SYSTEMATIC LITERATURE REVIEW ON USABILITY EVALUATION MODEL OF EDUCATIONAL GAMES : PLAYABILITY, PEDAGOGY, AND MOBILITY ASPECTS

[1]HANIF AL FATTA, [2]ZULISMAN MAKSOM, [3]MOHD HAFIZ ZAKARIA

[1]Faculty of Computer Science, Universitas AMIKOM Yogyakarta, Indonesia

[2,3] Faculty of Information and Communication Technology, Universiti Teknikal Malaysia Melaka, Malaysia

E-mail: [1]hanif.a@amikom.ac.id, [2]zulisman@utem.edu.my, [3]hafiz@utem.edu.my

ABSTRACT

Model for measuring usability evaluation for Mobile Game-Based Learning (m-GBL), has developed by many researchers. Since m-GBL had a unique purpose as a tool for the learning process, the model used for usability evaluation for this kind of application should address those individual characteristics such as playability, mobility, and pedagogy. This study presents the finding of literature review concerning to usability model to measure m-GBL. This study also highlights on the model from another application area which is considered to be relevant. Some papers even presented unique characteristics of measurement concerning to children context as a user. Another critical finding is the expert evaluation, and user testing is the most frequently used method for evaluating usability on m-GBL. And the last result shows some model to develop new dimension for usability evaluation is identified

Keywords: *Usability Evaluation, Playability, M-GBL, Mobility, Pedagogy, Heuristics Evaluation*

1.   INTRODUCTION

Led by the development of the smartphone market, mobile-based applications are proliferating. This also affects the form of Digital Game-Based Learning application, more applications published in the type of mobile app. Two primary operating system, i.e. Android and IOS also play an essential role in the development of mobile applications. Not only primary application developer but also indie developer could start their mobile application, publish it to those platforms and earnings income directly from the customer on Google Play store and Itunes[1]. And the quality of those mobile applications will determine the success of the sales, including educational game mobile apps. One of the things to ensure quality is to evaluate the usability of the product. Usability evaluation will have a significant role to produce a good quality software product and accepted by the user (the market).

An essential step to producing high-quality mobile based applications, especially in the educational game category is to perform usability evaluation. Conducting appropriate Usability evaluation, to ensure the quality of software products is not easy. Many methods and techniques have been proposed starting from following a very concise guideline, to supporting a popular model or framework of usability evaluation.

This literature review is s conducted according to Systematic literature review (SLR) procedure by [2] to find state of the art of the related research. This study reviews selected papers that presented usability evaluation on the m-GBL application. This study also focused on what model that has been used widely in this area and several models that are used in other domain outside m-GBL application area. Also, some papers discussing how to generate usability evaluation criteria based on existing ones are reviewed. And for the last, the methods to create new model also extracted from reviewed papers.

2.   BACKGROUND

2.1 Usability

The different definition of usability exists among the research of Human-Computer Interaction. Usability defined as the capability in human functional terms to be used efficiently, and effectively by a specified range of the user, given specified training and user support, to fulfill the

specified range of task, within a specified range of environment [3]. Nielsen argued [4] usability is usability is about learnability, efficiency, memorability, errors, and satisfaction. While the primary reference for usability definition is derived from ISO 9241-11 – "the extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency, and satisfaction in a specified context of use" [5].

2.2 Usability Heuristics

There are many ways to evaluate the user interface. One way to do this is by heuristically, where one looks directly at a user interface of a system, then provides an assessment based on personal opinion. Most user interface evaluations are based on the evaluation of heuristics using a set of usability heuristics. Heuristics is a usability criterion used to perform an assessment, in which the evaluator selects the criteria as a selected factor measured in a usability evaluation, more based on the "rule of thumb" than specific usability guidelines [6].

Two classical references to the highly popular usability heuristic proposed by [6] and [7]. And also, researchers for more complex and new heuristics determinations continue to be made by later researchers adapting to the proliferation of existing software variations[8].

2.3 Usability Evaluation

As the awareness of quality grows, the need for methods for conducting usability evaluations grows stronger. Usability evaluation methods consist of a set of well-defined processes for collecting information concerning to interaction between a software product and user, and contribution of specific properties of this software product to achieve a certain level of usability can be determined. Usability Evaluation Method helps the researcher to identify usability problem found in the user interface of the software product [9].

Usability evaluation method can be divided into two categories: "inspection methods" (without end users) and "test methods" (with end users). Heuristics Evaluation, Cognitive Walkthrough, and action analysis are very well known methods for implementing inspection methods. However, to get direct information about how a user uses our system and what problem they find during interacting system's user interface, test methods is fundamental usability methods to be employed. In this category thinking aloud, questionnaires and observation are the most

commonly used methods [10]. It is believed that the combination of those two methods can give more accurate and complete result in finding usability problem since there is no exact method considered to be the best method for usability evaluation [11].

2.4 Mobile Digital Game-Based Learning (mGBL)

Game-based learning is innovation in learning approach employing computer games equipped with educational value; it also refers to games software with the educational purpose such as learning support, teaching enhancement, assessment and evaluation of learners [12]. And [13] introduced digital game-based learning, refers to the usage of educational game software for game-based learning. And mobile game-based learning refers to Game-based learning application that runs under the mobile environment. Another term that considered to be the same concept is Mobile game for learning or mobile educational game (Game-Based Mobile Learning- GBML), a game utilized explicitly for learning which is also played on a mobile phone, smartphone, PDA or handheld devices [14].

3. RELATED WORK

Research on literature review concerning to usability evaluation has been done by several researchers. A review conducted by [8], using 75 papers to be reviewed, attempted to determine usability heuristics for usability evaluation and presented the methodology for creating new usability heuristics set. The main results of this research were: (1) there are several ways in creating the heuristic for usability: existing heuristics, Methodologies, literature reviews, usability problems, mixing processes, guideline/principle recommendation, interviews or theory. (2) Guideline for the researcher who wants to create heuristics: to determine specific feature, to identify existing usability heuristics, to specify a new set of heuristics based on standard, and to validate the new sets of heuristics. However, research focused on several domains such as games, mobile, system, websites, applications, interfaces, computers, and Learning but not specifically on the mGBL domain and only cites four papers on learning category.

Other research presented systematic mapping studies on serious games. The result from examination of 112 papers on defining the quality of serious games revealed the importance of usability. The characteristics of the quality product

model most frequently researched were usability (45.54%) distantly followed by functional suitability (8.93%) [15]. This study addressed the importance of usability as a quality factor. However, heuristics criteria and methods for conducting usability evaluation has not been presented.

Another research conducted by [16], reviewed usability evaluation on the mobile application. This paper is considered to be relevant since the mGBL also operates on the mobile environment. PACMAD usability model has proposed consists of three factors that affect usability evaluation in a mobile application: User, task and Context of use, and usability heuristics evaluation used in this model are Effectiveness, Efficiency, Satisfaction, Learnability, Memorability, Errors, and Cognitive load. A literature review has been conducted focused on papers published between 2008 – 2010. This review has highlighted the extent to which the attributes of the PACMAD model are considered within the mobile application domain. Nevertheless, this research focused on mobile application in general, didn't address usability factor for educational or learning content.

4. RESEARCH QUESTION

This literature review conducted according to Systematic literature review (SLR) procedure by [2] to find state of the art of the related research. This study of literature review has a major purpose in finding state of the art on the model used in usability evaluation for game-based learning application. And literature review analysis is conducted focused on several analysis questions (AQ) presented as follows:

AQ1. What usability evaluation models are available to evaluate mGBL?.

AQ2. What testing method is used to perform the usability evaluation in mGBL?

AQ3. What methods are used to determine or established heuristics for usability evaluation model for mGBL?

5. RESEARCH METHODS

5.1 Data Source

This study searched papers from selected database source that relevant to information technology, computer science, and engineering, including ScienceDirect, ACM Digital Library,

IEEE Xplore, Springer, Taylor and Franchise, and Emerald Insight.

5.2 Search Term

In this literature review the main search keywords used are the terms "usability evaluation" or "usability testing" combined with "model" and "mobile game-based learning." Besides, the keywords "usability test" and "educational games," "framework" are also used. For example, we used : ("usability testing" or "usability evaluation") and model and ("mobile educational games" or "mobile game-based learning" or mGBL), we also tried to search using

("usability testing" or "usability evaluation") And framework and ("mobile educational games" or "mobile game-based learning" or mGBL). Table 1 depicted the possible combination of search keyword to find relevance papers.

Table 1: Major Search Term And Alternative Terms

Inclusion Criteria	1. Papers that presented usability evaluation in the mobile application, educational games, mobile educational games, educational games for children and mobile learning.
	2. Papers that showed usability criteria with or without specifying the specific model.
	3. The paper that proposed methods to create new usability heuristics and how to validate them.
Exclusion Criteria	7. The Papers which does not focus on usability as mentioned in inclusion criteria for example usability for website, toys, business apps.
	8. Papers available only in the abstract
	9. Theses that have not been published.

The purpose of this SLR is to find out all papers about usability evaluation for mobile educational games, written in English, published until the end of 2017. Because we want to know since when usability evaluation model for mobile educational games research has existed, the start of publication period was not specified. The distribution of 26 papers containing usability evaluation model over the years is presented in figure 1.

Figure 1: Paper Distribution Of 26 Papers With M-GBL Usability Evaluation Themes Over Years

The selection process of the papers was done through several stages. The first stage, the papers generated from the search by keyword were analyzed by examining the title, abstract and keyword of the author. Articles related to usability evaluation in mobile educational games were then selected. The next step was to read the entire contents of the paper that has been selected from the previous stage. Inclusion and exclusion criteria for selecting documents are shown in table 2.

Table 2: Inclusion And Exclusion Criteria

No	Search String	
	Major Term	Alternative term
1	Usability Evaluation	("usability testing" or "usability test")
2	Mobile game based learning	("mobile educational game" or "mGBL" or " mobile learning")
3	Model	(framework or guideline)

5.3 Selection process

After conducting the reading process on the abstract and keyword from papers that appear from the database search, so far, until this research written by the author, 212 relevant articles were obtained. Of the 212 papers, the search continued by reading the entire contents of the document and found 30 papers considered the most pertinent.
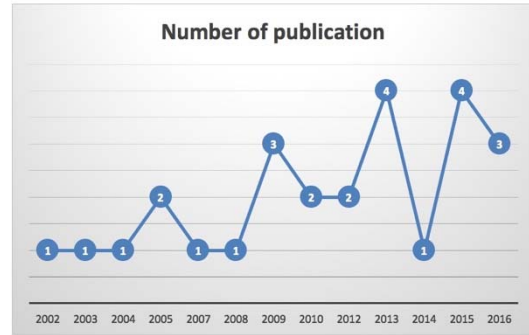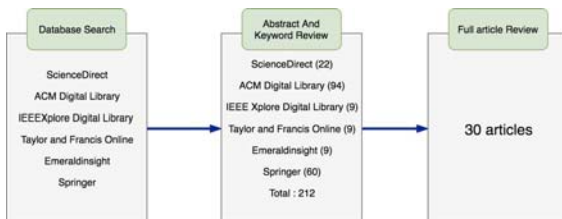

Figure 2. The Flow Of Paper Search

During conducting literature review 26 full reviewed paper will be used for analyzing AQ1 and AQ2, and more four papers will be added for finding the methodology to define the new set of heuristics in usability evaluation (AQ3).

The primary objective of this study is to identify the heuristics used to perform usability evaluation on the m-GBL application and to reveal some methods used to improve or augment usability heuristics from prior studies. All These papers are selected based on their content that is considered to be related to those two objectives. The first 26 papers are chosen because they showed the used of heuristic usabilities-from famous model or their modified model-concerning to m-GBL and other related application domains. These papers are critical literature to identify what has been explored by prior studies so far and what gap available to improve the quality of usability evaluation in term of defining the appropriate heuristics dimension. The last three papers indicated methods to improve or enhance available usability evaluation model. They were essential studies since their model can be used to perform new ways of strengthening existing evaluation model.

5.4 Findings

5.4.1 Usability Evaluation Models to Evaluate m-GBL

During conducting the literature review for finding prominent models affecting the mGBL usability evaluation, the author believed that proposed usability evaluation model for MGBL would be influenced by existing usability evaluation model for mGBL, mobile application, e-learning application, and game for children application.

The conceptual relation of those these models can be shown in figure 3. And ongoing literature review will be conducted to address those issues.

During literature review for answering AQ1, 26 papers were obtained concerning to evaluation on mGBL Distribution of reviewed journal based on application category/area is shown in figure 3. This study examined four categories of writing, that is considered to be relevant to the m-GBL application. They are the m-GBL application itself, games application, mobile application and Child-Elearning application.

During conducting the literature review, this study focussed on four categories of usability dimension namely usability interface, playability, pedagogy, and Mobility. All reviewed papers are examined to determine if those dimensions

Figure 3: The Proportion Of Reviewed Paper Based On Area/Application In Which Usability Evaluation Is Conducted

discussed in the reviewed papers. A paper can present more than one dimension. From the analysis of heuristic used on usability evaluation in each paper, the heuristics than categories into usability interface, playability, pedagogy, or Mobility dimension. These dimensions are believed having a high impact on measuring usability evaluation. This dimension could be found partially or entirely across the reviewed papers as depicted in figure 5.

After defining four categories of usability dimension, the next step is to identify the heuristics criteria used by selected papers and map them into an appropriate dimension. For an example how mapping process is conducted, a study from [17] is taken. This study proposed four aspects of usability evaluation comprising: Gaming experience, Learning experience, Adaptivity, and Usability.

Each dimension contains more specific heuristics as follows :

a. Gaming Experience: Challenge, competence, flow, immersion, negative affect, positive affect, and tension.

b. Learning experience: Learning goal, content appropriateness, integration, feedback, extensibility, and media matching

c. Adaptivity: Cognitive and motivational intervention, game pacing

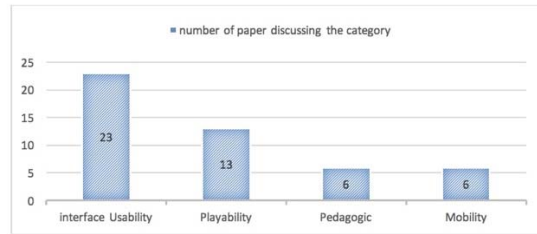d. Usability: interface and interaction.

Figure 5: Number Of Paper Discussing Interface Usability, Playability, Pedagogic And Mobility Dimension

Table 3: Usability Dimension Across Reviewed Papers

No	Year	Model	IU	Pla.	Ped.	Mob.
M-GBL Application						
1	2012	[18]	v	v	v	
2	2014	[19]	v	v	v	
3	2015	[20]	v	v	v	
4	2012	[17]	v	v	v	
5	2003	[21]	v			
6	2015	[22]	v	v	v	v
7	2009	[23]	v	v		
8	2010	[24]	v			
9	2013	[25]		v		
Game Application						
10	2004	[26]	v	v		
11	2005	[27]	v	v		
12	2008	[28]	v			
13	2002	[29]		v		
14	2009	[30]	v			
15	2015	[31]	v	v		
16	2005	[32]		v		
17	2007	[33]	v	v		v
Mobile application						
18	2016	[34]	v			
19	2013	[35]	v			
20	2015	[36]	v			
21	2016	[37]	v			v
22	2016	[38]	v			v
23	2009	[39]	v			v
24	2013	[40]	v			
25	2013	[16]				
Children E-learning						
26	2010	[41]	v	v	v	

IU   : Interface Usability
Pla.   : Playability
Ped.   : Pedagogy
Mob.   : Mobility

And finally, the mapping process is executed. From those four dimensions, usability will be mapped into interface usability dimensions in this study. Learning process and adaptivity will be outlined in pedagogy dimensions. And gaming experience is mapped into playability dimension. So study from [17] lacks in mobility dimensions. Complete mapping of those 26 paper is presented in table 3.

From the paper reviewed above, several models can be used exclusively to measure usability evaluation for mobile educational games. These models determined usability dimension that can be used to measure usability in term of interface usability, playability, mobility, and portability. However, this dimension did not present completely across all models.

a. m-GBL usability evaluation model

The model that proposed complete dimensions for measuring usability in m-GBL is proposed by [22]. Zaibon argued that for measuring usability for mGBL four issues must be taken into consideration; they are interface Usability, Mobility, Playability and learning content. Zaibon also proposed usability heuristics for conducting usability as depicted in table 2.4. It seems that Zaibon's model already proposed complete heuristic dimensions for usability evaluation for mGBL. However, Zaibon's model did not address unique characteristic of children learning and children behavior. Because how children learn is different from adults, the heuristics for mGBL for children should take attributes of children learning into account.

Other good examples of usability evaluation models for mGBL are USAECG [18] , UGALCO [19], PREMEGA [20] and Law & Sun Model [17]. These models defined good heuristics for interface usability, playability and learning content but these models were not built specifically for the mobile environment, so it lacks heuristics for mobility. [23] used two dimensions: interface usability and playability, while   [24] only focused on interface usability and [25] used game-flow model [32] for assessing playability for mGBL.

b. Game usability evaluation model

Since m-GBL is the subset of game `application, the basic characteristic of games–playability-will be a fundamental issue concerning to usability evaluation for m-GBL. Playability played importance rules for child learning because of the natural ways for children to learn through experience. The environment for children during using and evaluating the application, are very different from adult's. And this issue is essential during designing application for children [42]. So in this section, we will discuss several models concerning to how to assess usability for a game application. And hopefully, heuristics used to measure usability for the game application can be added to author's proposed model, especially on playability-for-children dimension.

An evaluation model for playability that put emphasize on children characteristics can be found on SEEM model [27]. SEEM model proposed predictive evaluation model for evaluating children computer games based on Norman's Theory of Action and Malone's Fun Concept [43]. This predictive test model is cheap to apply, and they can be conducted in the early process of design, as soon as the game is finished. Heuristics to measure usability and playability for children's computer games that address children's characteristics were well defined, however pedagogic and mobility issues weren't available for this model.

Another model is proposed by [29], explicitly measuring fun factor in children games. This model can be used as a reference for the model being proposed by the author because fun factor is a critical aspect in determining the quality of m-GBL designed for children.

c. Mobile usability evaluation model

Since M-GBL will run under the mobile environment, usability evaluation model for these applications will affect the proposed m-GBL model. An exciting model called USERBILITY model [34] argued that acceptance of the mobile app is influenced by the quality of UX as well as usability. Userbility proposed a model for evaluating UX and usability for mobile application. In assessing usability, the Userbility model proposed ten heuristics that appears very similar to Neilsen's model [4]. So, although Userbility was designed for the mobile application, It used generic heuristics and did not address the distinctive characteristics of the mobile environment. For UX evaluation, Userbility used 3E (Expressing Emotions and Experiences) model [44], and this model also a

generic model for measuring UX, not specially designed for the mobile application.

Another model, also based on Nielsen's model called SMASH, proposed by [37]. This model extended 11 heuristics from Nielsen by adding one heuristic: Physical interaction and ergonomics that addressed the characteristics of the mobile environment. This model could detect more usability problem, compared to the usage of Nielsen Model during usability evaluation.

A better framework for evaluating usability for the mobile environment is proposed by [39]. Heo et al. developed a framework for evaluating the usability of the mobile phone, not only the application that runs under the mobile environment but also the physical device of the mobile phone itself will affect the usability of the user. Several checklists to address usability problem is developed: Logical User Interface Checklist, Physical User Interface Checklist, and Graphical user Interface checklist. They are a suitable checklist for evaluating usability on mobile devices.

Other reviewed mobile application model is presented in table 2.6. Heuristics that measure mobility dimensions proposed in those models can be added to author's m-GBL proposed model for addressing mobility usability problems found on the m-GBL application.

d. Children E-learning usability evaluation model

In the development of usability evaluation model for m-GBL for the primary school, one important factor to consider is the dimension of usability for children and learning or pedagogy dimension. Therefore, studying the evaluation model for child E-learning will give essential benefits. One useful model for measuring the dimensions of playability for children as well as aspects of pedagogy named HECE, developed by [41]. To measure usability of the interface HECE used Nielsen models, and add Child usability to measure usability addressing unique characteristics of children. This model also developed a heuristic of the pedagogy to measure the usability problems that arise from the aspect of the delivery of learning material to children.

5.4.2   Testing Method To Perform Usability Evaluation In m-GBL

During evaluating usability on m-GBL, several evaluation methods can be used. Before discussing the methods used to conduct the evaluation, it is better to consider some terms related to usability evaluation. First, the author will discuss the use of the words usability testing and heuristics evaluation. The use of these terms emerged from the 1990s, with at least two studies - [45] and [46], tried to conduct the comparison between those two methods. Heuristic evaluation is an evaluation that conducted by evaluators (" expert") by inspecting or examining an interface and trying to give the opinion about what is good or bad about the interface [6]. During heuristics evaluation, experts/evaluators used certain rules or guidelines called " heuristics" to find any usability problem of an interface. At the end of the evaluation, evaluators came out with several recommendations to improve the usability of the evaluated system. Usability testing, on the other hand, involving representative users to test specific task in the specific environment. It is a comprehensive definition of usability testing includes testing prototype, screen mock-up, screen layout, " wizard of oz" technique and working version before officially released as well as already implemented version [47]. As stated by Nielsen, Usability testing is unique methods since it captured the genuine information on how user use the product and what problem they encountered while testing the interface. And this made Usability testing is the most fundamental methods for evaluating usability [4].

According to [48], usability evaluation methods can be divided into three categories: testing, inspection, and inquiry. In Usability testing, selected users use evaluated system or interface by testing chosen tasks and users give feedback on how useful interface/system on supporting users to reach their goals.

In Usability Inspection, usability expert/specialist checked the system/interface to find usability problems. And in Usability inquiry, usability specialists gather pieces of information from the user about user's understanding, like/dislike and experience by questioning users, observing users, or obtaining information verbally or in written form. Complete details of methods used in each usability evaluation methods are shown in figure 6. Among other reviewed papers, most common methods used by researchers is Usability Testing (User Testing), followed by Usability Inspection especially Heuristic Evaluation, accordance to [49] and some research used both usability testing and Heuristic Evaluation as seen on figure 4. Complete methods used in the reviewed paper is described in table 4.

5.4.3 Methods To Established Heuristics For Usability Evaluation Model For m-GBL

Methods for determining heuristics developed slowly for the last decade. A model proposed by [49] mentioned six steps for heuristics creation: Basic sets of heuristic, analysis of
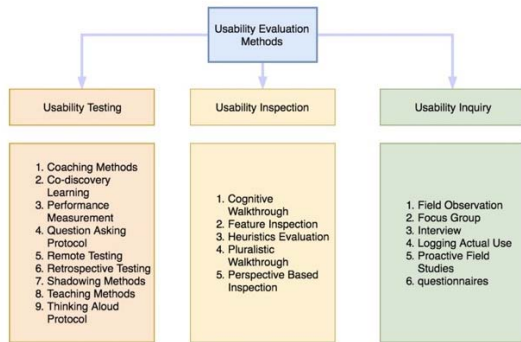

Figure 6: Usability Evaluation Methods Based On Zhang [48]

environment, preparation of heuristic creation, heuristic creation, evaluation of heuristics and final set of heuristics. This research also argued that methodology for developing a new model of usability should meet requirements as follows: Individual activities involved in developing new model should be systematically organized. A solved situation must be simply described. All part of the methodology should be homogenous. All fact is measurable; methodology result must be objective and free of ambiguity.

As the necessity of usability evaluation grew, it is essential to have usability heuristics for the specific application. [50] proposed a methodology for defining new usability heuristics for a specific application, comprising six steps : (a) exploratory stage to gather bibliography accordance to research: particular applications, characteristics of those applications, general or specific heuristics (if available). (b)Descriptive stage: choosing the most prominent characteristics of the formerly gathered information. (c) Correlational stage: definition of heuristics for the specific application based on traditional heuristics as well as case study analysis. (d) Explicative stage: the standard template for a set of proposed heuristics is formally specified. (e) Validation (Experimental) Stage: Experiment is conducted to compare new heuristics against traditional heuristics, using heuristics evaluations followed by user testing (usability testing). (f) Refinement stage, Heuristics from step 4 is refined

based on feedback. Iterative steps can be applied to stage 1 to stage 6, to find better heuristics. This model is the most frequently used methodology for generating heuristics set according to research conducted by [8]

Both [49] and [50] used user-arguably most crucial stake holder-as evaluator to verify the validity of proposed heuristic set, but they did not get involved in the development of the chosen heuristics.

Table 4: `Distribution Of Usability Evaluation Method

No	Year	Model	Usability Evaluation Methods.		
			UT	HE	FGD
1	2012	USAECG[18]		√	
2	2014	UGALCO[19]	√	√	
3	2015	PREMEGA[20]			
4	2012	[17]	√		
5	2003	[21]			
6	2015	[22]			
7	2009	[23]	√		
8	2010	[24]	√		
9	2013	[25]	√	√	

UT   : User Testing
HE   : Heuristic Evaluation
FGD   : Focus Group Discussion

A model that considered user contribution is valuable in adapting existing heuristics into the new set of heuristics is proposed by [51]. This research suggested two steps of heuristics validation: (1) Phase One: Usability experts participated in focus group discussion conducting two activities, rating existing heuristics and re-word-ing heuristics. (2) Phase 2: Real user was participating in rating existing heuristics, developing and validating new heuristics. At the end of the process, comparative analysis is carried out to compare heuristics rating performed by experts and real users.

Other research by [52] proposed methodology to established heuristics for the specific application by modifying generic heuristics. The method consisted of 2 main stage : (1) how to extend generic heuristic for particular heuristics and (2) how to validate those heuristics. In stage 1, most commonly used existing heuristics is mapped from documented resources on usability issues and design guidelines. In stage 2, validation conducted by expert and user to measure

thoroughness, validity, and effectiveness. In stage 1, the real user could get involved in the generation of heuristics by providing usability issues that can be documented by researchers. This usability issues, and then, can be completed by adding usability problem defined from guidelines or references.

6.   DISCUSSION
6.1 The Pros And Cons Of The Study
After Analyzing the answer of AQ1, AQ2, and AQ3 from the previous section, it was evident that this study is succeeding in finding small space among available model to be improved. Table 3 revealed 4 popular usability dimensions appeared on papers in selection, they are : Interface Usability, Playability, Pedagogy, and Mobility. However, dealing with educational application meant to be used by children, it is necessary to add one more dimension-children characteristic-to address uniqueness of children learning.   So the contribution of this literature review is discovering children learning dimension, augmenting the existing model to deliver more precise usability measurement on m-GBL for children. Other findings showed the famous usability evaluation methods employed in m-GBL usability testing. This study also described several methods for enhancing available usability model. These findings are essential for further research to defined model for enhancing children learning dimension into existing usability evaluation model and analyze the quality of a new proposed model.

However, this study still has a lot of things to do in the future. First, the definition of usability heuristics reflecting measurement on children learning characteristics has not been developed yet. A further systematic literature review is highly needed to discuss this part. Secondly, to conduct usability evaluation efficiently, a measurement metrics can be defined and employed. So another future work is to identify and to map papers containing usability measurement metrics suitable for m-GBL with children user.

6.4. Limitations
This study is also conducted within several limitation. The limitation is related to time periods of papers in selection's publication, the search term choosed and the selected database.

However, the selected paper in the current review provided contents of (a) usability heuristics used in m-GBL and related application, (b) the evaluation methods to conduct usability evaluation, and (c) the

methods to construct or modify new model for usability evaluations.

The most appropriated search term were employed to select suitable papers and prominent research database were taken as the source of the selected papers. And time period of paper publishing is within 2002 – 2016.

And finally, exclusion criteria is done by excluded the papers which does not focus on usability as mentioned in inclusion criteria for example usability for website, toys, business apps, papers available only in the abstract and theses that have not been published.

7.   CONCLUSION

After completing the full paper review, it is obvious to state that only a few models available for usability evaluation for m-GBL primarily designed for children. There are several issued that can be highlighted as follows :

1.   The model considered to be " best" according to the complexity of its dimension is proposed by [22], since Zaibon's model already proposed complete heuristic dimensions for usability evaluation for mGBL. However, Zaibon's model didn't address unique characteristic of children learning and children behavior, so there is a "small space" for the author to add children usability dimension for the model to address more usability issues as proposed by [27].

2.   Another valuable finding is that generation of heuristics for the specific application (m-GBL) can be derived from augmentation of traditional heuristics models-e.g. Nielsen's Model-as proposed by several researchers: [18], [21], [34] and [37].

3.   Generation of heuristics for m-GBL learning can be done by reviewing existing m-GBL model, but regarding the characteristics of m-GBL that also inherits mobility characteristics, playability from games characteristics and most notable characteristics, pedagogy, it is critical to review heuristics employed to measure usability on those applications.

4.   From reviewed papers, it is also shown that the most frequently used methods for conducting usability evaluation is user testing and heuristic evaluation and sometimes the combination of both methods is applied to the comparative analysis.

5.   And the last finding is about the methodology to develop the new model for usability evaluation. Model by [50] seems to be the

complete model to follow since each step of its methodology can be followed quickly, but in can be perfected by adding user involvement during heuristics generation stage as stated by [51] and [52]

REFERENCES:

[1]   M. Y. Ricky, B. Yulianto, and F. Purnomo, "Mobile Application Software Defect Prediction," pp. 307–313, 2016.

[2]   B. Kitchenham *et al.*, "Systematic literature reviews in software engineering-A tertiary study," *Inf. Softw. Technol.*, vol. 52, no. 8, pp. 792–805, 2010.

[3]   B. Shackel *et al.*, "Human Factors for Informatics Usability," 1st ed., S. B. and R. S.J, Eds. Cambridge: Cambridge University Press, 1991, pp. 24–25.

[4]   J. Nielsen, *Usability Engineering*. San Fransisco: Morgan Kauffmann Publisher, 1993.

[5]   ISO/IEC, "Ergonomic requirements for office work with visual display terminals (VDT)s - Part 14 Menu dialogues9241-14: 1998," 1998. [Online]. Available: https://www.iso.org/obp/ui/#iso:std:iso:9241:-14:ed-1:v1:en:sec:A.

[6]   R. Molich and J. Nielsen, "HEURISTIC EVALUATION OF USER INTERFACE," in *CHI 1990*, 1990, no. April, pp. 249–256.

[7]   B. Shneidermann and C. Plaisant, *Designing The User Interface : To Strategies For Effective Human-Computer Interaction*, 4Th ed. Pearson Education, 2005.

[8]   D. Quiñones and C. Rusu, "How to develop usability heuristics: A systematic literature review," *Comput. Stand. Interfaces*, vol. 53, pp. 89–122, 2017.

[9]   A. Fernandez, E. Insfran, and S. Abrahão, "Usability evaluation methods for the web : A systematic mapping study q," *Inf. Softw. Technol.*, vol. 53, no. 8, pp. 789–817, 2011.

[10]   A. Holzinger, "Usability Engineering Methods for Software Developers," *Commun. ACM*, vol. 48, no. 1, pp. 71–74, 2005.

[11]   F. Paz, F. A. Paz, D. Villanueva, and J. A. Pow-Sang, "Heuristic Evaluation as a Complement to Usability Testing: A Case Study in Web Domain," in *2015 12th International Conference on Information Technology - New Generations*, 2015, pp. 546–551.

[12]   S. Tang, M. Hanneghan, and A. El Rhalibi, "Introduction to Games-Based Learning," in *Games-Based Learning Advancements for Multi-Sensory Human Computer Interfaces : Techniques and Effective Practices*, 1st ed., K. Klinger, Ed. IGI Global, 2009, pp. 1–17.

[13]   M. Prensky, *Digital Game-Based Learning*, 1st ed. New York: McGraw Hill, 2001.

[14]   S. Šćepanović, "GAME BASED MOBILE LEARNING – APPLICATION DEVELOPMENT AND EVALUATION Mobile usability," no. September, pp. 24–25, 2015.

[15]   J. A. Vargas, L. García-Mundo, M. Genero, and M. Piattini, "A systematic mapping study on serious game quality," *CEUR Workshop Proc.*, vol. 1196, pp. 52–57, 2014.

[16]   R. Harrison, D. Flood, and D. Duce, "Usability of mobile applications: literature review and rationale for a new usability model," *J. Interact. Sci.*, vol. 1, no. 1, p. 1, 2013.

[17]   E. L. Law and X. Sun, "Evaluating user experience of adaptive digital educational games with Activity Theory," *Int. J. Hum. Comput. Stud.*, vol. 70, no. 7, pp. 478–497, Jul. 2012.

[18]   H. Mohamed-Omar, R. Yusoff, and A. Jaafar, "Quantitive analysis in a heuristic evaluation for usability of Educational Computer Game (UsaECG)," in *2012 International Conference on Information Retrieval & Knowledge Management*, 2012, pp. 187–192.

[19]   D. C. C. Peixoto, R. F. Resende, and C. I. P. S. Padua, "Evaluating software engineering simulation games: The UGALCO framework," *2014 IEEE Front. Educ. Conf. Proc.*, pp. 1–9, 2014.

[20]   L. Shoukry, C. Sturm, and G. H. Galal-Edeen, "Pre-MEGa: A Proposed Framework for the Design and Evaluation of Preschoolers' Mobile Educational Games," in *Innovations and Advances in Computing, Informatics, Systems Sciences, Networking and Engineering*, Springer International Publishing, 2015, pp. 385–390.

[21]   G. D. Magoulas, S. Y. Chen, and K. A. Papanikolaou, "Integrating layered and heuristic evaluation for adaptive learning environments," in *Proceedings of the Second Workshop on Empirical Evaluation*

of Adaptive Systems, held at the 9th International Conference on User Modeling UM2003, 2003, no. June 2014.

[22] S. B. Zaibon, "USER TESTING ON GAME USABILITY, MOBILITY, PLAYABILITY, AND LEARNING CONTENT OF MOBILE GAME-BASED LEARNING," *J. Teknol.*, vol. 77, no. 29, pp. 131–139, Dec. 2015.

[23] W. S. Yue and N. A. M. Zin, "Usability evaluation for history educational games," in *Proceedings of the 2nd International Conference on Interaction Sciences Information Technology, Culture and Human - ICIS '09*, 2009, pp. 1019–1025.

[24] J. Sánchez, M. Saenz, and J. M. Garrido, "Usability of a Multimodal Video Game to Improve Navigation Skills for Blind Children," *ACM Trans. Access. Comput.*, vol. 3, no. 2, pp. 1–29, 2010.

[25] K. Khanana and E. L. Law, "Designing Children ' s Digital Games on Nutrition with Playability Heuristics," in *CHI 2013: Changing Perspectives, Paris, France*, 2013, pp. 1071–1076.

[26] H. Desurvire, M. Caplan, and J. A. Toth, "Using heuristics to evaluate the playability of games," in *Extended abstracts of the 2004 conference on Human factors and computing systems - CHI '04*, 2004, p. 1509.

[27] Ester Baauw, M. M. Bekker, and Wolmet Barendregt, "A Structured Expert Evaluation Method for the Evaluation of," in *Human-Computer Interaction – INTERACT*, M. F. Costabile, Ed. rome: Springer International Publishing, 2005, p. 457.

[28] D. Pinelle, N. Wong, and T. Stach, "Heuristic evaluation for games: usability principles for video game design," *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, pp. 1453–1462, 2008.

[29] J. Read, S. Macfarlane, and C. Casey, "Endurability , Engagement and Expectations : Measuring Children ' s Fun," *Interact. Des. Child.*, vol. 2, no. May 2014, pp. 1–23, 2002.

[30] H. Desurvire and C. Wiberg, "Game usability heuristics (PLAY) for evaluating and designing better games: The next iteration," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 5621 LNCS, pp. 557–566, 2009.

[31] H. Desurvire and C. Wiberg, "User Experience Design for Inexperienced Gamers: GAP—Game Approachability Principles," in *Game User Experience Evaluation, Human-Computer Interaction Series*, 2015, pp. 169–186.

[32] P. Sweetser and P. Wyeth, "GameFlow: A Model for Evaluating Player Enjoyment in Games," *Comput. Entertain.*, vol. 3, no. 3, pp. 3–3, 2005.

[33] H. Korhonen and E. M. I. Koivisto, "Playability heuristics for mobile multi-player games," in *Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts - DIMEA '07*, 2007, p. 28.

[34] I. Nascimento, W. Silva, B. Gadelha, and T. Conte, "Userbility: A Technique for the Evaluation of User Experience and Usability on Mobile Applications Ingrid," *HCI 2016, Part I*, vol. 9731, pp. 372–383, 2016.

[35] J. Park, S. H. Han, H. K. Kim, Y. Cho, and W. Park, "Developing Elements of User Experience for Mobile Phones and Services: Survey, Interview, and Observation Approaches," *Hum. Factors Ergon. Manuf. Serv. Ind.*, vol. 23, no. 4, pp. 279–293, Jul. 2013.

[36] R. Upitis, P. Abrami, J. Brook, D. Pickup, and L. Johnson, "Usability Testing for DREAM," *Procedia - Soc. Behav. Sci.*, vol. 171, pp. 543–552, 2015.

[37] R. Inostroza, C. Rusu, S. Roncagliolo, V. Rusu, and C. A. Collazos, "Developing SMASH: A set of SMArtphone's uSability Heuristics," *Comput. Stand. Interfaces*, vol. 43, pp. 40–52, 2016.

[38] K. Moumane, A. Idri, and A. Abran, "Usability evaluation of mobile applications using ISO 9241 and ISO 25062 standards.," *Springerplus*, vol. 5, no. 1, p. 548, Dec. 2016.

[39] J. Heo, D. H. Ham, S. Park, C. Song, and W. C. Yoon, "A framework for evaluating the usability of mobile phones based on multi-level, hierarchical model of usability factors," *Interact. Comput.*, vol. 21, no. 4, pp. 263–275, 2009.

[40] W. Brown, P. Y. Yen, M. Rojas, and R. Schnall, "Assessment of the Health IT Usability Evaluation Model (Health-ITUEM) for evaluating mobile health (mHealth) technology," *J. Biomed. Inform.*, vol. 46, no. 6, pp. 1080–1087, 2013.

[41] A. A. Alsumait and A. Al-Osaimi, "Usability Heuristics Evaluation for Child E-learning Applications," *J. Softw.*, vol. 5, no. 6, pp. 654–661, Jun. 2010.

[42] A. Druin, B. Bederson, A. Boltman, A. Miura, D. Knotts-callahan, and M. Platt, "Chapter 3 : Children as Our Technology Design Partners +," no. Age 8, pp. 1–9, 1998.

[43] T. V Malone, "What Makes Things Fun to Learn? Heuristics for Designing Instructional Computer Games," in *SIGSMALL '80 Proceedings of the 3rd ACM SIGSMALL symposium and the first SIGPC symposium on Small systems*, 1980, pp. 162–169.

[44] M. Tähti and M. Niemelä, "3E-Expressing emotions and experiences," *WP9 Work. Innov. Approaches Eval. Affect. Syst.*, 2006.

[45] J. Prescott and M. Crichton, "USABILITY TESTING VS. HEURISTIC EVALUATION: A HEAD-TO-HEAD COMPARISON," in *Proceedings of the 27th annual ACM SIGUCCS conference on User services Mile high expectations - SIGUCCS '99*, 1999, pp. 176–179.

[46] R. Jeffries and H. Desurvire, "Usability Testing Vs Heuristics Evaluation : Was there a contest?," *SIGCHI Spec. Interes. Gr. Comput. Hum. Interact.*, vol. 24, no. 4, pp. 36–41, 1992.

[47] J. Lazar, J. H. Feng, and H. Hochheiser, *Research Methods in Human-Computer Interaction*, Second edi. Chichester, UK: Morgan Kauffmann Publisher, 2017.

[48] Z. Zhang, "Overview of usability evaluation methods. http://www.usabilityhome.com (retrieved February 14th, 2018)," *http://www.usabilityhome.com*, 2018. .

[49] M. Hub and V. Čapková, "Heuristic evaluation of usability of public administration portal," *Proc. Int. Conf. Appl. Comput. Sci.*, pp. 234–239, 2010.

[50] C. Rusu and S. Roncagliolo, "A methodology to establish usability heuristics," *ACHI 2011, Fourth …*, no. c, pp. 59–62, 2011.

[51] B. Lechner, S. Petter, A. Fruhling, and H. Siy, "The chicken and the pig: User involvement in developing usability heuristics," *19th Am. Conf. Inf. Syst. AMCIS 2013 - Hyperconnected World Anything, Anywhere, Anytime*, vol. 5, pp. 3263–3270, 2013.

[52] S. Hermawati and G. Lawson, "A user-centric methodology to establish usability heuristics for specific domains," *Contemp. Ergon. Hum. Factors 2015*, no. April, pp. 80–85, 2015.
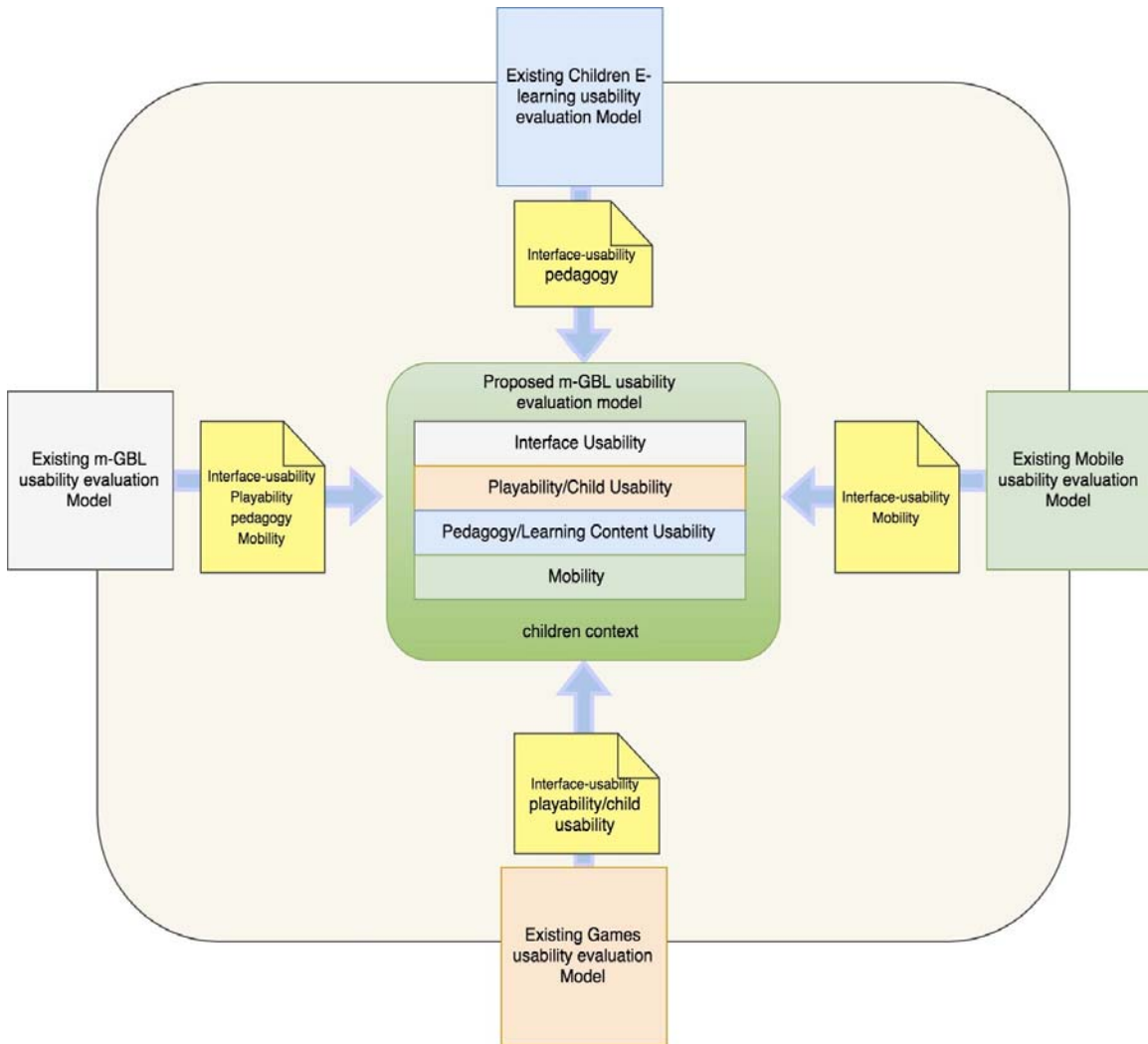
Figure 4: Generation Of Proposed Model Affected By Other Existing Models