, usability score and gameplay performance; and the correlation between mobile usage experience of children and their learning gain, usability score and gameplay performance, and usability issues faced by children when playing with the learning game for the first time. This study used mixed methods including pre/post-test, questionnaire, interview, game logs, observation checklist, and notes to analyse the learning gain with GBL and the effect of user characteristics (age group, learning modality preferences, and mobile usage experience) on the learning gain, usability, and game performance.

The contribution of this paper is two-fold. *First*, it proposes an integrated approach (LEAGUE-GQM) for planning a GBL evaluation. *Second*, it presents implications and design recommendations for effective learning game design based on the evaluation study results on the potential of using game-based language learning for teaching refugee children Arabic reading skills in an

informal learning setup. The rest of this paper is organised as follows: Section 2 presents an overview of the related work. Section 3 describes the methodology that includes the proposed integrated LEAGUE-GQM approach for planning GBL evaluation and its application to the user study (evaluation of Arabic language learning game-Feed the Monster) presented in this article. Section 4 presents the results from the quasi-experiment concerning five research questions. Section 5 discusses the results, design recommendations, and limitations of the study. Lastly, Section 6 concludes the article and gives directions for future research.

2. Related work

This section presents the related work describing the potential of GBL identified by relevant research studies, importance of GBL evaluation and challenges in educational game evaluation. Further it highlights the scarcity of empirical research concerning the effectiveness of digital game-based language learning (DGBLL) in general and a lack of DGBLL studies in Arabic language learning. Moreover, this section also underlines the need for research concerning learner characteristics to understand the mitigating factors (such as age, gender, learning styles, prior knowledge etc.) that influences learning with serious games.

2.1. Game-Based learning effectiveness evaluation

Game-based learning (GBL) refers to games for education and learning purposes (Tang, Hanneghan, and El Rhalibi 2009). Many researchers have investigated the potential of learning games, and mixed results are obtained regarding the evidence about its impact (Salah et al. 2016). Some research studies found positive effects, whereas others found no significant effect of using games for learning (Boyle et al. 2016; Wouters, Van der Spek, and Van Oostendorp 2009). López-Fernández et al. (2021) based on their research finding reported that students who used educational games for learning were more motivated and experienced fun. Moreover, majority of students preferred GBL over traditional teaching approach. Similarly, Eltahir et al. (2021) also found in their study that students using GBL showed more improved knowledge of concepts and higher motivation compared to students taught with traditional strategy. Akçelik and Eyüp (2021) investigated the effects of educational games on vocabulary knowledge and found that students enjoyed learning vocabulary with games and learned more easily. Although many researchers have reported that

games can increase motivation and interest, there is still a need for more empirical studies to assess the educational effectiveness of learning games as most of the studies base their claims on subjective judgment and personal encounters (Ariffin and Sulaiman 2014). According to the systematic review conducted by Kaloigiannakis, Papadakis, and Zourmpakis (2021) digital technologies such as gamification has the potential to heavily influence the learning process. However, the review outlined that mostly small longitudinal studies revealing mixed results have been conducted highlighting the need for more research exploring long-term effects to clarify the impact of such technology on education. Boyle et al. (2016) reported an increase in the empirical evidence concerning positive outcomes of playing games. However, they suggested that detailed experimental studies are crucial for future research to systematically explore the game features most effective in supporting learning. Behnamnia et al. (2020) investigated whether digital game-based learning application can improve creativity skills in preschool children. The study focused on the components of creativity and learning levels when using Digital Game-Based Learning (DGBL) for children under the age of six and found that DGBL can potentially affect young children's ability to develop creative skills, knowledge transfer, critical thinking, acquisition of digital experience skills and a positive attitude for deep and insightful learning. According to Bellotti et al. (2013), educational games must be able to show that necessary learning has occurred like any other education tool. Therefore, to affirm their impact, it is crucial to systematically evaluate them (Marciano, de Miranda, and de Miranda 2014). Moreover, the costly and time-consuming development of educational games demands for continued assessment of their efficacy and a need to identify principal criteria (De Freitas and Oliver 2006; De Freitas and Liarokapis 2011). The diverse GBL characteristics make its evaluation a difficult task (Djelil et al. 2014). Mohamed and Jaafar (2010) identified that establishing evaluation criteria and process are the main challenges in educational game evaluation. Although different aspects important for GBL have been highlighted by previous research, an overreaching approach is required to guide evaluation and design iterations (De Freitas and Liarokapis 2011; Van Staaldunin and De Freitas 2011; Oprins et al. 2015). Depending on the evaluation goal, educational games are evaluated at different development stages, and different criteria are selected for assessing different characteristics (Calderón and Ruiz 2015; Tahir and Wang 2017). Patton (2008) and other researchers pointed out that there is no single approach for designing an evaluation study (Diamond, Horn, and

Uttal 2016). However, All, Castellar, and Van Looy (2021) evaluated the feasibility of previously defined best practices for assessing DGBL effectiveness and focused on research design components providing insights into feasible experimental designs to further guide design of DGBL effectiveness studies. Dondi and Moretti (2007) pointed out that identifying criteria is a time-consuming and complex process, and not many approaches are available to guide the evaluating process of learning games (Becker 2011). Ak (2012) emphasised the need to define critical aspects of educational games that make them effective. These aspects could serve as evaluation criteria and guide the evaluation process. Calderón and Ruiz (2015) proposed that having a taxonomy of models could be helpful for evaluating different quality characteristics. Moreover, according to Zourmpakis and Kalogiannakis (Zourmpakis, Papadakis, and Kalogiannakis 2022), teachers play a key role to understand the individual needs of students and provide them with proper learning material and evaluate the complete learning process. Therefore, it is also important to explore how teachers design and integrate gamified environments into education and teaching.

2.2. Educational games in language learning, refugee context and arabic language

The use of learning games for language acquisition is not new, and many researchers have focused on digital games for language and second language learning (Hung et al. 2018; Poole and Clarke-Midura 2020; McKiddy 2020). Most studies have investigated the following categories of language acquisition using games: alphabets, listening, reading, speaking, writing, vocabulary, grammar, pronunciation, and mixed or integrated skills, focusing mainly on formal education (Hung et al. 2018; Kocaman and Cumaoglu 2014; Neville, Shelton, and McInnis 2009; Jalali and Dousti 2012; Chen and Yang 2013). Despite the increase in the emerging literature on digital GBL and its educational value, empirical evidence is scarce concerning its effectiveness in language education (Godwin-Jones 2014). Chen and Yang (2013) investigated the effects of using an adventure game for college students' foreign language learning. They found no significant difference in the vocabulary score of the two groups. However, the students perceived the game to be helpful in improving their motivation and language skills. Jalali and Dousti (2012) explored the impact of educational games on grammar and vocabulary gain of elementary students and found no significant differences between the two groups (experimental and control).

The devastating impact of the Syrian conflict on refugee children and their education led researchers to focus on literacy education interventions for serving Syrian refugee children (Wofford and Tibi 2018; Al Janaideh et al. 2020). A review study juxtaposing refugee needs with mobile learning apps characteristics found that mobile learning is beneficial for refugees (Drolia et al. 2020). The study concluded that mobile learning provides access to education and also improves the quality of education provided to refugees. Drolia et al. (2022) highlighted in their study that mobile learning research focusing on social groups such as refugees, learner with disabilities or learning difficulties is very limited. The review study focused on existing mobile learning applications for refugees and their characteristics, and the most important characteristics included refugees' cultural features and interwoven psychological and educational features. Many researchers have focused on language learning tools for refugees and identifying the language learning needs (Castaño-Muñoz, Colucci, and Smidt 2018; Abou-Khalil et al. 2019). Akçelik and Eyüp (2021) found that educational games are effective for teaching vocabulary and improved refugee students' Turkish vocabulary knowledge. According to Hung et al. (2018), English as a second language is most targeted in digital game-based language learning (DGBLL) literature, whereas other languages like Arabic are lacking. Although the research has highlighted the need for Arabic language and literacy skills for the academic achievement of refugee students (Baddour 2020), very few DGBLL initiatives focus on Arabic literacy skills (Czuderna and Guardiola 2019). Azizt and Subiyanto (2018) investigated the effects of digital GBL on high school students' Arabic reading skills. They found it useful to increase students' academic performance (experimental group scored significantly higher) in Arabic learning. Kenali et al. (2019) investigated the impact of using smartphone language games on Arabic speaking skills in non-native speakers by conducting an experiment with university students. The findings showed a significant positive effect on the speaking skill of students. Sahrir and Alias (2012) found a positive perception of learning Arabic using online games among university students. Moreover, Putri et al. (2021) in their study reported that educational games are effective in increasing Arabic vocabulary in the higher education by increasing learning motivation and making it easier for students to understand the content. Eltahir et al. (2021) also investigated the impact of GBL on an Arabic language grammar course in higher education and found that students using GBL showed more improved knowledge of Arabic grammar concepts and higher motivation compared to students taught with traditional strategy. According to a

review by Murtadho (2021), there is increased interest in digital tools among Arabic language educators. However, its availability and usage are still limited in Arabic classrooms and require more successful management, acquisition and use.

Moreover, as highlighted by Masrop et al. (2019), there is a lack of DGBLL studies in Arabic language learning, and only a few Arabic learning games are dedicated to children. According to bin Zainuddin et al. (2021) digital game-based language learning applications for Arabic Language are much needed for primary school children to improve their Arabic language proficiency. Ali Ramsi (2015) analysed the Arabic learning games for children and found that they are generally simplistic, revolve around the same trivial idea, lack a systematic design process, and do not have quality animation, colours, graphics, and voice-over. Similar results were reported by Masrop et al. (2019) from their analysis of existing games. In addition, they found that Arabic language learning games are mostly limited to alphabet content and lack engaging features in the game. The majority of studies in DGBLL focused on higher education/university students, and only a few studies have explored the individual differences of learners and how it affects their learning and content knowledge (Hung et al. 2018). The younger age-groups, informal learning setups, and learner characteristics are less explored areas and need attention.

2.3. Learner characteristics

Previous research indicates that all students do not perform equally in technology-enhanced learning environments due to factors related to learner characteristics (such as learning styles, prior knowledge) or/and the learning environment itself that influences student success (Terrell and Dringus 2000; Birch and Bloom 2002; Wojciechowski and Palmer 2005). With the growing use of learning technologies, there is an increased demand for research concerning learner characteristics (Nakayama, Yamamoto, and Santiago 2007; Wojciechowski and Palmer 2005). The review on serious game research by Wouters, Van der Spek, and Van Oostendorp (2009) identified the lack of understanding of mitigating factors (such as age, gender) that affect learning with serious games. Nakayama, Yamamoto, and Santiago (2007) emphasised the need to better understand learner characteristics that influence learning and an effective design that customises the learning activities to serve individual characteristics and learning needs of different users. Bontchev and Paunova (Bontchev, Terzieva, and Paunova-Hubenova 2021)

focused on principles for personalisation of gameplay and learning content in serious games based on player and learner-related aspects of the student's profile. The research findings highlighted the importance of characteristics such as student's age, goals, level of knowledge, and learning style to be included in the student model for personalisation of learning content in serious games for learning. Belay, McCrickard, and Besufekad (2016) emphasised the need to consider parameters such as technology exposure, computing literacy, and level of help required for user classification and cater to these differences in the design. Ariffin (2013) found that none of the 16 identified game evaluation frameworks concentrated on learner background (such as culture, spoken language, and ethnicity). However, Kanwar et al. (Jossan, Gauthier, and Jenkinson 2021) investigated the impact of culture (cultural integration and cultural associations) on students' views and acceptability of GBL while adjusting for exposure to video gaming and gender. The study found differences between individuals associating with different cultural groups providing insights into cultural considerations for GBL design and evaluation for international populations. The study suggested that culture should be assessed more broadly as culturally aware design of GBL can further support learning. Moreover, Osman and Bakar (2012) also stressed that factors related to learners' background should be included in the educational game design model. It helps to provide effective learning experiences and further refine the game design.

The learner characteristics researchers focus most on are gender, prior experience, age, background, and learning style (Nakayama, Yamamoto, and Santiago 2007; Chen and Huang 2013). Sundqvist and Sylvén (2014) explored the English language-related activities and digital game playing of learners outside of school and found significant gender differences in gaming habits between girls and boys, the latter being more frequent gamers. Sundqvist and Wikström (2015) investigated the relationship between the students' English learning performance in school and the frequency of out-of-school gameplay. The results indicated that, particularly for male students, the English learning measures are correlated with their gameplay experience. Similar results were found by Erfani et al. (2010). In addition to gender, they also found a significant influence of age on performance. Chen and Huang (2013) demonstrated that prior knowledge (prior digital games experience and digital games playing frequency) has positive effects on GBL but only for the context of declarative knowledge. On the contrary, results from (Kim and Chang 2010) showed that individual

differences in computer experience, prior mathematics knowledge, and English language skills had no significant effect on students' achievement when using a computer game. Ariffin and Sulaiman (2014) investigated the effectiveness of GBL in higher education, focusing on learner's background (culture, ethnicity, and language) and GBL environment. They found a strong correlation between learners' background and learners' motivation and between the learner's motivation and learner's performance. The research proposed that the learner's background parameters affect the learner's learning performance and should be integrated within the GBL environment. Numerous research studies focus on students' learning modality preferences and how they affect learning (Alkhasawneh et al. 2008; Aslaksen et al. 2020). Although many research studies show that modality does not affect learning, some researchers think that different media may afford different instructional methods. Therefore, different instructional media's distinctive characteristics and functional capabilities might be relevant to the learning process, which determines their effectiveness (Moreno 2006; Rummer et al. 2011). Moreover, based on the results from studies focusing on learning style, many researchers advocate the notion of multimodal learning (Aslaksen et al. 2020).

3. Materials and methods

This section presents the methodology adopted for this research study. To define the study dimensions and develop a GBL evaluation plan, the authors propose an integrated approach as a guide for planning a GBL evaluation. The approach is inspired by the LEAGUE (Learning, Environment, Affective–cognitive reactions, Game factors, Usability, UsEr) framework (Tahir and Wang 2020) and the GQM (Goal Question Metric) model (Caldiera and Rombach 1994). The LEAGUE-GQM approach is used to identify and select the GBL evaluation criteria with respect to evaluation purposes to verify the specified objectives of a learning game. In this way, the approach guides the researchers and practitioners interested in evaluating learning games in different domains. Calderón and Ruiz (2015). The next sections describe the approach and its application in a user study to develop an evaluation plan and demonstrate its use.

3.1. Integrated LEAGUE-GQM approach for GBL evaluation

The proposed approach facilitates the GBL evaluation process by providing GBL-specific evaluation criteria

Table 1. Integrated LEAGUE-GQM approach.

P1: Define evaluation type and data based on GBL development stage and evaluation rational

Evaluation Type	Evaluation Data
<i>Pre-prototype evaluation</i> (pre-production phase: evaluating learning game idea)	<i>Qualitative</i>
<i>Formative evaluation</i> (production phase: evaluating learning game prototype or versions)	<i>Quantitative</i>
<i>Summative evaluation</i> (post-production phase: evaluating developed or launched learning game)	<i>Qualitative and Quantitative</i>

P2: Develop a LEAGUE tree and evaluation plan with the three-step parallel process using the LEAGUE-GQM evaluation guide

1) Conceptual Level (objects/goals)	2) Operational Level (assessment/questions)	3) MeasurementLevel (subjective-objective/measures)
<i>Plan evaluation goal choosing dimensions with LEAGUE evaluation guide.</i>	<i>Plan evaluation questions choosing factors/sub-factors, relations, and select data sources with LEAGUE evaluation guide.</i>	<i>Plan evaluation measures using metric types and select analysis methods with LEAGUE evaluation guide.</i>
Integrated process of selecting LEAGUE dimensions to identify, define and finalise the GBL evaluation goal(s)	Integrated process of picking and choosing LEAGUE factors/sub-factors and relations (if targeted) to identify and formulate GBL evaluation questions/hypotheses and define data sources that outline the goal completely and quantifiably	Integrated process of selecting LEAGUE metric types (based on required subjective or objective data) to identify and define the specific measure(s) and analysis method(s) that need to be collected and analysed to answer the questions and track the goal
<i>LEAGUE tree</i> (chosen LEAGUE components (dimensions, factors, relations, metric types) for specific GBL evaluation).		
<i>Final Evaluation Plan</i> (finalised evaluation goal, questions, data sources, measures, and analysis methods).		

on three levels to create a strategy and plan for evaluating learning games. The proposed integrated LEAGUE-GQM approach is made up of two parts (Table 1) and provides a LEAGUE-GQM evaluation guide (Table 2) for guiding the steps in the approach. The two parts are as follows:

- (1) P1) Define the evaluation type and data based on GBL development stage and rationale behind the evaluation
- (2) P2) Develop a LEAGUE tree and evaluation plan using a three-step parallel process based on the LEAGUE-GQM evaluation guide.

Educational games are evaluated at different development stages with different purposes/rationale (Calderón and Ruiz 2015; Tahir and Wang 2017). The type of GBL evaluation is linked with the educational game

Table 2. LEAGUE-GQM evaluation guide.

Conceptual level (objects/goals)		Operational level (assessment/questions)	Measurement level (subjective objective/ measures)	
Plan evaluation goal choosing dimensions		Plan evaluation questions choosing factors/sub-factors, relations, and select data sources		Plan evaluation measures and select analysis methods choosing metric types
Dimensions	Factors	Sub-factors	Relations	Metrics types
Learning	Learning Objectives	Prior knowledge, learning, and retention, Potential transfer	Learning & Game Factors	Objective:
	Learning Strategies	Learning style, learning theory or model, learning task/activity	Game Factors & ACR	Scores
	Learning Content	Educational material, instructional support, difficulty levels	ACR & Learning (integration of gameplay and learning) & ACR	Time
	Learning Outcome	Knowledge/skills/attitudes enhancement competencies, performance	Usability & (Learning, Game factors, ACR)	Number of occurrences Rating Subjective: Reviews/ responses/ opinions
Game Factors	Game Definition	Game goals, Game Rules, Game tasks	Usability & Environment	Analysis Methods
	Game Narrative	Player Characters, storyline, fantasy/fiction	Usability & User	Qualitative analysis methods:
	Game Mechanics	Game Interactions, Game Controls	User & Environment	Measures of central tendency (mean, median, quartile, mode)
	Game Resources	Game tutorial and Help, Rewards & resources, Game Customisability	User & (Learning, Game factors, ACR)	Measures of dispersion (standard deviation, range)
	Game Aesthetics	Multimedia elements, Game visualisation	Environment & (Learning, Game factors, ACR)	Measures of dependency (Pearson correlation coefficient, Spearman rho)
	Game Play	Challenge, Strategy, Pace & adequate levels, Game feedback		Graphical visualisation (histograms, 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)
Affective-Cognitive Reactions (ACR)	Enjoyment	Immersion, social interaction, Challenge, Goal clarity, Feedback, Concentration, Control, Knowledge improvement	Data sources	Qualitative analysis methods
	Engagement Motivation	Immersion, Control, Challenge, Purpose, Interest Attention, Relevance, Confidence, (learner) Satisfaction	Questionnaires Observation with checklist	Content analysis
	Flow	Challenge, Clear goals, Feedback, Playability, control, Rewarding experience, Concentration, Loss of self-consciousness, Time distortion	Observation or field notes Interviews Pre/post-test Screen recording Game-logs Expert evaluation Video recording Challenges, tasks, or exercises Learner diaries/reports	Grounded theory approach
Usability	Interface	Feedback (interface), Metaphor and Objects, User control and settings, Consistency, Error presentation, Navigation, Adaptivity and Accessibility, Screen design, help, and support		
User	Learability Satisfaction			
	Learner Profile	Biodemographic, Experience, Personality (preferences, styles)		
	Cognitive Needs	Cognitive stage (Piaget's theory), Cognitive load (Mental effort)		
Environment	Psychological Needs	Psychosocial Stage (Erikson's Theory), Psychosocial Well Being		
	Technical Aspects	Technology type, Technology related issues, Meet technical requirements and specifications		
	Context	Place, Settings		

development stages and rationale behind the evaluation (Steiner et al. 2015; Connolly, Stansfield, and Hainey 2009; Zaibon and Shiratuddin 2010). Likewise, when designing an evaluation study, it is important to select the criteria and measures in line with the evaluation rationale, type, and data to guide the learning game's evaluation process (Calderón and Ruiz 2015) as there is no single approach for designing an evaluation study (Patton 2008; Diamond, Horn, and Uttal 2016). Each evaluation study is distinct, as highlighted by Dondi and Moretti (2007) and identifying criteria for GBL evaluation is a time-consuming and complex process. Hence, a guiding approach like LEAGUE-GQM outlines the important criteria in three levels that can guide planning and designing an evaluation strategy from articulating the goal, formulating questions and deciding data sources, and specifying measures and analysis methods. The process is useful for creating a GBL evaluation strategy and plan to guide the evaluation of learning games for both the analytical (single aspect) and global (holistic) evaluation process depending on the required evaluation type. The LEAGUE framework (Tahir and Wang 2020) serves as a GBL theoretical foundation in the proposed approach, providing core GBL components (Tahir and Wang 2020). It presents the GBL criteria listing six dimensions, twenty-two factors, seventy-four sub-factors, ten relations, and five metric types that are incorporated in the LEAGUE-GQM evaluation guide (see Table 2) for developing a LEAGUE tree. The LEAGUE tree provides the structure for the evaluation plan listing the selected GBL evaluation criteria. GQM, on the other hand, provides three measurement levels (conceptual, operational, and measurement) for defining the evaluation plan. Moreover, the evaluation guide also presents data sources (adopted from Petri and von Wangenheim 2017; Tahir and Wang 2019) and analysis methods (adapted from Petri and von Wangenheim 2017) for guiding the steps in the LEAGUE-GQM approach. The complete approach is presented in Table 1, and the evaluation guide is presented in Table 2.

In the *first part* (P1), the type of evaluation (Steiner et al. 2015) (pre-prototype, formative, summative) and the kind of data that will be collected (Petri and von Wangenheim 2017) (qualitative, quantitative, both) are determined depending on the learning game's development stage (Zaibon and Shiratuddin 2010) and evaluation rationale (the purpose behind the GBL evaluation) (Steiner et al. 2015; Connolly, Stansfield, and Hainey 2009). Three main types of evaluations can be conducted over the GBL development life-cycle, and identifying the type can help make better decisions by providing the right kind of data at the right time

(Aslan and Balci 2015; Faizan et al. 2019). *Pre-prototype evaluation* is conducted before the development stage begins for the purpose to incorporate and assess design ideas (for example, through participatory design Danielsson and Wiberg [2006] or user acceptance testing of pre-prototype Davis and Venkatesh [2004]) and to collect benchmark data for subsequent comparative analyses concerning the impact on game project outputs (Steiner et al. 2015). *Formative evaluation* is conducted during the development process for the purpose of highlighting any weaknesses or ambiguity in the learning game prototype or different versions (Steiner et al. 2015; Connolly, Stansfield, and Hainey 2009). *Summative evaluation* is conducted at the end of the game development process or after the game has been launched for the purpose of evaluating the potential of the end-product (Steiner et al. 2015; Connolly, Stansfield, and Hainey 2009).

In the *second part* (P2), a three-step parallel process is used to incrementally develop a LEAGUE tree and an evaluation plan, using the elements from the LEAGUE-GQM evaluation guide (see Table 2) and following the GQM levels, as described in Table 1. The evaluation guide for the integrated LEAGUE-GQM approach (presented in Table 2) specifies the GBL evaluation criteria of the three GQM levels, and evaluators can pick and choose components for each step. Depending on what is to be evaluated, dimension, factors/subfactors, relations, data sources, metric types, and analysis methods can be selected for pre-prototype evaluation to verify the idea, formative evaluation to identify issues to inform design changes, and summative evaluation for determining the effectiveness of the developed game. The three steps of the approach correspond to one of the three levels of GQM: conceptual, operational, and measurement. There are two simultaneous activities at each level: development of the LEAGUE tree and development of the evaluation plan. The development of the LEAGUE tree is done by selecting relevant dimensions, factors/sub-factors, relations, and metrics types from the evaluation guide for the specific evaluation study. The development of the evaluation plan is carried out by elaborating on the LEAGUE tree elements in each step (which act as a guide) for establishing the goal, questions, data sources, specific measures, and analysis methods.

The *first step* is related to the conceptual level in which the preliminary evaluation purpose is defined, selecting dimensions in the LEAGUE-GQM evaluation guide to outline the study dimensions into the final evaluation goal. The *second step* is the operational level in which this goal is broken down

using factors/ sub-factors and defined using relations (if required) in the evaluation guide for formulating quantifiable and assessable questions to assess the goal. The data source(s) for collecting the required data for evaluation is also specified at this level using the evaluation guide. The *third and final step* is the measurement level, where for each question, measures and analysis methods are defined by selecting metric types from the evaluation guide. Each metric type is elaborated with specific measures based on the selected data sources. The specific evaluation measures are easy to define considering the selected data sources and metric types. For example, the specific measures for metric type 'score' can be easily defined based on the selected data sources 'pre/post-test' and 'usability checklist' as pre/post-test score, usability checklist score (illustrated in the user study, see Table 3). Since the metric types are categorised into objective and subjective, it directs to select relevant analysis methods. The data analysis methods can be selected from the evaluation guide based on qualitative or quantitative data to provide

information, answer the formulated questions, and accomplish the goals set in the previous steps. At this point, the evaluation guide incorporates only two qualitative analysis methods based on our experience in GBL evaluation. However, it would be helpful to incorporate a list of different qualitative methods used for GBL evaluations in the LEAGUE-GQM guide in the future.

3.2. User study: evaluating a language learning game using integrated LEAGUE-GQM approach

This section illustrates how the integrated LEAGUE-GQM approach was used to develop an evaluation plan for the user study on evaluating the potential of the Feed the Monster (FTM) GBL application for reading skills of migrant refugee children.

3.2.1. Feed the monster: Arabic language learning game

Feed the Monster (FTM) is an Arabic language learning game developed as part of the EduApp4Syria project for

Table 3. User study: FTM.

Summative evaluation of FTM using quantitative and qualitative data		
1) Conceptual level (objects/goals)	2) Operational level (assessment/questions)	3) Measurement level (subjective objective/ measures)
Evaluation goal for this study	Evaluation questions and data sources for this study	Evaluation measures and analysis methods for this study
Chosen Dimensions: Learning, Game factors, Usability, and User	Chosen Factors/ Subfactors: Learning Outcome, Game Play, Interface (Feedback (interface), Metaphor and Objects, User control, and settings, Consistency, Error presentation, Navigation, Screen design, Help and support), Learnability, Satisfaction, Learner Profile (Bio-demographics, Experience, preferences) Chosen relation: User & (Learning, Game factors, Usability)	Chosen Metrics: Scores, Time, Number of occurrences, rating, Reviews/responses/ opinions
Final Evaluation Goal: The purpose of this study is to evaluate the effect of using GBL for Arabic reading skills with respect to learning, game factors, usability, and user characteristics from the point of view of migrant refugee children in the context of informal learning set up	Final Evaluation Questions: <i>RQ1:</i> What are the learning outcomes of GBL for Arabic reading skills of migrant refugee children in an informal learning setup? <i>RQ2:</i> What are the differences between younger and older children's learning outcome, usability, and gameplay performance using GBL for Arabic reading skills in an informal learning setup? <i>RQ3:</i> How is the learning modality preference of migrant refugee children related to their learning outcome, usability, and gameplay performance using GBL for Arabic reading skills in an informal learning setup? <i>RQ4:</i> How is the mobile usage experience of migrant refugee children related to their learning outcome, usability (learnability), and gameplay performance using GBL for Arabic reading skills in an informal learning setup? <i>RQ5:</i> What are the usability issues faced by migrant refugee children when playing with the language learning game for Arabic reading skills? Data sources: Pre/Post Test (EGRA), game log data, observation checklist for usability, observation notes and follow-up interview, demographic questionnaire and learning modality preference questionnaire (VARK)	Final Evaluation Measures: <i>Score:</i> Pre/post-test score, score in a play session, usability checklist score, VARK score. <i>Time:</i> Time played. <i>Number of occurrences:</i> Number of levels completed, number of wrong answers, number of sessions played, age group. <i>Rating:</i> rating for mobile usage experience, years of mobile use, frequency of mobile use, dependency in mobile usage, time spent on mobile. <i>Reviews/responses/opinions:</i> interview responses and observation reviews. Analysis method: Quantitative analysis: Wilcoxon Matched Pairs Test (RQ1), Mann-Whitney test (RQ2), Spearman correlation (RQ3&4) and, Qualitative analysis: grounded theory approach (RQ5)

refugee children to improve their literacy skills in Arabic. FTM is about helping the kind monsters grow and prosper by feeding them letters, words, and sentences. The storyline illustrates a world where an evil character 'Harboot' destroys and conquers the land of the friendly monsters who are sent to exile, and he cast a magic spell that turned these friendly monsters into eggs. The player needs to feed these eggs with Arabic letters, syllables, and words in each game level to help them grow and evolve. The storyline was designed to mirror the experience of refugee kids to nurture hope and harness their native language acquisition. The 'friendly monsters' are used as the game characters to cope with one's fears. In the game, the players act as the Monsters' main caregivers, helping them grow and manage their emotions. The players can unlock new monster friends' with in-game progress.

The main game mechanism is feeding the Monster with the correct answer (letters, syllables, words) based on the level activity. The game divides the Arabic alphabets into small clusters (letter grouping) of five to six alphabets. Each cluster introduces four or five visually distinct letters and one vowel. Every cluster starts with letters (shape and sounds); then players practice vowels variation (shape and sounds); letter in a syllable segment (written form and sound); letter sequence within a word; lastly, words (made of letters already learned in the cluster) and its sound. The small clusters make learning Arabic easy and practical for children because they start learning full words just after the first 5–6 letters, making it fun and engaging. In each task/activity, the player must feed the Monster the correct letter, vowel variation, syllable segment, or letter sequence in the correct spelling order, either based on matching the letter to a copy of the letter or matching the letter to its sound. Each cluster takes 15–20 min to complete, and then the player moves to the next cluster (letter grouping). The player can earn one to three stars in each level depending on performance (the number of correct answers and time taken to answer). Correct answers make the Monster happy. Players earn scores in each level, and periodically the Monster grows bigger. Incorrect answers make the Monster sad, but the child can pet the Monster to make it happy. Instructive feedback is provided by the Monster spitting out the incorrect answer. There are also three mini-games of drawing letters, a memory game, or collecting gifts.

This GBL application is designed for out of school children as a supplementary resource to play at home with minimum adult supervision. The GBL design aimed to provide effective literacy learning opportunities to Arabic-speaking refugee children to improve their literacy skills and psychosocial well-being.

Therefore, the game elements used in this GBL application focused on three pillars: ease of use and engaging game experience, Arabic reading acquisition and improvement in psychosocial well-being. The game design uses an intriguing storyline and character of friendly monster to engage children by providing a journey of friendship and discovery. Keeping refugee children in mind the game's storyline presents a fantasy world designed to nourish hope. The character of monster and simple interactions such as feeding are used to build association with children while keeping the game easy to use.

3.2.2. Evaluation plan using the league-GQM approach

We started by defining the evaluation type and data and planned to conduct a summative evaluation of the final released version of the FTM game with target users using quantitative and qualitative data to test the hypothesis and explore user behaviour and experience in-depth. Next, we used the integrated LEAGUE-GQM evaluation guide and followed the three-step parallel process to develop the LEAGUE Tree and evaluation plan. Table 3 presents the complete evaluation plan.

The rationale of this evaluation related to the Eduapp4syria project (Nordhaug 2019), was to evaluate the effectiveness of GBL approach (the FTM game) for teaching Arabic skills to refugee children in informal learning setup and without the need of parents' help. The dimensions in LEAGUE evaluation guide directed us to choose learning, game factors, usability and user characteristics that matched with initial rationale and defined our evaluation goal. Further, the LEAGUE evaluation guide made is easier to formulate assessment plan by outlining the factors, relations and metric types that were relevant for our selected dimensions and motivated by the employed game. Each level is elaborated below.

For the *conceptual level*, the dimensions learning, game factors, usability, and user were selected from the LEAGUE evaluation guide, which helped define the evaluation goal using the GQM template as shown in Table 3. For the *operational level*, the factors/subfactors selected from the evaluation guide are listed in Table 3. Besides, we also targeted the relation: User & (Learning, Game factors, Usability) to formulate five evaluation questions (see Table 3). The data sources for collecting the relevant data were selected keeping in mind the selected factors/subfactors: Pre/Post Test for obtaining children's' learning gain (learning outcome) with the game; game log data for recording game-play performance; observation checklist, observation

notes and short interview for collecting data related to the usability (interface, learnability, satisfaction); demographic questionnaire for collecting the bio-demographics and experience data, and learning modality preference questionnaire (VARK) for recording children preferences related to learner profile. Finally, for *the measurement level*, we selected four metric types from the evaluation guide: Scores, Time, Number of occurrences, and Reviews/responses/opinions focusing on both objective and subjective data. Each metric type was used to identify the specific measures for the defined data sources (from the previous step), as shown in Table 3. Finally, four analysis methods (quantitative and qualitative) were selected to answer the formulated questions.

3.2.3. Sampling

During Spring 2018, 30 children aged 5–10 years old from refugee migrant background, who speak Arabic but did not have reading or writing skills in Arabic, participated in our study. All participants had no previous experience with the FTM game. These participants were selected because FTM is an Arabic language learning game specifically designed for refugee children to improve their literacy skills in Arabic. Therefore, it was important that participants did not have Arabic literacy skills or previous experience with this game to get accurate results for effectiveness of GBL. The sample comprised 14 girls (mean age: 7.14, SD: 1.875) and 16 boys (mean age: 7.125, SD: 1.746).

In this study, the participants were recruited and contacted through teachers of the weekend class at Muslim Society Trondheim (MST). The study was organised in nine sessions over one month with migrant refugee children selected through this weekend class program at MST, Norway. The weekend program is an initiative organised at MST, a non-profit, religious, and cultural organisation in Trondheim, Norway. The sessions took place before the weekend class program, and 3–5 children participated in each session. A translator was also present in each session. The study was conducted in two rooms assigned by an MST representative.

3.2.3.1. Ethical issues in research with refugee children.

Some of the ethical issues when researching with refugee children included gaining access, privacy, language barrier and consent as participants are mostly reached through trusted NGOs qualification programs or, religious societies. The details are presented in our previous study (Tahir and Wang 2019). Petousi and Sifaki (2020) highlighted in their research the need for building trust and confidence. Therefore, the research study was thoroughly explained to the MST representatives and

teachers at weekend school to gain their trust before obtaining informed consent from children and parents. Later, the researcher contacted the participants' parents to obtain consent from the legal guardian for the data collection. The researcher also affirmed the child's consent to participate in the study.

The study had been notified to the Data Protection Official for Research, Norwegian Centre for Research Data (NSD) and ethical approval was obtained. The parents were required to give consent on behalf of their children if they are willing to let their children participate in the research, and they could request to see the observation checklist/pre/post questionnaire and interview guide. The parents were asked to sign the written informed consent form. The consent form provided the background and purpose of research along with the information to enable parents to understand what participation in the project implies and voluntarily decide whether to give permission to participate. If parents agree, assent was orally obtained from the child. The researcher explained the research activity and asked if they wish to participate, and the child decided whether the research (*as he/ she understands it*) is an activity in which he or she wanted to participate. In case the child said no, they were free to withdraw from the research. However, as the research involved young children and their decisional capacities may fluctuate, researcher came back to a child who said 'no' after some time to see whether he/she may feel differently later. The participation in the research was voluntary and they could withdraw their consent at any time without giving any reason and their data will be removed. Moreover, there were no negative consequences if they chose not to participate or later decided to withdraw. They were also informed that all information will be anonymised. The personal data will be processed confidentially and in accordance with data protection legislation. The personally identifiable information will be removed, re-written or categorised and participants will not be recognisable in the publication.

3.2.4. Study design

A quasi-experiment with a one-group pre–post-test design was used for this study. The experiment was designed to be one-week-long and comprised two parts: a playtest session (40–55 min duration) and one-week play at home. Children and their parents were invited to MST (for the playtest session and instructions for playing at home). Two rooms were specified for this study, where children played the FTM game using smartphones provided by the researchers. The children's demographic and learning modality preferences data were collected from parents

using questionnaires. Children played the game individually with one or two observers. The user study was conducted with children without (or with a minimum) Arabic reading and writing skills. A translator and two to three observers (two GBL experts and one novice) were present throughout the intervention focusing on observing, taking notes, and managing the user study's overall execution. A translator (with Arabic and English fluency) was required as most of the parents were not fluent in English.

The first part of the user study (playtest session) had three main sections: (1) A pre-test (10–15 min) to examine the previous knowledge of children with Arabic reading skills; (2) a gameplay session (20–25 min duration), where children will play the game; and (3) a short interview (10–15 min) to ask children follow-up questions regarding their game experience and issues encountered while playing. Each child individually played the FTM game with one observer. The observer was responsible for taking notes using an observation checklist and helping the child if he/she got stuck. An example observation checklist form and guidelines for filling out the observation checklist were provided to all observers before the study to have the same understanding. Children were free to leave playing the game before the session was finished if they did not like it. In the follow-up interview, children were asked to do few simple tasks (pause game, replay, turn music on/off, return to level screen), if they did not explore these features during gameplay session, and ask some questions related to their experience. After the playtest session, smartphones with FTM installed were handed to the parents to let children play the game at home daily for one week (for at least 20 min per day). Finally, a post-test was conducted with children when the parents returned the smartphone after one week.

3.2.5. Data sources

The data sources used in this study were (listed in Table 3):

- **Pre/Post Test:** Early Grade Reading Assessment (EGRA) was used (before and after playing the FTM game) to test Arabic reading skills focusing on the following five tasks: Arabic print orientation knowledge, Arabic letter knowledge, Arabic syllable knowledge, recognise initial sound in Arabic word and Arabic word segmentation. Considering the short duration (only one week) of the experiment, we only focused on the Arabic alphabets in the game's first two clusters (see Section 4.2.2 for game clusters). Therefore, the EGRA test was adapted to include only 10 Arabic letters, syllables, and possible words focusing on these clusters.
- **Game Log Data:** The main data in the game log files included the following details: Session Start, Profile ID, Level Start, Question Start, Answer Response, Wrong Answer, Level End, Rewards Gained, Level Score, Session End. The game logs data were recorded for both gameplay session and one week play at home.
- **Observation checklist:** An observation checklist was developed to record all the actions, behaviours, and expressions of children while observing them playing the game. The design of the checklist is adapted from the LEAGUE framework (Tahir and Wang 2020) to our research context. It contained 36 items (constructed using positive phrasing [Diah et al. 2010]) focusing on factors/subfactors of usability dimension mapped to the functionalities of the game and the order in which the players are supposed to perform activities/tasks in the ideal scenario (Diah et al. 2010).
- **Demographic questionnaire:** This included questions related to bio-demographics (such as age, gender, education of the child, level of education of parents, digital literacy, refugee background, country, and spoken language) and mobile usage experience (mobile usage expertise, dependency, frequency, time spent and years of use).
- **Learning modality preference questionnaire:** *VARK for littles* was used to collect the learning modality preference of children. VARK stands for visual (11 questions), audio (13 questions), read/write (11 questions), and kinesthetic (11 questions). The parents were asked to fill this form based on what matches their child's activities and preferences for each category. Each category has a score, and the highest score is considered the preferred modality.

3.2.6. Data analysis

For the quantitative data analysis (RQ1–4 in Table 3), the IBM SPSS Statistics v27 software was used. We conducted Wilcoxon Matched Pairs Test to test any potential difference in the Arabic reading assessment (EGRA) score of refugee children following a GBL intervention. To investigate differences between younger and older children, we split the sample using the age of six as a threshold following Piaget's theory of cognitive development (Huitt and Hummel 2003) and focusing on pre-operational (younger children) and concrete operational (older children) stages. We conducted the Mann–Whitney test to examine any potential differences in the children's age group and their learning gain, usability score, and gameplay performance.

Spearman correlation was used for identifying any potential correlations: between children learning modality preference and their learning, usability score, and gameplay performance, and between mobile usage experience and their learning, usability score, and gameplay performance. Furthermore, the grounded theory approach proposed by Gioia, Corley, and Hamilton (2013) was used to analyse and interpret the qualitative data (collected through observation notes and interviews) to present data structure for usability issues.

4. Results

This section presents the results from the quasi-experiment concerning the five research questions. The first four questions concern quantitative data analysis using statistical tests. We reported the differences in the learning outcome; looked at differences in children's age groups; identified the effect of learning modality preference and mobile usage experience on learning gain (LG), usability score, and gameplay performance. Lastly, the fifth question concerns qualitative analysis and looked at the usability issues faced by migrant refugee children when playing the game, the first time.

4.1. Learning outcome with game-based learning (RQ1)

This research question focused on the potential learning gain (difference in pre and post-test of Arabic reading assessment scores) of migrant refugee children after using a game-based language learning approach in an informal learning setup. Our null hypothesis was that there is no difference in the Arabic reading assessment score of migrant refugee children from playing the FTM language learning game for one week at home.

A Wilcoxon Matched Pairs Test was conducted to evaluate if there a difference in the pre and post-test scores of refugee children following a GBL intervention. The results revealed a statistically significant positive change in Arabic reading assessment score following playing the Arabic language learning game ('Feed the monster'), $z = -4.7821$; p -value is $< .00001$. The result

is significant at $p < .05$, with large effect size ($r = 0.873$) according to Cohen (1988) criteria.

4.2. Differences in age groups (learning gain, usability score, and gameplay performance) (RQ2)

This research question investigated how the children's learning gain (LG), usability score, and gameplay performance from using a learning game differs between two age-groups: younger (5–6 years old) and older children (7–10 years old). Our null hypothesis was that there are no differences between younger and older children's learning gain, usability score, and gameplay performance with the FTM language learning game.

A Mann–Whitney test was used to examine the differences between younger and older children. The independent variable was the children's age (younger or older). The dependent variables were learning gain, usability score, and gameplay performance parameters (total levels, total time, total wrong, total sessions, score in gameplay session). This test is appropriate for our quasi-experiment as it is a nonparametric test used to compare differences between two independent groups where the samples can be of different sizes and when the dependent variable is either ordinal or continuous. The results from the Mann–Whitney Test are shown in Table 4.

Table 4 shows significant differences in learning gain (with medium effect size $r = 0.4365$) and usability score (with large effect size $r = 0.5895$) of younger and older children. The younger children had higher learning gain compared to the older children, while the older children had better usability scores than the younger children. Observations from the playtest session confirm these differences between younger and older children. Most younger children had difficulty recognising some icons/buttons/concepts and performing gestures (such as extended drag). In comparison, the older children thought that the game was too easy with no option to increase the difficulty. The results showed a statistically significant difference in total levels played (with medium effect size $r = 0.4325$) by younger

Table 4. Mann-Whitney test results.

	Younger children Median ($n = 14$)	Older children Median ($n = 16$)	U	Z	P	$r = z/\sqrt{N}$
Learning gain	0.2	0.1	54.500	−2.391	0.0168*	0.4365
Usability score	25.5	33.5	34.500	−3.229	0.001*	0.5895
Total levels	75	131	55.000	−2.369	0.0178*	0.4325
Total time	184	264.5	69.000	−1.788	0.0738	0.3264
Total wrong	77.5	56.5	89.000	−0.956	0.3388	0.1745
Total sessions	10.5	9.5	106.500	−0.229	0.8186	0.0418
Score in game play session	20880.50	30151.00	69.000	−1.788	0.0738	0.3264

Significance level, * $p < .05$.

Table 5. Spearman correlation results for learning modality preference.

			LG	Usability score	Total levels	Total time	Total wrong	Total sessions	Score in gameplay
Spearman's rho	Visual	Correlation	-.077	.281	.361	.425*	.045	.318	.088
		Coefficient							
		Sig. (2-tailed)	.685	.133	.050	.019	.812	.087	.644
	Read/write	N	30	30	30	30	30	30	30
		Correlation	-.320	.570**	.527**	.397*	-.078	-.044	.350
		Coefficient							
		Sig. (2-tailed)	.084	.001	.003	.030	.682	.818	.058
	Kinesthetic	N	30	30	30	30	30	30	30
		Correlation	-.038	-.035	-.041	-.011	-.022	.117	-.052
		Coefficient							
		Sig. (2-tailed)	.841	.852	.831	.956	.906	.537	.786
	Aural	N	30	30	30	30	30	30	30
		Correlation	-.169	.357	.370*	.421*	.192	.243	.215
		Coefficient							
		Sig. (2-tailed)	.373	.052	.044	.021	.310	.196	.254
		N	30	30	30	30	30	30	30

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

and older children. The older children played more levels than younger children. However, there was no significant difference in the total time played, total sessions played, and total wrong answers between the two groups ($p \geq 0.05$).

4.3. Effect of learning modality preferences (learning gain, usability score, and gameplay performance) (RQ3)

This research question investigated the effect of learning modality preference on the learning gain, usability score, and gameplay performance of migrant refugee children when playing the language learning game (FTM). Descriptive statistics showed that 33.33% of children had multimodal learning preference, 30% preferred aural, 23.33% preferred kinesthetic, and 13.33% preferred visual. Interestingly, none of the children preferred read/write as a distinct preference, as it was always preferred within multimodal learning preference.

A series of spearman's correlations were conducted to identify any potential correlation between the children's learning modality preference parameters (visual, read/write, kinesthetic, and aural) and their learning gain (LG), usability score, and gameplay performance parameters (total levels, total time, total wrong, total sessions, score in gameplay session). Spearman correlation is about the strength and direction of the relation between two variables and is a nonparametric alternative of Pearson correlation. Spearman was used as the assumptions for Pearson correlation were not met. All results for spearman's correlations are presented in Table 5.

The two-tailed test of significance indicated no significant relationship between the four learning modality

preference variables and children's learning gain. It means that the preferred modality of learning does not affect children's learning with the educational game. However, regarding the association between the four learning modality preference variables and the usability score, spearman's test verified a relatively strong relationship between one of the four learning modality variables, as indicated in Table 5. There was a significant positive relationship between the preference for the read/write learning modality and children's usability score. It shows that children who had a higher preference for read/write modality had higher usability score. The observations from the playtest session also revealed that many children were uninterested in textual instruction provided in the game and would try to skip them without reading.

Lastly, the spearman's test for the correlation between the four learning modality preferences and the four gameplay performance variables indicated a significant positive relationship between three out of the four learning modality preferences with two out of four gameplay performance variables as shown in Table 3. Children's total time playing the learning game was positively related to their preference for visual, read/write, and aural learning modalities. Moreover, children's total levels were positively related to their preference for read/write and aural learning modalities. On the other hand, no significant relation was found between preference for kinesthetic learning mode and gameplay performance variables. Overall, the results showed that usability and gameplay performance was higher for children who prefer the read/write and aural modality for learning. These results are relevant in the employed game context (FTM) as the game highly emphasised these two modalities of learning.

4.4. Effect of mobile usage experience (learning gain, usability score, and gameplay performance) (RQ4)

This research question investigates the effect of mobile usage experience on the learning gain, usability score, and gameplay performance of migrant refugee children when playing the language learning game (Feed the Monster). Again, spearman's correlation was used to identify any potential correlation between the children's mobile usage experience parameters (mobile usage expertise, years of mobile use, dependency in mobile usage, frequency of usage, and time spent on mobile) and their learning gain (LG), usability score and gameplay performance parameters (total levels, total time, total wrong, total sessions, score in gameplay session). The results are presented in Table 6.

Spearman's test indicated no significant relationship between learning gain and any of the four mobile usage experience parameters. It means that their experience in using mobile technology does not affect children's learning with the educational game. On the other hand, spearman's test revealed a significant positive relationship between usability score and two of the four mobile usage experience variables, as indicated in Table 6. It shows that children who had more mobile usage expertise and more years of mobile use had high usability scores when playing the learning game for the first time. Regarding the association between mobile usage experience and gameplay performance parameters, spearman's test revealed a significant positive relation between mobile usage dependency and the total sessions played by children, and between years of mobile use and score in play session. Interestingly,

children who had higher mobile usage dependency (i.e. they do not use the mobile alone, but someone else mostly plays with them) played a greater number of sessions. In comparison, children with greater mobile experience (years of use) performed better in the game (higher score).

4.5. Usability issues faced by migrant refugee children (RQ5)

The last research question investigated the issues faced by migrant refugee children when playing the learning game (FTM) for Arabic reading skills. In addition to the usability observation checklist score, we also collected the qualitative data from observation notes and short interviews with children. As reported by Creswell and Creswell (2017), it is important to analyse the qualitative data of the research to deepen the understanding of research participants' critiques. Therefore, the results reported here (Figure 1) are based on the observations (children's behaviour in the playtest session) and the interview responses. The interview gathered children's feedback regarding what they learned from the game (learning objectives, retention, learning task), their understanding of the game (game definition, narrative, gameplay, rewards), what they liked or disliked about the game, favourite part of the game (game aesthetics, enjoyment, engagement motivation), what difficulties they faced (interface), and if they want to change anything (satisfaction and attractiveness).

We followed the procedure described by Gioia, Corley, and Hamilton (2013) for data analysis and presenting the issues found in the learning game. We came up

Table 6. Spearman correlation results for mobile usage experience.

			LG	Usability score	Total levels	Total time	Total wrong	Total sessions	Score in play session
Spearman's rho	Mobile usage expertise	Correlation	-.259	.362*	.243	.266	-.231	.123	.222
		Coefficient							
		Sig. (2-tailed)	.166	.049	.196	.156	.219	.516	.238
	Years of mobile use	N	30	30	30	30	30	30	30
		Correlation	-.251	.513**	.290	.306	-.040	-.126	.372*
		Coefficient							
	Dependency in mobile usage	Sig. (2-tailed)	.180	.004	.120	.100	.833	.506	.043
		N	30	30	30	30	30	30	30
		Correlation	-.018	-.069	.127	.219	.264	.527**	-.235
	Frequency of usage	Coefficient							
		Sig. (2-tailed)	.927	.719	.504	.244	.159	.003	.211
		N	30	30	30	30	30	30	30
	Time spent on mobile	Correlation	-.075	.237	.082	.049	-.079	.025	.115
		Coefficient							
		Sig. (2-tailed)	.694	.206	.667	.797	.679	.894	.547
		N	30	30	30	30	30	30	30
		Correlation	-.089	.157	.134	.137	-.225	.180	.102
		Coefficient							
		Sig. (2-tailed)	.641	.409	.480	.472	.232	.342	.592
		N	30	30	30	30	30	30	30

*Correlation is significant at the 0.05 level (2-tailed).

**Correlation is significant at the 0.01 level (2-tailed).

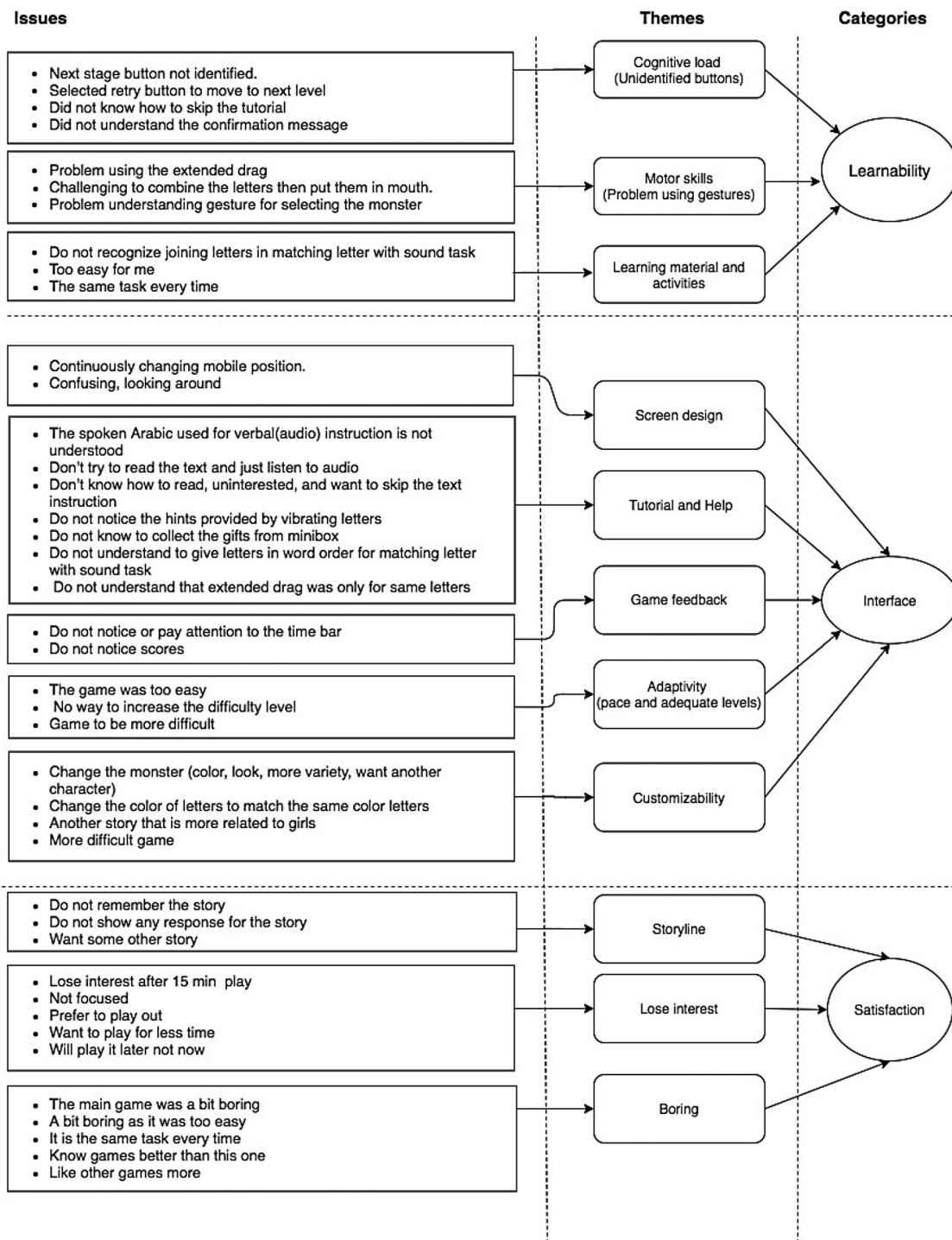


Figure 1. The qualitative analysis of observation and interview data.

with the data structure with three levels (shown in Figure 1): the first level with the issues highlighted in the raw data (from observations and children interview responses), the second level identified the themes for these issues and, the third level distilled these themes into categories of usability issues. The LEAGUE framework (Tahir and Wang 2020) and usability model (Tahir and Arif 2015) was used to analyse and combine

the groups of codes to generate themes and compared them against each other to form categories.

According to qualitative analysis (presented in Figure 1), the *learnability* category problems were related to cognitive/motor skills and learning material and activities. Many children could not recognise the icons/buttons/concept, or it was difficult for them to perform a gesture. Learning material and activities were not

suitable for all users, not comprehensive enough, and not many opportunities. Especially older children found them too easy. Deeper analysis showed that most children with cognitive or motor skills-related issues belonged to younger children age-group (5–6), which is in line with the quantitative results from RQ2. Moreover, issues related to learning material and activities and adaptivity were also age-related, where older children (7–10) wanted the game to be more difficult.

The *interface* category issues were related to screen-design, tutorial and help, game feedback, adaptivity, and customisability. For screen design, mostly younger children experienced some confusion and difficulty holding mobile while playing the game. For tutorial and help, the qualitative data from observation and interview showed that children most liked the visual and audio instruction; however, some children faced difficulties understanding the audio used in the game when they had a different dialect. On the other hand, many children were either uninterested or did not know how to read the textual instructions. Therefore, they would try to skip the instructions. The instructional support provided through hints such as vibrating letters was not noticed by many children. Moreover, some tasks and activities were not completely understood by children on their own as details were not provided with audio instruction. For game feedback, most children did not notice the time bar or score. For adaptivity, the game's pace was not suitable for many children (especially older), and they thought it was not challenging enough, and there is no variation in the game in terms of difficulty level and customisability. In addition to the need for different difficulty levels, many children wanted customisable options to change different characters, colours, and scenarios in the game.

Lastly, the *satisfaction* category issues were related to the following themes: storyline, lose interest, and boring. Many children were not engaged in the story of the game and did not remember it. For most children, the game kept their interest for a shorter duration only. Moreover, the game was boring for some children due to little variation in gameplay, and they thought they had other better options (games) to play.

5. Discussion

This section discusses the results from the previous section, presents design recommendations learned from the main findings and discusses some limitations of this evaluation study.

5.1. Discussion of the results

In this section, we discuss the results from our evaluation study with migrant refugee children (5–10 years

old), in which we examined the effectiveness of GBL approach for the Arabic reading skills when used in an informal learning context using quantitative and qualitative data (see Section 3.2.1 for evaluation plan).

5.1.1. Impact on learning gain

The findings from this research study suggest that GBL is an effective learning tool for informal learning setups to increase the Arabic reading skills of refugee children (RQ1). The test showed a statistically significant positive difference in the Arabic reading assessment score of children after playing with the game (Feed the Monster) for one week at home, which is in line with similar research by (Salah et al. 2016; Azizt and Subiyanto 2018; Kenali et al. 2019) conducted for formal teaching setup that showed an increase in learning gain.

5.1.2. Age-group differences and need for adaptivity in educational games

To investigate differences between children's age-groups when using GBL, we divided the sample into younger children (5–6 years) and older children (7–10 years) following (Huitt and Hummel 2003). The key motivation was to analyse the difference in learning gain, usability, and gameplay performance across age groups that can suggest design improvements for children's learning games. Surprisingly in this study, the younger children outperformed the older children in terms of learning gain. The qualitative analysis revealed that the main reason for this difference was linked to the educational material and difficulty level of the game, which was too easy for older children, and thus they had lower learning gain and felt boring. According to a study by Greenberg et al. (2010), the strongest motivator for 5th-grade children is the perceived challenge. Therefore, the learning game should provide children with challenges (related to the main task) balanced with their skill level (Kiili et al. 2012; Csikszentmihalyi 2000). The research findings indicate the need for adaptivity in the learning game (Streicher and Smeddinck 2016; Plass and Pawar 2020; Peirce, Conlan, and Wade 2008). The game should increase the difficulty level in relation to player skills, and the game designers should ensure that the challenges should become more difficult when the player's skill level increases (generate game levels tailored to player's knowledge) (Steiner et al. 2012; Kickmeier-Rust 2012; Lopes and Bidarra 2011; Andersen 2012).

However, as expected, the older children had a better usability score than the younger children. It led us to explore further any differences between age-groups in terms of their mobile usage experience, which could be the reason for better usability with older

children (supposedly having more mobile experience). Therefore, we conducted another Mann–Whitney test and found no significant differences in mobile usage experience between the two age groups, which confirmed that differences in usability score between younger and older children were due to age-related factors and not because of higher mobile usage experience. These results are also in line with the findings from qualitative data, where younger children faced difficulty in performing complex gestures (such as extended drag), recognising icons/buttons/concepts, and some also found the screen design to be a bit confusing. According to Piaget’s theory of cognitive development (Huitt and Hummel 2003), these issues are related to cognitive and motor skills that are not fully developed in younger children. Furthermore, the results revealed a statistically significant difference in the number of levels played between younger and older children, where older children played more levels than the younger. It can be linked to the nature of older children being more motivated by competition as claimed by Greenberg et al. (2010) and want to win the game by completing all levels.

To conclude, it appears that the children’s age can be associated with their learning gain, usability, and gameplay performance (RQ2). Therefore, adaptivity is important in learning game to adapt to each child individually for improved effectiveness (Andersen 2012; Vandewaetere et al. 2013).

5.1.3. Learning modality preferences and multimodal learning approach

We found no significant correlation between children’s learning modality preferences and their learning gain, consistent with the previous research in this area, which debunked the theory that presenting material in preferred modality can improve learning (Lodge, Hansen, and Cottrell 2016). Although many research studies showed that modality does not affect learning, some researchers think that different media may afford different instructional methods and be effective for the learning process (Moreno 2006).

We found positive and significant correlations between the children’s preference for the read/write learning mode and their usability score. The children who had a higher preference for read/write learning modality had higher usability score. According to the results, the older children (0.63) had a higher mean score for read/write modality than younger children (0.33). Therefore, we conducted a Mann–Whitney test and found that read/write preference in the older group was statistically significantly higher than, the younger children group ($U = 40.0$, $Z = -3.0139$,

$p = 0.00257$). The qualitative data from observation showed that most children (especially younger) skipped or tried to skip the textual instructions provided in the game because either they were uninterested or did not know how to read the instructions. On the other hand, some children also had difficulty understanding the audio instructions provided in the game, mainly because of the spoken dialect. However, the majority understood and followed the visual instructions. According to Harskamp, Mayer, and Suhre (2007), the principle of modality is more likely to occur when instructions aim to promote learner understanding. Therefore, the game instructions must be delivered through different media (Moreno 2006). The research by Harskamp, Mayer, and Suhre (2007) recommended accompanying graphics by concurrent narration instead of on-screen text. With the use of auditory media to process the words, the children are not forced to divide the limited visual working memory (between pictorial information and on-screen text), thus expanding the capacity of effective working memory (Moreno, Low, and Sweller 1995).

Finally, we also found that total time played was positively related to preference for visual, read/write, and aural learning modalities. Similarly, total levels completed was positively related to preference for read/write and aural modalities. These results were predictable since the game (FTM) was for Arabic reading skills and hence had more audio and text focusing on sound and written form of alphabets and words. Therefore, children with a higher preference for read/write, video, and aural modalities played the game more as opposed to those with a higher preference for kinesthetic. It was also evident from observation of playtest sessions where some children were bored after 15 min and instead wanted to play other more physical games. Therefore, it is recommended to provide different learning activities incorporating multiple modalities for the game to be effective for most children (Alkhasawneh et al. 2008; Ward et al. 2017). Overall multimodal learning is also supported by theoretical development as opposed to the modality-specific learning style theory (Aslaksen et al. 2020). Although learners might prefer a modality in certain situations but overall, a more multimodal approach is plausible. According to the results of this study, many children represented a multimodal learning preference, which is similar to the results by Alkhasawneh et al. (2008).

To conclude, children’s learning modality preferences are not associated with their learning gain, but it impacts their usability score and gameplay performance parameters (RQ3). It means that children’s learning with educational games is not affected by their preferred

modality of learning; however, different media's functional capabilities can be used to enable effective learning methods based on learning theories and research (Moreno 2006). The use of a multimodal approach can enhance the usability and gameplay experience for most users (Aslaksen et al. 2020).

5.1.4. Mobile usage experience and need for customisability

Similar to learning modality preferences, we found no significant relation between children's mobile usage experience and their learning gain using FTM, which is in line with the previous research by Kim and Chang (2010). However, these non-significant findings contrasted with Sylvén and Sundqvist (2012) that reported a positive relationship between the time spent playing games and language proficiency. In another study by Chen and Huang (2013), ANOVA results indicated significant differences in the learning performance of groups with different prior knowledge (frequencies of playing game and experience of playing game). However, the effect of prior knowledge (positive or negative) depended on the nature of knowledge (i.e. declarative or procedural knowledge) delivered by the game, and significant differences were found in only a few groups. Thus, one explanation of this could be that prior experience affects the learning outcome when that experience/knowledge is used in the game to solve tasks that induce learning.

Consistent with prior research (Orvis, Horn, and Belanich 2006), this study outlined that mobile usage experience parameters (usage expertise and years of use) are positively related to the usability score when playing the learning game for the first time. The children with more years of mobile use and who require less help to use mobile are already familiar with many of the interface characteristics from the prior experience shared in the current game (Salanova et al. 2000). Therefore, researchers argue that experience parameters (such as technology exposure, computing literacy, and level of assistance required) must be considered when designing the user interface (Belay, McCrickard, and Besufekad 2016). According to the research investigating expertise in various domains (Ericsson 2002), the performance difference between experts and novice is because of the deliberate practice accumulated over a period of time. According to MacDorman et al. (2011), it is important to note that an effective interface is not necessarily the one that is easiest to use for novices as an interface with a steep learning curve is much more efficient to perform the task fast after sufficient experience. Bunt, Conati, and McGrenere (2007) recommended providing customisation suggestions

tailored to user characteristics, expertise, use patterns, and features of the interface to maximise user performance. Many researchers supported user interface customisation, Burkolter et al. (2014) showed that it enhances user acceptance and reduces errors, whereas Jorritsma, Cnossen, and van Ooijen (2015) pointed that users work more efficiently when provided with customisation support.

Finally, we also found an association between mobile usage experience and gameplay performance parameters. Dependency in mobile usage is positively correlated to the total sessions played by children, and years of mobile use is positively correlated to score in the gameplay session. The latter is in line with the previous research on video game experience, which showed that learners with greater experience performed better (Orvis, Horn, and Belanich 2006, 2008). Previous research suggested that different instructional techniques are effective for novice and high experience users to maximise performance and motivation (Kalyuga 2009; Schnotz and Rasch 2005; Baig and Kavakli 2018). Therefore, to reduce the difference between novice and high experience, the game developers should allow the users to select the desired level of content and practice in the learning game according to different user needs and provide an adequate amount of preparatory practice (Orvis, Horn, and Belanich 2008). Customised GBL will lead to positive perceptions and enhanced performance (Ku, Hou, and Chen 2016). Moreover, allowing the players to customise game characters' behaviour increases user engagement that comes from the sense of personal ownership (Kleinsmith and Gillies 2013).

Surprisingly, the children with higher mobile usage dependency (need someone else to play with them) played more sessions. One plausible explanation for this could be that the additional engagement is driven by social interaction because these children are mostly playing with family or friends. As indicated by previous research that co-participation and availability of adults in play can increase the duration of children's play (Pursi and Lipponen 2018; Singer et al. 2014; Siraj-Blatchford 2009).

To conclude, children's mobile usage experience does not affect their learning gain, but it impacts their usability and gameplay performance (RQ4).

5.2. Design recommendations

In this section, we present design recommendations based on the qualitative analysis of observation notes from the playtest session and interview responses of children that identified usability issues (RQ5) and the

lessons learned from the main findings of the quantitative data (RQ1–4). The design recommendations are summarised as follows.

5.2.1. Age-appropriate design

It is essential to consider children's age-group differences in the learning game design as they impact their learning gain, usability, and gameplay experiences. Children in different age groups belong to different cognitive developmental stages (Huitt and Hummel 2003). The younger children have less developed cognitive/motor skills that can affect their educational game learnability when using it the first time. On the contrary, older children might find the learning material and activities too easy, wanting the game to be more challenging. Therefore, it is vital to design the interface (icons, buttons, concept, and gestures), learning material, and activities suitable for all users considering their cognitive and motor skills. It should be comprehensive enough for younger children but at the same time provide more opportunities for the older children.

5.2.2. Customisability

Educational games must provide customisation opportunities to different users based on their prior experiences.

Customisation is referred to a manual process of user-controlled adaptation that provides features for users to allow them to make adjustments and variations by themselves. This process usually occurs before the actual activity, where users adjust the parameters provided by the game to shape the experiences and content according to their choice (Feng et al. 2020; Orji, Oyibo, and Tondello 2017). In the present study, many children wanted customisable options to change different characters, colours, difficulty levels, and scenarios in the game. Children with more mobile usage experience are already familiar with many of the interface characteristics (Salanova et al. 2000). Therefore, there is a performance difference between experts and novices due to the practice and experience accumulated over time (Ericsson 2002). Therefore, the interface should provide the users with customisation options in the learning game (Orvis, Horn, and Belanich 2008).

Previous research has suggested different customisation suggestions. Providing options tailored to user expertise, characteristics, features of the interface, and use patterns can maximise user performance (Bunt, Conati, and McGrenere 2007). Allow users to select the level of content and practice according to their needs and desire and providing an adequate amount of basic practice to reduce the difference between novice and experienced users (Orvis, Horn, and Belanich

2008). As different instructional techniques might be effective for different users (novice and high experience) to maximise motivation and performance (Kalyuga 2009; Schnotz and Rasch 2005; Baig and Kavakli 2018). Many researchers supported user interface customisation. Burkolter et al. (2014) showed that reconfiguration user interface is promising for enhancing user acceptance. Jorritsma, Cnossen, and van Ooijen (2015) pointed that user interface customisation support (suggestions) was useful in a natural work environment. Moreover, researchers have highlighted many positive effects of providing customisation: lead to positive perceptions and enhanced performance (Ku, Hou, and Chen 2016), increases user engagement that comes from the sense of personal ownership (Klein-smith and Gillies 2013), users work more efficiently when provided with customisation support (Jorritsma, Cnossen, and van Ooijen 2015), enhances user acceptance and reduces errors (Burkolter et al. 2014).

5.2.3. Adaptivity

Adaptivity is an autonomous process of system-controlled adaptation requiring minimal to no effort from the user. It assesses users' learning progress and competencies by monitoring their performance to make adaptations dynamically. The most prevalent adaptation includes the balance of game difficulty (Feng et al. 2020; Kelley 1969; Kickmeier-Rust and Albert 2010). The GBL designers should ensure that the game challenges become difficult with increasing players' skills by generating game levels tailored to players' knowledge (Steiner et al. 2012; Kickmeier-Rust 2012; Lopes and Bidarra 2011; Andersen 2012; Malone 1981; Sweetser and Wyeth 2005). In the present study, the game pace was not suitable for many children. Especially older children thought that the game is not challenging and there is no variation in terms of the game's difficulty level.

5.2.4. Meaningful feedback

A learning game should provide meaningful feedback to the users (Sweetser and Wyeth 2005). Use engaging animations for providing hints and feedback to make them noticeable. In the present study, most children did not notice the time bar, score, and hints (provided by the correct letter's slight vibration). However, the instructional support, such as feedback on answers (provided by the monster spitting out the incorrect answer and eating the correct one), was understood and liked by most children. Therefore, it is crucial to design meaningful feedback. Visual feedback with hand gestures was most appropriate for children as most young children did not understand the written feedback. Use of

exciting rewards such as gifts and actions (such as monster growing bigger) was found more useful rather than simple score. Researchers have suggested that the use of adaptive feedback can facilitate learning, attitude, and immersion (Kickmeier-Rust et al. 2008).

5.2.5. Adequate help and tutorial

It is important to provide adequate help and tutorials in the educational game to support learners. According to Kiili (2005), it is possible to extend the flow channel and keep players interest in the game by providing guidance. The research has found that the use of tutorials increases playtime in more complex games (Andersen et al. 2012). Moreover, researchers argue that the experience parameters (level of assistance required, technology exposure, and computing literacy) must be considered when designing the user interface (Belay, McCrickard, and Besufekad 2016). An effective interface is not necessarily the easiest one, but the one with a steep learning curve is much more efficient to perform the task fast after sufficient experience (MacDorman et al. 2011). The designers should keep users' different needs in mind and provide the appropriate amount of preparatory practice and tutorials (Orvis, Horn, and Belanich 2008). It will reduce the difference between novice and high experience. Moreover, additional help and details are required for difficult and different game tasks at least the first time. In the present study, some children did not understand that for some game tasks, they had to give letters in the correct order when feeding the monster with letters.

5.2.6. Pedagogic feedback for reinforcement learning

Educational games are effective if they impart the required learning to the users. Therefore, educational games must provide opportunities to self-correct in order to reinforce learning. It can be done by providing pedagogic feedback and correct answers. In the present study, there were not many options provided to children to self-correct. The game does not demonstrate the correct answer when children gave an incorrect one and do not explain why the selected choice was incorrect. Children understood that the answer is correct or incorrect depending on whether the monster eats the letter or throw it. However, they did not know why their answer was incorrect or what is the correct answer. It is important to demonstrate the correct answer to reinforce learning. However, in the present game, children can progress in the game even if they feed incorrect letters to the monster. Researchers have found corrective feedback useful in educational games (Cornillie, Clarebout, and Desmet 2012). According to Tzetzis, Votsis, and

Kourtessis (2008), corrective feedback increases self-confidence and outcome scores over time. Moreover, cognitive feedback provides the account for cognitive immersion and learning as it relates to cognitive problem-solving. It stimulates game players to reflect on their solutions and experiences to further develop their playing strategies and mental model, focusing player attention on relevant information intended for learning objectives (Kiili et al. 2012; Van Merriënboer and Kirschner 2017).

5.2.7. Engaging game tasks and learning activities

Educational games should provide variation in the game and learning tasks and offer tasks and activities with adequate levels of difficulty for different users to keep their interest (Malone 1981). Research suggests providing users with different learning activities in the game incorporating multiple modalities to be effective for most children (Alkhasawneh et al. 2008; Ward et al. 2017). Users enjoy playful tasks and put more effort into learning and exploring the new tasks. Playful tasks also foster creativity (Prensky 2001). In the present study, some children were bored after 15 min and instead wanted to play other more physical games. The game was boring for some children due to little variation in gameplay activity. They did not like the same task every time and wanted something different. Moreover, most children would become active and engaged in the different mini-games (level breaks with different activities) and repeated them several times. Many researchers have focused on designing tasks with different difficulty levels (Ibrahim and Jaafar 2009). However, it is also vital to design interesting and engaging tasks (such as role-playing, interacting with multimedia, or gaming aids) as they keep students' focus (Abdul Jabbar and Felicia 2015).

5.2.8. Empathy and connection

It is important to create empathy and connection with users in the educational game. In the present study, children felt an association with the game character. They liked feeding the character (monster) and making it bigger. Therefore, educational game designers should use animations and game characters that connect with users. Moreover, it is also possible to create empathy through game story scenarios to emotionally connect players (Neuenhaus and Aly 2017). Bachen et al. (2016) suggest that educational game designers should prioritise focus on stimulating empathy and immersive presence in users. In-game empathy predicts increased interest in learning about the game topics (Bachen et al. 2016).

Research has found an association between game character and affective reactions. Sierra Rativa, Postma, and Van Zaanen (2020) reported that facial expressions and virtual animals' appearance in a simulated environment leads to higher levels of players' empathy and immersion. Therefore, the game character's appearance is important as it can influence the user experience concerning empathy and immersion. Bailey, Wise, and Bolls (2009) indicated that game avatar customisation could affect both psychophysiological indicators of emotion and subjective feelings of presence during gameplay, making children's gameplay experience more enjoyable. The customisation of the avatar can incite identification with the avatar (Birk et al. 2016). Researchers have also explored the use of player-avatar identification. It is a process where players view themselves as a game character and feel the same emotional attachment (Wang et al. 2020). Wang et al. (2020) found that player-avatar identification impacts user's behavioural intention. The greater avatar identification translates into motivated behaviour in a task and fosters immersion, experienced autonomy, enjoyment, invested effort, and positive affect (Birk et al. 2016). It also has implications for educational game design to help change players' behaviour. The options to facilitate identification include increasing similarity, sense of embodiment, and wishful identification with an avatar representing player's ambitions (Birk et al. 2016).

5.2.9. Effective use of multimedia for multimodal learning

GBL should incorporate effective use of multimedia by offering learning activities integrating multiple modalities for multimodal learning approaches (Ward et al. 2017) to enhance usability and gameplay experience (Aslaksen et al. 2020). Therefore, it is important to identify different media's functional capabilities and their relevance to the learning process to facilitate effective instructional and learning methods for the game to be effective for most users. In the present study, overall, most children liked the visual and audio instruction. The majority of children understood and followed the visual instructions. However, there were some difficulties in understanding the different spoken dialects in audio instruction. On the other hand, most younger children were uninterested in textual instructions or did not know how to read. Therefore, visual and audio instructions should be used more often than written instructions, especially when designing for children. It is important to use a clear voice and dialect familiar to the target users when using audio instructions in learning games. Moreover, sometimes it is important to use voice instructions within visual tutorials for providing

extra details, especially if the visual tutorial cannot emphasise the minor details and extra information is needed. In the present study, some tasks and activities were not completely understood by children, and they needed help from observers as details were not provided with audio instructions. Previous research also found that it is more effective to accompany graphics with concurrent narration rather than using on-screen text (Harskamp, Mayer, and Suhre 2007). The use of auditory media expands the capacity of effective working memory as children are not forced to divide the limited visual working memory between pictorial information and on-screen text (Moreno, Low, and Sweller 1995). It was also found that interesting sounds of Arabic letters attract attention and help children remember the letters. Children picked the letter sounds from the game and enjoyed repeating them. The instructions in learning games must be delivered through different media (Moreno 2006) as the principle of modality is more likely to occur when instructions aim to promote learner understanding (Harskamp, Mayer, and Suhre 2007). However, designers must consider the multimedia learning issues in correspondence with the cognitive load theory when incorporating different media in educational game design.

5.2.10. Intriguing storyline

It is important that the storyline of an educational game is interesting and engaging for different user groups. The story can embody fantasy and generate emotions (Malone 1981; Prensky 2001) to keep users interested by sustaining their intrinsic motivation, curiosity, and plausibility instigated by the game environment (Dickey 2011). In the present study, some children were not engaged in the game's story and did not remember it, which affected their satisfaction with the game. Park et al. (2010) discovered that players enjoy the game more when exposed to a pre-game story. Fantasy plays an important role in children's play to 'assimilate' experience in the existing structures in children's mind and can be very important in making the instructional environments more motivating, interesting and educational (Malone 1981; Piaget 1952). Therefore, it can be useful to incorporate fantasies in the storyline to make the educational game environment more interesting; however, it is crucial to carefully choose fantasies that appeal to the target users (Malone 1981). Moreover, narrative game mechanics can be used to create an engaging story focusing on evoking empathy through representation, creating moral dilemmas via player choices, and building tension via spatial conflict (Dubelman 2016).

5.3. Limitations of the study

We recognise that our work entails several limitations. One such limitation is that there was no control group to compare the learning gain. Since the employed learning game teaches Arabic reading skills to refugee children in an informal learning setup where no other means like traditional education are available, therefore, a quasi-experiment was used without any control group considering the context of use. However, to mitigate this, we performed a matched-pairs study with one group pre-test–post-test design using the Wilcoxon Matched-Pairs Signed-Ranks Test to examine the difference in value before and after the intervention from the same subjects (MacFarland and Yates 2016). Since the pre and post-test used similar questions, one could argue that the pre-test might have affected the post-test outcome. However, the post-test was conducted after a one-week gap and was not exactly the same, which counters the testing threat. Although the post-test followed the same pattern, and the questions were based on the same ten alphabets, syllables, and words, they were rearranged and reordered. The sample used in this evaluation study was not randomised. We used selective sampling since only children with refugee backgrounds (limited in number and availability) could serve as the primary data sources due to this study's objective. However, the selection was based on defined criteria, and the gender distribution was fairly equal, thus reducing the effect of selection bias (Sharma 2017). Finally, the results reported in this article were based on the evaluation of a particular learning game with migrant refugee children. Although the results should be applicable for GBL for various subjects, we acknowledge that the results might not be transferable to any user population. It is possible that similar analyses with other user groups might produce different results in terms of age differences. The proposed LEAGUE-GQM approach is still in early stages and will be further refined by applying in GBL evaluation studies at different development stages. One limitation of the proposed LEAGUE-GQM approach and evaluation guide is that it provides only two qualitative analysis methods in the current version and a link between first and second part of the approach is missing. For example, it does not guide which criteria to prioritise for different evaluation types, evaluators have to pick and choose themselves. The extended version will incorporate these changes.

6. Conclusion and future work

In this paper, we proposed and employed an integrated LEAGUE-GQM approach to plan and evaluate the

potential of a GBL approach for Arabic reading skills of migrant refugee children, discover any differences in their performance (learning outcome and gameplay performance) and experience (usability) concerning their user characteristics (age group, learning modality preferences, and mobile usage experience), and identify any usability issues in the learning games. We conducted a quasi-experiment using one group pre and post-test design with 30 children aged 5–10 and collected quantitative and qualitative data using multiple methods including pre/post-test, questionnaire, interview, game log, observation checklist, and notes. Based on our findings, we present the following conclusions:

- The GBL approach has great potential as an educational tool in informal learning setups to improve refugee children's Arabic reading skills.
- Educational game designers, educators, and researchers should consider the existence of differences between different age-groups of children and realise its impact on learning, usability, and gameplay.
- GBL should incorporate effective multimodal learning approaches (Ward et al. 2017) by offering different learning activities integrating multiple modalities to enhance usability and gameplay experience. Therefore, it is important to identify different media's functional capabilities and their relevance to the learning process to facilitate effective instructional and learning methods for the game to be effective for most users.
- Educational games should assess users' prior experiences and provide different users (novice and experienced) with customisation opportunities.

The paper presents some recommendations for the design of GBL applications for refugee children based on data analysis and the identified usability issues that are also applicable for the learning game design in general. The future work will focus on employing the integrated approach to plan various GBL evaluation studies, exploring the potential of the approach with different study designs, evaluation stages, and learning domains. The future work will also focus on exploring the relationship between user characteristics and affective reactions, such as focusing on the relation between parent–children play that increases engagement. Future studies could also compare the results of the GBL approach in an alternative context of use and obtain deeper insights from a longitudinal collection of learning and psychosocial data for language learning games.

Disclosure statement

No potential conflict of interest was reported by the author(s).

References

- Abdul Jabbar, A. I., and P. Felicia. 2015. "Gameplay Engagement and Learning in Game-Based Learning: A Systematic Review." *Review of Educational Research* 85 (4): 740–779.
- Abou-Khalil, V., S. Helou, B. Flanagan, N. Pinkwart, and H. Ogata. 2019. "Language Learning Tool for Refugees: Identifying the Language Learning Needs of Syrian Refugees Through Participatory Design." *Languages* 4 (3): 71.
- Ak, O. 2012. "A Game Scale to Evaluate Educational Computer Games." *Procedia-Social and Behavioral Sciences* 46: 2477–2481.
- Akçelik, D., and B. Eyüp. 2021. "The Effect of Intelligence Games on the Vocabulary Knowledge of Refugee Students Learning Turkish as the Second Language." *Education Quarterly Reviews* 4, Special Issue 1: Primary and Secondary Education: 527–541.
- Ali Ramsi, F. M. 2015. "A Game-Based Learning Model." Master of Science (MS)Theses, United Arab Emirates University, UAEU.
- Al Janaideh, R., A. Gottardo, S. Tibi, J. Paradis, and X. Chen. 2020. "The Role of Word Reading and Oral Language Skills in Reading Comprehension in Syrian Refugee Children." *Applied Psycholinguistics* 41 (6): 1283–1304.
- 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. N. P. Castellar, and J. Van Looy. 2021. "Digital Game-Based Learning Effectiveness Assessment: Reflections on Study Design." *Computers & Education* 167: 104160.
- Andersen, E. 2012. "Optimizing Adaptivity in Educational Games." In *Proceedings of the International Conference on the Foundations of Digital Games, FDG 2012, May 29–June 1, Raleigh, NC, USA*, 279–281.
- Andersen, E., E. O'Rourke, Y-E. Liu, R. Snider, J. Lowdermilk, D. Truong, S. Cooper, and Z. Popovic. 2012. "The Impact of Tutorials on Games of Varying Complexity." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 59–68. Austin, Texas.
- Ariffin, M. M. 2013. "Gad-em: An Adaptive Game Design Model for Malaysian Higher Education (HE)." *International Journal of Scientific ve Engineering Research* 4 (5): 100–103.
- Ariffin, M. M., and S. Sulaiman. 2014. "Evaluating Game-Based Learning (GBL) Effectiveness in Higher Education (HE)." In *International Conference on Advanced Computer Science Applications and Technologies*, 485–489. Kuching, Malaysia. doi:10.1109/Acsat.2013.101.
- 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* 5: 40.
- Aslan, S., and O. Balci. 2015. "GAMED: Digital Educational Game Development Methodology." *Simulation* 91 (4): 307–319.
- Azizt, A., and S. Subiyanto. 2018. "Digital Game-Based Learning in Arabic Language Learning and Its Effects on Students' Academic Performance." In *International Conference on Indonesian Technical Vocational Education and Association (APTEKINDO 2018)*. Atlantis Press.
- Bachen, C. M., P. Hernández-Ramos, C. Raphael, and A. Waldron. 2016. "How Do Presence, Flow, and Character Identification Affect Players' Empathy and Interest in Learning from a Serious Computer Game?" *Computers in Human Behavior* 64: 77–87.
- Backlund, P., and M. Hendrix. 2013. "Educational Games-are They Worth the Effort? A Literature Survey of the Effectiveness of Serious Games." In *2013 5th International Conference on Games and Virtual Worlds for Serious Applications (VS-GAMES), 11–13 September 2013*, 1–8. Poole, UK: IEEE.
- Baddour, J. 2020. "The Development of English and Arabic Language and Literacy Skills of Syrian Refugee Children and Youth in Canada." PhD diss., University of Toronto, Canada.
- Baig, M. Z., and M. Kavakli. 2018. "Analyzing Novice and Expert User's Cognitive Load in Using a Multi-Modal Interface System." In *2018 26th International Conference on Systems Engineering (ICSEng), 18–20 December 2018*, 1–7. Sydney, NSW: IEEE.
- Bailey, R., K. Wise, and P. Bolls. 2009. "How Avatar Customizability Affects Children's Arousal and Subjective Presence During Junk Food-Sponsored Online Video Games." *CyberPsychology & Behavior* 12 (3): 277–283.
- 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.
- Becker, K. 2013. "The Magic Bullet: A Tool for Assessing and Evaluating Learning Potential in Games." In *Developments in Current Game-Based Learning Design and Deployment*, edited by Patrick Felicia, 273–284. Ireland: Waterford Institute of Technology.
- Behnamnia, N., A. Kamsin, M. A. B. Ismail, and A. Hayati. 2020. "The Effective Components of Creativity in Digital Game-Based Learning among Young Children: A Case Study." *Children and Youth Services Review* 116: 105227.
- 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." In *International Conference on Cross-Cultural Design 17–22 July*, 409–417. Toronto, Canada : Springer.
- 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.
- bin Zainuddin, G., S. R. bt Ramlan, N. A. bin Mohamed Masrop, M. S. bin Sahrir, and E. bin Abdullah. 2021. "Teachers' Perspectives on Digital Game-Based Language Learning for Arabic Language in Malaysian Primary Schools." In *4th International Conference on Sustainable Innovation 2020–Social, Humanity, and Education (ICoSIHESS 2020), October 13–14 2020*, 436–441. Yogyakarta, Indonesia: Atlantis Press.
- Birch, S. A., and P. Bloom. 2002. "Preschoolers are Sensitive to the Speaker's Knowledge When Learning Proper Names." *Child Development* 73 (2): 434–444.

- Birk, M. V., C. Atkins, J. T. Bowey, and R. L. Mandryk. 2016. "Fostering Intrinsic Motivation Through Avatar Identification in Digital Games." In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems, May 07-12, 2016*, 2982–2995. San Jose, CA, USA.
- Bontchev, B. P., V. Terzieva, and E. Paunova-Hubenova. 2021. "Personalization of Serious Games for Learning." *Interactive Technology and Smart Education* 18 (1): 50–68.
- Boyle, E. A., T. Hainey, T. M. Connolly, G. Gray, J. Earp, M. Ott, T. Lim, et al. 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.
- Bradley, L., L. Bartram, K. W. Al-Sabbagh, and A. Algers. 2020. "Designing Mobile Language Learning with Arabic Speaking Migrants." *Interactive Learning Environments*, 12(2): 1–13.
- Bunt, A., C. Conati, and J. McGrenere. 2007. "Supporting Interface Customization Using a Mixed-Initiative Approach." In *Proceedings of the 12th International Conference on Intelligent User Interfaces, IUI '07, January 28–31, 2007*, 92–101. Honolulu, Hawaii, USA.
- Burkolter, D., B. Weyers, A. Kluge, and W. Luther. 2014. "Customization of User Interfaces to Reduce Errors and Enhance User Acceptance." *Applied Ergonomics* 45 (2): 346–353.
- 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.
- Caldiera, V. R. B. G., and H. D. Rombach. 1994. "The Goal Question Metric Approach." In *Encyclopedia of Software Engineering*, edited by John J. Marciniak, 528–532. New York, NY: Wiley.
- Castañó-Muñoz, J., E. Colucci, and H. Smidt. 2018. "Free Digital Learning for Inclusion of Migrants and Refugees in Europe: A Qualitative Analysis of Three Types of Learning Purposes." *International Review of Research in Open and Distributed Learning* 19 (2): 1–21.
- Chen, S. Y., and P.-R. Huang. 2013. "The Comparisons of the Influences of Prior Knowledge on Two Game-Based Learning Systems." *Computers & Education* 68: 177–186.
- Chen, H.-J. H., and T.-Y. C. Yang. 2013. "The Impact of Adventure Video Games on Foreign Language Learning and the Perceptions of Learners." *Interactive Learning Environments* 21 (2): 129–141.
- Cohen, J. 1988. *Statistical Power Analysis for the Behavioral Sciences*. 2nd ed. Hillsdale, NJ: Erlbaum Press.
- Connolly, T., M. Stansfield, and T. Hainey. 2009. "Towards the Development of a Games-Based Learning Evaluation Framework." In *Games-based Learning Advancements for Multi-Sensory Human Computer Interfaces: Techniques and Effective Practices*, edited by Thomas Connolly, Mark Stansfield, and Liz Boyle, 251–273. Paisley, Scotland: IGI Global.
- Cornillie, F., G. Clarebout, and P. Desmet. 2012. "Between Learning and Playing? Exploring Learners' Perceptions of Corrective Feedback in an Immersive Game for English Pragmatics." *ReCALL: Journal of Eurocall* 24 (3): 257–278.
- Creswell, J. W., and J. D. Creswell. 2017. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 5th Edition. Los Angeles, CA: Sage.
- Csikszentmihalyi, M. 2000. *Beyond Boredom and Anxiety*. San Francisco, USA: Jossey-Bass.
- Czauderna, A., and E. Guardiola. 2019. "The Gameplay Loop Methodology as a Tool for Educational Game Design." *Electronic Journal of e-Learning* 17 (3): 207–221.
- Danielsson, K., and C. Wiberg. 2006. "Participatory Design of Learning Media: Designing Educational Computer Games with and for Teenagers." *Interactive Technology and Smart Education* 3 (4): 275–291.
- Davis, F. D., and V. Venkatesh. 2004. "Toward Preprototype User Acceptance Testing of New Information Systems: Implications for Software Project Management." *IEEE Transactions on Engineering Management* 51 (1): 31–46.
- De Freitas, S., and F. Liarokapis. 2011. "Serious Games: A New Paradigm for Education?" In *Serious Games and Edutainment Applications*, edited by Minhua Ma, Andreas Oikonomou, and Lakhmi C. Jain, 9–23. London: Springer.
- 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.
- Diah, N. M., M. Ismail, S. Ahmad, and M. K. M. Dahari. 2010. "Usability Testing for Educational Computer Game Using Observation Method." In *2010 International Conference on Information Retrieval & Knowledge Management (CAMP), 16-18 March 2010*, 157–161. Shah Alam, Selangor, Malaysia: IEEE.
- Diamond, J., M. Horn, and D. H. Uttal. 2016. *Practical Evaluation Guide: Tools for Museums and Other Informal Educational Settings*. Lanham, MD: Rowman & Littlefield.
- Dickey, M. D. 2011. "Murder on Grimm Isle: The Impact of Game Narrative Design in an Educational Game-Based Learning Environment." *British Journal of Educational Technology* 42 (3): 456–469.
- 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." In *European Conference on Games-Based Learning, Oct 2014*. Berlin, Germany.
- 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., S. Papadakis, E. Sifaki, and M. Kalogiannakis. 2022. "Mobile Learning Applications for Refugees: A Systematic Literature Review." *Education Sciences* 12 (2): 96.
- 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.
- Dubbelman, T. 2016. "Narrative Game Mechanics." In *proceedings of 9th International Conference on Interactive Digital Storytelling, ICIDS 2016, 15–18 November, 2016*, edited by Frank Nack and Andrew S. Gordon, 39–50. Los Angeles, CA: Springer.
- Eltahir, M., N. R. Alsalihi, 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* 26 (3): 3251–3278.

- Erfani, M., M. S. El-Nasr, D. Milam, B. Aghabeigi, B. A. Lameman, B. E. Riecke, H. Maygoli, and S. Mah. 2010. "The Effect of age, Gender, and Previous Gaming Experience on Game Play Performance." In *Human-Computer Interaction. HCIS 2010. IFIP Advances in Information and Communication Technology*. Vol. 332, edited by P. Forbrig, F. Paternó and A. Mark Pejtersen, 293–296. Berlin, Heidelberg: Springer.
- Ericsson, K. A.. 2002. "Attaining Excellence Through Deliberate Practice: Insights from the Study of Expert Performance." In *The Pursuit of Excellence Through Education*, edited by M. Ferrari, 21–55. London: Lawrence Erlbaum Associates Publishers.
- Faizan, N., A. Löffler, R. Heining, M. Utesch, and H. Krcmar. 2019. "Classification of Evaluation Methods for the Effective Assessment of Simulation Games: Results from a Literature Review." *International Journal of Engineering Pedagogy* 9 (1): 19–33.
- Feng, Z., V. A. González, C. Mutch, R. Amor, A. Rahouti, A. Baghouz, N. Li, and G. Cabrera-Guerrero. 2020. "Towards a Customizable Immersive Virtual Reality Serious Game for Earthquake Emergency Training." *Advanced Engineering Informatics* 46: 101134.
- Gaved, M., and A. Peasgood. 2017. "Fitting in Versus Learning: A Challenge for Migrants Learning Languages Using Smartphones." *Journal of Interactive Media in Education* 1 (1): 1–13. doi:10.5334/jime.436.
- Gioia, D. A., K. G. Corley, and A. L. Hamilton. 2013. "Seeking Qualitative Rigor in Inductive Research: Notes on the Gioia Methodology." *Organizational Research Methods* 16 (1): 15–31.
- Godwin-Jones, R. 2014. "Games in Language Learning: Opportunities and Challenges." *Language Learning & Technology* 18 (2): 9–19.
- Gordano Peile, C., and A. Ros Hajar. 2016. "Immigrants and Mobile Phone Uses: Spanish-Speaking Young Adults Recently Arrived in London." *Mobile Media & Communication* 4 (3): 405–423.
- 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.
- Harskamp, E. G., R. E. Mayer, and C. Suhre. 2007. "Does the Modality Principle for Multimedia Learning Apply to Science Classrooms?" *Learning and Instruction* 17 (5): 465–477.
- 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.
- Ibrahim, R., and A. Jaafar. 2009. "Educational Games (EG) Design Framework: Combination of Game Design, Pedagogy and Content Modeling." In *2009 International Conference on Electrical Engineering and Informatics*, 5–7 Aug. 2009, vol. 1, 293–298. Bangi, Malaysia: IEEE.
- Jalali, S., and M. Dousti. 2012. "Vocabulary and Grammar Gain Through Computer Educational Games." *GEMA Online® Journal of Language Studies* 12 (4): 1077–1088.
- Jorritsma, W., F. Cnossen, and P. M. van Ooijen. 2015. "Adaptive Support for User Interface Customization: A Study in Radiology." *International Journal of Human-Computer Studies* 77: 1–9.
- Jossan, K. S., A. Gauthier, and J. Jenkinson. 2021. "Cultural Implications in the Acceptability of Game-Based Learning." *Computers & Education* 174: 104305.
- Ju, S. Y., and Z. Adam. 2018. "Implementing Quizizz as Game Based Learning in the Arabic Classroom." *European Journal of Social Science Education and Research* 5 (1): 194–198.
- Kafai, Y. B., M. L. Franke, and D. S. Battey. 2002. "Educational Software Reviews Under Investigation." *Education, Communication & Information* 2 (2-3): 163–180.
- Kalogiannakis, M., S. Papadakis, and A.-I. Zourmpakis. 2021. "Gamification in Science Education. A Systematic Review of the Literature." *Education Sciences* 11 (1): 22.
- Kalyuga, S. 2009. "The Expertise Reversal Effect." In *Managing Cognitive Load in Adaptive Multimedia Learning*, edited by S. Kalyuga, 58–80. Australia: University of New South Wales.
- Kelley, C. R. 1969. "What is Adaptive Training?" *Human Factors* 11 (6): 547–556.
- 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 (5): 972.
- Kickmeier-Rust, M. D. 2012. "An Alien's Guide to Multi-Adaptive Educational Computer Games." In *Informing Science*, edited by Michael D. Kickmeier-Rust and Dietrich Albert. Santa rosa, California, USA.
- Kickmeier-Rust, M. D., and D. Albert. 2010. "Micro-Adaptivity: Protecting Immersion in Didactically Adaptive Digital Educational Games." *Journal of Computer Assisted Learning* 26 (2): 95–105.
- Kickmeier-Rust, M. D., B. Marte, S. Linek, T. Lalonde, and D. Albert. 2008. "The Effects of Individualized Feedback in Digital Educational Games." In *Proceedings of the 2nd European Conference on Games Based Learning (ECGBL 2008)*, 16-17 October 2008, 227–236. Barcelona, Spain: Academic Publishing Limited.
- Kiili, K. 2005. "On Educational Game Design: Building Blocks of Flow Experience". PhD Thesis, Tampere University of Technology, Tampere, Finland.
- 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.
- Kim, S., and M. Chang. 2010. "Computer Games for the Math Achievement of Diverse Students." *Journal of Educational Technology & Society* 13 (3): 224–232.
- Kleinsmith, A., and M. Gillies. 2013. "Customizing by Doing for Responsive Video Game Characters." *International Journal of Human-Computer Studies* 71 (7-8): 775–784.
- Kocaman, O., and G. K. Cumaoglu. 2014. "The Effect of Educational Software (Denis) and Games on Vocabulary Learning Strategies and Achievement." *Egitim ve Bilim* 39 (176): 305–316.
- Ku, O., C.-C. Hou, and S. Y. Chen. 2016. "Incorporating Customization and Personalization into Game-Based Learning: A Cognitive Style Perspective." *Computers in Human Behavior* 65: 359–368.
- Lim, D. H., and H. Kim. 2003. "Motivation and Learner Characteristics Affecting Online Learning and Learning

- Application.” *Journal of Educational Technology Systems* 31 (4): 423–439.
- 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.
- Lopes, R., and R. Bidarra. 2011. “Adaptivity Challenges in Games and Simulations: A Survey.” *IEEE Transactions on Computational Intelligence and AI in Games* 3 (2): 85–99.
- López-Fernández, D., A. Gordillo, P. P. Alarcón, and E. Tovar. 2021. “Comparing Traditional Teaching and Game-Based Learning Using Teacher-Author Games on Computer Science Education.” *IEEE Transactions on Education* 64 (4): 367–373.
- Lou, N. M., and K. A. Noels. 2020. “Mindsets About Language Learning and Support for Immigrants’ Integration.” *International Journal of Intercultural Relations* 79: 46–57.
- MacDorman, K. F., T. J. Whalen, C.-C. Ho, and H. Patel. 2011. “An Improved Usability Measure Based on Novice and Expert Performance.” *International Journal of Human-Computer Interaction* 27 (3): 280–302.
- MacFarland, T. W., and J. M. Yates. 2016. “Wilcoxon Matched-Pairs Signed-Ranks Test.” In *Introduction to Nonparametric Statistics for the Biological Sciences Using R*, edited by Thomas W. MacFarland, 133–175. Switzerland: Springer.
- Malone, T. W. 1981. “Toward a Theory of Intrinsically Motivating Instruction.” *Cognitive Science* 5 (4): 333–369.
- 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* 2014. Article ID 139205.
- Masrop, N. A. M., H. Ishak, G. Zainuddin, S. R. Ramlan, M. S. Sahrir, and H. Hashim. 2019. “Digital Games Based Language Learning for Arabic Literacy Remedial.” *Creative Education* 10 (12): 3213.
- McKiddy, J. R. 2020. “Efficiency in Game Based Language Learning: A Meta-Analysis of Social-Network Gaming.” In *Program Booklet of the 3rd Pan-Pacific Technology-Enhanced Language Learning & Critical Thinking Meeting (PPTELL 2020)*, Jun 29–Jul 1, 2020, 82–84. Denton, Texas, USA.
- Mohamed, H., and A. Jaafar. 2010. “Challenges in the Evaluation of Educational Computer Games.” In *2010 International Symposium on Information Technology, ITSIm, 15–17 June 2010*, vol. 1, 1–6. Kuala Lumpur, Malaysia: IEEE.
- 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.
- Moreno, S., R. Low, and J. Sweller. 1995. “Cognitive Principles in Multimedia Learning: Applications to e-Learning.” *Journal of Educational Psychology* 91: 358–368.
- Murtadho, N. 2021. “Digital Resources and Their Use in Arabic Language Classroom: A Mini-Narrative Review.” In *Proceedings of the International Seminar on Language, Education, and Culture (ISoLEC 2021)*. Vol. 612, edited by Maria Hidayati, Yazid Basthomi Francisca, Maria Ivone, Nova Ariani, and Achmad Tohe, 242–248. Atlantis Press.
- 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.
- Neuenhaus, M., and M. Aly. 2017. “Empathy up.” In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, CHI ’17*, 6–11 May 2017, 86–92. Denver Colorado, USA.
- Neville, D. O., B. E. Shelton, and B. McInnis. 2009. “Cybertext Redux: Using Digital Game-Based Learning to Teach L2 Vocabulary, Reading, and Culture.” *Computer Assisted Language Learning* 22 (5): 409–424.
- Nordhaug, L. M. K. 2019. “Eduapp4Syria [Online].” norad.no: Norwegian Agency for Development Cooperation (Norad). Accessed April 25, 2019. <http://norad.no/eduapp4syria>.
- 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.
- Orji, R., K. Oyibo, and G. F. Tondello. 2017. “A Comparison of System-Controlled and User-Controlled Personalization Approaches.” In *Adjunct Publication of the 25th Conference on User Modeling, Adaptation and Personalization*, 9–12 July, 2017, 413–418. Bratislava, Slovakia.
- Orvis, K. A., D. B. Horn, and J. Belanich. 2006. *Videogame-based Training Success: The Impact of Trainee Characteristics-Year 2*. Alexandria, VA: Army Research Institute for the Behavioral and Social Sciences.
- 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.
- Orvis, K. A., D. B. Horn, and J. Belanich. 2009. “An Examination of the Role Individual Differences Play in Videogame-Based Training.” *Military Psychology* 21 (4): 461–481.
- Osman, K., and N. A. Bakar. 2012. “Educational Computer Games for Malaysian Classrooms: Issues and Challenges.” *Asian Social Science* 8 (11): 75.
- Park, N., K. M. Lee, S.-A. A. Jin, and S. Kang. 2010. “Effects of pre-Game Stories on Feelings of Presence and Evaluation of Computer Games.” *International Journal of Human-Computer Studies* 68 (11): 822–833.
- Patton, M. Q. 2008. *Utilization-focused Evaluation*. Thousand Oaks, CA: Sage.
- Peirce, N., O. Conlan, and V. Wade. 2008. “Adaptive Educational Games: Providing non-Invasive Personalised Learning Experiences.” In *2008 IEEE International Conference on Digital Game and Intelligent toy Enhanced Learning, DIGITEL ’08*, 17–19 November, 2008, 28–35. Banff, Canada: IEEE.
- Petousi, V., and E. Sifaki. 2020. “Contextualising Harm in the Framework of Research Misconduct. Findings from Discourse Analysis of Scientific Publications.” *International Journal of Sustainable Development* 23 (3–4): 149–174.
- Petri, G., and C. G. von Wangenheim. 2017. “How Games for Computing Education are Evaluated? A Systematic Literature Review.” *Computers & Education* 107: 68–90.

- Piaget, J. 1952. *Play, Dreams and Imitation in Childhood*. London: Routledge.
- 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.
- Poole, F. J., and J. Clarke-Midura. 2020. "A Systematic Review of Digital Games in Second Language Learning Studies." *International Journal of Game-Based Learning (IJGBL)* 10 (3): 1–15.
- Prensky, M. 2001. "Fun, Play and Games: What Makes Games Engaging." *Digital Game-Based Learning* 5 (1): 5–31.
- Prensky, M. 2003. "Digital Game-Based Learning." *Computers in Entertainment (CIE)* 1 (1): 21.
- 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.
- Putri, A. H., F. E. Permatasari, A. L. Hijriyah, and L. Mauludiyah. 2021. "Arabic Quizzes Game to Improve Arabic Vocabulary." *Tanwir Arabiyyah: Arabic as Foreign Language Journal* 1 (1): 45–54.
- Ros, A. 2010. "Interconnected Immigrants in the Information Society." In *Diasporas in the New Media Age. Identity, Politics and Community*, edited by A. P. Alonso and J. Oiarzabal, 19–39. Reno, Nevada: University of Nevada Press.
- Rummer, R., J. Schweppe, A. Fürstenberg, K. Scheiter, and A. Zindler. 2011. "The Perceptual Basis of the Modality Effect in Multimedia Learning." *Journal of Experimental Psychology: Applied* 17 (2): 159.
- Sahrir, M. S., and N. A. Alias. 2012. "A Study on Malaysian Language Learners' Perception Towards Learning Arabic via Online Games." *GEMA Online® Journal of Language Studies* 11 (3): 129–145.
- Sahrir, M. S., and G. Yusri. 2012. "Online Vocabulary Games for Teaching and Learning Arabic." *GEMA Online® Journal of Language Studies* 12 (3): 961–977.
- Salah, J., S. Abdennadher, C. Sabty, and Y. Abdelrahman. 2016. "Super Alpha: Arabic Alphabet Learning Serious Game for Children with Learning Disabilities." *Lecture Notes in Computer Science* 9894: 104–115. doi:10.1007/978-3-319-45841-0_9.
- 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.
- Schleyer, T. K., and L. A. Johnson. 2003. "Evaluation of Educational Software." *Journal of Dental Education* 67 (11): 1221–1228.
- Schnotz, W., and T. Rasch. 2005. "Enabling, Facilitating, and Inhibiting Effects of Animations in Multimedia Learning: Why Reduction of Cognitive Load Can Have Negative Results on Learning." *Educational Technology Research and Development* 53 (3): 47.
- Sharma, G. 2017. "Pros and Cons of Different Sampling Techniques." *International Journal of Applied Research* 3 (7): 749–752.
- Sierra Rativa, A., M. Postma, and M. Van Zaanen. 2020. "The Influence of Game Character Appearance on Empathy and Immersion: Virtual Non-Robotic Versus Robotic Animals." *Simulation & Gaming* 51 (5): 685–711.
- 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.
- Siraj-Blatchford, I. 2009. "Conceptualising Progression in the Pedagogy of Play and Sustained Shared Thinking in Early Childhood Education: A Vygotskian Perspective." *Education and Child Psychology* 26 (2): 77–89.
- Steiner, C., Hollins, P., Kluijfhout, E., Dascalu, M., Albert, D., Westera, W., and Nussbaumer, A. 2015. "Evaluation of Serious Games: A Holistic Approach." In *ICERI 2015 Proceedings: 8th International Conference of Education, Research and Innovation 18-20 November, 2015*, edited by L. Gómez Chova, A. López Martínez, and I. Candel Torres, 4334–4342. Seville, Spain: IATED Academy.
- Steiner, C. M., M. D. Kickmeier-Rust, E. Mattheiss, S. Göbel, and D. Albert. 2012. "Balancing on a High Wire: Adaptivity Key Factor of Future Learning Games." In *An Alien's Guide to Multi-Adaptive Educational Computer Games*, edited by Michael D. Kickmeier-Rust and Dietrich Albert, 43–88. California, USA: Informing science press.
- Streicher, A., and J. D. Smeddinck. 2016. "Personalized and Adaptive Serious Games." In *Entertainment Computing and Serious Games, Lecture Notes in Computer Science*, Vol. 9970, edited by Dörner, R., S. Göbel, M. Kickmeier-Rust, M. Masuch, and K. Zweig, 332–377. Cham: Springer.
- Sundqvist, P., and L. K. Sylvén. 2014. "Language-related Computer Use: Focus on Young L2 English Learners in Sweden." *ReCALL* 26 (1): 3–20.
- Sundqvist, P., and P. Wikström. 2015. "Out-of-School Digital Gameplay and In-School L2 English Vocabulary Outcomes." *System* 51: 65–76.
- Sweetser, P., and P. Wyeth. 2005. "GameFlow: A Model for Evaluating Player Enjoyment in Games." *Computers in Entertainment (CIE)* 3 (3): 3.
- Sylvén, L. K., and P. Sundqvist. 2012. "Gaming as Extramural English L2 Learning and L2 Proficiency among Young Learners." *ReCALL* 24 (3): 302–321.
- Tahir, R., and F. Arif. 2015. "A Measurement Model Based on Usability Metrics for Mobile Learning User Interface for Children." *The International Journal of E-Learning and Educational Technologies in the Digital Media* 1 (1): 16–31.
- Tahir, R., and A. I. Wang. 2017. "State of the Art in Game Based Learning: Dimensions for Evaluating Educational Games." In *Proceedings of European Conference on Games Based Learning, ECGBL 2017, 5-6 October 2017, Graz, Austria*, edited by M. Pivec et al., 641–650. Reading, UK: Academic Conferences International Limited.
- Tahir, R., and A. I. Wang. 2019. "How to Evaluate Educational Games With Refugee Children: Methodological Aspects and Lessons Learned From EduApp4syria." In *Proceedings of European Conference on Games Based Learning, ECGBL 2019, Academic Conferences International Limited, Reading*, 722–730.
- Tahir, R., and A. I. Wang. 2020. "Codifying Game-Based Learning: Development and Application of LEAGUE Framework for Learning Games." *Electronic Journal of e-Learning* 18 (1): 69–87.
- Tang, S., M. Hanneghan, and A. El Rhalibi. 2009. "Introduction to Games-Based Learning." In *Games-based Learning Advancements for Multi-Sensory Human Computer Interfaces: Techniques and Effective Practices*, edited by T. Connolly, M. Stansfield, and L. Boyle, 1–17. Hershey, PA: IGI Global.

- Terrell, S. R., and L. Dringus. 2000. "An Investigation of the Effect of Learning Style on Student Success in an Online Learning Environment." *Journal of Educational Technology Systems* 28 (3): 231–238.
- Tzetzis, G., E. Votsis, and T. Kourtessis. 2008. "The Effect of Different Corrective Feedback Methods on the Outcome and Self Confidence of Young Athletes." *Journal of Sports Science & Medicine* 7 (3): 371.
- 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.
- Van Merriënboer, J. J., and P. A. Kirschner. 2017. *Ten Steps to Complex Learning: A Systematic Approach to Four-Component Instructional Design*. New York: Routledge.
- Van Staalduinen, J.-P., and S. De Freitas. 2011. "A Game-Based Learning Framework: Linking Game Design and Learning." In *Learning to Play: Exploring the Future of Education with Video Games*, vol. 53, edited by MS Khine, 29–54. New York, USA: Peter Lang.
- Wang, X., A. H. Butt, Q. Zhang, M. N. Shafique, H. Ahmad, and Z. Nawaz. 2020. "Gaming Avatar Can Influence Sustainable Healthy Lifestyle: Be Like an Avatar." *Sustainability* 12 (5): 1998.
- Ward, N., E. Paul, P. Watson, G. Cooke, C. Hillman, N. Cohen, A. Kramer, and A. Barbey. 2017. "Enhanced Learning Through Multimodal Training: Evidence from a Comprehensive Cognitive, Physical Fitness, and Neuroscience Intervention." *Scientific Reports* 7 (1): 1–8.
- Wofford, M. C., and S. Tibi. 2018. "A Human Right to Literacy Education: Implications for Serving Syrian Refugee Children." *International Journal of Speech-Language Pathology* 20 (1): 182–190.
- Wojciechowski, A., and L. B. Palmer. 2005. "Individual Student Characteristics: Can Any be Predictors of Success in Online Classes." *Online Journal of Distance Learning Administration* 8 (2): 13.
- Wouters, P., E. D. Van der Spek, and H. Van Oostendorp. 2009. "Current Practices in Serious Game Research: A Review from a Learning Outcomes Perspective." In *Games-based Learning Advancements for Multi-Sensory Human Computer Interfaces: Techniques and Effective Practices*, Information Resources Management Association, edited by T. Connolly, M. Stansfield, and L. Boyle, 232–250. Hershey, PA, USA: IGI Global.
- Yazgan, P., D. E. Utku, and I. Sirkeci. 2015. "Syrian Crisis and Migration." *Migration Letters*, 12 (3): 181–192. doi:10.33182/ml.v12i3.273.
- Zaibon, S. B., and N. Shiratuddin. 2010. "Mobile Game-Based Learning (mGBL) Engineering Model as a Systematic Development Approach." In *Proceedings of Global Learn Asia Pacific 2010--Global Conference on Learning and Technology, May 17, 2010*, edited by Z. Abas, I. Jung and J. Luca, 1862–1871. Waynesville, USA: Association for the Advancement of Computing in Education (AACE).
- Zourmpakis, A.-I., S. Papadakis, and M. Kalogiannakis. 2022. "Education of Preschool and Elementary Teachers on the Use of Adaptive Gamification in Science Education." *International Journal of Technology Enhanced Learning* 14 (1): 1–16.