Informatics in Education, 2019, Vol. 18, No. 2, 427–442

DOI: 10.15388/infedu.2019.20

Heuristic Evaluation on Usability of Educational Games: A Systematic Review

Estela Aparecida Oliveira VIEIRA, Aleph Campos da SILVEIRA, Ronei Ximenes MARTINS

Departament of Education, Federal University of Lavras – Minas Gerais, Brazil
e-mail:{estela.ap.o.vieira; alephcampos; profroneimartins}@gmail.com

Received: Decenber 2018

Abstract. This work presents a systematic review whose objective was to identify heuristics applicable to the evaluation of the usability of educational games. Heuristics are usability engineering methods that aim to detect problems in the use of a system during its development and / or when its interface is in interaction with the user. Therefore, applying heuristics is an essential part of developing digital educational games. Search sources were articles available in all the databases present in the Capes / MEC / Brazil periodicals portal, in the available languages. The descriptors adopted were "educational games", "heuristic" and "usability" in Boolean search in titles, abstracts and keywords, with AND operator, for publications starting in 2014. The inclusion criteria were: (a) articles with a clear description of the methodology used in the usability analysis; (b) studies presenting primary data and (c) articles whose focus corresponds to the investigated question. Two examiners conducted the searches in the databases and a third the evaluation and general review of the data. Initially, 93 articles were identified, of which 19 were repeated, 5 were literature reviews. Of the 69 that remained, 57 were elected as not eligible with only 12 selected for full studies, of which 6 entered the final review. With this review we can deduce that the field of heuristics and usability for educational games is still little explored, with few specific evaluations validated or in the process of validation, requiring greater investment in the area. Through this review, we found at least one heuristic that meets the usability evaluation of educational software: Game User Experience Satisfaction Scale (GUESS).

Keywords: education, educational games, heuristic evaluation, usability.

1. Introduction

The concept of usability is not only applied to software but also related to the ease of use and accomplishment of certain tasks when interacting with the object through its interface. Usability proposes basic requirements for the functionality of an object in use, i.e that it works properly when used by an individual with reasonable skill and experience (Cheng and Mustafa, 2015).

Authors like Hermawati and Lawson (2016) emphasize the relevance of usability, which is the key to ensure the success of a system. They depart from the concept of usability defined by ISO-9241-11[1], which is also explained in the Brazilian standardization in the following words: "the product must respond to specific objectives with effectiveness, efficiency and satisfaction in a specific context of use"(ABNT[2], 2002, p. 3).

In order to sort usability from heuristics it is necessary to understand that usability refers to the ability of a product to be used by a minimally skilled user and can be projected within the scope of creation of a product and evaluated by means of inspections or tests performed by specialists and by potential users. Heuristic evaluation is a usability inspection method widely used to find structural and/or heuristic problems from the interface review, taking into account aspects of the user experience (Quinõnes and Rusu, 2017). Heuristics can be specific or general and applied during its development or at its ends.

Although general heuristics are suitable for evaluating most user interfaces, it is still necessary to establish heuristics for specific domains (Phan *et al.*, 2016; Quinõnes and Rusu, 2017). Even with specific heuristics, they present some gaps, such as absence of validation after the heuristic proposition and the lack of robustness and rigour of the validation method adopted. Hermawati and Lawson (2016) reiterate this gap by emphasizing that the lack of validation quality also affects the effort to improve existing heuristics applied to a specific domain because its weaknesses are not inspected.

In the case of specific domains we can point educational games, which, because they present attractive contexts, immersive and interactive activities contribute to obtain more space in the educational games market and consequently the need for content validation, gameplay, motivation among other elements. However, game designers have problems when confronted with fantasy content and context integration in serious games, since there is a lack of viable design and evaluation models for them. Two factors are pointed out for the failure of these games (Yeni and Cagiltay, 2017): the relation to academic aspects and entertainment. Content and entertainment integration alone does not guarantee that the game is effective in terms of entertainment or motivation, or in meeting its educational or business goals. Games that can handle both instructional content and fantasy components and entertain players can be considered successful in terms of effective design.

Yeni and Cagiltay (2017) in the study of heuristic validation of their mathematics educational game used the heuristic evaluation proposed by Nielsen and by the end of the article they point out the limitations of the research, since the obtaining of opinions and research scores were carried out with 15 participants who work in different specialized areas in Istanbul and who are interested in video games. The authors pointed out to the need to reapply a similar study with different educational games in order to increase the

[1] Defines usability and explains how to identify the information needed to be considered in the specification or usability assessment of a visual interaction device in terms of performance measures and user satisfaction.

[2] Associação Brasileira de Normas Técnicas, or *Brazilian Association of Technical Standards.*

reliability of the findings presented to game designers to design and build educational games more integrated, both academically and in entertainment.

Thus, even with the creation of new sets of usability heuristics for specific domains, they do not yet evaluate particular features of the different types of software and applications (Quinões and Rusu, 2017). Even with several studies that focus on the design of new heuristic sets that can identify usability problems for specific application domains, those are conducted in an informal process. There are also studies that apply methodologies to define, validate and refine the set of proposed heuristics, but they do not formalize the proposal, and therefore do not contribute to the process of building specific usability heuristics.

Regarding the step for its application, the heuristic evaluation can be performed at the end of the development or in a modular way, along with the development process. If the team opts for evaluation in a cascading approach such a *Analyze Design Develop Implement and Evaluate* (ADDIE) this can lead to re factoring issues[3], since the usability will be realized only at the end of the development of the game. Therefore, some authors (Busch *et al.*, 2015; Hermawati and Lawson 2016) propose a modular evaluation approach. In other words, they propose a comprehensive iterative development methodology, both in terms of usability/user experience as in its portability of its standards to other projects.

Most recurrent approaches conducted by Quinões and Rusu (2017) demonstrate that the development of heuristics are based on pre-existing heuristics and many are established as an extension or adaptation of Nielsen's heuristics (1994, 1995), as well as the new sets of heuristics that focus on the specific domain.

Heuristic evaluation is still the preferred method to evaluate usability in games, when experts conduct this evaluation, being the most used guideline to evaluate the usability of games is still the proposal of Nielsen, which is focused on generic software (Yanez-Gomez, 2019) thus not covering important aspects in games such as mobility, multiplayer interactions, fun and gameplay, etc. If on the one hand the fragmentation and dispersion of the new proposals leads to perfectly evaluate an aspect of usability or particularities of the game genre, on the other they may ignore other aspects that could potentially be covered by other heuristics. Thus, the authors propose metaheuristics, where non-applicable heuristics are minimized and the coverage of usability aspects and game specificities is maximized.

Yanez-Gomez and his colleagues present the MUSE (*Meta-heUristics uSability Evaluation*) tool for gaming, which allows the reconstruction of heuristic guidelines based on the selection of metadata to obtain a customized list for each real evaluation case. MUSE is a tool in development that presents itself as promising and had its first prototype approved. In this tool, criteria drawn from the literature are compiled based on game reviews or expert opinions under specific criteria determined by the researchers related to the characteristics of the game being evaluated, the usability aspects that they want to evaluate and other circumstances of the evaluation, thus developing a tool for automatic creation of custom usability heuristic guidelines for game evaluation.

[3] process of modifying a software system seeking its optimization, without changing its functionality.

As we can see, the meta-data has gained space in the game market, consequently, the use of this data in the usability evaluation. If initially the development of games is marked, among other factors, by two important aspects, modeling and prediction of individual behavior, in the case of educational games is not the rule. Its application is often extremely difficult due to the vast wide field of actions that can be created by educational games.

For this reason, according to Hooshyar *et al.* (2017), data-based approaches have shown promise, in part because they do not depend entirely on specialized knowledge. In their review study they scored three major movements: 1) the goal of data-based approaches in modeling educational games, such as behavior modeling, goal recognition, and procedural content generation; 2) approaches used in such modeling; 3) current challenges of using data-based approaches in modeling educational games. In this study the authors point out challenges in the area such as the lack of adequate and rich data publicly available to the researchers, the lack of a data-based method to identify conceptual log data resources, hybrid player modeling approaches and mining technical data for individual forecasting.

Another point to be remarked in educational games is the development of custom games, structured by pre-made algorithms demands a restructuring of the game's own heuristics. As is the case in the study by Hooshyar *et al.* (2018) that brings to the center of this discussion the adaptation of content to individual needs through the technique of Procedural Content Generation (PCG). With the support of algorithms are offered games with contents and levels of personalized difficulties. The advantage is that the proposed method does not depend on the intuition of the designer in the application of the game content to suit the skills of the player. In the case of previous approaches, they predominantly target the generation process through goals or heuristics defined by the designer, so the author is obliged to dictate specific constraints or evaluative heuristics to evaluate the suitability of the game content, as well as estimation of the effect of game content on the learning process. In contrast, the proposed approach does not require any intuition on the part of the author or instructor to connect the content of the game to the abilities of the players or to the learning objectives. Instead, a data-driven approach is employed to refine the appropriateness function for content evaluation and prediction of player capabilities. To assert the efficacy of the proposed method, the authors evaluated the data-based approach and a heuristic approach leading to two main outcomes: users perceived higher performance gains by playing content tailored to their abilities compared to non-customized game content; and the data-based approach was more effective in generating content that closely matches a specific goal of player performance than the heuristic-based approach.

Our interest is to identify subsidies for a clear and formal usability process of educational games. Thus, by applying usability heuristics in educational games based on existing methodologies and clearly defined phases, we will be able to specify, validate and refine products in development and to be developed in our research laboratory of innovative pedagogical practices. In this sense, the challenge we set ourselves was to seek a heuristic that would respond to the demand for educational game projects.

1.1. *Clarification on Heuristic Evaluation and Usability Testing*

During the review, some doubt was identified in the terms "heuristic evaluation" and "usability tests" with some authors considering them as synonyms (Marciano *et al.*, 2014) while others approach them in a different way (Busch *et al.*, 2015; Malliaraki *et al.*, 2014; Aleem *et al.*, 2016). It was carried out a brief research on the subject in order to clarify such doubt, assuming that the approach is epistemological and not on the efficiency and utility of each one.

Heuristic evaluation, in Nielsen's definition, is an inspection technique that aims to identify usability problems. Taking into consideration aspects of the user experience, a heuristic evaluation can be defined as a revision of its interface to identify structural and/or heuristic problems (Nielsen, 1995). A heuristic evaluation of usability should not, however, replace a usability test and "although heuristics refer to criteria that affect the usability of your site, the problems identified in a heuristic assessment are different from those found in a usability test" (Moraes and Rosa, 2010).

Now, "Usability tests" are performed in conjunction with the user, where the main objective is to observe the product as a product (be it a website, a web application, and so forth) before or after its official launch. In practical terms, every "heuristic evaluation" is a "usability test", but not every usability test is necessarily a heuristic evaluation because it must be in accordance with Nielsen and Molich's "10 Usability Heuristics."

2. Methodology

According to Kitchenham and Charters (2007), systematic literature review (SLR) is a means of identifying, evaluating and interpreting the available scientific results relevant to a particular research problem or phenomenon of interest, in this case, the evaluation of usability of educational games through heuristics.

Therefore, the focus of this SLR is to gather subsidies for a formal process of evaluation of usability of educational games through heuristics, since it was verified through preliminary Boolean search that there are no SLR papers of this nature.

In order to perform the systematic review, the method described by Kitchenham and Charters (2007) was applied from the following research question: "What software usability heuristics can be applied to the evaluation of educational games?". So the descriptors adopted were "educational games", "heuristic" and "usability" in Boolean search in titles, abstracts and keywords, with the AND operator, for publications starting in 2014.

The search for papers was carried out in the meta-search made available by the CAPES / Brazil Portal of Periodicals[4] that offers access to the complete texts selected in approximately 21,500 journals international journals in 126 databases with document summaries in all areas of knowledge, including Scopus, Science Direct, Web of Science, Springer Link, ACM Digital Library, ERIC, Engineering Research Database, Directory of Open Access Journals. As a date limit, the period from 2014 to 2019 was adopted.

[4] `http://www.peri_odicos.capes.gov.br`

The protocol for papers inclusion was applied by three reviewers, one from education's field and another from information technology and an evaluator responsible for the final review, with the following criteria: (a) articles with a clear description of the methodology used in the analysis of usability; (b) studies presenting primary data and (c) articles whose focus corresponds to the application of heuristics. Fig. 1 shows the Flowchart used to select the articles.

In the initial search, performed by double-blind researchers, 93 articles were found, of which 24 were repeated or systematic reviews, remaining 69 articles as shown in the flowchart (Fig. 1). With the Title and Summary reading, 17 were excluded because they did not meet the pre-established criteria. 40 studies were considered relevant but did not meet the criteria of clear description of the methodology used in the usability analysis and / or of working with primary data and 12 articles considered to correspond to the research question: Designing Educational Games for Computer Programming A holistic Framework (Malliaraki *et al.*, 2014) ; Evaluating multiple aspects of educational computer games literature review and case study (Marciano *et al.*, 2014); Game development software engineering process life cycle a systematic review (Aleem *et al.*, 2016); Critical Success Factors to Improve the Game Development Process (Aleem *et al.*, 2016); The Development and Validation of the Game User Experience Satisfaction Scale (GUESS) (Phan *et al.*, 2016); Designing educational games through a conceptual model based

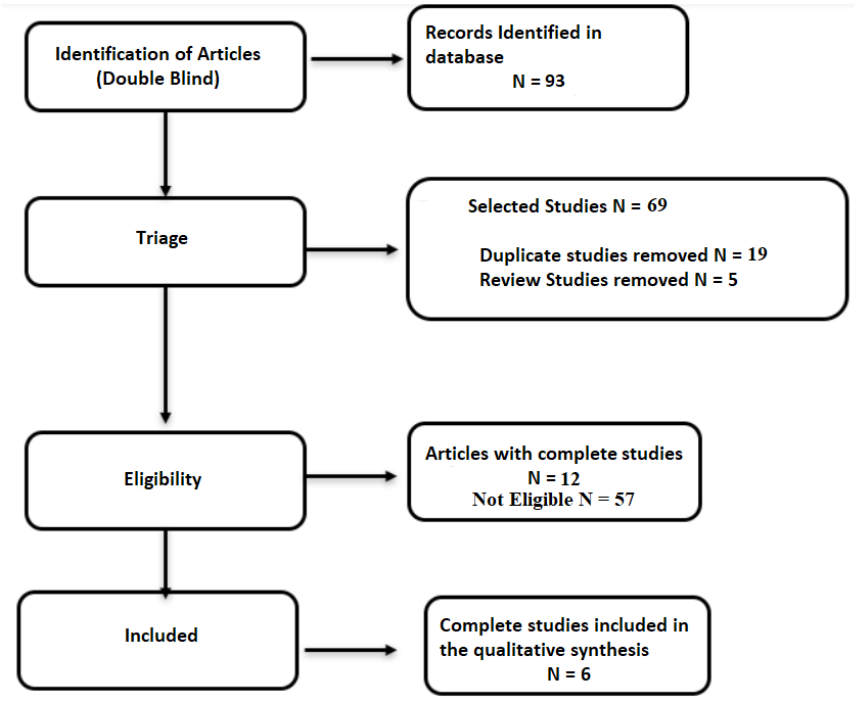

Fig. 1. Flowchart selection of articles. Source: Hooshyar *et al.* 2017.

on rules and scenarios (Zarraonandia *et al.*, 2014); Playability And Social Experiences Acceptance Study Of Interactive Video Puzzle (HUI *et al.*, 2016); GameFlow in Different Game Genres and Platforms (Sweetser *et al.*, 2017); Software Quality Requirement Analysis on Educational Mobile Game with Tourism Theme (Trisnadoli *et al.*, 2016); Agile human centered methodologies to develop educational software (González *et al.*, 2015); Gamification in theory and action Asurvey (Seaborn and Fels, 2015); Mapping quality requirements for pervasive mobile games e Software assurance practices for mobile applications (Valente *et al.*, 2015).

After the complete reading of the 12 articles, six were identified as being capable of detailed analysis, since they included not only an overview of several usability evaluation frameworks by heuristics, but also to propose and validate their own methodology, that is, they present primary data.

3. Results

For the analysis and synthesis of the selected articles, in order to answer that software usability heuristics can be applied to the evaluation of educational games, we adopted the Line of Argument Synthesis approach described by Kitchenham and Charters (2007), based on Noblit and Hare (1988). This approach is used for inferences about a topic as a whole from a set of selective studies that look at a part of the issue. For the organization of the inferences, three categories were established, which emerged from the frequency analysis of the content established by Bardin (2010). These categories identify the stages in which the heuristic evaluations were carried out in the research, they are: (a) concomitantly to the development; (b) at the end of the project and finally (c) using metadata to facilitate the supply and demand of the product.

3.1. *First Category: Approaches During the Development*

Marciano *et al.* (2014) provides a literature review on assessments of multiple aspects of software, followed by a more focused review of educational games and finally a case study where one of these methods is validated with a previously developed game. Several evaluation models were analyzed by the authors, such as HEP (Heuristics to Evaluate Playability[5]), developed by Desurvire and collaborators (2004 apud Marciano *et al.*, 2014), WIMP (Windows, Icons, Menus and Pointers[6]) by Ivory and Hearst (2010) apud Marciano *et al.*, 2014), Playability Heuristics for Educational Games (PHEG) (2008 apud Marciano *et al.*, 2014), among others. The last one , developed by Omar and Jaffar (2010 apud Marciano *et al.*, 2014), was chosen as a model for application in the case study because it covers multiple aspects of interface exclusively linked to educational games.

[5] Heuristic for Gameplay Evaluation.

[6] Windows, Icons, Menus and Pointers.

The game used in this case study is called Karuchã Ships Invaders CALL Game, an educational application inspired by Space Invaders[7] intended for teaching the Japanese language. The PHEG heuristics emphasizes five main aspects:

1. Interface – Navigation, interactivity, design and consistency.
2. Pedagogy / Education – objectives, challenges, feedback, control.
3. Content – Aspects related to the educational material made available.
4. Multimedia – Use of texts, audio, video, animations et cetera.
5. Gameplay – Player control over the game, levels, pitch, balance, and so on.

Marciano *et al.* (2014) highlights that methodologies are and should be adapted to the need and context that they are applied, thus, the heuristics used in the analyzed article are adaptive and can be performed for validation of games by both developers and end users, so the results can be used during the period of development or post-launch. The evaluation revealed it cover interest points like interface, education, content, multimedia and gameplay. As noted by the authors, PHEG proved overly technical and did not evaluates user satisfaction or learning metrics, not considering emotional responses to the game. Thus, it was considered limited by not identifying all aspects needed for heuristic evaluation, revealing the need for evaluations by other methods.

The second article analyzed (Busch *et al.*, 2015) advocates a comprehensive iterative development methodology. The authors propose it to either evaluate terms of usability/ user experience or portability of mechanics to other projects. This proposal is a counterpoint to ADDIE[8], or "waterfall", popular in the construction of traditional software that suffer from problems reworking when applied to game development.

To validate this methodology, with each new development phase, five users performed evaluations based on lessons learned previously. In other words, in order to overcome the problems of traditional methodologies, the evaluation phase was repeated at each new iteration, allowing short and concise software rework. Economically solid, this approach rules out the need for a greater number of test users because when performed in smaller parts, the number of problems and bugs are more easily detected by fewer evaluators.

Busch *et al.*, (2015) then proposed two questionnaires to be answered: ISONORM (Pataki *et al.*, 2010) and User-Experience-Questionnaire (Laugwitz *et al.*, 2008). Over the course of two days, tests were performed at each development iteration: a pretest, followed by a post test after 25 minutes of play, and finally post testing after completion of development.

The first questionnaire, ISONORM, developed by Prümper (1993), aims to evaluate the software according to the recommendations of ISO 9241, known as "Ergonomics of Human-System Interaction". This norm brings a set of usability heuristics that apply to people's interaction and information systems. The User-Experience-Questionnaire (UEQ) is a usability test (Laugwtiz *et al.*, 2008) developed to look for a fast and efficient UX (User Experience) measurement. Initially containing 229 potential items related to

[7] Game released in 1978 by Talto Corporation. It consists on hitting several targets that advance against the fixed position of the player.

[8] Analysis, Design, Development, Implementation, Evaluation.

user experience, it was later reduced to 80 items in order to guarantee the minimum effort required with maximum analysis potential. Combined, the heuristic assessment of usability by professionals and the usability test with users brought positive results that helped to conclude that if used separately problematic interface or usability objects would not have been identified.

The third article analyzed (Malliarakis *et al.*, 2014) proposes, from the abstraction of models previously studied by the authors, to merge all the important points in a single framework without redundancies that was validated in the construction of a game developed by them (CMX). This proposed framework aims to provide support for creating games that are motivational and interesting enough to support computer programming courses. Fig. 2 presents a representation of the CMX Framework, adopted by the authors of the analyzed article.

CMX[9] is a proposed model that aspire to include concepts that need to be presented in any game, such as: infrastructure, learning objects, pedagogy, learning outcomes, user, scenario and activities. Even though it is an ad hoc mechanism, it is satisfying abstracted tool that can be adapted in the development of other games and applications.

These articles analyzed so far have the differential of carrying out heuristic and/or usability testing approaches concomitantly with the development itself, instead of being applied exclusively to the end of the project as a method of evaluation and/or validation.

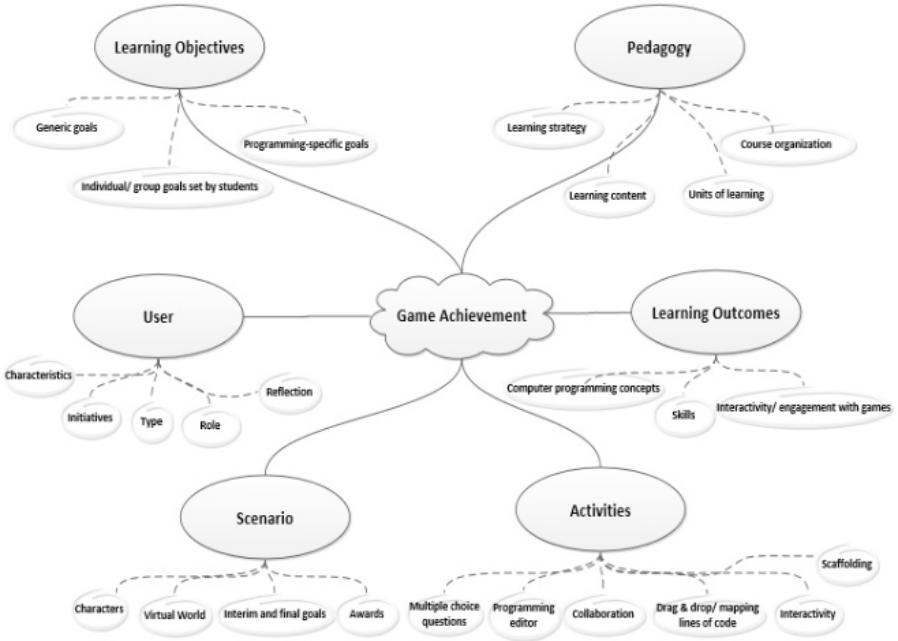

Fig. 2. Representation of the CMX Framework. source: Malliarakis *et al.*, 2014.

[9] Named CMX Design Framework, built with the support of the framework, is a Massively multiplayer online role-playing game (MMORPG) that seeks to familiarize high school students with computer programming.

Such perspective is important because it diverge from ADDIE or "cascade/waterfall" models, commonly used in software development, which as stated earlier, suffer from problems when implemented in projects that do not follow sequential developmental ow, resulting in difficulties when problems are identified later.

3.2. *Second Category: Post-Development Approach*

As in other articles, the model studied here was adapted to the context used, this time focusing on mobile games, bringing 17 requirements that should be evaluated. From a case study of an educational game about tourism, the authors attempted to validate their research. (Trisnadoli *et al.*, 2016). The main objective of this article was to inspect and propose quality requirements for tourism educational games, based on the analysis of the "A Quality Model Proposal for Mobile Games" developed by Trisnadoli *et al.*, (2015 apud Trisnadoli *et al.*, 2016) and Quality Requirements for Mobile Education Games based on User Perspective (Trisnadoli *et al.*, 2016).

The authors identified the educational potential of games, assigning high importance to the role of "non-formal" training instead of formal education received in classrooms. Two evaluations were carried out: the first one was applied through interviews with 130 people from different contexts, not necessarily game users. The second one, more specific, was applied in an on-line forum involving 35 people. Merging the results, a list of 17 Quality Requirements for Mobile Educational Games was proposed for the theme of Tourism, presented in Fig. 3.

Trisnadoli *et al.* (2017) concluded that quality requirements are strictly dependent on the type of games, i.e. a model must always be adapted to its context.

The article from Phan *et al.* (2016) aimed to develop and validate psychometrically a new instrument to comprehensively measure the satisfaction of video game

No.	Quality Requirements
1	Game provides a clear goals
2	Game is easy to understand
3	Game can be played in a short time
4	Game have complete features
5	Control of the game has been consistently in according with the standards
6	Screen layout is comfortable to be seen
7	Audio of the game was comfortable and supportive with a good game
8	The mobile device screen is appropriate to the game
9	The experience of playing is Fun
10	Storyline or scenario of the game is interesting
11	No repetition that make the game boring
12	Game control is appropriate and flexible
13	Players gain knowledge from the game
14	Game is simple but meaningful
15	Content on tourism must be update
16	Language usage appropriate to target of player
17	Having educate messages behind the entertainment given

Fig. 3. Quality Requirements for Mobile Educational Games for the Tourism Theme.
Source:Trisnadoli *et al.*, 2016.

users based on key factors such as commercial and universal appeal as well as originality (Chalker, 2008; Shelley, 2001; Totilo; 2012) or personal interest, attention and immersion (Chen *et al.*, 2011 apud Phan *et al.*, 2016). The pretext presented by the authors were that it is difficult to obtain quality feedback from test sessions without a game quality assessment tool. Therefore, it is needed to obtain a psychologically valid and comprehensive game scale that is appropriate for game testing and evaluation purposes.

Regarding the methodology in the development process and validation of this new scale, the authors argue that they followed the current best practices of development and validation, with Exploratory Factor Analysis (EFA) followed by Factorial Confirmatory Analysis (CFA) (Cabrera-Nguyen, 2010; Hinkin, 1998; Worthington and Whittaker, 2006 apud Phan *et al.*, 2016). Based on these selected practices, a mixed method was Consisting on generating sets of items, expert review, pilot questionnaire study, exploratory factorial analysis (N = 629) and confirmatory factorial analysis (N = 729) that could be used both in the post-production phase, when estimating user satisfaction, as well as development tool, allowing an insight into the state of game construction.

As a research product, was presented a new instrument called GUESS, with nine sub scores emerged:

1. Usability / Reproduction.
2. Narratives.
3. Playing Involvement.
4. Pleasure.
5. Creative Freedom.
6. Audio Aesthetics.
7. Personal Gratification.
8. Social Connectivity.
9. Visual Aesthetics.

GUESS demonstrated content validity, internal consistency, and convergent and discriminant validity. The authors concluded that GUESS was developed and validated based on the evaluation of more than 450 video game titles in various popular genres. Thus, it can be applied to various types of video games in the industry, both as a way to assess which aspects of a game contribute to user satisfaction and as a tool to help introduce users to their gaming experience.

Despite the method has demonstrated strong psychometric properties, high content validity and internal consistency, the authors of this new scale have emphasized there is little information about their score pattern in larger varieties of games since it has been applied only in popular commercial games well accepted by the public, and there is a shortage of evaluations outside that entertainment market. Therefore, as a result, there is little evidence of its application in educational games (Phan *et al.*, 2016).

The articles in this category are similar to the of methods based on the construction of standardized requirements (in the form of list or scale) and used to evaluate products already developed. This approach is the traditional and most applied in software development.

3.3. *Third Category: Items of Interest and Acquisition of Games by Users: Metadata*

Opposing the previous articles, the study proposed by Lee *et al.*, (2016) aims to understand the needs and behaviour of video game players and stakeholders. The proposal is to understand how they choose video games and what information is needed in the search and selection of video games. This work points out the importance of prior research of the target audience and its context so that recommendations can be elaborated based on their opinions. The proposal indicates the aid of the organization of the information in search sites, libraries and even the cover itself generated from the metadata supply.

The authors begun with two questions: What information and behaviors of video game players and stakeholders related to the gaming industry exhibit when it comes to discovering, accessing, and maintaining video game information? What is the implication of these behaviors and information for the design and development of video game information systems? During the work's development the authors quote relevant research on the desire and behavior of video game users during the game. However, it stresses the need for studies that directs the needed information and behavior about the game before the acquisition, in other words, the information contained in the previous search and what leads them to choose one game instead of other.

The empirical research developed by Lee *et al.*, (op. Cit.) that proposes a user-centered metadata and coding scheme for describing and organizing videogames lasted four years. The article, which reports part of the research, included 56 semi-structured interviews conducted in two phases, with participants over 18 years of age whom searches , play, buy, collect or recommend video games. In the initial phase 24 young players and parents were interviewed, which added an important variable, since these expressed different needs. Thus, a second phase was elaborated, in which the scope of research was expanded. A snowball-type sample was stratified with 32 respondents, including casual and hardcore or professional gamers, who are more serious and committed to playing games than casual gamers, country, collectors, industry professionals, librarians, educators and scholars.

The interview protocol included approximately 40 sample questions about (a) preferences, motivations, and experiences related to user games; (b) attitudes towards physical and digital game formats; (c) collection behavior and organization; (d) game-related information needs and research behavior; (e) comments on current metadata; and (f) demographic information. Interviews for specific user groups have included additional questions (for example, asking librarians about the collection of library games, parents about how they select games for their children, and so on). The interviews were conducted live, via telephone or web conferencing, lasting approximately 1 hour.

The final version of the resulting code set contained categories related to six different aspects: game experience (eg, user game history, game behavior), collection and organization of games (eg collection behavior, preference of physical or digital formats) information needs and search behavior (such as, game information resources used and game related information needs), 43 types of metadata, 17 resource factors and user

characteristics (eg, age , gender, profession) plus a "other" category to capture potentially interesting information that they did not fit into any of those categories.

From that empirical data, the researchers presented nine design implications in choosing games: exibility in design; new approaches to gender classification; metadata of composition / technical function and responsible corporate body; information about the relationships between games and additional content; visual metadata (eg, trailers, gameplay videos) are crucial to making quick purchase and gaming decisions; information about the duration of the game sessions; and finally the similarity and appeals/motivations may work well for recommendations for successful games and potential new access points. Authors still need a greater automatic organizational variety to improve access to collections of personal games and to games distributed digitally.

As we can see, the approach of this category is different from the previous categories. It incorporates part of the of evaluation of developed product, but with the intention of identifying characteristics to be inserted in projects to be developed, before prototyping. This combination of post-development testing to obtain the target audience as precursors to the design of a new educational game can allow the prevention of problems in the projects of educational games that do not follow sequential ow of development, predominant characteristic in this type of product.

4. Final Considerations

The results presented in this article present a clipping that aimed to identify possible heuristics that are applicable to the evaluation of the usability of educational games. With this review, it was verified that the field of heuristics and usability for educational games is still little explored, with few specific validated or validation evaluations, and it is necessary to foster this area of research.

Before presenting possible answers to the question that led to this SLR "What software usability heuristics can be applied to the evaluation of educational games?", We present considerations from the analysis of the papers included in the study. The first is confusion on the use of the terms "Heuristic Analysis" and "Usability Analysis". There is no standard that indicates the conceptual consolidation of heuristic analysis as a method for usability measurement. It was observed that in the terms are used as synonyms, although they characterize distinct methods of evaluation. This conceptual overlap impairs the search and understanding of research reports on the usability of educational games. Therefore, there is a need for the determination of standardized descriptors from articles that consolidate theoretical / conceptual references for Usability Analysis and for Usability Heuristic Analysis.

The second is the lack of research with specific evaluations for electronic educational games from sets of specific heuristics. This may be due to the absence of specific software engineering frameworks or models for educational game development. There is therefore a field of research related to the framework and / or development of software for educational games, as well as heuristics for usability analysis of these games. Future heuristic modeling works for educational game will be developed by the team that elaborated this article.

Therefore, the answer to the question that guided the SLR is that, at least in the research published between 2014 and 2019 on heuristics for analysis of educational games, there is no general or specific heuristic that can be adopted. The scarce identified models lack validation, demonstrating that this is a field of knowledge still in the beginning of development, with fragmented frameworks specific or very limited to specific applicability, being a general recommendation the adaptation of the general models for particular application. It was observed the absence of abstraction that allows the applicability in generalized contexts, being this a relevant subject for future works.

In relation to the evaluation and validation process, it is concluded that heuristics must be elaborated to be applied in three moments. The first, prospecting, represented by the third category of this study, which will allow the obtaining of subsidies obtained with the target public as precursors for the design of the new game. The second one for procedural approach concomitant to all stages of development and the third for validation, to be applied at the end of the project. The first and second should be carried out from a more comprehensive iterative development methodology, which implies extensive modular and procedural evaluations to identify possible architectural, mechanical, design and usability flaws. The third heuristic, proposed for the end of the game, should contemplate an approach with specialists and another one with the target public. This specialized heuristic should approach to identify specific problems of usability of the educational game that probably are not found through general heuristics, in this case the recommendations of Nielsen (1994). A complementary heuristic could be directed to identify general problems of the domain with the target public, according to the Game User Experience Satisfaction Scale (GUESS) instrument, for its consistency and validity, being the method most commonly used at the present time for this purpose.

5. References

Aleem, S., Capretz, L. F., Ahmed, F. (2016). Critical success factors to improve the game development process from a developer's perspective. Journal of Computer Science and Technology, 31(5), 925–950.

Bardin, L. (2010). Análise de conteúdo. Lisboa, PO: Edições 70,. 281 p.

Busch, C., Claßnitz, S., Selmanagic, A., Steinicke, M. (2015). Developing and Testing a Mobile Learning Games Framework. *Electronic Journal of e-Learning*, 13(3), 151–166.

Cheng, L. C; Mustafa, M. (2015). A reference to usability inspection methods. In: International Colloquium of Art and Design Education Research (i-CADER 2014). Springer, Singapore. p. 407–419.

Desurvire, H.; Caplan, M.; Toth, J. A. (2014). Using Heuristics to Evaluate the Playability of games. in Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI '04), pp. 1509–1512, April 2004. apud Marciano, J. N., Miranda, L. C. D., Miranda, E. E. C. D. (2014). Evaluating multiple aspects of educational computer games: literature review and case study. *International Journal of Computer Games Technology*, 2014, 14.

Gonzáles-Gonzáles, C. S; Toledo-Delgado, P; Muños-Cruz, V. (2015). Agile human centered methodologies to develop educational software. Dyna, v. 82, n. 193, p. 187–194.

Hooshyar, D; Yousefi, M; Lim, H. A systematic review of data-driven approaches in player modeling of educational games. (2017). Artificial Intelligence Review, p. 1–21.

Hooshyar, D; A data-driven procedural-content-generation approach for educational games (2018). Journal of Computer Assisted Learning, v. 34, n. 6, p. 731–739.

Kitchenham B, Charters S (2007). Guidelines for performing systematic literature reviews in software engineering. Keele University and Durham University Joint Report.

Laugwitz, B., Held, T., Schrepp, M. (2008). Construction and evaluation of a user experience questionnaire. A. Holzinger (Edit.), USAB 2008, LNCS 5298 (p. 125–134). Springer

Lee, J. H., Clarke, R. I., Perti, A. (2015). Empirical evaluation of metadata for video games and interactive media. *Journal of the Association for Information Science and Technology*, 66(12), 2609–2625.

Lee, J. H., Clarke, R. I., Rossi, S. (2016). A qualitative investigation of users' discovery, access, and organization of video games as information objects. Journal of Information Science, 42(6), 833–850.

Ivory, M. Y.; Hearst, M. A. (2001). The state of the art in automating usability evaluation of user interfaces. ACM Computing Surveys, vol. 33, no. 4, pp. 470–516, apud Marciano, J. N.,

Miranda, L. C. D.; Miranda, E. E. C. D. (2014). Evaluating multiple aspects of educational computer games: literature review and case study. *International Journal of Computer Games Technology*, 14.

Malliarakis, C., Satratzemi, M.; Xinogalos, S. (2014). Designing Educational Games for Computer Programming: A Holistic Framework. *Electronic Journal of e-Learning*, 12(3), 281–298.

Marciano, J. N.; Miranda, L. C. D.; Miranda, E. E. C. D. (2014). Evaluating multiple aspects of educational computer games: literature review and case study. International Journal of Computer Games Technology, 14.

Moraes, A; ROSA, J. G. (2010). Avalia_c~ao e projeto no design de interfaces. 2AB Editora. ISBN-10: 8586695467. ISBN-13: 978-8586695469.

Noblit, G.W. and Hare, R.D. Meta-Ethnography: Synthesizing Qualitative Studies. Sage Publications, 1988.

Nielsen, J. (1994). Usability inspection methods. In Conference companion on Human factors in computing systems (pp. 413–414). ACM.

Nielsen, J. (1995). How to Conduct a Heuristic Evaluation, 1 de Janeiro.

Omar, H. M., Jaffar, A. (2010). Development and potential analysis of heuristic evaluation for educational computer game (PHEG)," in Proceedings of the 5th *International Conference on Computer Sciences and Convergence Information Technology (ICCIT '10)*, pp. 211–219, Seoul, Republic of Korea, November.

Phan, M. H., Keebler, J. R., Chaparro, B. S. (2016). The Development and Validation of the Game User Experience Satisfaction Scale (GUESS). *Human factors*, 58(8), 1217–1247.

Prümper, J. (2010). ISONORM 9241/110-S Evaluation of software based upon International Standard ISO 9241.

Quinõnes, D; Rusu, C. (2017). How to develop usability heuristics: A systematic literature review. Computer Standards & Interfaces, v. 53, p. 89–122.

Seaborn, K; Fels, D. I; (2015). Gamification in theory and action: A survey. International Journal of human-computer studies, v. 74, p. 14–31.

Trisnadoli, A., Muslim, I., Novayani, W. (2016). Software Quality Requirement Analysis on Educational Mobile Game with Tourism Theme. *JSW*, 11(12), 1250–1257.

Valente, L; Feijó, B., Do Prado Leite, J. C. S. (2017). Mapping quality requirements for pervasive mobile games. Requirements Engineering, v. 22, n. 1, p. 137–165.

Yanez-Gomez, Rosa *et al.* (2019). Heuristic usability evaluation on games: a modular approach. Multimedia Tools and Applications, v. 78, n. 4, p. 4937–4964.

Yeni, S., Cagiltay, K. (2017). A heuristic evaluation to support the instructional and enjoyment aspects of a math game. Program, v. 51, n. 4, p. 406–423.

Zarraonandia, T., Diaz, P., Aedo, I., Ruiz, M. (2015). Designing educational games through a conceptual model based on rules and scenarios. Multimedia Tools and Applications, 2015, Vol.74(13), pp.4535–4559

E.A.O. Vieira: professor of the Department of Education and coordinator of the projects and training at the Distance Education Center of the Federal University of Lavras, Brazil. Vieira holds a bachelor's degree in Dentistry, a specialist in public health and a Ph.D. in Social and Human Sciences from Aix Marseille University, France, and graduated from the Federal University of Minas Gerais. Has experience in education, focusing on educational technology, working mainly on the following topics: distance education, adult learning and computerized evaluation.

A.C. Silveira, Student of the Master program of the Department of Education of the Federal University of Lavras. Graduated in Information Systems from the same university, current student of Philosophy. Enthusiast of games, educational games and educational technologies

R.X. Martins, professor at the Education Department and the coordinator of the Centre for Distance Education at Federal University of Lavras, Brazil. Martins is graduated in mathematics, a specialist in distance education, and doctorate at psychology from São Francisco University, Brazil. He has experience in education, focusing on educational technology, acting on the following subjects: distance education, learning management systems and computerized assessment.

Appendix A. NIELSEN HEURISTICS

1. Visibility of system state.
2. Equivalence between the system and the real world.
3. Freedom and user control.
4. Consistency and standards.
5. Error prevention.
6. Recognize instead of remembering.
7. Flexibility and efficiency of use.
8. Aesthetic and minimalist design.
9. Help users recognize, diagnose, and recover wrong actions.
10. Help and documentation.