State of the art in Game Based Learning: Dimensions for Evaluating Educational Games

Rabail Tahir and Alf Inge Wang
Dept. of Computer Science, Norwegian University of Science and Technology, Norway
rabail.tahir@idi.ntnu.no
alfw@idi.ntnu.no

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

Keywords: game based learning, educational games, evaluation framework, evaluation dimensions

1. Introduction

Computer games are played for many reasons including enjoyment, entertainment as well as educational purposes. Computer games prove to be an effective educational tool as they can provide enjoyment to learners in the learning process, allowing them to engage in education while having fun (Mohamed and Jaafar, 2010). Educators began to acknowledge the power of computer games for educational purposes back in 1980s. Games that embody educational objectives are considered to make learning more enjoyable, interesting, fun and more learner-centered thus making it more effective (Wang et al., 2011). The term GBL refers to the use of computer games or software applications that utilize games, for educational or learning purposes. Now days, mobile phones have also been widely used for GBL under the label "mobile game-based learning". Although traditional practices for education remain in use, GBL has extensively been implemented in various courses (Alfadhli and Alsumait, 2015).

The rapid increase in the use of educational games makes it essential to verify these tools to provide the learners with a suitable learning environment. This verification is made through the evaluation of these tools from multiple dimensions (De Freitas and Oliver, 2006). However, the study of games has dearth of a consistent research paradigm. To realize the potential of educational games, there is a need to have a scientifically sound approach to evaluate their effectiveness. (Eagle, 2009). Most articles on educational games emphasize on whether to use games for learning or to explore the potential of games in providing effective learning (Wang et al.,2011). The study on educational games evaluation are deficient. What is missing in GBL literature is the dearth of empirical evidence on the validity of the approach (Wang et al.,2011). Generally, the development of any software specifically educational games is a very demanding process involving costly resources and time. Therefore, evaluation is vital to remove imperfections and improve efficiency.

Mohamed and Jaafar (2010) highlight three challenges in the evaluation of educational computer games. These challenges are evaluation criteria, evaluation process and the evaluators. Some researchers (Petri and von Wangenheim, 2016) have examined the evaluation process for educational games. Mostly in the past evaluation of educational games use questionnaires, observations, log files or interviews in an ad-hoc manner without defining any criteria for evaluation. However, there were very few attempt in developing a framework that specify criteria to evaluate educational games and research to explore this area is deficient (Mohamed and Jaafar, 2010). There is a need to define the aspects of educational games (Ak, 2012). The process for identifying criteria is more complex and time-consuming than one would think (Dondi and Moretti, 2007).

This research work try to fill this gap by carrying out a SLR specifically focusing only on the evaluation of educational games and identification of the key dimensions for evaluating GBL. The paper is organized as follows.

Section 2 presents the related work, Section 3 describes the method used for conducting SLR, Section 4 illustrates the search results, Section 5 presents analysis, and lastly Section 6 concludes the paper.

2. Related work

Some literature reviews have been carried out on educational games and GBL. Marciano, Miranda, and Miranda (2014) presented a literature review regarding evaluation of various aspects of software, and describes the evaluation methods and applications. The study intended to understand the context of use of different evaluation technique both general and specific with the aim to be able to select and adapt the method to be used in the specific context of language learning with computer games. Abdul Jabbar and Felicia (2015) reviewed papers to investigate that how the design of game-based activities influence engagement and learning. They developed a set of some general recommendations for the instructional design of GBL. Based on a review of literature, Djelil et al. (2014) proposed to organize an evaluation process in design and experimental phases, and use empirical and analytical evaluation methods to lower the risk of a poor designed learning game. The paper highlighted four criteria classes (ludopedagogical environment, learner affective and cognitive reactions, training context, and learner profile) that effect the evaluation process of learning game. The measurement and analysis criteria are introduced before linking them with the three evaluation dimensions' usability, usefulness and acceptability to evaluate learning games in a training context. However, the research study does not provide a developmental basis or method for the selection of these three dimensions. The review by (Arttu et al., 2017) especially focused on exploring the meaning of flow within the context of serious games in addition investigating the relationship between learning and flow, factors influencing the occurrence of flow and operationalization of flow. The review mainly showed that there are only conceptual considerations and no robust empirical evidence exist about the meaning of flow.

Petri and von Wangenheim (2016) presented an SLR on systematic evaluation of educational games focusing on the evaluation process. The study results are based on 11 relevant articles describing 7 approaches to systematically evaluate educational games. The focus was on how the approaches are defined, operationalize, developed and evaluated. The study confirmed that only a few approaches are available to systematically evaluate educational games. However, the research results are based on only 7 encountered approaches where no clear pattern emerged on which factors are essential to evaluate educational games. This showed that further research is required on educational game evaluations to obtain more valid and uniform results. Another study by Petri and von Wangenheim (2017) presented an SLR that is specific for computing education games and explored how evaluations on computing education games are defined, executed and analyzed. According to the results of this study, most evaluations use a simple research design where the game is used and subsequently a subjective feedback is obtained through questionnaires. Most of the evaluations are carried out without replication, using qualitative methods for data analysis without using a well-defined evaluation framework. Thus, although several reviews exist on educational games, the focus of these studies is either on the design of educational games or on the evaluation process and methodology used. There is not a single review that presented a detailed and complete overview of studies focused on GBL evaluation. Thus, the question of what criteria is important for evaluating educational games remains open.

3. Method

This research is conducted as a Systematic Literature Review (SLR) based on the work by (Kitchenham et al., 2009). The goal of the review is to present state of the art in GBL evaluation and identify dimensions of GBL evaluation. The steps of SLR method are described below.

3.1 Research questions

In accordance with the goal of this research work, we performed an SLR focusing on the following questions.

R1. What are the current trends in GBL and/or educational games evaluation?
R1 further includes: R1.1 How much GBL evaluation research has been carried out? R1.2 Which research topics are being addressed? R1.3 Who is leading GBL evaluation research?
R2**.** What are the evaluation practices in GBL or educational games?
This includes: R2.1 Which evaluation approaches (frameworks/models/guidelines etc.) exist for GBL? R2.2 What criteria has been used for evaluation of educational games?
R3. What are the different dimensions for evaluating GBL and/or educational games?

3.2 Search strategy and process

A systematic literature review was conducted in March and April 2017 from a data pool consisting of four open databases (Google scholar, IEEE Xplore, ACM Digital Library and Directory of Open Access Journals (DOAJ)) and four journal proceedings (International Journal of Game-Based Learning (IJGBL), International Journal of Game Theory and Technology (IJGTT), International Journal of Serious Games (IJSG), and Computers & Education. The journal proceedings were selected based on their relevance in the field of GBL and educational games.

The core concepts include educational games, evaluations, evaluation frameworks and children. Several search strings were constructed using the keywords (including synonyms) based on the research questions. The search strategies were formed and adapted according to the specific syntax of each of the selected data sources however, search terms included the keywords "educational games" or "game based learning" or "serious games" or "educational games for children" with "evaluation", "assessment", "evaluation framework", "evaluation criteria," "assessment criteria," "or "metrics for evaluation". Manual search was also conducted for IJGBL, IJGTT and IJSG. However, the journal proceedings of Computers & Education were searched using search strings due to the extensive set of papers. Initially, we wanted to focus on evaluation of educational games for children, but since there is very little literature on this user group we decided to focus on evaluation of GBL in general.

After the initial search results were obtained the selection of primary studies was conducted as a two-stage process described by (Brereton et al, 2007). In the first stage, the title and abstract (abstract was read in case title did not provide clear idea) of articles were reviewed and all irrelevant papers were rejected and duplications were removed. In the second stage, full copies of all the selected papers were reviewed against inclusion/exclusion criteria to obtain relevant studies for this research.

3.3 Inclusion and exclusion criteria

The inclusion and exclusion criteria were defined according to research objectives and presented in Table 1.

3.4 Quality criteria

To ensure the quality of studies reviewed, search of literature was limited to journal articles, conference proceedings and book chapters. Any unpublished article or grey literature was not included. Only articles written in English were considered. The articles were excluded if full text was not available. Furthermore, after review only articles that provided considerable information on GBL evaluation were considered.

Table 1: Criteria for inclusion/ exclusion

Inclusion criteria	Exclusion criteria
Evaluation approach (frameworks/model etc.) for GBL or educational games Review studies for evaluation of GBL Evaluation/Assessment of one or more dimensions of GBL Guidelines or criteria for evaluating GBL Design guidelines/models for GBL useful for evaluation purpose Case studies, empirical studies etc. of evaluating any educational games (using some evaluation framework or predefined criteria).	Analysis studies of GBL acceptability/applicability in education. Comparison of GBL with traditional learning approaches Effectiveness of games in general for education purpose Evaluation of video/leisure games Evaluation of educational software's (m-learning/e-Learning) i.e. not game. Evaluation of Serious games other than education domain. Evaluation methods/process used to carry out evaluation instead of criteria for evaluation (how to evaluate rather than what to evaluate) Any duplications Different versions of same paper

3.5 Data extraction

The data was systematically extracted for each research question. The selected papers were thoroughly read and data was extracted by the first author and reviewed by the second author. The data extracted for **R1.1** is title, year, resource type and resource name, for **R1.2** research topic and description, for **R1.3** no of citations, country and references of each article, for **R2.1** name of evaluation approach, type of evaluation approach and description. for **R2.2** criteria used for evaluation, and for **R3** dimension(s) for evaluation.

4. Search results

In the initial search, we found a total of 1929 articles (see Table 2). The aim of this research was to include all possible relevant articles on GBL evaluation, and therefore the search queries were not restricted by the year. Almost all search strings retrieved results from between 2000 to 2017. A total of 28 search results were older than 2000; from two data sources; google scholar (3 results) and journal of computer & education (25 results) that ranged from 1940 to 1995. They were specially reviewed and were found not relevant to the topic of GBL and hence excluded.

The first stage of selection resulted in 232 potentially relevant articles which were further reviewed. After the second stage, a total of 162 articles were excluded based on inclusion/ exclusion and 12 based on quality criteria (for 8 papers full text was not available and four papers were not in English). Therefore, resulting in 58 articles that were selected for this research study. All the selected studies are listed in references.

Table 2: Initial search results

Data source	Search results	Year(range)
Computers & Education	712	2003-2017
Google scholar	390	2002-2017
IEEEXplore	311	2003-2017
ACM Digital Library	263	2000-2017
DOAJ	237	2005-2017
IJSG	11	2014-2017
IJGBL	5	2011-2017
IJGTT	0	-
Total	1929	2000-2017

5. Analysis

5.1 R1: What are the current trends in game base learning(GBL) or educational games evaluation?

The analysis of 58 selected studies showed that relevant publications were all from 2004 onwards indicating that the research in GBL evaluation is relatively a new field of study. Results indicated an increasing trend in GBL research with most number of studies in 2009 and 2015. Figure 1 presents how much GBL evaluation research has been carried out per year (R1.1).

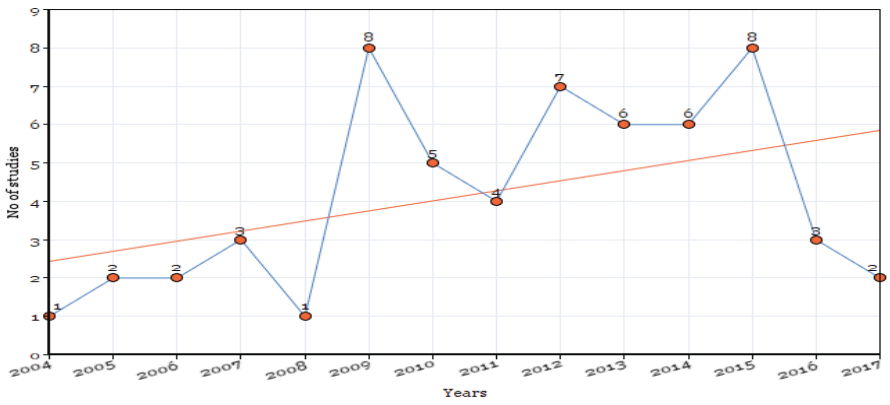

Figure 1: Year wise distribution of GBL evaluation studies (R1.1)

Most the selected studies were journal papers (51.7%) followed by conference papers (44.8%), and then book chapters (3.4%). Most of the research in GBL is published by Elsevier (13 studies) and IEEE (13 studies), followed by ACM (8 studies), Hindawi (5 studies), springer (3 studies), IJSE (3 studies), IGI Global (2 studies) and remaining 11 were from different resources (1 study per resource).

Regarding the research topics being addressed (R1.2), the selected research articles were categorized into five categories: (1) *evaluation approach (32.8%)*: studies presenting some GBL evaluation approach including framework, models, guidelines etc., (2) *development focus approach (5.2%)*: articles presenting an approach with focus on GBL development but can also be used for evaluation of GBL, (3) *design focus approach (25.9%)*: studies presenting any design model, guidelines etc. that can be also used for evaluation, (4) *review Studies*

(17.2%): review articles in GBL evaluation, and (5) *educational game evaluation (18.9%)*: this category includes all the articles that present case studies, empirical evaluation or any type of educational game evaluation studies.

The year-wise distribution of research topics/purpose for the selected relevant studies is presented in Figure 2. According to the results, there is a gradual increase in no of studies for all research topics, with 2009 to 2015 being the peak years of research. Although almost all research topics span over the years except review studies that were not seen in earlier years but have been trending from 2013 onwards with 10 studies in last five years. Moreover, the figure also shows that in the earlier years the design focus approach was the target research topic whereas in the latter years this trend has shifted towards evaluation approaches with 11 studies on this research topic from 2012 to 2016 whereas, only 4 in design focus approach.

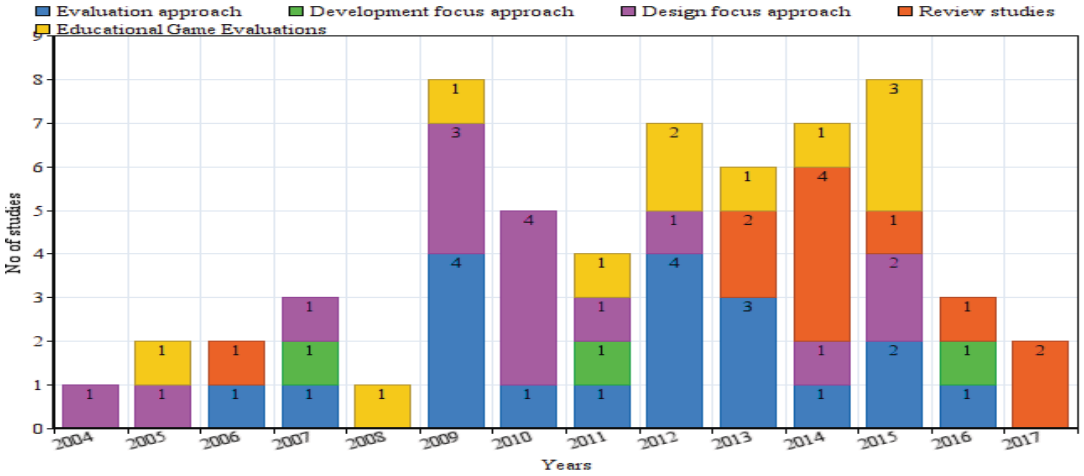

Figure 2: Year wise distribution of research topics

The results on who is leading GBL evaluation research (R1.3) is presented as country-wise distribution of studies, no of citations, and highly cited studies. The country-wise distribution of studies is presented in Figure 3, the results showed that major contribution of GBL evaluation research comes from Malaysia (8 studies).

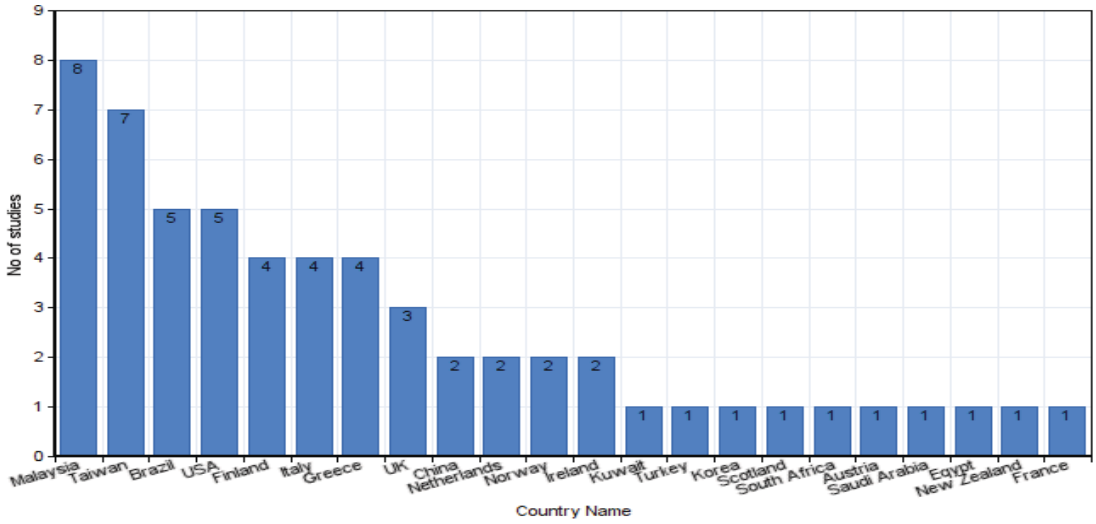

Figure 3: Country wise distribution of studies

Google Scholar was used for the citation counts for article as it indexes and finds more cited references. The results for citation counts are presented in Figure 4. The 10 papers with most citations were further analyzed for research topics, authors names and number of citation, and the results are shown in Table 3.

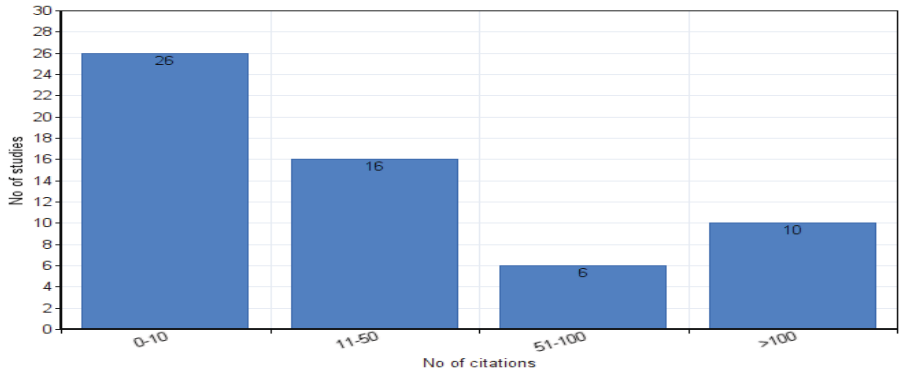

Figure 4: Distribution of studies by no of citation

5.2 R2: What are the evaluation practices in GBL or educational games?

To answer this question, we first wanted to find which evaluation approaches exist for GBL (R2.1). After thorough analysis of the 58 selected studies, 19 evaluation approaches were identified for the evaluation of GBL or educational games

Table 3: Highly cited articles

Author	Research Topic	Country	Citation
(Kiili, K., 2005)	Design Approach	Finland	1070
(Papastergiou, M., 2009)	Educational Game Evaluation	Greece	1043
(De Freitas, S. and Oliver, M., 2006)	Evaluation Approach	UK	623
(Fu, F.L., Su, R.C. and Yu, S.C., 2009.)	Evaluation Approach	Taiwan	376
(Amory, A., 2007)	Development Approach	South Africa	253
(Annetta, L.A., 2010)	Design Approach	USA	238
(Bellotti et al,2013)	Review Study	Itlay	237
(Virvou, M. and Katsionis, G., 2008)	Educational Game Evaluation	Greece	189
(Wouters, P. and Van Oostendorp, H., 2013)	Review Study	Netherland	108
(Mitgutsch, K. and Alvarado, N., 2012)	Evaluation Approach	USA	103

Out of 19 approaches, 8 presents a framework, 4 presents heuristics/guidelines, 2 presents a model, 2 presents a scale, 1 presents a method, 1 presents a standard, and another one presents evaluation constructs. Only 3 approaches (De Freitas and Oliver, 2006), (Su, Chen, and Fan, 2013) and (Mitgutsch and Alvarado, 2012) focus on overall evaluation of GBL. All the other approaches deal with the evaluation of either one or two specific dimensions for evaluating GBL e.g. flow framework for flow dimension, EGameFlow for user enjoyment, framework of UX etc. Table 4 presents the evaluation approaches and their description.

Table 4: Evaluation approaches in GBL

Evaluation Approaches	Description
Four-dimensional framework	Evaluate the potential of using games and simulation based learning in their practice (De Freitas and Oliver, 2006)
Flow Framework	Describes the dimensions of flow experience that can be used to analyze overall quality of playing experience (Kiili et al.,2014)
Playability Heuristic for Educational Games (PHEG)	Heuristic Evaluation for finding usability problems in educational computer games, (Mohamed, Yusoff, and Jaafar, 2012)
Evaluation framework for effective GBL	GBL evaluation with focus on pedagogical perspective. (Connolly, Stansfield and Hainey, 2009)
EGameFlow	Assess user enjoyment of e-learning games (Fu, Su, and Yu, 2009)
Evaluation framework of UX	Evaluate of user experience for adaptive digital educational games (DEGs). (Law and Sun, 2012)
Evaluation Framework for GBL	Guide GBL evaluation from learning perspective. (Wang, Liu, Lin and Xiang, 2011)
Game scale to evaluate Educational games	Evaluate quality in educational computer games in terms of learning and enjoyment characteristics. (Ak, 2012)
Quality Evaluation Standard	Identify quality evaluation elements of educational serious games both technical and non-technical elements. (Yoon and Park, 2013)
Heuristic Evaluation for Educational Games(HEEG)	Heuristic for evaluating educational games in terms of usability and game experience. (Marcelo, Andreza and Igor, 2015)
Heuristics Evaluation Strategy	Evaluate specifically for mGBL. (Zaibon and Shiratuddin, 2010)

Evaluation Approaches	Description
Game-based learning evaluation model(GEM)	Measure the effectiveness of serious games in a practical way. (Oprins et al., 2015)
Guidelines for evaluating games	Identify promising games for teaching computer science based on topics taught, easy to install, engaging, time to use. (Gibson and Bell, 2013)
Evaluation framework for assessing games	Focused on quality aspect in selecting and assessing learning games. (Dondi and Moretti, 2007)
Framework for serious game design evaluation	Evaluate the effectiveness of evaluation model and provide design criteria for multimedia game design educators. (Su, Chen and Fan, 2013)
Quality evaluation model	ISO quality model for mobile games. (Alhuhud and Altamimi, 2016)
Usability evaluation constructs	Present six evaluation constructs for usability evaluation for history educational game design. (Yue and Zin, 2009)
Methodology for interface evaluation	Heuristics based usability evaluation that describe usability factors to evaluate interface of educational games. (Omar and Jaafar, 2009)
serious game design assessment framework	Identified six essential components of the formal conceptual structure underlying a serious game. (Mitgutsch and Alvarado, 2012)

Further, we looked into what criteria has been used for evaluation of educational games (R2.2). The selected papers were classified into three categories for criteria used for evaluation of educational games: (1) *evaluation approach* (framework/model etc.), (2) *predefined criteria* (ad hoc), and (3) *not specified*. Most of the studies did not use any well-defined existing framework or model to conduct the evaluation. From the total of 11 studies on educational game evaluation; 72.7% (8 studies) used some predefined criteria (ad hoc). Most of the studies just outlined the dimensions (goals) of evaluation without explicitly defining the basis for selection or the factors and measures used for evaluation except one study (de Lima, de Lima Salgado and Freire, 2015) that stated the use of game experience questionnaire and intrinsic motivation inventory (IMI) for predefined dimensions of user experience and intrinsic motivation. On the other hand, 27.3% of the studies (3 studies) used some existing evaluation approaches not specific for GBL. Such as Nielsen's heuristics was used by (Mei, Ku and Chen, 2015), and USE scale (Lund, 2001) by (Tseloudi and Tsiatsos, 2015) for evaluating usability. Flow (Csikszentmihalyi, 1992), and a taxonomy of intrinsic motivations for learning (Malone and Lepper, 1987) was used by (Pourabdollahian, Taisch and Kerga, 2012) for measuring engagement. Only one study (Tseloudi and Tsiatsos, 2015) used EGameFlow scale for measuring enjoyment that is a scale developed specifically for educational games.

5.3 R3. What are the different dimensions for evaluating GBL or educational games?

To identify the different dimensions for evaluating educational games, we analyzed the selected studies for the goals of GBL evaluation. The analysis highlighted two critical issues: *first*, there is a wide diversity of elements considered for GBL evaluation and are defined inconsistently across studies; and *second*, the terms such as evaluation dimensions, factors, sub factors and metrics/measures are themselves defined inconsistently across studies and therefore not allowing the proper categorization of these elements and identification of a clear pattern. For example, some studies consider feedback in a broader scope as a dimension (goal) for evaluation whereas other studies use feedback as a factor to achieve a goal (usability). There is no distinction between macro and micro level elements. Therefore, we take the first step towards making this distinction by defining the terms of use. For this research work the term "dimension" is used in a broader scope referring to elements essential for educational game; the main goals/aim of GBL evaluation. "Factors/sub factors" are the elements considered important for achieving a dimension and "metrics/measures" is the gauge to assess that factor. In terms of scope this can be shown as: Dimension > factors> sub factors> metrics/measures. Therefore, the first step is to identify the dimensions. =A total of 37 dimensions were identified in the analysis of 58 studies. Learning is the most widely used dimension (19 studies) followed by usability (12 studies) and game factors (10 studies) including game design, game story and game mechanics. Only one study (Alfadhli and Alsumait, 2015) presents GBL design guidelines that focus on children requirements. The identified dimensions are shown in Table 5 along with the number of studies using them. Dimensions with same frequency (no of studies) are listed in a single row.

Table 5: Dimensions for GBL evaluation

Evaluation dimensions	No
Learning/Pedagogical	28
Usability	12
Game factors (design, story, mechanics)	10

Evaluation dimensions	No
User experience (UX), Motivation	5
Enjoyment, Flow, Engagement	4
Playability	3
Gameplay, Cognitive load, Instructional design, Immersion, Challenges/increased complexity	2
Child requirements, Likeability, feedback, understandability, relevance, interactivity, embedding, transfer, adaptation, naturalization, identity, informed teaching, fidelity, context, learner specification, mode of representation, technical verification, social collaboration, emotional, instructional support, collaborative learning, acceptability, usefulness, Learning Content	1

6. Conclusion

The previous review studies on GBL provided insights on the design of GBL and focused either on exploring either one or two dimensions of GBL or on the evaluation process and methodology (research design, instruments used, data collection and analysis etc.). The existing research fail to provide the state of the art in GBL evaluation. This research fills this gap by exploring the issue from directions such as the trends in GBL evaluation (amount of research in GBL evaluation, research topics, highly cited articles), current practices in GBL evaluation (approaches and criteria), and the dimensions for evaluating GBL.

The main findings of this paper includes: (1) an increasing trend in GBL research within past few years with most studies from 2009 to 2015 (2) Elsevier and IEEE are the two major resources for GBL evaluation research with more journal papers, (3) the research topic/ purpose of most studies focus on an evaluation approach followed by design focused approaches for GBL evaluation, (4) the review studies for GBL evaluation increased over the past few years and there is also a shift in research topics from design to evaluation, (5) most studies focused only on one or two dimensions of GBL and very few focused on overall evaluation specifying all the dimension essential for GBL evaluation, highlighting the need for a comprehensive evaluation framework, (6) current evaluation approaches in GBL does not cater children needs, only one out of all the reviewed studies(design focused approach) considered children requirements, (7) majority of the studies for educational game evaluations do not use existing GBL evaluation frameworks instead they mostly employ pre-defined criteria(ad hoc) for evaluation or few use general guidelines/approaches, and (8) evaluation dimensions, factors/sub factors and metrics are defined inconsistently across the studies and a wide diversity of elements are considered for GBL evaluation, however most extensively used dimension in GBL evaluation are learning, usability and game factors.

For future work, research can be extended to discuss the factors/sub factors that need to be evaluated for each GBL dimensions, why they are important, interrelation of factors and further exploring the metrics for quantifying these factors/sub factors. In sum, this study can help supplement connections with previous studies and forms an important reference base for future research in GBL evaluation.

References

A. Lund, “Measuring usability with the USE questionnaire,” STC Usability SIG Newsletter, vol. 8 (2), 2001.

Abdul Jabbar, A.I. and Felicia, P., 2015. Gameplay engagement and learning in game-based learning: A systematic review. Review of Educational Research, 85(4), pp.740-779.

Ak, O., 2012. A game scale to evaluate educational computer games. Procedia-Social and Behavioral Sciences, 46, pp.2477-2481.

Aleven, V., Myers, E., Easterday, M. and Ogan, A., 2010, April. Toward a framework for the analysis and design of educational games. In Digital Game and Intelligent Toy Enhanced Learning (DIGITEL), 2010 Third IEEE International Conference on (pp. 69-76). IEEE

Alfadhli, S. and Alsumait, A., 2015, December. Game-Based Learning Guidelines: Designing for Learning and Fun. In Computational Science and Computational Intelligence (CSCI), 2015 (pp. 595-600). IEEE.

Alhuhud, G. and Altamimi, W., 2016. Quality Evaluation of Mobile Game: Miftah Alfasaha. Mobile Information Systems, 2016.

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

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

Arttu et al, 2017.Flow experience in game based learning – a systematic literature review. International Journal of Serious Games. 2017, 4(1), pp.57-72

Brereton, P., Kitchenham, B.A., Budgen, D., Turner, M. and Khalil, M., 2007. Lessons from applying the systematic literature review process within the software engineering domain. Journal of systems and software, 80(4), pp.571-583.

Catalano, C.E., Luccini, A.M. and Mortara, M., 2014. Best practices for an effective design and evaluation of serious games. International Journal of Serious Games, Volume 1, Issue 1.

Chorianopoulos, K. and Giannakos, M.N., 2014. Design principles for serious video games in mathematics education: from theory to practice. The International Journal of Serious Games, 1(3), pp.51-59.

Connolly, T., Stansfield, M. and Hainey, T., 2009. Towards the development of a games-based learning evaluation framework. Games-based learning advancements for multisensory human computer interfaces: Techniques and effective practices. Hershey PA: IGI Global.

Csikszentmihalyi, M. and Csikszentmihalyi, I.S. eds., 1992. Optimal experience: Psychological studies of flow in consciousness. Cambridge university press.

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

De Lima, Luis Gustavo Rotoly, André de Lima Salgado, and André Pimenta Freire. "Evaluation of the user experience and intrinsic motivation with educational and mainstream digital games." Proceedings of the Latin American Conference on Human Computer Interaction. ACM, 2015.

Djelil, F., Sanchez, E., Albouy-Kissi, B., Lavest, J.M. and Albouy-Kissi, A., 2014, October. Towards a Learning Game Evaluation Methodology in a Training Context: A Literature Review. In European Conference on Games Based Learning (Vol. 2, p. 676). Academic Conferences International Limited

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

Eagle, M., 2009, April. Level up: a frame work for the design and evaluation of educational games. In Proceedings of the 4th International Conference on Foundations of Digital Games (pp. 339-341). ACM.

El-Sattar, A. and Hussein, H.K., 2016. Learning Islamic Principles with Serious Games. In Proceedings of 13th International Conference on Advances in Computer Entertainment Technology (p. 30). ACM.

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

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

Ibrahim, R. and Jaafar, A., 2009, August. Educational games (EG) design framework: combination of game design, pedagogy and content modeling. In Electrical Engineering and Informatics, 2009. ICEEI'09. International Conference on (Vol. 1, pp. 293-298). IEEE.

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

Kiili, K.et al., 2014. Flow framework for analyzing the quality of educational games. Entertainment Computing, 5(4), pp.367-377

Kitchenham et al,2009. Systematic literature reviews in software engineering–a systematic literature review. Information and software technology, 51(1), pp.7-15.

Law, E.L.C. and Sun, X., 2012. Evaluating user experience of adaptive digital educational games with Activity Theory. International Journal of Human-Computer Studies, 70(7), pp.478-497.

Liao, Y.H. and Shen, C.Y., 2012, March. Heuristic evaluation of digital game based learning: a case study. In Digital Game and Intelligent Toy Enhanced Learning (DIGITEL), 2012 IEEE (pp. 192-196). IEEE.

Malone, T.W. and Lepper, M.R., 1987. Making learning fun: A taxonomy of intrinsic motivations for learning. *Aptitude, learning, and instruction*, *3*(1987), pp.223-253.

Marcelo, B., Andreza B. and Igor, de., 2015. HEEG: Heuristic Evaluation for Educational Games. Proceedings of SBGames, 2015, XIV SBGames, pp.224-227.

Mei, S.Y., Ku, O. and Chen, S.Y., 2015, July. Interface Evaluation of a Game-Based Learning System: Experts vs. Novices. In Advanced Applied Informatics (IIAI-AAI), 2015 IIAI (pp. 313-318). IEEE.

Mitgutsch, K. and Alvarado, N., 2012. Purposeful by design? a serious game design assessment framework. In Proceedings of International Conference on the foundations of digital games (pp. 121-128). ACM.

Mohamed, H. and Jaafar, A., 2010, June. Challenges in the evaluation of educational computer games. In Information Technology (ITSim), 2010 International Symposium in (Vol. 1, pp. 1-6). IEEE.

Mohamed, H. and Jaafar, A., 2010, November. Development and potential analysis of heuristic evaluation for educational computer game (PHEG). In Computer Sciences and Convergence Information Technology (ICCIT), 2010 5th International Conference on (pp. 222-227). IEEE.

Mohamed, H., Yusoff, R. and Jaafar, A., 2012, March. Quantitive analysis in a heuristic evaluation for usability of educational computer game (UsaECG). In Information Retrieval & Knowledge Management (CAMP), 2012 International Conference on (pp. 187-192). IEEE.

Omar, H.M. and Jaafar, A., 2009, November. Conceptual framework for a heuristic based methodology for interface evaluation of educational games. In Proceedings of the 2009 International Conference on Computer Technology and Development-Volume 01 (pp. 594-598). IEEE Computer Society.

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

Pappa, D. and Pannese, L., 2010. Effective design and evaluation of serious games: The case of the e-VITA project. Knowledge Management, Information Systems, E-Learning, and Sustainability Research, pp.225-237.

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

Petri, G. and von Wangenheim, C.G., 2016. Journal of Universal Computer Science, vol. 22, no. 7, 992-1021.

Petri, G. and von Wangenheim, C.G., 2017. How games for computing education are evaluated? A systematic literature review. Computers & Education.

Ronimus, M., Kujala, J., Tolvanen, A. and Lyytinen, H., 2014. Children's engagement during digital game-based learning of reading: The effects of time, rewards, and challenge. Computers & Education, 71, pp.237-246.

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

Shi, Y.R. and Shih, J.L., 2015. Game factors and game-based learning design model. International Journal of Computer Games Technology, 2015, p.11.

Shiratuddin, N. and Zaibon, S.B., 2011, November. Designing user experience for mobile game-based learning. In User Science and Engineering (i-USEr), 2011 International Conference on (pp. 89-94). IEEE.

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

Tan, P.H., Ling, S.W. and Ting, C.Y., 2007. Adaptive digital game-based learning framework. In Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts (pp. 142-146). ACM.

Thomas, S., Schott, G. and Kambouri, M., 2004. Designing for learning or designing for fun? Setting usability guidelines for mobile educational games. Learning with mobile devices: A book of papers, pp.173-181.

Tseloudi, C. and Tsiatsos, T., 2015, July. Panic in the gallery: An online educational game for art history: Design and evaluation of a matching game. In Information, Intelligence, Systems and Applications (IISA), 2015 6th International Conference on (pp. 1-6). IEEE.

Virvou, M. and Katsionis, G., 2008. On the usability and likeability of virtual reality games for education: The case of VR-ENGAGE. Computers & Education, 50(1), pp.154-178.

Wallner, G. and Kriglstein, S., 2011. Design and evaluation of the educational game DOGeometry: a case study. 8th International Conference on Advances in Computer Entertainment Technology (p. 14). ACM

Wang, Y., Liu, X., Lin, X. and Xiang, G.,2015, An Evaluation Framework for Game-Based Learning. 15th Global Chinese Conference on Computers in Education.

Wouters, P. and Van Oostendorp, H., 2013. A meta-analytic review of the role of instructional support in game-based learning. Computers & Education, 60(1), pp.412-425.

Yoon, S.J. and Park, H.S., 2013. Development of a Quality Evaluation Standard for Educational Serious Games. Journal of information and communication convergence engineering, 11(2), pp.103-111.

Yu, K.C., Hsiao, H.S. and Tsai, F.H., 2005, June. The implementation and evaluation of educational online gaming system. In Information Technology: Research and Education, 2005. ITRE 2005. 3rd International Conference on (pp. 338-342). IEEE.

Yu, S.C., Fu, F.L. and Su, C.H., 2009. EGameDesign: guidelines for enjoyment and knowledge enhancement. In International Conference on Hybrid Learning and Education (pp. 35-44). Springer Berlin Heidelberg.

Yue, W.S. and Zin, N.A.M., 2009. Usability evaluation for history educational games. 2nd international Conference on interaction Sciences: information Technology, Culture and Human (pp. 1019-1025). ACM.

Zhang, H.F., Fan, X.Y. and Xing, H.F., 2010, October. Research on the design and evaluation of educational games based on the RETAIN model. In Knowledge Acquisition and Modeling (KAM), 2010 3rd International Symposium on (pp. 375-378). IEEE.