Journal of Positive School Psychology
2022, Vol. 6, No. 3, 3049–3063

Hybrid User Evaluation Methodology for Remote Evaluation: Case study of Educational games for children during Covid-19 Pandemic

Marwan Alshar'e[1], Ali Albadi[2], Malik Mustafa[3,] Noman Tahir[4], , Maria Al Amri[5]

[1,3,4,5] *Faculty of Computing Sciences, Gulf College, Muscat, Sultanate of Oman*
[2] *Centre for Postgraduate Studies and Research, Gulf College, Muscat, Sultanate of Oman*
Corresponding author: [1]marwan@gulfcollege.edu.om,

Abstract

Games learning (G-learning) eases educational process for teachers and learners. In Oman, this technology-based learning makes education more convenient, but its acceptance among students is affected by the students' awareness of it. The appropriateness of games used by Omani children has yet to be assessed. Furthermore, most of these games were developed and presented in foreign languages, and were following western culture. Meanwhile, usability assessments are generally done face-to-face with participants, but due to COVID-19 pandemic, such method is impracticable. Thus, new approaches must be developed for Human Computer Interaction (HCI), in usability evaluations especially. For assuring sufficient usability evaluation at least, a hybrid methodology comprising combined techniques was proposed in this study. The efficiency, effectiveness, and appeal of this methodology were qualitatively tested by selected experts. The obtained results show that the methodology is successful.

Keywords— User evaluation method, remote evaluation, educational games, games learning, Covid-19 pandemic

I. INTRODUCTION

Usability evaluation methods measure users' experience with a system. Hence, these methods could determine any concerns impeding users from finishing their tasks, delaying their progress, or otherwise worsening their user experience. The purpose of evaluations is to assure the effectiveness of the extensively used systems but test procedures differ and their accurateness has been frequently disputed. Countless usability evaluation methods are currently available and researchers have been debating over the best ones [1]. Also, the existing approaches have been undergoing improvement but these approaches have not been fully proven appropriate for use during Covid-19 pandemic. As such, a study should be carried out to determine the most appropriate usability assessment during the pandemic.

In usability evaluations, there is no universal method that usability specialists would utilize. However, Covid-19 pandemic has compelled researchers to figure out the most appropriate methods for conducting evaluation in the filed of Human Computer Interaction and in usability evaluations specifically. Meanwhile, social distancing is among the measures recommended by World Health Organization (WHO) in dealing with the pandemic, and considering that usability assessments are generally carried out face-to-face with participants but is not feasible during such time, remote usability is currently the main method in evaluation techniques.

Various usability evaluation methods were thus examined in this study in the resolution of usability evaluation issues with obstacles especially when children are involved, during Covid-19 pandemic. Additionally, the cultural impact of Omani children with educational games were examined. The findings of this study are expected to facilitate the

determination of the most appropriate methodology to be applied during this pandemic. Alternately, the findings could lead to the recommendation of a new methodology, followed by a recommendation of a framework that could potentially measure the cultural behaviour intention of Omani children toward the use of educational games.

The two questions below are thus presented in this study:

RQ1: How are educational games received by Omani children (1st to 3rd grade students)?

RQ2: How can we remotely evaluate the usability of educational games (Designed for children)?

The present study attempted to determine the most appropriate methods that could sufficiently and remotely evaluate educational games particularly concerning the usability of these games for children in Oman, specifically those aged 6 to 8 years old. The methods would allow the evaluators to still carry out the usability evaluations during the presence of hurdles such as the COVID-19 pandemic.

The present study thus attempted to achieve the specific objectives as follows:

1. To investigate the usability evaluation methods that are suitable for children's education games.
2. To propose a usability evaluation methodology.
3. To empirically evaluate the proposed methodology.

II. RESEARCH BACKGROUND

2.1 Educational Games

In academic institutions, the application of Electrocnic Learning (E-Learning) in the process of education is common today, as E-learning has been proven to facilitate learners at different levels, while also facilitating learners of specific groups such as special needs learners and young learners [2] [3] [4].

Meanwhile, mobile applications have significantly evolved, that now mobile applications also include games, allowing users to perform various types of tasks. In general, games in mobile applications are for entertainment and enjoyment, but currently, games have expanded their functions as they are now being applied in different domains including health care, education, and business.

Mobile educational games are generally driven by digital educational game played on mobile devices by the the generations today. However, the functionality of educational games and the mobile devices have brought challenges in the measurement of the usability of these games for the mobile platform [5]. Educational games encompass the incorporation of educational material into games in teaching the player certain subjects to achieve certain learning outcomes [6]. On the other hand, game-based learning relates to games commonly utilized for learning purposes [5].

Educational games could stimulate students' desire to learn and for this reason, educational games have been regarded as an efficient tool for learning [6] [7] while also improving the students' educational performance [8] [9]. Mobility is among the features of educational games. Hence, this mode of learning allows learner to learn at any place and time of their convenience. Still, the efficacy of mobile educational games is difficult to determine. In fact, gauging the usability of any mobile platform application is not a simple task considering that mobile devices today have limitations, for instance, inappropriate screen size, inefficient control or input interface, and interruptions, and all of these can hamper usability [10]. Furthermore, usability of educational games is not easy to determine as it requires the incorporation of both educational and gaming requirements [11].

Within apps marketplaces, the available educational games are mostly those relating to mathematics, science and language [12]. As mentioned in the literature, serious games encompass those with a purpose that is beyond fun and enjoyment [13] and are executed according to research on educational games, and produced math games rather than observing math games on the apps marketplaces [14] [15]. App marketplaces have been the main place for people to find and install apps [16]. As such, the observation of available games gives insight

into the user experience of coming across different types of educational games.

2.2 Usability evaluation

Remote usability evaluation is a method of HCI evaluation that has been commonly applied in both research and real world situation. Notably, the currently available established user-centred assessment methods have been considered as usability evaluation methods. The currently available HCI evaluation methods are still being used today, but considering the Covid-19 pandemic at present time, these methods are being questioned in terms of how the methods are being impacted, in their usability assessment particularly. Among the lurking questions are:

1. What are the challenges faced by experts in their execution of usability evaluation using the "common" methods?
2. What are the practices that remain unchanged but successful, true, and reliable during the Coronavirus pandemic?
3. What are the latest usability assessment requirements now?

Usability evaluations are generally carried out inside a usability evaluation lab, and the participants are asked to the test lab in which they perform the evaluation. In essence, the lab is a special room comprising intricate and sophisticated audio and visual recording and analysis apparatus, and the evaluation is independently executed [17].

Remote usability testing entails a situation in which the involved evaluators and participants are in different location, and remote usability evaluation can be carried out either using synchronous method or asynchronous method, with specific tools applied for each method. In the former method, the evaluators would obtain the information and remotely carry out the session of evaluation in real-time with the participant. In this situation, the participant may be at home, work, or other locations. Video conferencing software or remote viewing are among the tools used, and this allows sharing of computer screens between the evaluator and the participant, so that the evaluator can view the participant's screen to see what is going on [18] [19] [20].

Meanwhile, the asynchronous method does not provide real-time data access to evaluator. Also, there is no communication between evaluator and participant during data collection. The asynchronous methods have been associated with the automated approaches whereby clickstreams of users are spontaneously attained. Among the key advantages of this approach is that it allows more participants to take part in the evaluation, without extra cost or just a minute amount of cost incurred per participant. In the execution of asynchronous studies, there have been many methods suggested. Among the solutions is to let the participants download and utilize an instrumented browser that will gather the clickstreams and screenshots of the users, and then, the participants would send the data to the evaluator's site for review purposes [17].

In addition, in the use of asynchronous method, observational data and unprompted verbalization records cannot be produced during remote evaluation sessions. Hence, the qualitative data are recorded through post-evaluation questionnaires or self-report. Notably, this method allows the recording of participants in large number.

It being comparable to laboratory research and its ability in facilitating the capture of qualitative data are among the reasons why synchronous approach is preferred among some scholars [18][21][22]. As opposed to laboratory user test, synchronous remote testing is economical, especially in terms of travel expenses incurred to participants residing in different regions of the country, but in certain situations, the costs can be comparable to those of laboratory research. Additionally, remote synchronous approach does not require the use of facilities, especially for an online application testing or when the product or software is deliverable electronically. In addition, the approach is rather speedy in process, that is, it does not consume much time. For these two major reasons, the approach is preferred to the traditional user evaluation approach. Somehow, as opposed to the traditional laboratory user

evaluation, synchronous remote monitoring may seem more invasive [23].

Past extant studies are demonstrating the need to explore these aspects more comprehensively. Also, the studies were demonstrating variances in the types of applications under study, the procedures used, the instructions presented to the study participants, the coding of the participants' actions, and the methods of data analysis used.

This study thus attempted to establish a methodology that fits with the study objectives in evaluating a given situation notwithstanding the unfeasibleness of direct contact with the target audiences in educational games evaluation. An appropriate methodology that could facilitate the accomplishment of the study objectives while also responding to the research problem was thus proposed in section three.

III.PROPOSED METHODOLOGY

The main aim of this study was to evaluate the effectiveness of educational games among Omani children in their learning. Somehow, the outbreak of COVID-19 has made it unsafe for researcher to freely reach the students at their schools or homes. As a solution, a technique that facilitates the execution of the required evaluation with accurate result was created. Figure 1 accordingly displays the proposed methodology.

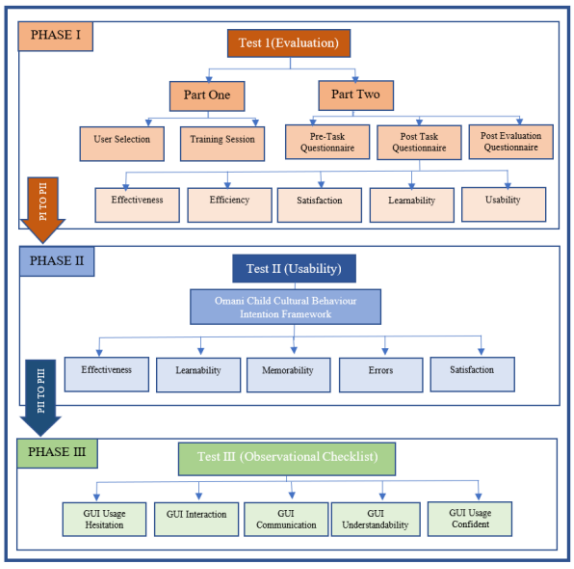

Figure 1: Proposed framework of Hybrid user evaluation methodology

The proposed framework encompasses a hybrid user evaluation methodology comprising several available techniques of usability evaluation. Notably, the investigation and selection of methods considered the motive underpinning the study.

The framework proposed involves three phases as detailed below:

Phase I: Test 1 (Evaluation)

Five primary elements of usability that are essential to the process of evaluation are evaluated in this phase. These elemets include the following: Effectiveness, Efficiency, Satisfaction, Learnability and Usability. There are two parts to this stage as follows: Part 1 includes User Selection, and Training Session for the facilitator, while Part 2 includes Post-evaluation, Post-task questionnaire, and Post-Task Evaluation. Part 2 follows the literature investigation where number of studies have used the mentioned procedures to evaluate closely related situations.

During part 1, the evaluator is replaced with a carefully selected facilitator as proposed by the methodology. Teacher and parents are the mostly likely facilitator. The facilitator will undergo training session to help them collect and record accurate data, and this helps them perform the evaluation sufficiently.

During this phase, direct evaluation is carried out on certain games by direct gathering of data by trained facilitator. At the end of this phase, the acceptance of certain game by children and the presence of major issues (if any), will be understood.

Phase II: Test 2 (Usability)

The second phase involves the evaluation of usability of certain games. This involves the evaluation of five main quality components of the games namely Efficiency, Learnability, Memorability, Errors, and Satisfaction. The evaluation will facilitate the assurance of reliable and accurate evaluation results.

According to [24] "child usability heuristk-Evoke Child Mental Imagery Explanation, they are drawn from the child's own culture". It is therefore believed that the success of evaluation

of educational games is significantly affected by the Cultural Behavior Intention of a child. In view of this, a new framework was proposed in this study. Accordingly, a hypothesis reflecting the impacts of the five factors on the Cultural Behaviour Intention of Omani Children was formed. Through the framework proposed in Phase 1, the facilitators (parents/teachers) will obtain the data on the observers' views on the usage of children of specific games. For the purpose, a questionnaire will be used. The obtained results will be then be statistically analysed and presented.

Phase III: Test 3 (Observational Checklist)

During the third phase, the Behavior Intention of using the Educational Games (EGs) among Omani children is evaluated. For the purpose, an observational checklist will be used. As highlighted in [25], user success can be determined from the observation checklist utilized by the researcher in his/her execution of usability test. Meanwhile, the satisfaction measurement is obtained from post questionnaire with the children. GUI's usage hesitation, interaction, communication, understandability, and confident are among the elements examined during this phase. This study adopts the observational checklist items from [25].

The results of each phase will become inputs for supporting the subsequent phases. Accordingly, the proposed framework will present clear picture, and accurate and consistent results reflecting the actual information on usability evaluation that is specifically constructed for remote evaluation, or in a situation where a professional observer executes a face-to-face usability evaluation. This consequently can reveal the effectiveness of this method, especially in regards to remote evaluation.

IV. MODEL EVALUATION

Upon the completion of the design process and the affirmation of the ability of the methodology proposed, the quality of the proposed methodology will be evaluated and affirmed in terms of its usability. The evaluation will determine and assure that the phases are appropriate for guiding the evaluation process of a given educational game.

For a design theory, or for a guideline, model, and so forth, preferability becomes a key concern, and preferability relates to the degree to which a method is "better" than other known methods to achieve the sought after result [23]. Accordingly, the **Criteria for Evaluating Research on Generalizable Design Knowledge** comprises three values that should be considered in the theory design, and they are as follows [26] [27] [28]:

1. Effectiveness: This value primarly concerns the degree to which the theory (or guideline or method) application achieved the goal in a particular situation.
2. Efficiency: This value involves two elements namely effectiveness and cost, which are summed up in the phrase of "bang for the buck" whereby a measure of "bang" relates to effectiveness and a measure of the "buck" relates to cost incurred either in money or time, or other types of cost, or a mix of costs. For instructional-design theory, time, effort, and energy required from human, in addition to the cost of additional required resources for instance, materials, equipment or other required setting requirements for instruction, are taken into consideration.
3. Appeal: This value encompasses the matter concerning how enjoyable the resulting designs are for all individuals involved.

The evaluation hence will determine if the proposed methodology has all the three mentioned values in order to affirm if the methodology has the appropriate design.

A usability study was carried out by [29] and two lists covering all issues of usability were presented as follows. The first list contains the key heuristics to describe all problems associated with usability which include consistency, language that user is familiar with, clear user input, making the actions of user discernible, a simple and minimalist design, shortcuts and connection to the real life situation, facilitates error recognition and

recovery, allowing wrong actions to be reversed, and presenting sophisticated set of evaluable actions. Another list comprises the key heuristics for the expression of serious usability problems and these include making the actions of user discernible, having feedbacks that are both consistent and accurate, having a sophisticated set of evaluable actions, allowing wrong actions to be reversed, the use of solid analogies, having clear user input, attempts to prevent errors, and having one goal per mechanism).

In the evaluation of the target system outside of the actual domain, the discounted usability inspections usually would utilize the usability-expert evaluators set. Somehow, the identification between under an end-user persona is no simple task for usability-experts. Equally, it is no easy task to maximally imitate the context of the original domain and workplace. However, it is hardly considred as an imitation of the original settings. For this reason, there have been reports that these techniques have missed the usability problems associated to the actual task settings [30] [31].

Taking into account the nature of this research, an expert review evaluation technique has been selected for evaluating the appropriate design and usefulness of the methodology proposed. In [32], the notion "expert review" was elaborated as an informal method that is utilized by at least one expert usability professional in user interface evaluation. The method is based on the notion that experts could provide clear judgment and opinions as they are experienced and knowledgeable in their domains.

In discussing the notion of individual expert review, [33][34][35] mentioned that an individual expert review involves one expert examining some data relating to a given product or service. In this situation, the the expert acts like a user for the purpose of finding out problems in a system through the application of many techniques including individual walkthrough, interviews with users, product review in opposition to some heuristics and product scrutiny from various outlooks. Individual expert review execution may include the aspects of different methods of evaluation as in heuristic evaluations, cognitive walkthroughs, perspective-based inspections, and informal usability testing [33][36][37][38].

In this study, the individual expert review evaluation was carried out utilizing a questionnaire executed following a walkthrough process whereby the creator of the proposed methodology walks the participants through the process. In the process, the participants critically review the proposed methodology and try to find errors or domains of special concern.

4.1.1 Evaluation procedure

There are three stages to the evaluation session procedure applied in this study, as follows:

1. Evaluation Planning: During this stage, the experts are selected, contacted and then invited to take part in the study, and prepare a set of questions.
2. The evaluation: This is the stage in which the evaluation is carried out. During this stage, the researcher will present to the experts, namely the individuals who have agreed to take part in the study, the complete particulars and account on the proposed methodology for review purposes. The questionnaire is provided as well, so that the participants could provide their feedback on the surveyed model.
3. Post evaluation: This stage is the final stage in which the researcher will analyse the gathered data, present the results, and make recommendations as necessary/ appropriate.

4.1.2 Evaluation Questionnaire

The validity of the gathered data was assured through the use of close-end questions in the questionnaire. These questions were for obtaining the opinions from the partaking esperts. There were five questions, and these questions were equipped with five-point likert scales with indicators of strongly agree, agree, neutral, disagree, and strongly disagree. Through the evaluation procedure, the experts were provided with the documentation and specifics of the proposed methodology. The experts then examined the methodology, and

presented their opinion and observations through the questionnaire provided.

The following portion of the paper details the asked questions and the attained results:

a. Efficiency

(i.) The suggested phases based on the described purpose of the new framework would help to collect accurate data to support an evaluation process.

The above statement was to affirm that the proposed three phases were appropriately chosen and designed. These phases are intergral in the acquirement of the needed output from an evaluation process.

(ii.) The designed order and flow of the suggested phases is rational, and would provide the necessary feed-forward data to the successor phases.

The above statement was to affirm that the overall structure of the proposed methodology was formulated correctly. The statement was also to affirm the reliability of the results generated in each phase.

(iii.) The suggested Phases are reasonable in terms of the required efforts, cost, and time.

The above statement was to measure the agreement of the partaking experts towards the efficiency of the proposed methodology concerning the proposed factors of time, cost and efforts.

b. Effectiveness

(iv.) The chosen UI evaluation procedures within each of the suggested phases is adequate and acceptable.

The above statement was to affirm the appropriateness of the UI evaluation techniques selected within each phase.

(v.) The proposed methodology evaluates the necessary factors that directly contribute to the evaluation of educational games UI and produces reliable and accurate evaluation results.

The above statement was to affirm the coverage of the correct and necessary factors of certain product features, with respect to UI evaluation.

c. Appeal

(vi.) Overall, the proposed methodology is accepted and useful for the purpose of its existence.

The above statement was to obtain general affirmation by the experts of their belief and trust towards the ability of the new methodology in the remote evaluation of educational games for children.

4.1.3 Evaluation Result

A total of eighteen (18) experts took part in the evaluation, the reults for each question was as following:

- *Result From Question One*

The various responses obtained from the participants were analysed and results were obtained. From the results, it was clear that the viewpoints of the experts/ participants affirmed what this study had brought forth in highlighting the importance of the selected phases; the phases are integral for achieving the sought-after output from the evaluation process. The results generated from data analyses can be viewed in Table 1.

Table 1: Respondents statistics for Question 1.

Q1. The suggested phases based on the described purpose of the new framework would help to collect accurate data to support an evaluation process.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Neutral	5	27.8	27.8	27.8
	Agree	9	50.0	50.0	77.8
	Strongly Agree	4	22.2	22.2	100.0
	Total	18	100.0	100.0	

The obtained results as displayed in Table 1 indicate that 27.8 percent of the experts were neutral about the idea discussed in Question 1 whereby 50.0 percent expressed “Agree,” and

22.2 percent expressed "Strongly Agree." Based on the collected data and opinions, the researcher concluded that **the suggested phases based on the described purpose of the new framework are valid and approved to collect accurate data to support an evaluation process.**

- *Results from Question Two*

The results obtained from the responses for Question Two were in agreement with the purposes of and affirmed this question. Specifically, the analysis for all responses to this question confirmed the correct formulation of the general structure of the proposed methodology, and the reliability of the results obtained from each phase. The viewpoints of the experts are as displayed in Table 2.

Table 2: Respondents statistics for Question 2.

Q2. The designed order and flow of the suggested phases is rational, and would provide the necessary feed-forward data to the successor phases.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Neutral	3	16.7	16.7	16.7
	Agree	8	44.4	44.4	61.1
	Strongly Agree	7	38.9	38.9	100.0
	Total	18	100.0	100.0	

The obtained results as displayed in Table 2 indicate that 16.7 percent of the experts were neutral about the idea discussed in Question 2 whereby 44.4 percent expressed "Agree," and 38.9 percent expressed "Strongly Agree." Based on the collected data and opinions of these experts, the researcher concluded that **The designed order and flow of the suggested phases are rational, and would provide the necessary feed-forward data to the successor phases, which shall provide expected results.**

- *Results from Question Three*

The results of Question Three, based on the responses from the participants, affirmed the agreement of the experts towards the methodology efficiency with respect to the suggested factors of time, cost and efforts. The obtained opinions from the experts can be observed in Table 3.

Table 3: Respondents statistics for Question 3.

Q3. The suggested Phases are reasonable in terms of the required efforts, cost, and time

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	12	66.7	66.7	66.7
	Strongly Agree	6	33.3	33.3	100.0
	Total	18	100.0	100.0	

The obtained results as displayed in Table 3 demonstrate that 66.7 percent of the experts showed agreement towards the idea discussed in Question 3 by expressing "Agree," while 33.3 percent expressed "Strongly Agree." As such, taking into account the gathered data and opinions of these experts, the researcher concluded that **the suggested Phases are reasonable in terms of the required efforts, cost, and time which shall prove that the new framework is useful when it is adopted.**

- *Results from Question Four*

The results attained from the responses of the participants for Question Four signify their affirmation of the the appropriateness of the chosen UI evaluation techniques within each phase. Table 4 accordingly displays the opinions gathered from the partaking experts.

Table 4: Respondents statistics for Question 4.

Q4. The chosen UI evaluation procedures within each of the suggested phases is adequate and acceptable

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	9	50.0	50.0	50.0
	Strongly Agree	9	50.0	50.0	100.0
	Total	18	100.0	100.0	

The obtained results as displayed in Table 4 demonstrate that 50.0 percent of the experts showed strong agreement towards the idea discussed in Question 4 by expressing "Strongly Agree," while the remaining 50.0 percent expressed "Agree." As such, taking into account the gathered data and opinions of these experts, the researcher concluded that **the chosen UI evaluation procedures within each of the suggested phases is adequate and acceptable. Therefore, it shall produce correct and accurate data when it is being used.**

- *Results from Question Five*

For Question Five, the achieved results affirmed the viewpoint of the experts concerning the coverage or inclusion of the correct and needed factors of the given product features with respect to UI evaluation. Table 5 can be referred.

Table 5: Respondents statistics for Question 5.

Q5. The proposed methodology evaluates the necessary factors that directly contribute to the evaluation of educational games UI and produce reliable and accurate evaluation results

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	14	77.7	77.7	77.7
	Strongly Agree	4	22.2	22.2	100.0
	Total	18	100.0	100.0	

The obtained results as displayed in Table 5 demonstrate that 77.7 percent of the experts expressed agreement towards the idea discussed in Question 5 by expressing "Agree" while 22.2 percent of the experts were showing strong agreement by expressing "Strongly Agree." As such, based on the gathered data and opinions of these experts, the researcher came to a conclusion that **the proposed methodology evaluates the necessary factors that directly contribute to the evaluation of educational games UI and produce reliable and accurate evaluation results when it is used. This means that when the techniques are appropriately applied, and when the mechanisms are correctly evaluated, the outcomes would be correct and useful.**

- *Results from Question Six*

The responses of the partaking experts for Question Six signify the final confirmation of these experts on their belief and trust towards the ability of the new methodology in the remote evaluation of educational games for children. Table 6 provides the details.

Table 6: Respondents statistics for Question 6.

Q6. Overall, the proposed methodology is accepted and useful for the purpose of its existence.

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Agree	12	66.7	66.7	66.7
	Strongly Agree	6	33.3	33.3	100.0
	Total	18	100.0	100.0	

The obtained results in Table 6 demonstrate that 66.7 percent of the experts expressed agreement towards the idea deliberated in Question 6 by expressing "Agree" while 33.3 percent were showing strong agreement by expressing "Strongly Agree." As such, based on the gathered data and opinions of these experts, the researcher came to a conclusion that **Overall, the proposed methodology is accepted and useful for the purpose of its existence. This shows that correct and accepted usage of the framework for the purpose of its existence, and proper evaluation mechanisms shall lead to correct and useful outcomes.**

4.1.4 Discussion

This section discusses the adequacy of the proposed framework. In this regard, eighteen experts had evaluated the framework and provided their opinions on the framework after observing it. Using the study questionnaire, these experts provided their opinions and views on the proposed methodology. In essence, the framework was tested on three main features namely efficiency, effectiveness, and appeal, and the evaluation of these features were through six questions. Based on the responses gathered from the experts, the researcher concluded that the researchers were positively supporting the proposed framework, and they also confirmed the tested features of the framework. The research team thus contemplated the success and adequacy of the proposed framework as a valid framework for evaluating educational games for children in Oman. The proposed framework may thus be viable for real life applications..

4.2 Research Evaluation:

There are three parts to the section. The first part discusses the measurement model, followed by the second part that presents the study's descriptive analysis, and finally, the third part of this section presents the study's structural model. The details are as follows:

4.2.1 Measurement Model

In general, the evaluation of the measurement model involves the examination of reliability, convergent validity and discriminant validity of the model. In this study, reliability refers to the internal consistency.

Table 7: Sample characteristics

Measure	Frequency (%)	Percentage (%)
Gender		
Male	11	61.1%
Female	7	38.9%
Age		
<25	6	33.3%
25-35	8	44.5%
More than 35	4	22.2%
Educational level		
Diploma Degree	7	38.9%
Bachelor's Degree	5	27.8%
Master's Degree	6	33.3%
Experience		
<4 Years	4	22.2%
4-9 Years	3	16.7%
10-15 Years	7	38.9%
More than 15 Years	4	22.2%
Total	18	100%

4.2.2 Descriptive Analysis

The research data underwent a descriptive analysis. This was to evaluate the perceptions of the participants towards the degree of importance of the seven variables in the determination of their Intention to use Educational games. Accordingly, Table 8 presents the obtained means, standard deviations, the minimum and maximum of the respondents' feedback on the research variables. There are five levels of agreement, and to ease the interpretation, the levels are divided into three categories of low level (mean scores lower than 2.33), high level (mean scores higher than 3.67), and moderate level (mean scores between 2.33 and 3.67).

Table1. 8: Descriptive statistics for Mean, Standard Deviation, and coefficients of correlations variables (N = 18)

Construct	**M** [a]	**SD** [b]	**1**	**2** [c]	**3**
Efficiency	3.74	3.643	0.75	0.55**	0.43**
Effectiveness	3.82	3.849		0.47**	0.42**
Appeal	3.74	3.767			0.38**
Intention to use Educational games	3.88	3.829			

4.2.3 Structural Model

The measurement model (outer model) is evaluated through the determination of its convergent and discriminant validities. Specifically, convergent validity is about the degree to which differing instruments could measure similar construct [39] [40]. It also means the level to which these instruments are concurring. Reliability of construct measurement is part of convergent validity. Evaluation of reliability involves the evaluation of composite reliability and internal consistency. The Cronbach's Alpha coefficient is a common measure of internal consistency whereby the value greater than 0.7 denotes reliability. Table 9 accordingly shows that all constructs scored Composite Reliability (CR) of greater than 0.7.

In this study, all constructs were affirmed in terms of their Internal Consistency (ICR) of the scales. As for convergent validity, its measurement involved the determination of the factor loadings of the items on the constructs of the proposed model. It was accordingly mentioned in [41] that items with loadings of 0.70 or more should be retained. There were four (4) constructs involved as follows: Efficiency (EF), Effectiveness (Eff), Appeal (App) and Intention to use Educational games (ItU), and none of the constructs had loading value of smaller than 0.7 (value is displayed in bold). Those scoring at least the threshold value were kept for further analysis.

Table 9: Result of construct assessment

Constructs	**Items**	**Factor loading**	**Mean±SD**	**CR**	**Cronbach's α**	**AVE**
Efficiency (EF)	**EF1**	0.787	3.815±0.926	0.928	0.815	0.521
	EF 2	0.786	3.816±0.921			
	EF 3	0.787	3.817±0.936			
	EF 4	0.785	3.715±0.826			
Effectiveness (Eff)	**Eff 1**	0.783	3.972±0.971	0.814	0.981	0.576
	Eff 2	0.882	3.712±0.825			
	Eff 3	0.784	3.713±0.974			
	Eff 4	0.881	3.854±0.923			
Appeal (App)	**App1**	0.787	3.716±0.923	0.834	0.971	0.586
	App 2	0.787	3.717±0.932			
	App 3	0.786	3.815±0.822			
	App 4	0.785	3.612±0.975			
	App 5	0.784	3.775±0.921			
	App 6	0.783	3.676±1.082			
Intention to use	**ItU 1**	0.874	3.647±1.072	0.834	0.991	0.556
Educational games	**ItU 2**	0.758	3.624±1.024			
(ItU)						

AVE: Average variance extracted, SD: Standard Deviation

4.3 Research finding

The findings of this study innovatively add to the literature of intention to use, specifically towards Educational games for children during Covid-19 Pandemic. This study presents to the extant literature the new theoretical knowledge that is of value for the new proposed model that comprises various variables.

The present research highlights the necessitated considerations in the application of educational games for children during Covid-19 Pandemic. This study specifically provides understanding of Educational games in Oman, while also highlighting the specific factors impacting the intention to use Educational games for children, during Covid-19 Pandemic, in Oman. Additionally, the measurements and the conceptual framework were presented in this study, through which, the relationship among students Efficiency (EF), Effectiveness (Eff) and Appeal (App) as independent variables, and Intention to use Educational games (ItU) as dependent variable, were accordingly described.

4.4 Research Contributions

Several issues salient to the present situation that calls for the application of Hybrid User Evaluation Methodology for Remote Evaluation of intention to use Educational games for children during Covid-19 Pandemic were discussed in this study. Accordingly, this study discusses its contribution to the extant domain as follows:

- The present study adds to the enduring debate on the Impact factors of adopting **Hybrid User Evaluation Methodology for Remote Evaluation**
- The present study enriches the knowledge of the advantages and requirements for Improving the **User Evaluation Methodology for Remote Evaluation**.
- The present study presents a model that could deliberate the processes and factors impacting the remote usability evaluation.

The present study also looked into the impact factors that influence the intent of students in schools in Oman to use Educational games (ItU) and such scrutiny significantly adds to the related literature. Additionally, the research model in this study examines the Factor Analysis of **Hybrid User Evaluation Methodology for Remote Evaluation**, and its relation to the Intention to use Educational games (ItU) in developing countries.

V. CONCLUSION AND RECOMMENDATIONS

Education games are an innovative method used for enhancing the process of teaching and learning, and children have been those that benefit greatly from such teaching method. Somehow, the adequateness of the games needs to be be evaluated based on countries or cultures.

In essence, this research attempted to evaluate educational games, in terms of usability, for children in Oman. However, Covid-19 pandemic has compelled many countries all over the world including Oman to impose lockdown as a safety measure. For this reason, it was impossible to gather data from the children face-to-face. Hence, some special tools or method are needed to remotely obtain such data.

This study proposed a methodology for user evaluation of educational games in Oman for children of six to eight years old in age. For the purpose, several existing methods were examined and applied in the construction of a hybrid methodology that could perform appropriate evaluation using the right data.

It was clear that the evaluation of the selected experts of the proposed methodology denotes their acceptance towards the proposed methodology in the gathering of appropriate data. The proposed methodology demonstrated the ability in the execution of valid evaluation result of the examined games.

Based on the findings, the researcher recommends the Hybrid User Evaluation Methodology for Remote Evaluation (framework) detailed in Section 3 onwards in this paper for Remote User Evaluation who want to evaluate the usability of a system or educational ganme remotely to produce more valid and accurate results.

Acknowledgment

The research leading to these results has received funding from the Ministry of Higher Education, Research & Innovation (MoHERI) of the Sultanate of Oman under the Block Funding Program (Block Funding Agreement No.: TRC/BFP/GULF/01/2019, Project Code: BFP/RGP/ICT/19/292).

REFERENCES

1. F. Paz, D. Villanueva, C. Rusu, S. Roncagliolo, and J. A. Pow-Sang, "Experimental evaluation of usability heuristics," in *2013 10th International Conference on Information Technology: New Generations*, 2013, pp. 119–126.
2. Marwan Alshar'e, "Mobile Appointment System for UniversitiesNo Title," University Utara Malaysia, 2008.
3. M. Mustafa, S. Alzubi, and M. Alshare, "The Moderating Effect of Demographic Factors Acceptance Virtual Reality Learning in Developing Countries in the Middle East," in *Advances in Computing and Data Sciences*, 2020, pp. 12–23.
4. M. Alshare and M. Mustafa, "Evaluation of autistic children's education in Oman: the role of eLearning as a major aid to fill the gap," *Elem. Educ. Online*, vol. 20, no. 5, pp. 5531–5540, 2021.
5. N. M. V. Senap and R. Ibrahim, "A Review of Heuristics Component for Usability Evaluation of Mobile Educational Games," *Open Int. J. Informatics*, vol. 7, no. 2, pp. 190–199, 2019.
6. R. Ibrahim, K. Khalil, and A. Jaafar, "Towards educational games acceptance model (EGAM): A revised unified theory of acceptance and use of technology (UTAUT)," *Int. J. Res. Rev. Comput. Sci.*, vol. 2, no. 3, p. 839, 2011.
7. M. Papastergiou, "Digital game-based learning in high school computer science education: Impact on educational effectiveness and student motivation," *Comput. Educ.*, vol. 52, no. 1, pp. 1–12, 2009.
8. G. T. Perry, C. C. Kulpa, E. Pinheiro, and M. L. Eichler, "Lessons from an educational game usability evaluation," *Int. J. Interact. Mob. Technol.*, vol. 6, no. 2, pp. 23–28, 2012.
9. Y.-L. Huang, D.-F. Chang, and B. Wu, "Mobile game-based learning with a Mobile app: Motivational effects and learning performance," *J. Adv. Comput. Intell. Intell. Informatics*, vol. 21, no. 6, pp. 963–970, 2017.
10. A. Hussain, A. Saleh, A. Taher, I. Ahmed, and M. Lammasha, "Usability evaluation method for mobile learning application using agile: A systematic review," *J. Teknol.*, vol. 77, no. 5, 2015.
11. H. Mohamed and A. Jaafar, "Challenges in the evaluation of educational computer games," in *2010 International Symposium on Information Technology*, 2010, vol. 1, pp. 1–6.
12. T. Hainey, T. M. Connolly, E. A. Boyle, A. Wilson, and A. Razak, "A systematic literature review of games-based learning empirical evidence in primary education," *Comput. Educ.*, vol. 102, pp. 202–223, 2016.
13. T. Susi, M. Johannesson, and P. Backlund, "Serious games: An overview." Institutionen för kommunikation och information, 2007.
14. M. F. Young *et al.*, "Our princess is in another castle: A review of trends in serious gaming for education," *Rev. Educ. Res.*, vol. 82, no. 1, pp. 61–89, 2012.
15. P. Wouters, C. Van Nimwegen, H. Van Oostendorp, and E. D. Van Der Spek, "A meta-analysis of the cognitive and motivational effects of serious games," *J. Educ. Psychol.*, vol. 105, no. 2, p. 249, 2013.
16. S. Hyrynsalmi, T. Mäkilä, A. Järvi, A. Suominen, M. Seppänen, and T. Knuutila, "App store, marketplace, play! an analysis of multi-homing in mobile software ecosystems," *Jansen, Slinger*, pp. 59–72, 2012.
17. J. M. C. Bastien, "Usability testing: a review of some methodological and technical aspects of the method," *Int. J. Med. Inform.*, vol. 79, no. 4, pp. e18–e23, 2010.
18. S. Dray and D. Siegel, "Remote possibilities? International usability testing at a distance," *interactions*, vol. 11, no. 2, pp. 10–17, 2004.
19. K. E. Thompson, E. P. Rozanski, and A. R. Haake, "Here, there, anywhere: remote usability testing that works," in *Proceedings of the 5th conference on*

Information technology education, 2004, pp. 132–137.
20. J. Jeong, N. Kim, and H. P. In, "GUI information-based interaction logging and visualization for asynchronous usability testing," *Expert Syst. Appl.*, p. 113289, 2020.
21. R. Martin, M. A. Shamari, M. E. Seliaman, and P. Mayhew, "Remote asynchronous testing: A cost-effective alternative for website usability evaluation," *Int. J. Comput. Inf. Technol.*, vol. 3, no. 1, pp. 99–104, 2014.
22. K. Saarinen, "Experiences on remote synchronous usability testing with patients: A case cancer care," Aalto University, 2020.
23. T. W. Frick and C. M. Reigeluth, "Formative research: A methodology for creating and improving design theories," *Instr. Theor. Model. A new Paradig. Instr. theory*, vol. 2, pp. 633–652, 1999.
24. A. Alsumait and A. Al-Osaimi, "Usability heuristics evaluation for child e-learning applications," in *Proceedings of the 11th international conference on information integration and web-based applications \& services*, 2009, pp. 425–430.
25. M. Ismail, N. M. Diah, S. Ahmad, N. A. M. Kamal, and M. K. M. Dahari, "Measuring usability of educational computer games based on the user success rate," in *2011 International Symposium on Humanities, Science and Engineering Research*, 2011, pp. 56–60.
26. P. C. Honebein and C. H. Honebein, "Effectiveness, efficiency, and appeal: pick any two? The influence of learning domains and learning outcomes on designer judgments of useful instructional methods," *Educ. Technol. Res. Dev.*, vol. 63, no. 6, pp. 937–955, 2015.
27. T. W. Frick and C. M. Reigeluth, "Verifying instructional theory through analysis of patterns in time," in *Annual Conference of the American Educational Research Association*, 1992, pp. 1–28.
28. W. D. Dwiyogo and C. L. Radjah, "Effectiveness, efficiency and instruction appeal of blended learning model," 2020.
29. J. Nielsen, "Enhancing the explanatory power of usability heuristics," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, 1994, pp. 152–158.
30. C. M. Barnum, *Usability testing essentials: ready, set... test!* Morgan Kaufmann, 2020.
31. M. M. Hassan, M. Tukiainen, and A. N. Qureshi, "(Un)Discounted Usability: Evaluating Low-Budget Educational Technology Projects with Dual-Personae Evaluators," in *Proceedings of the 2019 8th International Conference on Software and Information Engineering*, 2019, pp. 253–258.
32. P. Tsai, "A SURVEY OF EMPIRICAL USABILITY EVALUATION METHODS," 2006. .
33. C. Wilson, *User Interface Inspection Methods*. 2014.
34. M. Alshar'e, A. Zin, R. Sulaiman, and M. R. Mokhtar, "Evaluation of the TPM user authentication model for trusted computers," *J. Theor. Appl. Inf. Technol.*, vol. 81, pp. 298–309, 2015.
35. M. I. Alshar'e, R. Sulaiman, M. R. Mokhtar, and A. MohdZin, "DESIGN AND IMPLEMENTATION OF THE TPM USER AUTHENTICATION MODEL," *J. Comput. Sci.*, vol. 10, no. 11 SE-Research Article, Dec. 2014.
36. K. Alkhatib, A. Al-Aiad, M. Mustafa, and S. Alzubi, "Impact Factors Affecting Entrepreneurial Intention of Jordanian Private Universities Students: A Mediation Analysis of Perception Toward Entrepreneurship BT - Sustainable and Energy Efficient Computing Paradigms for Society," M. A. Ahad, S. Paiva, and S. Zafar, Eds. Cham: Springer International Publishing, 2021, pp. 53–65.
37. K. Venkatesan, S. Nelaturu, A. J. Vullamparthi, and S. Rao, "Hybrid ontology based e - Learning expert system for children with Autism," in *2013 International Conference of Information and Communication Technology (ICoICT)*, 2013, pp. 93–98.
38. A. Silva *et al.*, "Experts Evaluation of Usability for Digital Solutions Directed at Older Adults: a Scoping Review of Reviews," in *9th International Conference on Software Development and Technologies for Enhancing Accessibility and Fighting Info-exclusion*, 2020, pp. 174–181.
39. L. G. Portney and M. P. Watkins, "Statistical measures of reliability," in *Foundations of clinical research: applications to practice*, 2nd ed., New Jersey: Prentice Hall, 2000, pp. 557–584.

40. M. Mustafa, M. Alshar'e, A. Shariah, M. Al-Alawi, and A. Mohammad, "Managing and analyzing factors influencing Saudi college students' entrepreneurial intention during the Covid-19 pandemic," *Turkish J. Physiother. Rehabil.*, pp. 7486–7496, 2021.
41. D. Barclay, C. Higgins, and R. Thompson, *The partial least squares (PLS) approach to casual modeling: personal computer adoption ans use as an Illustration*. 1995.