

Codifying Game-Based Learning: The LEAGUE framework for Evaluation

Rabail Tahir and Alf Inge Wang

Dept. of Computer Science, Norwegian University of Science and Technology, Trondheim, Norway
rabail.tahir@ntnu.no, alf.inge.wang@ntnu.no

Abstract Educational games are now seen as effective learning tools. However, there is a gap in literature regarding core dimensions of Game-based learning (GBL) for comprehensive evaluation due to inconsistent use of elements. Literature on GBL evaluation reports an extensive diversity of evaluation elements used for GBL without any categorization of micro and macro-level elements. Hardly any studies systematically decompose these aspects to derive factors/sub-factors, obstructing identification of any clear pattern. The problem is not scarcity of GBL evaluation research but inconsistency in terminology, scope, definition and usage of elements leading to the absence of a holistic view of GBL-evaluation. This study bridges the gap by outlining terminology and scope with four conceptual levels, and based on that systematically categorizing GBL-evaluation elements in terms of scope, definition and usage. The methodology used is directed content analysis of GBL-evaluation literature collected through a previous systematic literature review. Dimensionalization of GBL and further decomposition into factor/sub-factors based on theoretical constructs has resulted in a consistent and clear pattern delineating the structure of educational game evaluation. Further codifying metrics and mapping the relationship among GBL dimensions deduce into a conceptual framework that facilitates greater insight into the process of learning with educational games, where to focus and what to evaluate.

Keywords: Game-based learning, educational games, evaluation framework, dimensions, factors, relations

1. Introduction

Game-based learning (GBL) is an innovative educational paradigm that utilizes games as a mode for transferring learning (Tan, Ling et al. 2007). Educational games are considered to have the potential of deeply engaging learners with any topic, allowing active participation in learning process (Wallner and Krigstein 2011). Bellotti, Kapralos et al. (2013) stated that educational games, like any educational tool, must be able to show that necessary learning has occurred. Therefore, it is crucial to systematically evaluate them to affirm their impact (Marciano, Miranda et al. 2014). The diverse characteristics of learning games makes their evaluation a difficult task. However, evaluation is the only means to verify that educational goals are achieved and spot any functional vulnerability (Djelil, Sanchez et al. 2014). Dondi and Moretti (2007) highlighted two important issues related to GBL evaluation. *First*, construction of a general framework is extremely difficult unless it is an abstract one. *Second*, it is difficult to differentiate between single and holistic aspects of GBL evaluation.

Many researchers have attempted to explain the important elements of serious games for education. The review of these theories draws only one conclusion: There is no consensus among researchers about the terminology and the comparable importance of GBL elements (Oprins, Visschedijk et al. 2015). Furthermore, it has been a constant challenge to understand the relationships between the different aspects embedded in GBL (Ahmad, Rahim et al. 2015). Our previous review study on GBL evaluation highlighted the following problem areas: *1) Most evaluation frameworks and studies focus on exploring any single aspect of GBL, making it difficult to identify all core dimensions; 2) Use of a wide diversity of elements for evaluating educational games does not allow the identification of any clear pattern; 3) Very few studies systematically decompose GBL aspects based on their theoretical construct, not allowing the hierarchical decomposition in terms of scope; and 4) The inconsistency in definition, usage, scope and terminology (e.g. dimensions, factors, etc.) of evaluation elements in GBL literature.* Therefore, to systematically analyze GBL-evaluation concepts, there is a need for proper categorization of the wide variety of elements available in literature (Petri and von Wangenheim 2017).

This study attempts to overcome the identified problems by performing directed content analysis on the data set of existing GBL-evaluation literature collected through a systematic literature review (Tahir and Wang 2017). The educational game evaluation elements extracted from systematic review are hieratically decomposed (using operations such as coding, categorization, abstraction, comparison and integration) into core dimensions, factors and sub-factors based on scope, frequency of occurrence, relationship between codes, underlying meaning across codes and mapping to existing theoretical frameworks and constructs defined by researchers in the domain of GBL. The metrics and relations between core dimensions is also detailed using similar process for a complete analysis. The result is a conceptual framework named LEAGUE (Learning, Environment, Affective-cognitive reactions, Game factors, Usability, UsEr) that list the core GBL elements in a hierarchy of scope. The paper is organized as follows: section 2 discusses related work, section 3 describes the method, section 4

presents the results of directed content analysis in the form of LEAGUE conceptual evaluation framework and finally, section 5 concludes the study with discussion and future research.

2. Related work

The multidimensionality of GBL demands to consider a number of aspects important for its evaluation (Furió, D. et al. 2013). However, there is still a debate around which aspects to consider (Oprins, Visschedijk et al. 2015).

2.1 Systematic reviews and evaluation studies on educational games

Many review studies in GBL (Perttula, Kiili et al. 2017; Djelil, Sanchez et al. 2014; Petri and von Wangenheim 2017; Tahir and Wang 2017) have reported the use of a wide diversity of evaluation aspects for educational games. These aspects are inconsistently defined, and most studies do not systematically decompose into their constituents (Oprins, Visschedijk et al. 2015; Petri and von Wangenheim 2016). For example, some studies consider the concept interactivity as one of the main dimensions of GBL (Annetta 2010), while other studies use interactivity in a narrow scope as a factor to achieve a GBL dimension such as usability (Djelil, Sanchez et al. 2014). Moreover, some others use it as a sub-factor of a factor interface (Omar and Jaafar 2010). There is no clear distinction between micro and macro level elements.

Our insight into evaluation studies showed a similar trend where most researchers used predefined ad-hoc criteria selecting different aspects for evaluating educational games. Moreover, existing GBL model/framework are less used in empirical research (Tahir and Wang 2017). Virvou and Katsionis (2008) evaluated usability and likeability in the VR-ENGAGE game for education. Pourabdollahian, Taisch et al. (2012) employed flow dimensions for measuring learner engagement in serious games for manufacturing education. Papastergiou. (2009) focused on evaluating learning effectiveness and motivation of GBL in computer science education. Giannakos. (2013) and Yu, Hsiao et al. (2005) used learning performance as a measure in their evaluation studies. The aspect of usability has also been widely used for evaluating educational games, but different studies used different factors to access this aspect (Liao and Shen 2012; Mei, Ku et al. 2015; Wallner and Kriglstein 2011). De Lima, de Lima Salgado et al. (2015) evaluated user experience and motivation in educational games.

2.2 Game-based learning concepts in existing frameworks

Although several GBL evaluation models/frameworks exist, it is important to highlight that each of these models/theories focuses on analyzing and understanding educational games using different aspects, where most researchers focus only on one specific aspect. Thus they could supplement one another but individually these studies are fairly narrow and may account only for a portion of a complete picture of GBL evaluation (Fu, Su et al. 2009); (Tan, Ling et al. 2007). Here are some main aspects explored in various GBL frameworks/models:

Learning: Most of the researchers in GBL mainly focus on education/learning aspects. Four-dimensional framework by (De Freitas and Oliver 2006) focuses on learning to help tutors evaluate the potential of employing simulation/GBL in practice. Connolly, Stansfield et al. (2009) describe an evaluation framework that focuses on pedagogical aspect, introducing attributes to measure GBL environment with attention on the learner and learning. Another evaluation framework proposed by (Wang, Liu et al. 2015) also emphasizes on learning perspective with respect to learning result, learner motivations and learner experience.

Flow: Conversely, some researchers focus on flow and enjoyment aspects in educational games. Kiili. (2005) introduced an experiential gaming model to facilitate flow experience serving as a link between game design and educational theory, but not offering complete game design. EGameFlow proposed by (Fu, Su et al. 2009) is a scale for assessing the level of enjoyment delivered by e-learning games. Kiili, Lainema et al. (2014) presented a flow framework to analyze overall playing experience of educational games through dimensions of flow.

Game design: Serious game design assessment framework implemented by (Mitgutsch and Alvarado 2012) structures different game design elements to analyze the formal conceptual design of serious games. It gives recommendation on how to shape serious games assessment in terms of design. Chorianopoulos and Giannakos (2014) and Shi and Shih (2015) also focused on game design aspect presenting design principles for serious games in mathematics, proposing 11 game factors for GBL design that described a thinking process to design and evaluate educational games using game elements.

Usability: Some researchers focus on usability (Mohamed, Yusoff et al. 2012; Rêgo and de Medeiros 2015; Omar and Jaafar 2010) and present heuristics for evaluating usability of educational games. These researchers incorporated concepts of learning, game play, interface and enjoyment within heuristics for evaluating GBL usability. Yue and Zin. (2009) proposed six usability evaluation constructs for design of history educational game.

Pedagogy and game design: Some researchers set a combined focus on learning and game design as two critical aspects for educational game design and evaluation. Some of the frameworks include: educational games design framework by Ibrahim and Jaafar (2009), framework for the analysis and design of educational games by Aleven, Myers et al. (2010), adaptive digital GBL framework proposed by Tan P.-H. et al. (2007), RETAIN model presented by Zhang, Fan et al. (2010), GBL evaluation model (GEM) by Oprins, Visschedijk et al. (2015), and Game object model (GOM) proposed by Amory. (2007). Rooney. (2012) presented a framework consisting of play, fidelity and pedagogy for serious game design.

2.3 Content analysis

Qualitative content analysis is a data reduction and sense-making effort that requires data samples to comprise of purposively selected texts that can inform research objectives and attempts to identify meanings and core consistencies (Hsieh and Shannon 2005, Loh, Sheng et al. 2015). Three different approaches exist for content analysis: Conventional, Directed, and Summative. In directed approach, analysis begins with relevant research findings/theory as guidance for the initial codes, and the goal is either to validate or conceptually extend a theory/ framework. Depending on research question, it has two strategies to begin coding. If aim is to identify and categorize all possible instances of any specific phenomenon, then it might be useful to first read and highlight the text representing the instances of that phenomenon and then start coding. The second strategy immediately begins coding with predetermined codes (Zhang and Wildemuth 2005).

3. Method

This study applied directed content analysis based on the work of (Hsieh and Shannon 2005). The general objective of this study is both to validate and conceptually extend the existing research on GBL evaluation by analyzing, interpreting, and organizing the evaluation aspects to fill the gap in current literature regarding inconsistency in systematic categorization, and use of features for GBL evaluation.

The content analysis was guided by the following research questions based on problem statements identified in introduction: *RQ1* What are the core dimensions for evaluating educational games? ; *RQ2* Which factors are important for achieving each core dimensions? ; *RQ3*. Are there any sub-factors for assessing these factors based on theoretical constructs? ; *RQ4* What metrics can be used to quantify these factors/ sub-factors? ; and *RQ5* Are the GBL dimensions interrelated?

3.1 Directed content analysis

The process of content analysis followed the steps defined by (Zhang and Wildemuth 2005). The steps included: preparing data, defining unit of analysis, developing coding strategy, testing coding strategy, coding all data, assessing coding consistency, drawing conclusion from coded data, and reporting method and findings. The directed content analysis was an iterative process involving progressing through extracted data to further analysis using following set of operations: coding, categorization, comparison, abstraction, integration, and iteration (Spiggle 1994, Engl and Nacke 2013) in such a way that preceding operations shaped the following ones. The analysis was not performed in a linear manner but moving back and forth between stages.

3.2.1 Corpus for analysis

The data set for directed content analysis comprised of data extracted from 58 articles on GBL evaluation literature from our previous systematic literature review (Tahir and Wang 2017). The selected articles comprised of GBL evaluation frameworks, evaluation studies, and reviews. The corpus is completely focused on GBL evaluation literature and not integration of gaming and learning field to be in line with (Loh, Sheng et al. 2015). The data items extracted from selected papers include: Dimensions, factors, sub-factors, metrics, interrelated dimensions/factors/sub-factors, relation type or description, and definitions of dimensions/factors/sub-factors. All the information was entered in excel spreadsheet.

3.2.2 Defining Unit of Analysis

To remove the inconsistency in terminology used in varying scope across studies; we introduced and defined four conceptual levels with regard to scope (dimensions, factors, subfactors and metrics) for analysis of GBL evaluation components that results in a hierarchy. Hierarchy is important when defining attributes for a specific application domain (Kececi and Abran 2001). The scope of terminology is defined as follows: the term "**Dimension**" refer to a broader concept but isolated within its kind and not a composition of different aspects; representing the main goals/objects of GBL. Each dimension represents one specific aspect of GBL. The term "**Factor**" refer to the elements important for achieving a specific dimension and the term "**Sub-factor**" refer to further categorized elements that constitute that specific factor. The term "**Metrics**" is the gauge to measure a factor/sub-factor either through objective or subjective data. This can be depicted (high to low level) as: Dimension > Factors> Sub-factors> Metrics.

3.2.1 Coding Strategy

We adopted the first strategy for coding (see Section 2.3) because the aim is to identify and categorize all possible instances of GBL evaluation components in the selected corpus systematically and consistently. Therefore, before starting analysis, we read the text data and extracted the text for each of the four conceptual levels (dimensions, factors, subfactors and metrics) in excel spreadsheet that appeared to represent them on first impression (as used in each study), and then started coding for each level (top-down). Hence, the analysis starts with identifying core dimensions and proceeds with factors & sub-factors (with reference to level above). The predetermined or initial categories used for coding dimensions were learning, game design, flow and usability (see Section 2.2), and were further analyzed, using set of operations mentioned in Section 3.1, until the final core dimensions were attained. The sub-sequent analysis focused on analyzing the sub-categories including factors for each dimension, sub-factors for each factor, metrics and relations using similar operations. The analysis of sub-factors mostly resulted in integration of constructs where possible by using existing concrete theories/models (e.g. sub-factors of flow were integrated by Csikszentmihalyi's flow model) for aiding the process and enhancing the validity of final GBL components which are theoretically grounded.

The coding was checked for consistency at each level where both authors discussed and finalized the categories formed. The analysis and findings resulted in a conceptual framework and is presented in the next section.

4. The LEAGUE conceptual framework

This section presents the results of directed content analysis in the form of a hierarchical integrated conceptual evaluation framework called LEAGUE (see Figure 2). The framework defines the core components of GBL. It distinguishes four levels called dimensions, factors, sub-factors and metrics. The highest level of this structure is dimensions and the lowest level is metrics. Each dimension is systematically decomposed into factors and then sub-factors. It also clarifies the relationship between high-level components (dimensions) of GBL.

4.1 Dimensions

Figure 1 shows the six dimensions in LEAGUE identified as key constituent of GBL evaluation.

Learning/Pedagogical	
Characteristics of an educational game that promote and facilitate learning.	(De Freitas and Oliver 2006; Yoon and Park 2013; Rêgo and de Medeiros 2015; Pappa and Pannese 2010; Omar and Jaafar 2010)
Game Factors	
Elements and features of a game environment (the game world) of a educational game that stimulate the users.	(Oprins, Visschedijk et al. 2015; Yue and Zin 2009; De Freitas and Oliver 2006; Yoon and Park 2013; Aleven, Myers et al. 2010; Zaibon and Shiratuddin 2010; Tan, Ling et al. 2007; Rooney 2012; Djelil, Sanchez et al. 2014; Liao and Shen 2012)
Affective Cognitive Reactions (ACR)	
Set of emotions, attitudes and feelings triggered by educational game.	(Rêgo and de Medeiros 2015; Zhang, Fan et al. 2010; Rooney 2012; Djelil, Sanchez et al. 2014; Annetta 2010; Pourabdollahian, Taisch et al. 2012)
Usability	
The extent to which an educational game is usable by the learners to accomplish specific goals (including how to learn, understand, control, and user satisfaction).	(Yue and Zin 2009; Rêgo and de Medeiros 2015; Zaibon and Shiratuddin 2010; Djelil, Sanchez et al. 2014)
User	
Characteristics of the target users/user group of an educational game to investigate if the game matches the indented users.	(Tan, Ling et al. 2007; De Freitas and Oliver 2006; Oprins, Visschedijk et al. 2015; Tan, Ling et al. 2007; Djelil, Sanchez et al. 2014)
Environment	
Conditions for practical use of GBL and accessibility of educational game in terms of its environment	(Zaibon and Shiratuddin 2010; Djelil, Sanchez et al. 2014; Pappa and Pannese 2010)

Figure 1. Dimensions in the LEAGUE framework

4.2 Factors and Sub-Factors

Each dimension in the framework has a set of factors. Factors are intermediate level concepts and the framework entails such 22 factors (see Figure 3). Factors in the framework are further systematically categorized *into sub-factors* based on their theoretical construct, allowing a hierarchical decomposition. The sub-factors are easier to quantify and also serve to reduce the subjectivity often associated with assessing the factors. However, the choice of components for evaluation should depend on the overall evaluation objective and type of data required. Sub-factors are mostly devised by integration and mapping of conceptual elements using well-developed and widely excepted models/theories in areas where researchers had consensus in literature. Figure 2 presents the complete hierarchy and association including sub-factors to each factor.

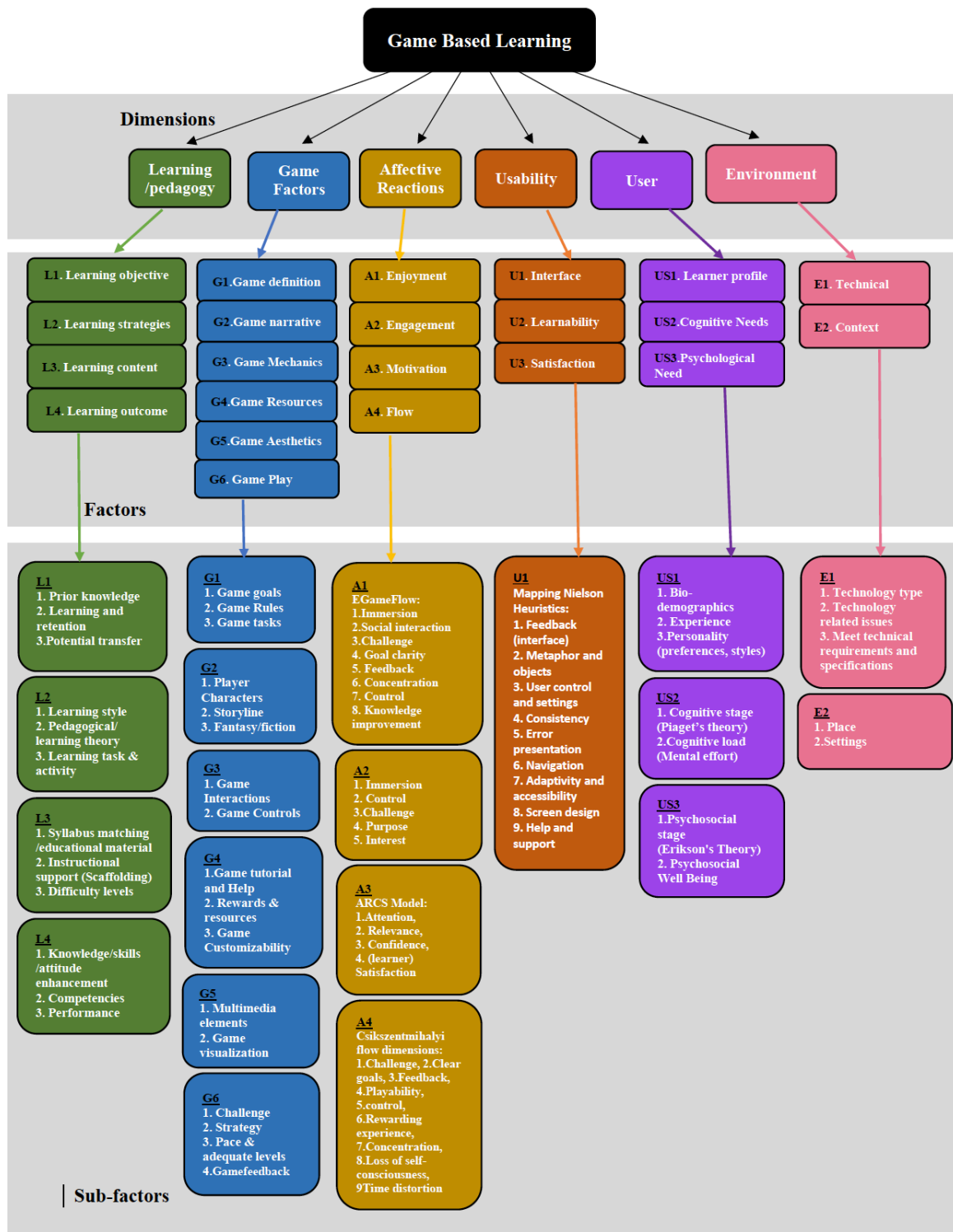


Figure 2: LEAGUE hierarchal structure and components

The sub-factors of learning objectives (L1, Figure 2) comes from (Aleven, V., et al. 2010). Learning task/activity (L2-3, Figure 2) is the specific task (designed in line with desired learning objectives and employed learning theory) that outline the interaction of learners, using specific game characteristics, orientated at specific outcomes (El-Sattar and Hussein 2016). The sub-factors of learning outcomes (L4, Figure 2) are adapted from the GEM model (Oprins, Visschedijk et al. 2015). The sub-factors of enjoyments (A1, Figure 2) are assimilated from EGameFlow (Fu, F.-L., et al. 2009) which uses flow as structural foundation, therefore has some common sub-factors as flow (Rêgo and de Medeiros 2015; Tseloudi and Tsiatsos 2015). The sub-factors of engagement (A2, Figure 2) are adopted from the framework by (Pourabdollahian, Taisch et al. 2012). The sub-factors of motivation

(A3, Figure 2) are adapted from the well-established ARCS model (Su, Chen et al. 2013). The sub-factors of flow (A4, Figure 2) are adapted from flow framework (Kiili, Lainema et al. 2014) and presents the original component of flow presented by Csikszentmihalyi (Perttula, Kiili et al. 2017). The sub-factors of interface (U1, Figure 2) are integrated by mapping the factors found in GBL literature to Nielsen’s heuristics (Yue and Zin 2009). This resulted in 9 sub-factors, one heuristic “help users recognize, diagnose, and recover from errors” could not be mapped to GBL literature. The analysis further clarified that in educational games the focus is on error prevention and confirmation messages rather than error messages. The review of GBL literature showed the lack of psychosocial indicators used for evaluating educational games. Although the importance of psychological needs and psychosocial stages is highlighted in (Tan, Ling et al. 2007), further details are not provided.

Learning Objective: Goals and purpose for transfer of knowledge and skills by educational games to users.	(Law and Sun 2012; Alfadhli and Alsumait 2015; Aleven, Myers et al. 2010; Marciano, Miranda et al. 2014)
Learning Strategies: Pedagogical theories, learning models and approaches to achieve learning objectives.	(De Freitas and Oliver 2006; Alfadhli and Alsumait 2015; El-Sattar and Hussein 2016)
Learning Content: Educational material (facts, data and information) provided used for learning in educational games.	(Marciano, Miranda et al. 2014; Alfadhli and Alsumait 2015)
Learning Outcome: Desired learning output (student achievements) from educational games.	(Alfadhli and Alsumait 2015; Wang, Liu et al. 2015; Oprins, Visschedijk et al. 2015)
Game Definition: Game goals depending on appropriate set of game rules defining the gaming tasks in games.	(Fu, Su et al. 2009; Oprins, Visschedijk et al. 2015; Djelil, Sanchez et al. 2014)
Game Narrative: Describes what happens in the virtual world including character development and plot.	(Omar and Jaafar 2010; Shi and Shih 2015; Mitgutsch and Alvarado 2012)
Game Mechanics: Defines the functioning, operation and interaction with the game world.	(Yue and Zin 2009; Omar and Jaafar 2010; Mitgutsch and Alvarado 2012; Shi and Shih 2015)
Game Resources: All resources provided to the player (should have both utility and scarcity).	(Alfadhli and Alsumait 2015; Wallner and Kriglstein 2011; Tan, Ling et al. 2007)
Game Aesthetics: Audio-visual language selected, conceptualized and used (images, text, audio, video, animation and multimedia).	(Mitgutsch and Alvarado 2012; Omar and Jaafar 2010)
Game Play: Process by which user reaches the goal which defines the set of challenges of problems the user face to win the game.	(Omar and Jaafar 2010)
Flow: State of absolute absorption in an activity which represents an optimal experience.	(Kiili, Lainema et al. 2014; Su, Chen et al. 2013; Fu, Su et al. 2009)
Motivation: Level of user involvement in participating and using an educational game by devoting additional time and effort.	(Djelil, Sanchez et al. 2014; Oprins, Visschedijk et al. 2015)
Engagement: User activity absorption and interest in activity or task, and users’ subjective acceptance of games reality combined with degree of focus on this realism.	(Oprins, Visschedijk et al. 2015)
Enjoyment: The extent of playing the educational game is considered pleasurable for an individual user.	(Giannakos 2013)
Interface: Component through which users interact with a game (elements utilized by user to interact with the educational game that influences its usability).	(Yue and Zin 2009; Omar and Jaafar 2010)
Learnability: Capability of an educational game to enable the user to easily learn to use specific functions.	(Yoon and Park 2013; Tseloudi and Tsiatsos 2015)
Satisfaction: Comfort and acceptability of an educational game to its users.	(Yoon and Park 2013; Tseloudi and Tsiatsos 2015; Ibrahim and Jaafar 2009)
Learner Profile: Attributes of a particular learner/ group.	(De Freitas and Oliver 2006; Djelil, Sanchez et al. 2014)
Cognitive Development: Games should suit learners’ cognitive development level.	(Tan, Ling et al. 2007)
Psychological Needs: Games should suit users’ psychological needs.	(Tan, Ling et al. 2007)
Technical: The accessibility and easiness of a user to enter the game world with the used technology and specified technical criteria.	(Zaibon and Shiratuddin 2010; Pappa and Pannese 2010)
Context: The particular context where GBL will take place using the educational game.	(De Freitas and Oliver 2006; Perttula, Kiili et al. 2017)

Figure 3. Factors in the LEAGUE framework

Therefore, the psychosocial well-being indicators (US3-2, Figure 2) are obtained from PSWBI (Negovan 2010). The PSWBI scale is validated with students for psychometric properties, construct validity, reliability and internal consistency, however, its use for educational games is to be explored. The sub-factors of technical (E1, Figure 2) include: *technology type* and *technology issues* (technology used for GBL e.g. issue of mobility in mobile technology), meet technical requirements (Zaibon and Shiratuddin 2010), (Pappa and Pannese 2010). The sub-factors of context (E2, Figure 2) are adopted from the framework by (De Freitas, S. and M. Oliver 2006). The sub-factors not directly integrated by using existing theories/models were analyzed using set of operations mentioned in Section 3.1, and selected based on scope, frequency of occurrence, relationship between codes and underlying meaning across codes.

4.3 Metrics

The metrics represent the lowest level in hierarchy which are used to collect evaluation data (Figure 1). The output of a metric interprets the status of sub-factor/factor: the degree to which the educational game possesses a given attribute. The choice of metrics depends on the type of data required, either subjective or objective, qualitative or quantitative. We identified 83 metrics from the corpus of analysis which were then coded and categorized into 5 types. The complete exhaustive list of metrics for each factor/subfactor is not provided here, rather the aim is to give guidance on the key metrics types used in GBL evaluation that can be utilized and adapted for different evaluation studies depending on the evaluation goals and selected factors/sub-factors. As a result, GBL metrics are coded into five main categories presented in Figure 4. First three metrics will result in objective data, while the last two will be useful for collecting subjective data. To illustrate, we introduce some common examples for each category in Figure 5.

Metrics	References
Scores	(Wang, Liu et al. 2015; Petri and von Wangenheim 2017; Papastergiou 2009; Tseloudi and Tsiatsos 2015; Yue and Zin 2009; Giannakos 2013; Yu, Hsiao et al. 2005)
Time	(Connolly, Stansfield et al. 2009; Eagle 2009; Rêgo and de Medeiros 2015; Virvou and Katsionis 2008; Ronimus, Kujala et al. 2014)
Number of occurrences	(Connolly, Stansfield et al. 2009); Mohamed, Yusoff et al. 2012; Eagle 2009; Yoon and Park 2013; Rêgo and de Medeiros 2015; Marciano, Miranda et al. 2014; Virvou and Katsionis 2008; Wallner and Kriglstein 2011)
Rating	(Mohamed, Yusoff et al. 2012; Yu, Hsiao et al. 2005; Fu, Su et al. 2009; Tseloudi and Tsiatsos 2015; Kiili, Lainema et al. 2014; Shiratuddin and Zaibon 2011; Yoon and Park 2013; Zaibon and Shiratuddin 2010; Oprins, Visschedijk et al. 2015; Marciano, Miranda et al. 2014; Shiratuddin and Zaibon 2011; de Lima, de Lima Salgado et al. 2015; Mei, Ku et al. 2015; Pourabdollahian, Taisch et al. 2012; Shi and Shih 2015; Giannakos 2013; Ronimus, Kujala et al. 2014)
reviews/responses/opinions	(De Freitas and Oliver 2006; Connolly, Stansfield et al. 2009; Wang, Liu et al. 2015; Omar and Jaafar 2010)

Figure 4. Metrics in the LEAGUE framework

Metrics	Examples
Scores	pre/post test score (Petri and von Wangenheim 2017), game performance/score, retention score (Wang, Liu et al. 2015) etc.
Time	Game session time (Rêgo and de Medeiros 2015), learning time, task completion time (Connolly, Stansfield et al. 2009), Time spent on each problem (Eagle 2009) etc.
Number of occurrences	No of usability problems, no of navigation problem (Virvou and Katsionis 2008), no of failures (Yoon and Park 2013), no of levels played (Connolly, Stansfield et al. 2009), no of overall attempts (Wallner and Kriglstein 2011), no of distraction occurrences (Virvou and Katsionis 2008) etc.
Rating	3,4,5,6,7point Likert scale ((Kiili, Lainema et al. 2014; Shiratuddin and Zaibon 2011; Fu, Su et al. 2009; Zaibon and Shiratuddin 2010; Yu, Hsiao et al. 2005), severity ratings (Mohamed, Yusoff et al. 2012), User task ranking (Connolly, Stansfield et al. 2009) etc.
Reviews/responses/opinions	expert review (Connolly, Stansfield et al. 2009), self-assessment (Bellotti, Kapralos et al. 2013), Mood and attitude statements (Connolly, Stansfield et al. 2009), game review based on selected (sub)factors (De Freitas and Oliver 2006), comments on positive/negative game aspects (Wang, Liu et al. 2015) etc.

Figure 5. Examples of the Metrics

4.4 Relations: High-level abstraction of Game-based Learning

The dimensions in GBL (depending on the evaluation objective) might be considered in isolation (picking and selecting components) for specific evaluation study. However, as a whole GBL dimensions are linked to each other in terms of cause and effect and can be viewed as a collective whole to understand the process. The highest abstraction of the framework is displayed in Figure 7. We identified 10 key relations from directed content analysis presented in Figure 6 (see Figure 7 for direction of relations).

The structure of GBL is depicted by the hierarchal layout presented in Figure 2. The high-level abstraction of the LEAGUE framework describes the internal operation of GBL and classifies the six dimensions into generic and

domain-specific. Learning, Game Factors, and Affective-Cognitive Reactions are the core domain-specific dimensions which represent the GBL phenomenon and process. Environment, Usability and Users are the core generic dimensions that influence the domain-specific dimensions and are essential for any software application to be effective for its users. An educational game is a game for education purposes that imparts learning by involving learners in the learning process. Game Factors generate Affective-Cognitive Reactions that absorb users in playing the game and positively influence the Learning. The main trick for an effective GBL approach is to keep generic dimensions in line while tweaking the Learning and Game Factors dimensions to integrate, create a balance and work in accordance with each other for enhancing the Affective Cognitive Reactions in order to meet the purpose of the educational game. We have introduced a term T-relation (see Figure 7) for the association between Learning, Game Factors and Affective-Cognitive Reactions as the core process of GBL, where integration of game and learning enhance affective reactions (Kiili, K. 2005). The generic dimensions not only influence the GBL phenomenon (domain-specific dimensions) but are also linked with each other. Usability should address the intended users and also cater the technical and context related specifications of environment. The technical specification and context (environment) should also map to the learner profile and capabilities (user). The overall process of GBL is a complex phenomenon and requires multidisciplinary approach. There is another view point to the LEAGUE framework which divides it into two views: technology-centric and human-centric. The technology-centric view includes three dimensions (Game Factors, Usability and Environment) related to technological aspects of game-based learning and human-centric view also include three dimensions (Learning, Affective-Cognitive Reaction and User) related to human aspects (identity, cognitive, behavioral) of game-based learning. The idea here is to model the technology centric dimensions in such a way that they facilitate human-centric dimensions.

Dimensions	Relation	References
Learning & Game Factors	Integration/ Balance	(De Freitas and Oliver 2006; Alevin, Myers et al. 2010;Zhang, Fan et al. 2010; Wallner and Kriglstein 2011; El-Sattar and Hussein 2016; Pappa and Pannese 2010)
ACR & Learning	Positive effect	(Kiili, Lainema et al. 2014; Thomas, Schott et al. 2004; Oprins, Visschedijk et al. 2015);Perttula, Kiili et al. 2017; Bellotti, Kapralos et al. 2013; Yu, Fu et al. 2009; Tseloudi and Tsiatsos 2015; Giannakos 2013)
Game Factors & ACR	Generate	(Rooney 2012;Abdul Jabbar and Felicia 2015; Ronimus, Kujala et al. 2014)
(Integration of Game Factors and Learning) & ACR	Enhance/facilitate /increase	(Kiili 2005; Zhang, Fan et al. 2010; Rooney 2012)
Usability & (Learning, Game factors, ACR)	Influence	(Gibson and Bell 2013; Shi and Shih 2015)
User & (Learning, Game factors, ACR)	Influence	(De Freitas and Oliver 2006; Abdul Jabbar and Felicia 2015;Mei, Ku et al. 2015;Yu, Hsiao et al. 2005;Mitgutsch and Alvarado 2012; Shi and Shih 2015)
Environment & (Learning, Game factors, ACR)	Influence	(De Freitas and Oliver 2006; Oprins, Visschedijk et al. 2015; Dondi and Moretti 2007; Virvou and Katsionis 2008)
Usability & Environment	Address/Cater	(Virvou and Katsionis 2008)
Usability & User	Address/ Cater	(Virvou and Katsionis 2008; Mei, Ku et al. 2015)
User & Environment	Map	(De Freitas and Oliver 2006)

Figure 6: Relations in LEAGUE

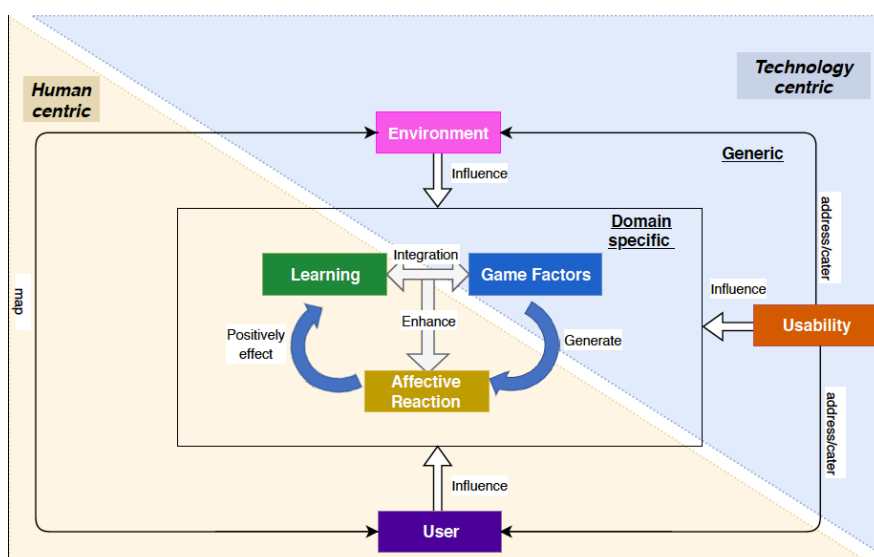


Figure 7 High level abstraction of LEAGUE

5. Discussion and Conclusion

This study expands on GBL evaluation literature to overcome the shortcomings in current research (problem areas highlighted in introduction) by conducting directed content analysis. The results of analysis are translated into a conceptual hierarchical framework LEAGUE, which shows that multidimensionality of GBL requires evaluation of several aspects referred as six core dimensions (addressing RQ1) including: Learning, Game Factors, Affective-Cognitive Reactions, Usability, User and Environment. Each dimension focuses on certain factors and sub-factors that constitute that aspect, and metrics are required to assess them. The framework presents 22 factors (addressing RQ2), 74 sub-factors (addressing RQ3) and 5 metrics categories (addressing RQ4). The dimensions of GBL are related to each other and it is essential to assess the relations presented as high abstraction of LEAGUE for greater insights into educational games (addressing RQ5). The framework provides a detailed picture of GBL that will guide not only researchers and evaluators, but also designers and developers of educational games. The proposed framework is built on components grounded in theory. Each component has strong basis for formation that is supported by theoretical constructs in GBL literature and not merely based on suspicion. Most of the existing GBL frameworks focus on some specific elements which makes them difficult to use in practice when the target genre differs from default game genres used in research (Shi and Shih 2015). Dimensions presented in LEAGUE are higher-level concept and not restricted by the game genre. The framework can be used from top to bottom as well as from bottom to top depending on evaluation goals.

The proposed framework can be useful for different stages (planning, design, analysis) of GBL evaluation and can aid at the following steps: 1) Evaluation plan/objective: Formulate research questions and determine the main objectives to evaluate GBL application using LEAGUE dimensions; 2a) Evaluation design/strategy: use LEAGUE factors/sub-factors to identify the components that are of particular interest in accordance with evaluation type (qualitative/quantitative) and method; 2.b) Design data collection using LEAGUE metrics to select relevant metric for each factor/ sub-factor based on required type of data (subjective/objective); and 3) Conduct evaluation and Analyze results: select appropriate evaluation design and also draw conclusions by considering the relations in high level abstraction of LEAGUE.

The proposed framework is employed to evaluate an educational game for reading skills of children. The complete process of using the framework for planning and conducting evaluation study and the results of evaluation will be presented in another paper, and framework components will be further validated and developed. Future research will focus on automating or partially automating GBL evaluation using proposed framework and game data logs.

References

- Ahmad, M., et al. (2015). Towards an Effective Modelling and Development of Educational Games with Subject-Matter: A Multi-Domain Framework. IT Convergence and Security (ICITCS), 2015 5th International Conference on, IEEE.
- Aleven, V., et al. (2010). Toward a framework for the analysis and design of educational games. Digital Game and Intelligent Toy Enhanced Learning (DIGITEL), 2010 Third IEEE International Conference on, IEEE.
- Amory, A. (2007). "Game object model version II: a theoretical framework for educational game development II." *Educational Technology Research and Development* 55(1): 51-77.
- Annetta, L. A. (2010). "The "I's" have it: A framework for serious educational game design." *Review of General Psychology* 14(2): 105.
- Bellotti, F., et al. (2013). "Assessment in and of serious games: an overview." *Advances in Human-Computer Interaction* 2013: 1.
- Chorianopoulos, K. and M. N. Giannakos (2014). "Design principles for serious video games in mathematics education: from theory to practice." *International Journal of Serious Game* 1(3): 51-59.
- Connolly, T., et al. (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.
- De Freitas, S. and M. Oliver (2006). "How can exploratory learning with games and simulations within the curriculum be most effectively evaluated?" *Computers & Education* 46(3): 249-264.
- De Lima, L. G. R., et al. (2015). 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.
- Djelil, F., et al. (2014). Towards a learning game evaluation methodology in a training context: A literature review. European Conference on Games Based Learning, Academic Conferences International Limited.
- Dondi, C. and M. Moretti (2007). "A methodological proposal for learning games selection and quality assessment." *British Journal of Educational Technology* 38(3): 502-512.
- El-Sattar, A. and H. K. Hussein (2016). Learning Islamic Principles with Serious Games. Proceedings of the 13th International Conference on Advances in Computer Entertainment Technology, ACM.
- Engl, S. and L. E. Nacke (2013). "Contextual influences on mobile player experience—A game user experience model." *Entertainment Computing* 4(1): 83-91.
- Fu, F.-L., et al. (2009). "EGameFlow: A scale to measure learners' enjoyment of e-learning games." *Computers & Education* 52(1): 101-11.
- Giannakos, M. N. (2013). "Enjoy and learn with educational games: Examining factors affecting learning performance." *Computers & Education* 68: 429-439.
- Hsieh, H.-F. and S. E. Shannon (2005). "Three approaches to qualitative content analysis." *Qualitative health research* 15(9): 1277-1288.

Ibrahim, R. and A. Jaafar (2009). Educational games (EG) design framework: combination of game design, pedagogy and content modeling. *Electrical Engineering and Informatics, 2009. ICEEI'09. International Conference on*, IEEE.

Kececi, N. and A. Abran (2001). "ANALYZING, MEASURING & ASSESSING SOFTWARE QUALITY WITHIN A LOGIC-BASED GRAPHICAL FRAMEWORK." *Universite du Quebec a Montreal*.

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

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

Liao, Y. H. and C.-Y. Shen (2012). Heuristic evaluation of digital game based learning: a case study. *Digital Game and Intelligent Toy Enhanced Learning (DIGITEL), 2012 IEEE Fourth International Conference on*, IEEE.

Loh, C. S., et al. (2015). *Serious games analytics: Theoretical framework*. Serious games analytics, Springer: 3-29.

Marciano, J. N., et al. (2014). "Evaluating multiple aspects of educational computer games: literature review and case study." *International Journal of Computer Games Technology* 2014: 14.

Mei, S.-Y., et al. (2015). *Interface Evaluation of a Game-Based Learning System: Experts vs. Novices*. *Advanced Applied Informatics (IIAI-AAI), 2015 IIAI 4th International Congress on*, IEEE.

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

Mohamed, H., et al. (2012). Quantitative analysis in a heuristic evaluation for usability of educational computer game (UsaECG). *Information Retrieval & Knowledge Management (CAMP), 2012 International Conference on*, IEEE.

Negovan, V. (2010). "Dimensions of students' psychosocial well-being and their measurement: Validation of a students' Psychosocial Well Being Inventory." *Europe's Journal of Psychology* 6(2): 85-104.

Omar, H. M. and A. Jaafar (2010). "Conceptual framework for a heuristics based methodology for interface evaluation of educational games." *Computer and Information Science* 3(2): 211.

Oprins, E., et al. (2015). "The game-based learning evaluation model (GEM): measuring the effectiveness of serious games using a standardised method." *International Journal of Technology Enhanced Learning* 7(4): 326-345.

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

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

Perttula, A., et al. (2017). "Flow experience in game based learning—a systematic literature review." *International Journal of Serious Games* 4(1).

Petri, G. and C. G. von Wangenheim (2016). "How to Evaluate Educational Games: a Systematic." *Journal of Universal Computer Science* 22(7): 992-1021.

Petri, G. and C. G. von Wangenheim (2017). "How games for computing education are evaluated? A systematic literature review." *Computers & Education* 107: 68-90.

Pourabdollahian, B., et al. (2012). "Serious games in manufacturing education: Evaluation of learners' engagement." *Procedia Computer Science* 15: 256-265.

Rêgo, M. B. B. A. B. and I. de Medeiros (2015). "HEEG: Heuristic Evaluation for Educational Games."

Rooney, P. (2012). "A theoretical framework for serious game design: exploring pedagogy, play and fidelity and their implications for the design process."

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

Spiggle, S. (1994). "Analysis and interpretation of qualitative data in consumer research." *Journal of consumer research* 21(3): 491-503.

Su, C.-H., et al. (2013). "Rough set theory based fuzzy TOPSIS on serious game design evaluation framework." *Mathematical Problems in Engineering* 2013.

Tahir, R. and A. I. Wang (2017). *State of the art in Game Based Learning: Dimensions for Evaluating Educational Games*. *European Conference on Games Based Learning, Academic Conferences International Limited*.

Tan, P.-H., et al. (2007). *Adaptive digital game-based learning framework*. *Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts*, ACM.

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

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

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

Wang, Y., et al. (2015). "An Evaluation Framework for Game-Based Learning."

Yu, K.-C., et al. (2005). *The implementation and evaluation of educational online gaming system*. *Information Technology: Research and Education, 2005. ITRE 2005. 3rd International Conference on*, IEEE.

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

Zaibon, S. B. and N. Shiratuddin (2010). *Heuristics evaluation strategy for mobile game-based learning*. *Wireless, Mobile and Ubiquitous Technologies in Education (WMUTE), 2010 6th IEEE International Conference on*, IEEE.

Zhang, H.-f., et al. (2010). *Research on the design and evaluation of educational games based on the RETAIN model*. *Knowledge Acquisition and Modeling (KAM), 2010 3rd International Symposium on*, IEEE.

Zhang, Y. and B. M. Wildemuth (2005). "Qualitative Analysis of Content by." *Analysis* 1(2): 1-12.