

Heuristics and Think-aloud Method for Evaluating the Usability of Game-based Language Learning

Kashif Ishaq^{1*}, Fadhilah Rosdi², Nor Azan Mat Zin³

Faculty of Information Science and Technology
Universiti Kebangsaan Malaysia, Bangi, Malaysia

Adnan Abid⁴

School of Systems and Technology (SST)
University of Management and Technology, Lahore, Pakistan

Abstract—Digital learning environments are increasingly becoming popular in recent years. The rising usage of cell phones has invited researchers to design and develop learning applications and games for mobile phones. Specifically, game-based language learning is being promoted by researchers in many parts of the world. “Language Learning Serious Game (LLSG)” is based on a theoretical model constructed by the researcher that supports children learning English as a second language in a cultural context. The usability of such games is evaluated based on well-defined heuristics and other standard methods. This research aims to appraise the usability of LLSG through heuristics and think-aloud approaches while involving all essential stakeholders, including language experts, students, teachers, and game developers. The researcher proposed the heuristics in a cultural context, whereas the think-aloud review is compiled from the rigorous discussion session involving these stakeholders to evaluate the LLSG. The findings obtained from the heuristics evaluation reveal that the usability of LLSG is acceptable. On the other hand, various interesting suggestions and reviews were gathered from the discussion between experts and students. This evaluation will further improve the future versions of the game.

Keywords—Engagement; game-based; heuristic evaluation; language learning; motivation; think aloud; usability

I. INTRODUCTION

Digital learning provides the ability to incorporate skills and knowledge into a wide variety of academic scenarios. Several options are available for digital learning, which involves game-based learning that employs interactive video games with content. Learning through gamification has recently been established in collaboration with technical developments to improve gameplay, modern equipment, and new methods of motivating people to engage and communicate together. Video and mobile phone games are perceived to be the perfect venue for education nowadays. Games help learners achieve a more robust understanding by getting more fun, becoming more appealing, and placing learners in roles that enable them to reflect on their performance [1][2].

Game-based learning is a creative instructing tool that may benefit children specifically, even those left behind in their studies. These games are an entertainment tool with educational objectives, in which the players gain awareness and develop their abilities when playing [3]. Game-based learning in business, science, heritage, education, math, and languages has thrived [4][5][6][7]. Besides, a game-based

approach is used for assessments, learning, collaborations, individual learning, and creativity. A survey was conducted in which educational area was prominent to determine the significant gamified zones. It was also discovered that game-based learning encouraged the users [8][9].

A lot of games for language learning according to the cultural context have been developed and implemented successfully, i.e., Arab [10][11] Iran [12], Sudan [13], Singapore [14], China [15][16], Taiwan [17], Denmark [18], Italy [19][20], Greece [21], Korea [22], Romania [23], Spain [24][25]. Similarly, a mobile application, ‘Literacy and Numeracy Drive,’ was developed and implemented in the public sector schools of Punjab, Pakistan [26], but it was less effective in obtaining the intended learning outcomes [9] [27]. Thus, the researcher constructed a theoretical model for language learning in a cultural context [28]. Then to validate and evaluate the usability of the model, a game-based language learning application, ‘LLSG,’ was designed and developed according to the cultural context of public sector schools. The content and requirements for the game development were gathered from the extensive literature review and the stakeholders [9][27][93][95]. There are different methods to evaluate a game developed for language learning [29], but the usability testing method is commonly used for its evaluation [30] [31].

The usability assessment introduced by [30] [31] makes it easier for programmers to consider and enhance the usability and efficiency of the user interface for software applications. Improving the system's usability will strengthen the system's utilization. Additionally, a usability classical testing method is think-aloud, where one user works separately by expressing their decisions, expectations, thoughts, and feelings during application interaction. The evaluators can understand the reasons behind the action performed by the user with the system [32]. For usability assessment, researchers have used a heuristic assessment method devised by [30]. This assessment method is suitable for software evaluation since it is low-cost and realistic. Usability problems can be perceived better from the viewpoint of users and professionals as both assessment approaches are used. A variety of research has used both assessment approaches to measure the usability of game-based and e-learning systems. The findings suggested positively that the two methods were used concurrently [33-37]. This research aims to analyze the use of a game-based language learning application to understand its interface's positive and negative aspects to increase the quality and performance of its usage. The researcher followed both the usability evaluation

*Corresponding Author

methods: Experts' heuristics and think-aloud from teachers and students to identify the game's usability issues. The stakeholders appraised the game-based language learning application in the cultural context, suggesting minor changes to improve its quality for more effective usage. Incorporating Nielsen's ten indices with proposed heuristics [94] is another significant contribution of the study [38].

The rest of the article has been structured in the following manner: the related research work has been discussed in Section 2, whereas the evaluation process, including data collection for a language learning game, has been presented in Section 3. Results for the evaluation process have been discussed in Section 4, while the article concluded with its future work in Section 5.

II. LITERATURE REVIEW

The game-based learning and heuristic evaluation has been addressed in this section, as it is the main dominant element of this paper.

A. Game-based Learning

The study of [39] demonstrated that games could stimulate learners' interest and significantly improve one's learning. When gamification is applied to a game, it tends to make the game more affectionate, interactive, and progress [40][41][42]. Through modern technologies, immersive learning is enhanced with vivid interactive learning displays, which help to improve academic achievement [43]. If various gaming elements were integrated into digital instruction, students would pay attention to the multimedia material and experiences for a more extended period [44]. The study by [45] demonstrated that when multimedia games were integrated with gamification in a learning system, students understood the content, enhancing their willingness to take on the learning material. Other studies have shown that game-based learning with gamification could promote learners' motivation, participation, and success in the process [46-50]. There is compelling proof that a game-based learning experience may be helpful to learning acquisition.

The author in [51] developed a gamified e-learning system for blackboard, and an experiment was conducted to measure the learning outcome and performance of the experimental and control group, resulting in outperformed by the experimental group. Games are becoming increasingly popular in this modern age and are called "Education-al Games" or "Game-based learning." Game-based learning is a creative method of educating, improving learning, and excitement to cater to the requirements of numerous children, particularly those who are left behind. According to [52-56], the game-based application has been designated an instructional entertainment tool to encourage players to learn the skills during the play.

B. Game-based Language Learning

Electronic devices, such as smartphones, laptops, tablets, iPods, can serve as an effective and productive teaching resource for learners and lecturers in academia, especially when it comes to language learning through digital games [57]. Various studies have shown that gaming-based learning is valuable and revolutionary in the context of learning and

teaching. Thus, participants can help them increase their efficiency for language learning, improve their collaboration and maintain their effective outcomes. The findings showed that game-based learning had generated a highly efficient interactive learning environment, promoting student language skills [58]. Even though digital language learning games are generally represented by excitement, creativity, interest, surprise, domination, and immersive elements, they can boost students' attention, achievement, motivation, learning, and commitment [2] [58].

The author in [59] explored the effectiveness of the digital language learning game and found it more effective than the conventional teaching method in learning English. Similarly, [60] found that MMORPGs improved the dialogue ability in the English language. Teamwork, mutual understanding, and support were the elements in the dialogue. The study of [61] explored digital game-based language learning among learners and found its positive impact on language learning. The author in [62] used digital games in the Chinese context for the English language in the university. It was found that game-based learning reduced anxiety, reservedness and improved academic engagement. An experimental study was conducted to see the effectiveness of games used in English language class and got significant improvement output for the experimental group [63].

C. Theory of Usability

The author in [64] defined usability as a user who operates a particular object efficiently, quickly, and enjoys its operation. It has five features: satisfaction, memorability, error rate, performance, and learnability [30]. Moreover, usability was classified into satisfaction, effectiveness, and performance by International Organization for Standardization (ISO), but there was a lack of emphasis on user interface design, particularly for educational applications [65-66]. Furthermore, usability is divided into effectiveness, fastness, and safety [67]. Safety in this research was similar to the error rate recommended in the study of Nielsen. In contrast, fastness in the same analysis was identical to the effectiveness recommended by ISO and Nielsen. In addition to operating features, Preece and ISO concentrated on effectiveness, but the characteristics were not adequate. The effectiveness mentioned above in the Technology Acceptance Model (TAM) by [68] shows the functionality of a system and the possible impact accomplishment.

1) *Think-aloud theory*: The think-aloud technique involves the researchers speaking out their ideas to grasp further how the system functions. "Doing, thinking, and speaking" is another term to relate to the strategy as this method gives assessors the ability to offer their input by using the application. This data can yield essential knowledge to enhance the system's effectiveness because this technique has been one of the popular usability assessment approaches often used [69]. The benefits are that evaluators may efficiently conduct a thinking-aloud method, document intrinsically uncontrollable cognitive tasks, clearly grasp the direct reasons for utilizing challenges, and interact specifically with system

operations [70]. A diversity of experiments that used the think-aloud method in usability was successful [71-75].

2) *Heuristic evaluation:* A heuristic analysis [76] was performed on developing a user experience for an application and assessed usability based on previously learned skills. This assessment approach requires a limited sample of specialists to evaluate an interface to decide if the interface satisfies a series of standard operating indices and assess the product's technological suitability. The apparent benefit of this approach is that it does not require objective decisions, while assumptions and discussions are produced based on verified experience indices. With time constraints, making experts conduct the analysis would have doubled the impact, but half the time. This assessment is cost-effective and realistic, which eliminates the cognitive burdens of the evaluator, as well as heuristic assessment, which has proved to be a fruitful method for designing guidance architecture. During an assessment test, evaluators criticize and give recommendations for making the interface easy to use.

Similarly, this approach concentrates on usability problems and provides suggestions for the change of a system. Moreover, for the early stage of the usability life cycle, heuristic assessment by experts was suitable [76] since their more significant organizational expertise and advanced skills are required to conduct such evaluations. In the assessment process, experts recognize the experience of new participants to provide them with encounters that can arise between general users. For heuristic assessment, [94] proposed heuristics for LLSG given in Appendix A.

D. Current Studies for Heuristic Evaluation

In addition to the latest research findings on game-based learning, there are also usability tests for educational games. For usability review, three to five experts are more than appropriate for recognizing most usability issues. The general usability evaluation approach is ineffective because game design is separate from the application or system design. Some study has shown that computer video games require their heuristic structures. Therefore, several heuristics concerning video games were developed, who collected a series of heuristics for games from a case study and tested using Nielsen's heuristic and new guidance in the game industry [77][78][79]. Federoff's heuristics suffer from a lack of validation, clarity, and consistency. Additionally, these heuristics have minimal use throughout the design process [80][81].

A game-based application CAMEG was developed to teach students effectively and measure usability by taking a management information course. The researcher collected the data from a usability survey by taking student reviews and proposed an application suitable for learning [9]. A heuristic analysis was conducted with five human-computer interaction experts for the MOSAD application developed for the system, analysis, and development in a science subject. The findings revealed that MOSAD is a reliable and helpful application for revision purposes in higher education [5]. Similarly, a usability test was conducted for the primo discovery tool to detect user behavior patterns for library research. For this

purpose, gestures, verbal, and display behavior were analyzed through diagnostics usability evaluation to identify problems faced by users [9].

Video games are analyzed by the Playability Heuristic (PLAY), which involves three constraints [82], and this heuristic is required for three kinds of interactive games: first-person shooter, basic strategy, and action fantasy [78]. The heuristics are developed on a low generality case study by [82], but these heuristics are sometimes contradictory and vague [80]. The heuristics in PLAY are not appropriate for all game styles since each game style has its characteristics, structure, and usages [81]. The study of [83] analyzed the usability issues in interactive instructional games, giving particular attention to three aspects: the design (e.g., button, navigation), the process (e.g., ease of access, power, and learnability), and the gameplay (e.g., responsiveness). Their research utilized observation methods and interviews to examine usability issues from both users and non-users. The analysis of [84] measured the effectiveness of a virtual reality learning environment utilizing the widely recognized usability scale, and the findings revealed users' ratings of their perceptions and interests.

An investigation was performed for the effectiveness of an interactive video game on teaching digital engineering concepts in [85]. Students found the interactive tool to be a valuable educational platform and of strong usability. In the study of [71], the author analyzed the usability of a virtual environment health game about their feelings, decisions taken in the game, and responses to questions about player experiences through user interviews utilizing think-aloud analysis. In [86], the author conducted usability research using qualitative assessment and quantitative methodology to acquire users' impressions of college nursing subjects. Still, none of the aforementioned analyses consider experts reviews.

A review by [36] tested the usability of a writing pal application by utilizing various analysis approaches, including focus groups, vivo testing, module tests, and internal testing. The research contained student and teacher evaluations. Gamification instruction is a practical aspect in writing pal, which is meant to enhance students' proficiency in writing. The study [37] analyzed three interactive features: user interface acquaintance, navigation initiative, and VR environmental disturbances in instructional virtual reality technology games for geographical education. The experiment was performed on subjects like beginner players, advanced players, and professional players. The author of [33] conducted a usability analysis to test the usability of medical instructional games and simulations. A computer programmer built a specialist machine-dependent heuristic. A range of evaluation tools was used, such as examination, interviews, polls, and think-aloud. The study of [34] measured the usability of a web game-based learning application to teach users facial movements by heuristic testing utilizing the Nielson scale and user-think-aloud system to see user opinions, emotions, and viewpoints throughout the play. As outlined in this portion, various researchers have utilized different approaches and methods to test the usability of game-based educational resources. This study used both

heuristic evaluation and feedback through the think-aloud method for usability testing of LLSG.

III. METHODOLOGY

In this research, a heuristic evaluation and think-aloud method were used to decide which issues in the Graphical User Interface (GUI) of LLSG were inappropriate for language learners. The outcome allowed for developing a better-designed product, and the particular experts observed elements to recognize usability concerns.

Instructors, students, and a game developer were involved in the evaluation of the game. The instructors have good knowledge of assessing game content, and the designer has a strong understanding of the ability to determine the correct game elements. They evaluated a system to decide if it correctly followed known usability standards named "heuristics." The Heuristic evaluation process consisted of three stages (Fig. 1 and Fig. 2). The first step was an analysis phase in which evaluators separately tested the game's user interface by playing it on tablets. The second was a planning phase in which evaluators independently compile their list of identified issues for aggregation. In contrast, in the third phase, evaluators cooperate to produce a standard summary of usability problems. A prioritized list of usability issues was aggregated, compiled, and after review by the researcher, forwarded to the game developer for modification.

A. Sample of Study

According to [30] [87], approximately five to eight evaluators were required to conduct a heuristic evaluation. In this research, five professionals in the evaluation phase were involved with expertise and knowledge of English language teaching and game development. Moreover, twenty public sector school students were also part of this evaluation. The participants in the study were selected based on the convenience sampling technique. Table I describes the profile of experts.

Evaluator 1 was a female of 38-year-old with a Master in Education degree. She was familiar with mobile technology usage, videos, multimedia, and design and had experience in teaching and administration of 12 years. Evaluator 2 was a 43-year-old female with the degree of Bachelor in English language and known mobile technology and the experience of using it for 15 years at the school level. Evaluator 3 was a 33-year-old female with the degree of Bachelor in English language and has an understanding of computer technology with vast experience of teaching at various grades for 16 years. Evaluator 4 was a 31-year-old female with a Master in Education degree and attained a computer diploma from a professional institute. She had experience with computer and mobile technology teaching at grade ninth and tenth for 08 years. Lastly, Evaluator 5 was a male 29-year-old Game developer with a Bachelor in Computer Science degree and an expert in game and mobile application development. He was the senior developer in a Software house and played the role of team leader for the last five years. Twenty students (fifteen female and five male) from grade three of Government Girls High School were selected randomly for usability evaluation.

The mean age of the student evaluators was 11.6. Fig. 3a-3c shows the evaluation process of teachers and students.

B. Research Instrument

Usability is an important and emerging area in smartphone applications that cannot be avoided by proper software design. The researcher developed a language learning serious game in this study after constructing a theoretical model to measure its usability [5] [9]. For the usability perspective of a game-based application, a heuristic evaluation [30] was conducted from the stakeholders. The following instrument was used to conduct this research:

1) *Language learning serious game (LLSG):* LLSG is a mobile-based and standalone game for English language learning consisted of Eight modules; "Sound," "Singular/Plural," "Uses (is/am/are)," "Action Words," "Parts of Speech," "Sentences," "W Family," and "Comprehension."



Fig. 1. Heuristic Evaluation Process-1.

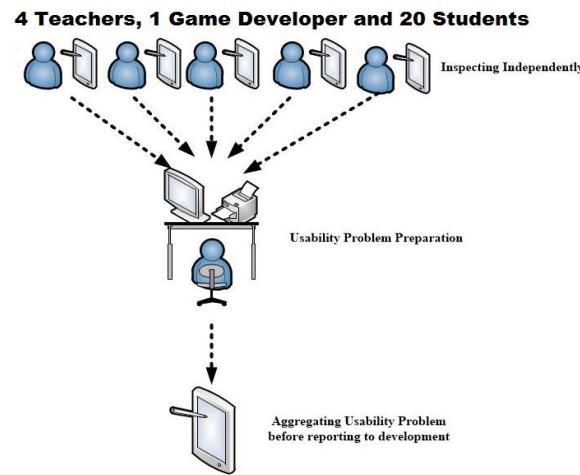


Fig. 2. Heuristic Evaluation Process-2.

TABLE I. PROFILE OF EVALUATORS

Sr. No.	Professional Role	User Experience (in Years)		
		Administration	Teaching at School	Games
1	Senior Headmistress	8	4	-
2	Teacher	5	15	-
3	Teacher	-	16	-
4	Teacher	-	08	-
5	Game Developer	-	-	8
6	Students	-	-	-



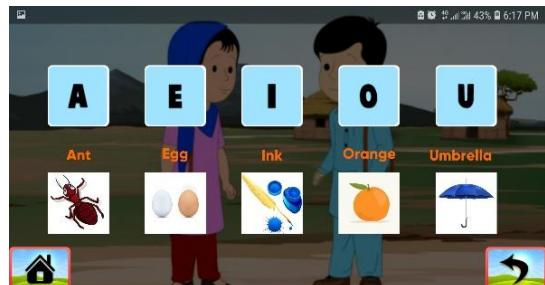
(a)



(b)



(c)



(b)



(c)



(d)



(e)



(f)

Fig. 3. (a). High-Fidelity Teacher Evaluation, (b). High-Fidelity Student Evaluation-1, (c). High-Fidelity Student Evaluation-2.



(a)



Fig. 4. (a) Main Menu, (b) Vowels, (c) Singular/Plural, (d) Singular/Plural Assessment, (e) Parts of Speech, (f) Parts of Speech Assessment, (g) Instruction after Mistakes, (h) Reward.

Fig. 4a shows the overview of the LLSG application containing all components offered in the game. For the learning, a user can click on any element, and the teacher will guide them to start with the first module, ‘sounds,’ in which a student will learn vowels (Fig. 4b), diagraphs, two/three letter sounds, and so on. Students can learn pronunciation by pressing the word/picture in the learning phase. After its completion, students can pick the assessment based on the complexity (Easy, Medium, and Advanced) to evaluate its progress. Fig. 4c and 4d are related to learning singular/plural and its assessment. Fig. 4e and 4f show the screenshot of parts of speech in which users need to click on the desired sub-module to learn noun, adjective, verb, pronoun, preposition, and assessments. As a user clicks on the assessment, they will proceed to their relevant exercises based on their complexity. Fig. 4g shows the instructions, as users make three mistakes while solving an exercise, will directly move back to its learning page to revise it. Finally, Fig. 4h shows the screenshot related to the assessment in which students can get correct or wrong attempts along with the remarks (Perfect/Good/Fair).

2) *Questionnaire:* For usability evaluation of LLGS, a questionnaire was administered based on a list of heuristics principles for interface design proposed by [94]. The questionnaire was categorized into three sections: demographic information in the first, heuristic for LLSG [94] for interface design at second, and expert comments in the third section were asked. When the evaluations were complete, the researchers compiled, interpreted, and evaluated the assessment findings. The heuristics in Appendix A for LLSG is illustrated.

3) *Technical tools (hardware & software):* A laptop and a mouse as hardware equipment, unity as the primary development kit to develop the android package (.apk) of the game application, was used. A heuristic evaluation was performed by offline tasks for the application launched on the tablet provided by the school. The assessment phase was composed of many steps:

4) *Permission:* A departmental permission was required for the experiment and to collect data from a public sector school. The researcher wrote a letter to Chief Education Officer to get permission for the purpose under the supervision of the supervisor. After receiving approval from the stakeholders, the researcher set up a meeting with the professional evaluators and described the intention of the evaluation whereas, the researcher arranged a meeting with the game developer through a call to consult and clarify the evaluation objectives.

5) *Application demonstration:* The researcher explained how the applications work before presenting them to the experts then the questionnaire survey was handed over to the evaluators.

6) *Feedback:* The expert analyzed the LLSG application depending on the questions asked in the questionnaire. The next step was to collect a questionnaire from experts then compile data. The expert advice to strengthen the application and also commented on the issues. When the assessment was concluded, the review of the data proceeded immediately.

C. Evaluation Process

The current research used a heuristic evaluation methodology and think-aloud method approach, which allowed for low cost and simplicity. It made it more effective than other usability studies and effectively guided the experts to evaluate the game. The questionnaire was completed by five experts and twenty school students who practiced LLSG during the evaluation process. The evaluation was implemented in a classroom for students and the principal’s office for teachers. Students completed their questionnaires with the help of a teacher.

The evaluation process took two hours for the teacher and three hours for students. Before the evaluation process, each game function was explained to them, and the researcher provided prompt responses to fix their issues.

Each evaluator independently assessed the language learning game using [76] and [88] heuristic assessment processes. Since each expert finished their assessment, they were allowed to speak openly amongst themselves on their experiences. Similarly, every student evaluator used LLSG and accomplished the questionnaire separately. Following usability testing, the evaluators reviewed the application to verify its reliability.

TABLE II. STEPS OF THINK-ALOUD APPROACH

Item #	Steps	Explanation
1	Explanation	The evaluators were briefed on the think-aloud research process and the most relevant usability concepts.
2	Practice	The evaluators experienced "speaking while thinking" to become familiar with the behavior model.
3	Execution	Shared the thoughts and suggestions with the other evaluators while the language learning game was running. The procedures of the activities were captured on film, and their viewpoints were documented; and
4	Protocol analysis	The author transcribed the voice samples.

The evaluators checked the application at least twice, the first time for thorough comprehension and the second time for ease of usage. The evaluators also used think-aloud methods to collect data when evaluating and using the game. Following the phases of the think-aloud method are defined in Table II.

This triangulation method was employed to enhance the validity and credibility of findings [89]. Data was gathered in different ways, and results were analyzed independently, but they needed to be compared. The researcher and co-researcher coded and evaluated the document using heuristics proposed by [94]. The evaluator checked the coding in-depth and updated it, focusing on the triangulation method for internal and external validity. It was further reviewed by the evaluator and the co-researcher for the efficiency of data transcription. Eventually, the study findings were combined with the participant's questionnaire responses and observations.

D. Validity and Reliability

This research employed the questionnaire after verifying the reliability and validity [3] [38] [90]. As determined by Cronbach's alpha, the reliability factor of each construct was higher than 0.8 based on the findings of 20 students. It showed perfect truthfulness for every structure and a sufficient internal consistency between all elements inside the construct. Furthermore, the following four factors were considered to enhance the confidence of the think-aloud method: validity, reliability, transferability, and credibility [91].

Triangulation of analysts was utilized in this research to ensure its reliability [92]. Three researchers individually examined data and evaluated their results. Additionally, a qualitative research technique specialist was present throughout the data processing process. Transparency was guaranteed by providing transcripts of a participant's think-aloud procedure and demonstrating how the data were coded and classified. Sampling techniques were also used in this research to increase transferability. Each think-aloud process was administered under identical circumstances to guarantee the reliability, and the researcher transcribed the data consistently. In addition, details have been provided on data collection and analysis methods. If required, the findings may be verified, as all the tape recordings, translations, and coding are preserved. As a result, truthfulness was achieved as well.

IV. RESULTS AND DISCUSSION

The results and discussion section were categorized into 1) Heuristic Evaluation results; and 2) teacher evaluators' comments and feedback.

TABLE III. HIGH-FIDELITY PROTOTYPE HEURISTIC EVALUATION RESULTS

Component	Item	Teacher Experts			Students		
		Yes	No	Not Sure	Yes	No	Not Sure
I	LHI1	100			100		
	LHI2	100			90		10
	LHI3	100			90		10
	LHI4	100			100		
	LHI5	100			100		
	LHI6	100			100		
	LHI7	100			100		
	LHI8	100			100		
	LHI9	100			100		
	LHI10	100			100		
	LHI11	100			100		
	LHI12	100			100		
GP	LHGP1	100			90		10
	LHGP2	100			100		
	LHGP3	100			100		
	LHGP4	100			100		
	LHGP5	100			100		
GM	LHGM1	100			100		
	LHGM2	100			100		
C	LHC1	100			90		10
	LHC2	100			100		
	LHC3	100			100		
	LHC4	100			100		
	LHC5	80	20		100		
	LHC6	100			100		
	LHC7	100			100		
	LHC8	100			100		
F	LHF1	100			100		
	LHF2	100			90		10
	LHF3	100			100		
LL	LHLL1	100			100		
	LHLL2	100			100		
	LHLL3	100			100		
	LHLL4	100			100		
	LHLL5	100			100		
	LHLL6	100			100		
CC	LHCC1	100			90		10
	LHCC2	100			100		
	LHCC3	100			100		
	LHCC4	100			100		
	LHCC5	100			80		
	LHCC6	100			100		
	LHCC7	100			100		
	LHCC8	100			100		
	LHCC9	100			100		
	LHCC10	100			100		
	LHCC11	100			100		

A. Heuristic Evaluation Result

The questionnaire was distributed between teachers and students to evaluate the high-fidelity prototype of LLSG, which was consisted of seven components and 47 elements. A descriptive analysis method to analyze the results of the heuristic evaluation questionnaire was used through Microsoft Excel. The authors [64] [88] specified that a heuristic evaluation result was a table or a list of usability issues presented in Table III. The individual agreement for each element was evaluated in terms of 'Yes,' 'No,' or 'Not sure,' and the frequency rates were calculated. The term 'Not sure' referred to any possibility that the expert is unsure about the item's answer.

The demographic information showed the teachers' and students' gender and age group, the study's respondents. Four (80%) female teachers and fifteen (75%) female students, whereas only one (20%) male (game developer) and five (25%) male students were the evaluators for the game. From the age perspective, one evaluator belonged to the age group of 25-30, two belonged to 30-35, and one was 35-40 and above 40 each. One student evaluator belonged to the age group of 7-8; the majority belonged to 9-10, and nine belonged to the age group of above 10.

(Language Heuristics-LH)

This section presented the results obtained from the heuristic evaluation and discussed with the feedback of the

experts. Table III shows the percentage score results from the descriptive data analysis. The majority of the heuristic evaluation elements scored 100% on average, whereas one element scored 80% from teachers' experts, and five elements scored 90% with 'not sure.' In the Interface (I) component, as shown in Table IV, all the elements by all the teachers' experts agreed 100%, and 90% of student evaluators were agreed with all the items except (LHI2, LHI3) as they were not sure about them. Therefore, the high-fidelity prototype interface design has no usability issues.

In the term gameplay (GP), all the teacher experts agreed with all the elements, whereas 90% of student evaluators were agreed with all the items except (LHPG1) as two evaluators responded with 'Not sure' comment. Regarding the game mechanics (GM), all the teachers and student evaluators were agreed 100% on all the elements. In content (C), all the elements were accepted by the evaluators except (LHC5) by one teacher's expert and (LHC1) by two student evaluators. The experts agreed that game design content was adapted from the book (English curriculum) approved by the concerned authority. Furthermore, the content in the game was suitable for learning vocabulary and enhancing English comprehension. All the elements in the feedback (F) component was agreed by all teachers and student experts, with 100% mentioning that game provided instant feedback after solving the exercises with correct and wrong answers status.

TABLE IV. HIGH-FIDELITY PROTOTYPE HEURISTIC EVALUATION RESULTS BASED ON ITEMS

Components	Teachers' Evaluators			Student Evaluators		
	No. of experts	Items	Marks	No. of experts	Items	Marks
I	5	(LHI1, LHI2, LHI3, LHI4, LHI5, LHI6, LHI7, LHI8, LHI9, LHI10, LHI11, LHI12)	Yes	18	(LHI1, LHI4, LHI5, LHI6, LHI7, LHI8, LHI9, LHI10, LHI11, LHI12)	Yes
				2	LHI2, LHI3,	Not Sure
GP	5	((LHPG1, LHPG2, LHPG3, LHPG4, LHPG5))	Yes	18	(LHPG2, LHPG3, LHPG4, LHPG5)	Yes
				2	(LHPG1)	Not Sure
GM	5	(LHGM1, LHGM2)	Yes	20	(LHGM1, LHGM2)	Yes
C	4	(LHC1, LHC2, LHC3, LHC4, LHC6, LHC7, LHC8)	Yes	18	(LHC2, LHC3, LHC4, LHC5, LHC6, LHC7, LHC8)	Yes
				1	(LHC5)	Not Sure
F	5	(LHF1, LHF2, LHF3)	Yes	20	(LHF1, LHF2, LHF3)	Yes
LL	5	(LHLL1, LHLL2, LHLL3, LHLL4, LHLL5, LHLL6)	Yes	20	(LHLL1, LHLL2, LHLL3, LHLL4, LHLL5, LHLL6)	Yes
CC	4	(LHCC1, LHCC2, LHCC3, LHCC4, LHCC5, LHCC6, LHCC7, LHCC8, LHCC9, LHCC10, LHCC11)	Yes	18	(LHCC2, LHCC3, LHCC4, LHCC5, LHCC6, LHCC7, LHCC8, LHCC9, LHCC10, LHCC11)	Yes
				1	LHCC1	Not Sure

In the language learning section (LL), all the elements were agreed upon by all the teachers and student experts with 100% by expressing that English language learning becomes easy with the help of the game. It further helped students to improve their vocabulary and English comprehension in grade three. Lastly, in the cultural context (CC), all the elements by the teacher evaluators were agreed with 100%. In contrast, only one element (LHCC1) by two student experts were 'not sure' by the student evaluators. All the elements are considered effective for the game in a cultural context by the teachers and student evaluators. The experts' responses and comments who responded with 'Not sure' were considered, and the score of seven constructs of the LLSG (I, GP, GM, C, F, LL, and CC) was 90%.

After analyzing all the data from the stakeholders, the finding from the above tables shows that the game-based learning application developed for grade three students of public sector schools is useful and usable. Overall, the results obtained from the heuristic evaluation were very positive, indicating that the high-fidelity prototype had most of the required language learning, educational, and cultural context elements. All the items provided in the sections of heuristics were acceptable by the teacher and student evaluators, but in one section, teachers and students were not sure about the component of LLSG. After compiling the identified issues, it was found that the color scheme in the game interfaces needed minor revision. Some shortcuts of common actions were not available in the game. So, it is necessary to follow a standard for the interface's color scheme by adding common touchpad button shortcuts in the game where necessary.

B. Think-Aloud Method (Expert Review)

In addition to studying the observable factors, evaluators have given their input and reflected on what they considered the games' positive and negative features. The comments and feedback are given in Table V.

The comments of teacher evaluators and students focused on the game usability that makes sure it is efficient and useful for language learning. The modules provided in the LLSG were quite important and helpful for learning the English language in public sector schools. The sounds module helped to learn vowel, short vowel, and long vowel sounds along with its exercises. Singular/plural and uses of

is/am/are/has/have/was/were the second and third modules that provide the pictorial presentations of the topic with pronunciation. Similarly, parts of speech, action words, sentences, w family, and comprehension are the further modules available in the LLSG enriched with easy and interesting learning material and with pronunciation that helped to learn these topics efficiently. After learning the desired topic, an assessment with its difficulty level could be made to evaluate the performance and progress with the defined reward.

According to the feedback, there were minor corrections of the color scheme on various interfaces that needs to be changed as well as the 'Home' and 'Back' button needs alignment in the appropriate place of interfaces. The 'Hint' button is to get help from the students while attempting medium and advanced level exercises in the game was also missing on some interfaces that need to be added. The font size of the text has a vital part in any game application, whereas font size on some interfaces of LLSG required some corrections. Furthermore, cultural pictures in the text for practicing a topic to learn a language are very helpful, but in LLSG, these pictures from the text were missing that will be added while pronunciation was also missing in language learning practicing some exercises that needed to be resolved. Similarly, in some exercises, right and wrong attempts and reward system that motivates students to learn a language effectively were not working properly and required some corrections. However, the game developers' remarks are meant to change the graphical user experience of the device to enhance functionality that renders the system complete.

For effectiveness, the teacher evaluators and students reported that the LLSG is helpful, easy, and effective for learning English and achieving the desired learning outcome. The evaluators noted that the learning material (content) provided in the game is easy, understandable, and logical, especially the pictorial presentation which could help to understand the topic efficiently. Lastly, it was reported that LLSG is easy to use at home because it is a standalone application that does not require internet access, and it could be used at home with the help of parents with interest. The quick response after solving exercises helped and passionate to see the progress of the desired topic that causes motivation and more engagement with the game.

TABLE V. FEEDBACK AND COMMENTS

Construct/ Modules	Evaluation Comments		Action to be taken
	Expert	Students	
Sounds	The sound module allowed learning vowels, short vowels, and long vowels with pronunciation and colorful pictures.	This section helped to learn the sounds of vowels with pronunciation.	-
Singular/Plural	This module guided me to learn singular/plural with pictorial representation and assessment.	This section helped me to learn singular/plural with pictures and pronunciation.	-
Uses	The uses module helped to learn: is, am, are, has, have, had, was, were with pictures and pronunciations.	This module guided to use is, am, are, etc., in the sentences with pronunciation.	-
Action Words	Action words guided to identify the activity using pictures, pronunciations, and assessment.	In this section, the sentence is represented in pictorial form and can be pronounced.	-

Parts of Speech	This module helped to learn nouns, pronouns, verbs, adjectives, and prepositions separately with its easy, medium, advanced assessment.	This section helped to learn with pronunciation and pictures the noun, pronoun, verb, adjective, and preposition.	-
Sentences	This section guided to write a proper sentence structure with its assessment.	This section taught the sentence structure with capitalization and full stop.	-
W Family	This module helped to learn W's family (Why What, Where, Who, Whom) with pronunciation and assessment.	This section guided learning: why, what, where, who, whom with pronunciation and different assessment levels.	-
Comprehension	This module guided to understand the English long sentences with pictorial representation and pronunciation for easy understanding.	This section taught the understanding of long sentences with pictures and pronunciation.	-
Usability	<ul style="list-style-type: none"> 1. The 'Home' and 'Back' buttons were not aligned on all interfaces, which caused a delay in using them. 2. The 'Hint' button was missing in the medium level of exercise that supports the user. 3. The font size was different in a few interfaces which is less readable. 	<ul style="list-style-type: none"> 1. The 'Home' button was different on one screen and another on other screens. 2. 'Hint' button that guided to solve a query was missing on some screens. 3. The font size of the text was different on some screens. 	<ul style="list-style-type: none"> 1. The home button will be aligned in the bottom right corner, and the Back button will be aligned in the bottom left corner. 2. A hint button will be added to the appropriate screens (where missing). 3. Font size in the game will be used according to a standard for the ease of users.
Content	<ul style="list-style-type: none"> 1. Pictures were missing in the practicing topic. 2. The alignment of questions in some exercises had disturbed, which was the cause of less efficient reading. 3. Pronunciation was missing in some practicing exercises. 	<ul style="list-style-type: none"> 1. Some pictures were not available while practicing. 2. Some questions were not in-line that was creating difficulty during reading. 3. On some screens, the pronunciation was missing 	<ul style="list-style-type: none"> 1. Necessary pictures will be added to the missing places. 2. The alignment will be made after making corrections in the code of the game. 3. The pronunciation issue will sort out by reviewing the code.
Assessment and Reward	<ul style="list-style-type: none"> 1. After solving the exercises, the correct and wrong attempt was missing, which could help a user's progress. 2. Reward with appropriate remarks was also missing, which could passionate the user for solving the exercises efficiently. 	<ul style="list-style-type: none"> 1. Correct and wrong attempts were not available to evaluate a topic. 2. Stars were missing in some exercises that could create the interest of the user to learn efficiently. 	<ul style="list-style-type: none"> 1. A screen of correct and wrong attempts will be added after solving all the questions. 2. Stars will be rewarded with an appropriate comment after solving an exercise.
Effectiveness	<ul style="list-style-type: none"> 1. This game-based language learning application helped to learn the language. 2. This could help to achieve the learning outcomes effectively. 	<ul style="list-style-type: none"> 1. The game was very easy, interesting, and helpful for learning the English language. 	-
Learnability	<ul style="list-style-type: none"> 1. The material provided for learning was easy to understand and logical. 2. The text and pictorial representation of topics made learning easy. 	<ul style="list-style-type: none"> 1. The content was very easy and understandable. 2. Pictures in the game were helpful to understand the topic clearly and easily. 	-
Efficiency	<ul style="list-style-type: none"> 1. The color scheme used in the game is not following a standard, and it looks less attractive. 2. The keypad was not functional to solve some exercises in the medium and advanced levels. 	<ul style="list-style-type: none"> 1. The color scheme of some screens is different from each other. 2. On some screens, there was a problem while solving questions. 	<ul style="list-style-type: none"> 1. The color scheme for the game will follow a standard to be the same in all interfaces. 2. The problem will be fixed after reviewing the code.
Satisfaction	<ul style="list-style-type: none"> 1. Instant feedback after solving the exercises developed the interest to use the game more and more. 2. Due to standalone, it is easy to use anytime at home on parents' mobile with interest. 	<ul style="list-style-type: none"> 1. The quick response for solving the exercise helped to see the progress and passion for using the game for a long time. 2. This game is easy to use at home with the help of parents and with interest. 	-
Cultural Context	The cultural context is very important for developing a language learning game and focusing on graphic symbols related to gender, age, sex, and religion. More cultural context pictures will engage the students for language learning with interest.	The game has cultural context, including icons, symbols, and images that helped to use it without any hesitation.	Few more pictures could be added where required to represent the culture.

V. CONCLUSION AND FUTURE WORK

This study aimed to evaluate the usability of LLSG, a language learning serious game comprising eight modules developed for learning English as a secondary language. Each module was enriched with learning content, pronunciation of words & sentences, and evaluation. To this end, two prominent methods, namely, heuristic evaluation and the think-aloud methods, were used while engaging different stakeholders, including language experts, students, teachers, and the game developers. As far as the heuristic evaluation is concerned, the researcher proposed heuristics, used to assess LLSG. At the same time, the think-aloud method was based on thorough discussion sessions held by the stakeholders.

The evaluation demonstrated that most of the domains were ranked above average and received positive scores, whereas two domains of the questionnaire were rated below average. The findings through both methods were very appreciative for LLSG. Teachers and students felt satisfied and accepted the effectiveness of game-based teaching and learning methods for language learning. Similarly, the findings from the think-aloud method were encouraging for LLSG, and feedback provided by the evaluators required slight changes to enhance game application according to the expectations and needs of the users. These minor concerns were about the look and feel, including the color scheme, font size, labels, and buttons. The teachers' remarks were primarily based on the subject material, and the improvement of gaming functionality was more significant. Apart from this, the primary assessment was covered in this game, where right and wrong attempts were recorded, but the detailed assessment might be recorded to see each student's progress level.

Moreover, this game was a standalone application that worked only on a tablet and recorded results in the local database. LLSG might be moved to the network model to increase its scope and connect it with a centralized database to store each student's question bank and grades for each assessment activity. In future work, the enhancements of the game-based learning may be carried out, and the network model may also be adopted from the current practice.

ACKNOWLEDGMENT

This research was funded by Universiti Kebangsaan Malaysia under the grant number: FRGS/1/2019/ICT01/UKM/01/1. The authors would like to thank the school education department (Sheikhupura) to provide special permissions for experimenting in the school.

REFERENCES

- [1] Malas, R. I.; Hamtini, T. A Gamified E-Learning Design Model to Promote and Improve Learning. *Int. Rev. Comput. Softw. (IRECOS)* 2016, 11 (1), 8.
- [2] Ishaq, K.; Zin, N. A. Mat; Rosdi, F.; Jehanghir, M.; Ishaq, S.; Abid, A. Mobile-Assisted and Gamification-Based Language Learning: A Systematic Literature Review. *PeerJ Comput. Sci.* 2021, 7 (e496), e496.
- [3] Abidin, S. R. Z.; Fadzilah, S.; Sahari, N. Heuristic Evaluation of Serious Game Application for Slow-Reading Students. *Int. J. Adv. Comput. Sci. Appl.* 2019, 10 (7). <https://doi.org/10.14569/ijacsa.2019.0100764>.
- [4] Ahmad, A.; Zeshan, F.; Khan, M. S.; Marriam, R.; Ali, A.; Samreen, A. The Impact of Gamification on Learning Outcomes of Computer Science Majors. *ACM trans. comput. educ.* 2020, 20 (2), 1–25.
- [5] Ishaq, K.; Rosdi, F.; Azan, N.; Abid, A. Usability and Design Issues of Mobile Assisted Language Learning Application. *Int. J. Adv. Comput. Sci. Appl.* 2020, 11 (6). <https://doi.org/10.14569/ijacsa.2020.0110611>.
- [6] Ishaq, K.; Azan, N.; Rosdi, F.; Abid, A.; Ali, Q. Usefulness of Mobile Assisted Language Learning in Primary Education. *Int. J. Adv. Comput. Sci. Appl.* 2020, 11 (1). <https://doi.org/10.14569/ijacsa.2020.0110148>.
- [7] Nielsen, J.; Landauer, T. K. A Mathematical Model of the Finding of Usability Problems. In Proceedings of the SIGCHI conference on Human factors in computing systems - CHI '93; ACM Press: New York, New York, USA, 1993.
- [8] Barata, G.; Gama, S.; Fonseca, M. J.; Gonçalves, D. Improving Student Creativity with Gamification and Virtual Worlds. In Proceedings of the First International Conference on Gameful Design, Research, and Applications; ACM: New York, NY, USA, 2013.
- [9] Ishaq, K.; Azan, N.; Rosdi, F.; Abid, A.; Ali, Q. Usability of Mobile Assisted Language Learning App. *Int. J. Adv. Comput. Sci. Appl.* 2020, 11 (1). <https://doi.org/10.14569/ijacsa.2020.0110145>.
- [10] Elaish, M. M.; Ghani, N. A.; Shuib, L.; Al-Haiqi, A. Development of a Mobile Game Application to Boost Students' Motivation in Learning English Vocabulary. *IEEE Access* 2019, 7, 13326–13337.
- [11] Sahrir, M. S., & Alias, N. A. A study on Malaysian language learners' perception towards learning Arabic via online games. *GEMA Online Journal of Language Studies* 2011, 11(3), 129-145.
- [12] Ashraf, H.; Motlagh, F. G.; Salami, M. The Impact of Online Games on Learning English Vocabulary by Iranian (Low-Intermediate) EFL Learners. *Procedia Soc. Behav. Sci.* 2014, 98, 286–291.
- [13] Ibrahim, A. Advantages of using language games in teaching English as a foreign language in Sudan basic schools. *American Scientific Research Journal for Engineering, Technology, and Sciences (ASRJETS)*, 2017, 37(1), 140-150.
- [14] Wong, L. H. Analysis of Students' After-School Mobile-Assisted Artifact Creation Processes in a Seamless Language Learning Environment. *J. Educ. Technol. Soc.* 2013, 16(2), 198-211.
- [15] Chen, X. Evaluating language-learning mobile apps for second-language learners. *Journal of Educational Technology Development and Exchange (JETDE)*. 2016, 9(2), 3.
- [16] Wang, Z., & Han, F. Developing English language learners' oral production with a digital game-based mobile application. *PLoS ONE*. 2021, 16(1), 1–11. <https://doi.org/10.1371/journal.pone.0232671>.
- [17] Hwang, W. Y., Ma, Z. H., Shadiev, R., Shih, T. K., & Chen, S. Y. Facilitating listening and speaking with game-based learning activities in situational context. In *Emerging issues in smart learning*. Springer, Berlin, Heidelberg, 2015, 193-200.
- [18] Meyer, B. Learning English through serious games—reflections on teacher and learner performance. In *Transactions on edutainment iii*. Springer, Berlin, Heidelberg. 2009, 82-92.
- [19] Cervi-Wilson, T., & Brick, B. ImparApp: Italian language learning with MIT's TaleBlazer mobile app. Innovative language teaching and learning at university: integrating informal learning into formal language education. 2018, 49.
- [20] Roccetti, M., Salomoni, P., Loiseau, M., Masperi, M., Zampa, V., Ceccherelli, A., Cervini, C., & Valva, A. On the design of a word game to enhance Italian language learning. 2016 International Conference on Computing, Networking and Communications, ICNC. 2016, 2–6. <https://doi.org/10.1109/ICNC.2016.7440546>.
- [21] Vasileiadou, I., & Makrina, Z. Using Online Computer Games in the ELT Classroom: A Case Study. *English Language Teaching*. 2017, 10(12), 134-150.
- [22] Lee, K., Kwon, S. O., Lee, S., Noh, H., Lee, J., Lee, J., ... & Lee, G. G. Effects of language learning game on Korean elementary school students. In *Speech and Language Technology in Education*. 2011.
- [23] Supuran, A. The opportunity of introducing serious games in teaching English for specific purposes. A study case on playing “simplycycle” serious game. *Journal of Teaching English for Specific and Academic Purposes*. 2017, 5(3), 459-466.
- [24] Casañ-Pitarch, R. An approach to digital game-based learning: Video-games principles and applications in foreign language learning. *Journal of Language Teaching and Research (Online)*. 2018, 9(6), 1147-1159.

- [25] Rachels, J. R., & Rockinson-Szapkiw, A. J. The effects of a mobile gamification app on elementary students' Spanish achievement and self-efficacy. *Computer Assisted Language Learning*. 2018, 31(1–2), 72–89.
- [26] Ishaq, K., Zin, N. A. M., Rosdi, F., Abid, A., & Farooq, U. Effectiveness of literacy & numeracy drive (LND): A students' perspective. In 2019 International Conference on Innovative Computing (ICIC). IEEE. 2019, 1–10.
- [27] Ishaq, K., Zin, N. A. M., Rosdi, F., Abid, A., & Ali, Q. Usefulness of mobile assisted language learning in primary education. *International Journal of Advanced Computer Science and Applications*. 2020e, 11(1), 384–395. <https://doi.org/10.14569/ijacsa.2020.0110148>.
- [28] Ishaq, K., Rosdi, F., Zin, N. A. M., & Abid, A. Requirements elicitation for game-based language learning application (Unpublished). In 2021 International Conference on Innovative Computing (ICIC). IEEE. 2021, 1–10.
- [29] O'Neil, H. F., Wainess, R., & Baker, E. L. Classification of learning outcomes: Evidence from the computer games literature. *Curriculum Journal*. 2005, 16(4), 455–474.
- [30] Nielsen, J. Usability engineering. Morgan Kaufmann. 1994a.
- [31] Nielsen, J. Designing websites to maximize press relations: guidelines from usability studies with journalists. Silicon Valley, CA: Nielsen Norman Group. 2003.
- [32] Adebesin, T. F., De Villiers, M. R., & Ssemugabi, S. Usability testing of e-learning. Proceedings of the 2009 Annual Conference of the Southern African Computer Lecturers' Association. 2009, 4, 1–1.
- [33] Albu, M., Atack, L., & Srivastava, I. Simulation and gaming to promote health education: Results of a usability test. *Health Education Journal*. 2015, 74(2), 244–254.
- [34] Isleyen, F., Gulkesen, K. H., Cinemre, B., Samur, M. K., Zayim, N., & Sen Kaya, S. Evaluation of the Usability of a Serious Game Aiming to Teach Facial Expressions to Schizophrenic Patients. *Studies in Health Technology and Informatics*. 2014, 205, 662–666.
- [35] Lim, C., Song, H. D., & Lee, Y. Improving the usability of the user interface for a digital textbook platform for elementary-school students. *Educational Technology Research and Development*. 2012, 60(1), 159–173.
- [36] Roscoe, R. D., Allen, L. K., Weston, J. L., Crossley, S. A., & McNamara, D. S. The writing pal intelligent tutoring system: Usability testing and development. *Computers and Composition*. 2014, 34, 39–59.
- [37] Virvou, M., & Katsionis, G. On the usability and likeability of virtual reality games for education: The case of VR-ENGAGE. *Computers and Education*. 2008, 50(1), 154–178.
- [38] Nielsen, J., & Molich, R. Heuristic evaluation of user interfaces. Conference on Human Factors in Computing Systems – Proceedings. 1990, 249–256.
- [39] Prensky, M. Digital Game-based Learning Prensky. Games2train. 2003, 1(1), 1–4.
- [40] Brangier, E., & Marache-Francisco, C. Measure of the Lived and Functional Effects of Gamification: An Experimental Study in a Professional Context. In *Advances in Intelligent Systems and Computing*. Springer International Publishing. 2020, 955.
- [41] Osipovskaya, E., & Miakotnikova, S. Using Gamification in Teaching Public Relations Students. In *Advances in Intelligent Systems and Computing*. Springer International Publishing. 2020, 916.
- [42] Tundjungsari, V. Mobile Learning Design Using Gamification for Teaching and Learning in Algorithms and Programming Language. *Advances in Intelligent Systems and Computing*. 2020, 916, 650–661.
- [43] Mayer, R. E. The promise of multimedia learning: using the same instructional design methods across different media. *Learning and Instruction*. 2003, 13(2), 125–139.
- [44] Garris, R., Ahlers, R., & Driskell, J. E. Games, motivation, and learning: A research and practice model. *Simulation and Gaming*. 2002, 33(4), 441–467.
- [45] Mayer, R., & Johnson, C. Adding instructional features that promote learning in a game-like environment. *Journal of Educational Computing Research*. 2010, 42(3), 241–265.
- [46] Couceiro, R. M., Papastergiou, M., Kordaki, M., & Veloso, A. I. Design and evaluation of a computer game for the learning of Information and Communication Technologies (ICT) concepts by physical education and sport science students. *Education and Information Technologies*. 2013, 18(3), 531–554.
- [47] Gee, J. P. What video games have to teach us about learning and literacy. *Computers in Entertainment*. 2003, 1(1), 20–20.
- [48] Giannakos, M. N. Enjoy and learn with educational games: Examining factors affecting learning performance. *Computers and education*. 2013, 68(246016), 429–439.
- [49] Klisch, Y., Miller, L. M., Wang, S., & Epstein, J. The Impact of a Science Education Game on Students' Learning and Perception of Inhalants as Body Pollutants. *Journal of Science Education and Technology*. 2012, 21(2), 295–303.
- [50] Steinkuehler, C., Squire, K., & Barab, S. A. Games, learning, and society: learning and meaning in the digital age. Cambridge University Press. 2012.
- [51] Domínguez, A., Saenz-De-Navarrete, J., De-Marcos, L., Fernández-Sanz, L., Pagés, C., & Martínez-Herráiz, J. J. Gamifying learning experiences: Practical implications and outcomes. *Computers and Education*. 2013, 63, 380–392.
- [52] Abidin, S. R. Z., Noor, S. F. M., & Ashaari, N. S. Low-fidelity prototype design for serious game for slow-reading students. *International Journal of Advanced Computer Science and Applications*. 2019, 10(3), 270–276.
- [53] Acquah, E. O., & Katz, H. T. Digital game-based L2 learning outcomes for primary through high-school students: A systematic literature review. *Computers and Education*. 2020, 143, 103667.
- [54] Hamari, J., Shernoff, D. J., Rowe, E., Coller, B., Asbell-Clarke, J., & Edwards, T. Challenging games help students learn: An empirical study on engagement, flow and immersion in game-based learning. *Computers in Human Behavior*. 2016, 54, 170–179.
- [55] Vee Senap, N. M., & Ibrahim, R. (2019). A review of heuristics evaluation component for mobile educational games. *Procedia Computer Science*. 2019, 161, 1028–1035.
- [56] Zhonggen, Y. (2019). A Meta-Analysis of Use of Serious Games in Education over a Decade. *International Journal of Computer Games Technology*. 2019, 2019(3).
- [57] Suo, Y. M., Suo, Y. J., & Zalika, A. Implementing Quizizz as game based learning in the Arabic classroom. *European Journal of Social Science Education and Research*. 2018, 12(1), 208–212.
- [58] Eltahir, M. E., Alsalhi, N. R., Al-Qatawneh, S., AlQudah, H. A., & Jaradat, M. 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*. 2021.
- [59] Tsai, Y. L., & Tsai, C. C. Digital game-based second-language vocabulary learning and conditions of research designs: A meta-analysis study. *Computers & Education*. 2018, 125, 345–357.
- [60] Peterson, M. The use of massively multiplayer online role-playing games in CALL: An analysis of research. *Computer Assisted Language Learning*. 2016, 29(7), 1181–1194.
- [61] Hung, H. T., Yang, J. C., Hwang, G. J., Chu, H. C., & Wang, C. C. A scoping review of research on digital game-based language learning. *Computers & Education*. 2018, 126, 89–104.
- [62] Ho, J. Gamifying the flipped classroom: how to motivate Chinese ESL learners?. *Innovation in Language Learning and Teaching*. 2020, 14(5), 421–435.
- [63] Mifsud, C. L., Vella, R., & Camilleri, L. Attitudes towards and effects of the use of video games in classroom learning with specific reference to literacy attainment. *Research in Education*. 2013, 90(1), 32–52.
- [64] Nielsen, J., & Loranger, H. Prioritizing web usability. Pearson Education. 2006.
- [65] ISO. Ergonomic requirements for office work with visual display terminals (VDTs)-Part II Guidance on usability. ISO/IEC 9241-11. 1998.
- [66] ISO. ISO/IEC 25010. 2011.

- [67] Preece, J., Benyon, D., & University, O. A guide to usability: Human factors in computing. Addison-Wesley Longman Publishing Co., Inc. 1993.
- [68] Davis, F. D. Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly: Management Information Systems*. 1989, 13(3), 319–339.
- [69] Nielsen, J., Clemmensen, T., & Yssing, C. Getting access to what goes on in people's heads? Proceedings of the Second Nordic Conference on Human-Computer Interaction - NordiCHI, 101. 2002.
- [70] Cotton, D., & Gresty, K. Reflecting on the think-aloud method for evaluating e-learning. *British Journal of Educational Technology*. 2006, 37(1), 45–54.
- [71] Brown-Johnson, C. G., Berrean, B., & Cataldo, J. K. Development and usability evaluation of the mHealth Tool for Lung Cancer (mHealth TLC): A virtual world health game for lung cancer patients. *Patient Education and Counseling*. 2015, 98(4), 506–511.
- [72] Chon, Y. V. The electronic dictionary for writing: A solution or a problem. *International Journal of Lexicography*. 2009, 22(1), 23–54.
- [73] Currie, S. L., McGrath, P. J., & Day, V. Development and usability of an online CBT program for symptoms of moderate depression, anxiety, and stress in post-secondary students. *Computers in Human Behavior*. 2010, 26(6), 1419–1426.
- [74] Granić, A., & Ćuković, M. Usability testing and expert inspections complemented by educational evaluation: A case study of an e-learning platform. *Educational Technology and Society*. 2011, 14(2), 107–123.
- [75] Hamel, M. J. Testing aspects of the usability of an online learner dictionary prototype: A product- and process-oriented study. *Computer Assisted Language Learning*. 2012, 25(4), 339–365.
- [76] Nielsen, J. Heuristic Evaluation: How-To: Article by Jakob Nielsen. Nielsen Norman Group. <https://www.nngroup.com/articles/how-to-conduct-a-heuristic-evaluation>. 1994b.
- [77] Federoff, M. A. Heuristics and usability guidelines for the creation and evaluation of fun in video games (Doctoral dissertation, Indiana University). 2002.
- [78] Hermawati, S., & Lawson, G. Establishing usability heuristics for heuristics evaluation in a specific domain: Is there a consensus? *Applied Ergonomics*. 2016, 56, 34–51.
- [79] Omar, H. M., & Jaafar, A. Heuristics evaluation in computer games. *Proceedings - 2010 International Conference on Information Retrieval and Knowledge Management: Exploring the Invisible World, CAMP'10*. 2010, 188–193.
- [80] Paavilainen, J. (2010). Critical review on video game evaluation heuristics. *Proceedings of the International Academic Conference on the Future of Game Design and Technology*, 56–65. <http://dl.acm.org/citation.cfm?id=1920787%5Cnhttp://portal.acm.org/citation.cfm?doid=1920778.1920787>.
- [81] Korhonen, H., & Koivisto, E. M. I. Playability heuristics for mobile multi-player games. *Proceedings of the 2nd International Conference on Digital Interactive Media in Entertainment and Arts – DIMEA*. 2007, 28.
- [82] Desurvire, H., & Wiberg, C. Game usability heuristics (PLAY) for evaluating and designing better games: The next iteration. *International Conference on Online Communities and Social Computing*, 5621 LNCS. 2009, 557–566.
- [83] Law, E. L. C., & Sun, X. Evaluating user experience of adaptive digital educational games with Activity Theory. *International Journal of Human-Computer Studies*. 2012, 70(7), 478–497.
- [84] Lin, H. K., Hsieh, M., Wang, C., Sie, Z., & Chang, S. Establishment and Usability Evaluation of an Interactive Ar. *Turkish Online Journal of Educational Technology*. 2011, 10(4), 181–187.
- [85] Xenos, M., & Velli, V. A serious game for introducing software engineering ethics to university students. *International Conference on Interactive Collaborative Learning*. 2019, 579–588.
- [86] da Silva, G. C., da Silva, L. P. F., Jofilsan, N. C., Correia, W. F. M., Gomes, A. S., & Campos Filho, A. S. Satisfaction analysis for using educational serious games for teaching wound treatment. In *International Conference on Applied Human Factors and Ergonomics*, Springer, Cham. 2018, 673–682.
- [87] Barnum, C., Bevan, N., Cockton, G., Nielsen, J., Spool, J., & Wixon, D. The “magic number 5”: Is it enough for web testing? *Conference on Human Factors in Computing Systems – Proceedings*. 2003, 5, 698–699.
- [88] Nielsen, J. Enhancing the explanatory power of usability heuristics. *Conference on Human Factors in Computing Systems – Proceedings*. 1994c, 152–158.
- [89] Denzin, N. K. Introduction: Entering the field of qualitative research. *Handbook of qualitative research*. 1994.
- [90] Preece, J. A guide to usability: human factors in computing. Wokingham, England/Reading MA: Addison-Wesley. 1998.
- [91] Lincoln, Y. S., & Guba, E. G. Naturalistic inquiry. Newbury Park, CA: Sage. 1985.
- [92] Patton, M. Q. Qualitative research and evaluation methods. Thousand Oaks, CA: Sage. 2002.
- [93] Ishaq, K., Azan, N., Rosdi, F., Abid, A., & Ijaz, M. The Impact of ICT on Students' Academic Performance in Public Private Sector Universities of Pakistan. *International Journal of Innovative Technology and Exploring Engineering*. 2020d, 9(3), 1117–1121.
- [94] Ishaq, K., Rosdi, F., Zin, N. A. M. & Abid, A. 2021b. Heuristic and Think aloud method to evaluate the Low-Fidelity prototype of Game-Based Language Learning Application. *4th International Conference on Innovative Computing, ICIC 2021*. (Accepted).
- [95] Ishaq, K., Rosdi, F., Zin, N. A. M. & Abid, A. 2021c. Requirements Elicitation for Game-based Language Learning Application. *4th International Conference on Innovative Computing, ICIC 2021*. (Accepted).

APPENDIX

Appendix A. Heuristics for LLSG		
Constructs	ITEMS	
Interface (I)	LHI1	Aesthetic and minimalist design
	LHI2	Maximize consistency and matches standards
	LHI3	Color, text, and space follow the principles of screen design.
	LHI4	Text, color, and font follow the readability principles.
	LHI5	The quality of text, images, and sound elements is acceptable.
	LHI6	The use of multimedia elements support meaningfully the text provided.
	LHI7	The integration of presentation means is well-coordinated.
	LHI8	The game speaks with words phrases and concepts.
	LHI9	The game helps me to navigate from one screen to another easily.
	LHI10	Pronunciation helps to understand the concept easily.
	LHI11	Consistent errors take back to learning screen.
	LHI12	Provide support (Hint) during assessment
Game Play (GP)	LHGP1	The control keys in game follow standard conventions.
	LHGP2	The game provides score after completion of stage.
	LHGP3	The game rewarded player after completion of stage.
	LHGP4	The game is interesting and engaging.
	LHGP5	The game is enjoyable to replay.
Game Mechanics (GM)	LHGM1	The game should behave in consistent, exciting and challenging way to players' action.
	LHGM2	The game controller actions have consistently mapped and learnable responses.
Content (C)	LHC1	The game has reliable and proven content with correct flow.
	LHC2	The game has clear goal, structure and learning objectives of content.
	LHC3	The content of game has main topic and subtopics.
	LHC4	Navigation is easy and accurate.
	LHC5	Supporting materials are sufficient and relevant (exercises).
	LHC6	Materials are interesting and engaging me.
	LHC7	The content helps to improve vocabulary.
	LHC8	The content helps to improve English comprehension.
Feedback (F)	LHF1	The game provides instant feedback on the progress.
	LHF2	The game notify me on the mistakes.
	LHF3	The game provide information on success or failure after completion of the stage.
Language Learning (LL)	LHLL1	The game helps to improve language learning.
	LHLL2	The game confident me after learning language.
	LHLL3	The game helps me to enhance my vocabulary.
	LHLL4	The game helps me to learn English comprehension easily.
	LHLL5	The game helps me to enhance English comprehension.
	LHLL6	The information is understandable conveyed to the users of game.
Cultural Context (CC)	LHCC1	The game should speak the language of the user with words, phrases and concepts.
	LHCC2	The game objects should be related to culture such as images, colors and familiar objects in order.
	LHCC3	The game should provide emergency exit to leave the state.
	LHCC4	The game should not the user think of similar actions, situations, or word mean the same.
	LHCC5	The game should minimize the memory burden with objects, actions and visible options.
	LHCC6	The game should provide interface without distracter elements.
	LHCC7	The game should provide the equal access to new user and expert.
	LHCC8	The error message in game should indicate to solve the problem.
	LHCC9	The game should provide help to user with less documentation.
	LHCC10	The game should provide diverse access to its provided options.
	LHCC11	The game should uses graphic symbols related to gender, age, sex, and religion where they have greater significance.