

HEURISTIC USABILITY EVALUATION APPLIED TO EDUCATIONAL GAMES

Anara Olimpio and Pollyana Notargiacomo
Mackenzie University, *São Paulo, Brazil*

ABSTRACT

The article analyzes the success of interactive methods for evaluating digital tools and digital games with a focus on usability. To do so, it maps out the game's heuristic evaluations of usability, called playability. In this context, a bibliographic review was performed, seeking to encompass the most prominent authors as well as the articles that help in understanding the best way to apply software evaluation and the best timing for such evaluation. Heuristic evaluation of usability is a useful tool for developers to improve interactive tools. Many developers have adopted interactive game-based learning because of its characteristics, and have been able to motivate mobile learning systems. After the consolidation of these papers and heuristics, three digital games geared toward music teaching were selected. The results show that it is possible to identify and analyze problems of playability, allowing designers to make corrections that can improve user experience, such as the possibility of contributing to the learning process. The qualitative study was beneficial, as it helped to understand usability problems that occur in interactive tools.

KEYWORDS

Playability, Usability Heuristics, Digital Games, Interactive Tools

1. INTRODUCTION

These days, games stand out on the market for having characteristics such as learning, skill development, and concepts combined with pleasure. Learning while playing is an effective and attractive strategy. (Ibrahim, 2014). According to Federoff (2002), companies in other sectors can benefit by seeking to understand and apply the reasons why games are successful and benefit from constructing motivational tools. Learning and implementing the techniques that games use in order to make software fun can help tools make a business interesting, and improve their purchasing power.

An important point to consider is the aspect of why interactive game-based learning fails to obtain the motivation and the space to be used in learning systems. There is little consensus regarding game resources that truly aid in education and engage students; there is a risk of designing systems that are unable to instruct or engage the student to the point where learning actually occurs, and difficulty in designing games that are both appealing and instructional at the same time. On the one hand, there is concern about violent themes that constitute certain games, as well as the intensity of involvement and the amount of time young people devote to games. On the other hand, they are thought to be quite instructive and enlightening and just as appealing as games (Garris, Ahlers and Driskell, 2002).

For a meaningful experience, it is important to be concerned with interactive learning as well as the user interface presented. The interface should favor the use of the software so that the user can concentrate on learning and having fun, i.e., the purpose that the software proposes and that enables the user to have a positive experience. (Korhonen e Koivisto, 2007). One way to assess whether the user interface is appropriate and includes features that captivate users is through the use of usability evaluation methods. According to Nielsen (1994), the methods for evaluating user interface specifications are the following: Heuristic evaluation (the most informal method, involving usability specialists who assess whether each element of the dialogue follows the predetermined guiding principles); Cognitive walkthrough (more focused on the process of solving problems that a user encounters at each stage of the dialogue, simulating the user's goals; the content of the memory should lead to the next correct action); Formal inspections (uses clearly defined functional procedures that combine heuristic evaluation and a simplified form of cognitive guidance); Pluralistic walkthrough

(meetings in which users, developers and human factors go through a scenario, discussing each element of the dialogue); Features inspection (a list of sequences of features intended to carry out typical tasks, verify long sequences, complex steps that would not be easy for users to evaluate, and steps that require broad knowledge and experience to evaluate a set of features); Consistency inspection (validation made by different designers who inspect an interface to see if they do things the same way as their own designs); and standards inspection (inspects the user interface's compliance with current standards).

Among these methods, the characteristic of being fast and cost-effective for evaluating tool usability problems is by means of heuristic evaluation. Usability inspection is a method that evaluates user interfaces to find usability problems. Evaluators analyze the user interface based on a set of pre-established rules called heuristics. Heuristic evaluation has been generalized and many evaluators have chosen to create their own sets of heuristics to be used in their analyses, so there are several lists of heuristics. This creates doubt and confusion when defining which list to use (Nielsen, 1994). Also, according to the Brazilian Association of Technical Standards (ABNT) (2011), usability is the ability of a software package to be used by specific users and meet all three characteristics: efficiency, effectiveness and satisfaction in a given context. Moreover, heuristics are guidelines for creating and evaluating easy-to-use software; game heuristics should include design elements that ensure user satisfaction. Additionally, the heuristics should not be obvious or irrelevant (Fedoroff, 2002).

Usability evaluation of the game is also known as playability. However, according to Desurvire (2004), playability is much more than usability evaluation. It encompasses four categories: 1) The game, which is the combination of difficulties and provocations that a user must face to overcome challenges; 2) Game mechanics involve programming that provides the framework whereby the units interact with the environment; 3) The story of the game is the story of the plot and persona; 4) Game usability includes the interface and covers the elements that the user can use to interact with the game (e.g. mouse, keyboard, controller, game shell, heads-up display).

This article focuses on the usability heuristics used in the process of evaluating games and tools. Therefore, according to the literature review, relevant studies related to usability heuristics and those focused on interactive learning applied to digital games were compiled and listed. Subsequently, a description of how these heuristic evaluations can be applied in educational tools is presented.

2. RELATED STUDIES

Drawn by the reasons why computer games are interesting, and features of the games that can be used to captivate users, Malone (1982) developed his work to provide a list of usability heuristics for the purpose of evaluating educational tools. His work was developed based on an elementary math game called Darts. It was found that the heuristics identified also served for other interfaces and could be used in the design of toys and tools. The features that make games enjoyable can help one learn. This is why Malone (1982) suggests dividing usability heuristics into three categories: challenge (activities with objectives, the outcome of which is uncertain), fantasy (mental images of physical objects or social situations that are not present in reality and that are embedded in the systems), and curiosity (requires the environment not to be too complex or too simple, because it should surprise and bring novelty without being incomprehensible, which can be done using visual and audio effects).

It is noteworthy that Jakob (1994) conducted a usability study and consequently obtained two lists that cover all usability problems: main heuristics to explain all usability problems (including consistency, language familiar to the user, clear user input, making user actions visible, a simple and minimalist design, shortcuts and connection to the real world, contributing to error recognition and recovery, allowing one to reverse wrong actions, and offering advanced set of evaluable actions) and main heuristics for expressing serious usability problems (which make user actions visible, have consistent and accurate feedback, have an advanced set of evaluable actions, allow one to reverse incorrect actions, use concrete analogies, have clear user input, care to avoid errors, and each mechanism has only one goal).

A few years later, in 2002, Federoff (2002) sought implicit and explicit heuristics from a leading game development company to study what developers are concerned about when designing user entertainment, as well as what usability evaluation process was used in the development cycle. The article argues that software builders in general could benefit from understanding what to consider when designing a system for fun and thereby make their products more attractive.

Among the heuristics, Fedoroff (2002) highlights aspects related to fast and responsive user involvement, the use of visual and audio effects to pique player interest (including to provide meaningful feedback), the presence of intuitive and mapped actions, the ability to check one's score at any time, pay attention to the learning curve, avoid mistakes, and save the game.

Desurvire (2004) found that heuristic evaluation is most highly recommended for use in the early stages of building a game. The list was divided into four categories: game, history, mechanics and usability. However, this work will focus on usability heuristics geared toward games and will work with immediate responses, ability to quit, save or shut down the game at any time, consistency, mandatory use of an intuitive menu, sufficient instructions to start the game, concern about sound characteristics, non-intrusive interface, quick and effective user engagement, and allowing players to recognize art in the game.

Korhonen and Koivisto (2007) presented a list of heuristics designed to evaluate games for mobile phones. After the heuristics survey, validation was performed on five games. For the evaluation, three heuristic categories were considered: usability, gameplay, and mobility. Finally, they concluded that mobility and usability problems are more noticeable, since the evaluation is similar to those already used in software evaluation. However, gameplay problems are more complex, which hinders the perception and resolution thereof, requiring greater concentration by evaluators. These authors surveyed heuristics that deal with visual and audio effects, as well as interface efficiency, which should provide visible indicators and clear vocabulary. It also addressed consistency in similar situations and controls, as well as immediate feedback, error prevention and absence of any need to memorize actions, and system help.

Other authors who worked with heuristic evaluation were Pinelle et al (2008). They presented a set of usability heuristics that can be used to evaluate the usability of video games. The heuristics list was created based on usability issues found in the game ratings of a popular games website (Gamespot). They also point out that usability heuristics are best applied early in development and the prototypes. The rating was applied to 108 games that contained six different genres. In the end, a list of 10 heuristics was compiled, and the authors concluded that simple usability problems can be avoided through this kind of evaluation, and if the evaluation had not been done, these problems would not even be remembered. Pinelli et al. (2008) highlight heuristics such as consistency of actions, settings of elements (audio, sound, speed and difficulty), player-accessible views, ability to skip frequently repeated content, ease of control, system layout for obtaining help, and ease of visual interpretation.

Soomro, Ahmad and Sulaiman (2012), through an interview and questionnaire applied to college students and game users, compiled a list of usability heuristics for mobile games. In all, 100 questionnaires were distributed, each containing three sections with 20 questions. The goal was to identify usability problems and create a list of heuristics that would help game developers during the design phase. Among the heuristics are the aspects of skipping non-reproducible content, as well as the customization of the difficulty level.

Ibrahim et al. (2014) presented a set of usability heuristics regarding educational video games based on educability, with a focus on the characteristics required to achieve an appropriate design of the video game. Through heuristic evaluation, it is possible to achieve an efficient level of usability, which ensures player motivation. During the design phase, educational usability attributes were identified, which were grouped into sets of attributes and rated to assess whether the game offers fun and learning. The heuristic sets were divided as follows: game objective (defines that there must be interaction with the game content, whether entertaining or educational, in order to make it more attractive by keeping the player immersed, interested and motivated, i.e. engaging, simple and enjoyable, paying attention to the effectiveness of the educational component); balance (regarding the balance between fun and education); game challenges (aimed at maintaining motivation and providing solutions to remain in the game); feedback (as to the level and amount achieved in relation to the entertainment and educational aspects); interactivity (whether the user feels part of the game); adaptability (difficulty levels are adapted to the user's performance); game control (ease of control to build confidence and self-esteem); ethics (acceptable in the context of the game, in view of the beliefs and attitudes included therein); realism (ability to simulate real-life situations that facilitate learning); game reward (bonuses and advantages aimed at improving user performance); structuring (structure related to motivational content and the introduction of challenges at each level of the game); and player knowledge (encouraging and activating the player's previous knowledge).

These approaches presented above are consolidated in Table 1 below:

Table 1. Summary of the papers cited

Authors	Year	Device	Games or Tools
Malone	1982	Personal Computer	Tool
Nielsen	1994	Personal Computer	Both
Federoff	2002	Video games	Games
Desurvire	2004	Video games, personal computers, board games.	Games
Korhonen and Koiviso	2007	Mobile devices	Games
Pinelle et al.	2008	Video games	Games
Soomro et al.	2012	Mobile devices	Games
Ibrahim et al.	2014	Video games	Tool

3. HEURISTIC EVALUATION APPLICATION

Based on the studies presented, it is proposed the application of heuristic evaluation in music education applications aimed at mobile devices.

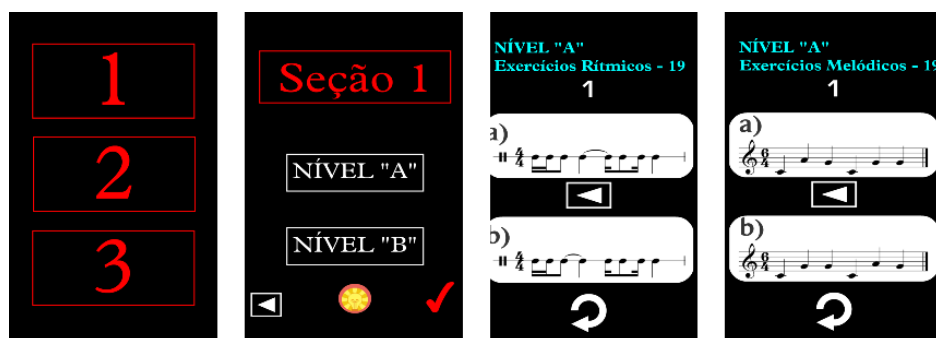
The present study was developed following the following steps:

- Selection of applications to be evaluated;
- Selection of heuristics that will be used for application evaluation;
- Heuristic evaluation of the application;
- Evaluation of results.

3.1 Selecting Apps to Evaluate

For this study, we chose to evaluate apps aimed at teaching music. These apps are available online at the Google Play Store. The apps were selected by searching the phrase "Music Education" on October 30, 2019 through the Google Play Store. The apps were selected according to the following criteria: be free-of-charge, be interactive, and require the use of audio to perform the activities. Based on these criteria, three apps were chosen at random.

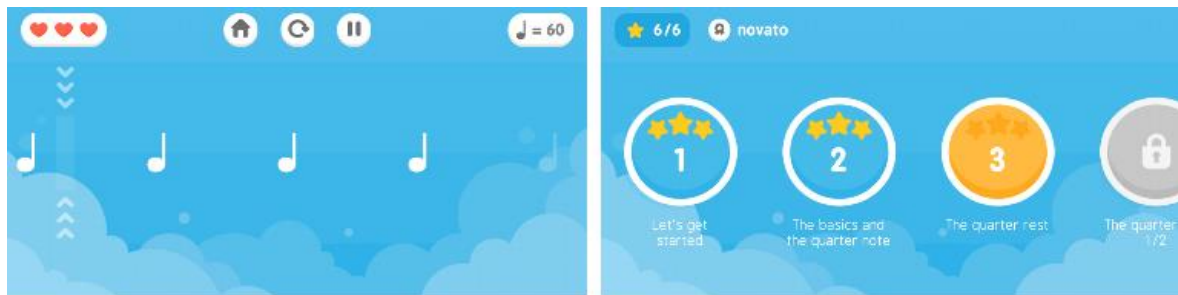
The first one, Ear Training (Figure 1), is an app focused on the teaching of musical perception, aimed at differentiating the duration and pitch of sounds. It is divided into three sections with rhythmic and melodic lessons, and the difficulty level of the exercises increases with each exercise. It is worth noting that it is possible to choose the desired exercise, and users can skip forward to the more advanced exercises, depending on their prior knowledge.

Figure 1. Section and Levels of the *Ear training*

The second one was Jungle Music (Figure 2), an app designed to learn the musical notes on the staff. This system allows the user to choose the clef they want to exercise. The story of the game consists of a mission that takes place in the jungle of a desert island, where the user must feed a chameleon by pressing the notes. As the user progresses in the game, the chameleon increases in size.

Figure 2. Screens of the *Jungle Music*

The last one, Beat the Rhythm (Figure 3), is an app that allows users to learn rhythm. The exercises are time based, and the system detects the touch on the screen and checks whether the user is counting the notes correctly and at the right tempo.

Figure 3. Screens of the *Beat the Rhythm*

3.2 Heuristic Evaluation of Audio, Sound and Educational Aspects

Based on the literature review presented, a list of heuristics was established to assess the quality of audio, sound and interactive learning issues in the selected apps. Since apps require the audio capability in order to perform the exercises, it was considered essential to have high-quality sound capabilities.

Based on these assumptions and the literature review, a list of heuristics was established:

- The game uses audio effects to kindle interest. (Malone, 1982; Federoff, 2002; Korhonen & Koivisto, 2007);
- Use of different tones for hits and errors (Malone, 1982);
- The game's sounds are significant (Desurvire, 2004);
- The system uses various features to provide feedback (music, vibration, sound effects etc.) (Ibrahim, 2014);
- The system uses sound to provide meaningful feedback (Federoff, 2002);
- The system allows for audio and sound configuration (Pinelle, 2008);
- The system facilitates the learning process (Ibrahim, 2014);
- The system allows user involvement in the challenges (Ibrahim, 2014);
- The system enhances the user's skills and experience (Ibrahim, 2014);
- The system evaluates and seeks to recognize the user's prior knowledge (Ibrahim, 2014);
- There is a balance between fun and education (Ibrahim, 2014);
- Educational content is consistent with the needs of users, stages and status of the System (Ibrahim, 2014);
- The reward is on par with the player's progress (Ibrahim, 2014).

In addition to the heuristics found in the studies analyzed, it was noted that two problems involving audio may interfere with system quality, so two heuristics were added to the list:

- The system informs the user that the audio is turned off when using the app;
- The sound used is pleasant.

3.3 App of Heuristic Evaluation and Results

To classify the heuristics throughout the qualitative evaluation, we used the Scale of Severity Ratings for Usability Problems, developed by Nielsen (1995).

0 = The heuristic was satisfied, so the system has no problems;

1 = Simple problem: simple problem, does not need to be fixed urgently;

2 = Minor usability problem: Medium problem, should be fixed with low priority;

3 = Severe usability problem: Serious usability problem, important to fix, so high priority should be given;

4 = Usability catastrophe: a catastrophic usability problem needs to be fixed before the product can be released.

In this regard, Table 2 is shown with the heuristics and the scores attributed to the apps evaluated.

Table 2. Heuristic Evaluation Results

Heuristics	Ear Training	Jungle Music	Beat the Rhythm
The game uses audio effects to arouse interest ^{1,2,3}	0	0	0
Use of different sound tones for errors and hits ¹	0	0	2
Game sounds are significant ⁴	0	0	1
The system uses various features to provide feedback (music, vibration, sound effects etc.) ⁵	1	1	2
The system uses sound to provide meaningful feedback ²	0	0	1
The system allows audio and sound configuration ⁶	2	1	2
The system facilitates the learning process ⁵	3	3	3
The system allows user involvement in the challenges ⁵	0	1	0
The system enhances user skills and enhances user experience ⁵	0	0	2
The system evaluates and seeks to recognize the user's prior knowledge ⁵	0	0	1
There is a balance between fun and education ⁵	2	0	0
Educational content is consistent with the needs of users, stages and status of the System ⁵	0	1	3
The reward is adherent to the progress of the player ⁵	1	0	2
Sum of Grades	9	7	19

¹[Malone 1982]; ²[Federoff 2002]; ³[Korhonen & Koivisto 2007]; ⁴[Desurvire 2004]; ⁵[Ibrahim 2014]; ⁶[Pinelle 2008]

Considering the bibliographic study conducted, as well as the analysis and evaluation of the apps (*Ear Training*, *Jungle Music*, and *Beat the Rhythm*), it can be said that heuristic evaluation not only contributes to the perception of different aspects of the apps, but also is a relevant, well-structured and efficient instrument for analyzing different types of digital games, such as those related to music education (covered in this article). Although heuristic evaluation is straightforward and simple, it allows one to verify the improvements that can be made and the criticality of the problems encountered. When experimenting with digital games, one can see the importance of using sound, ranging from immediate feedback to indication of hits and errors.

Regarding the tests and evaluations performed, it is worth noting that none of the three games alert the user when the audio is off and there is relevant information, since the activities presented depend on the sound feature. The system could – as with the WhatsApp instant messaging app – detect when the audio is off and give a warning, or when the user accesses the app, generate a warning that the device requires the sound feature to be enabled.

It was also found that the *Beat the Rhythm* game had more problems related to the sound feature (Ibrahim, 2014). The system could – like the other two apps – make better use of sound features to provide feedback or even differentiate users' hits and misses (Federoff, 2002; Pinelle, 2008).

Another point concerning sound refers to the sound setting options; the three apps could provide options for inserting different sounds and ringtones. *Jungle Music* offers sound configuration, but it needs to have improvements in this aspect (Federoff, 2002).

The system could use solid resources to promote learning. All three systems could enhance the issue of providing an interactive learning service (Ibrahim, 2014). However, they fail to do so because they are designed to support music lessons, when they could (for example) have explanations so that the user can understand the meaning of each feature offered and learn from the system, and not just serve as reinforcement for learning.

Also with regard to learning, it was noted that the *Ear Training* app is not fun, but it is able to engage the user with its exercises and also allows the user to skip ahead to more advanced steps, which makes it interesting (Ibrahim, 2014). On the other hand, there is *Jungle Music*, which is a fun app and has an interesting plot, but fails by requiring that the user repeat the same exercise three times in each step and does not allow the user to skip steps (Ibrahim, 2014). Lastly, the *Beat the Rhythm* app has simple content and minimalist design, but fails to provide user learning (Ibrahim, 2014).

4. CONCLUSIONS AND FURTHER WORKS

This study mapped the usability heuristics of digital games available in the literature, in order to gather significant studies in this area. Subsequently, heuristics were compiled that are directly linked to the issue of the evaluation of digital games for music teaching. Finally, heuristic evaluation was performed in three music education apps, aimed at evaluating the use of sound and audio capabilities to improve learning.

The study showed how important it is to use a form of evaluation of educational tools; such evaluation using heuristics contributes to the detection of system problems and can show parts of the system that are working well and need not be changed. The suggested heuristics, after the literature review, can be used to evaluate other apps geared toward teaching music, since these apps must be concerned with sound and audio quality. The issue of learning must also be carefully reviewed.

It is recommended to use heuristic evaluation in the software development process – i.e., prior to launching the app – so that problems can be found and fixed in a timely manner, thereby ensuring high-quality delivery.

For further studies, we intended to extend the evaluation by expanding the number of apps for the top 100 digital music education games that received the highest ratings in the Google Play store. Similarly, we intend to verify the mirroring of this set for the Apple Store, making it possible to establish a state of the art for this type of educational app. Moreover, artificial intelligence mechanisms will be established to automate the heuristic evaluation process. Likewise, a digital game for music education will be developed, based on the heuristics identified for further testing with users.

ACKNOWLEDGEMENT

The authors thank MackPesquisa for the financial support, and acknowledge the support of JAS3 Laboratory (Laboratory of Games, Learning, Simulation, Systems and Signal).

REFERENCES

- Desurvire, H. et al., 2004. Using heuristics to evaluate the playability of games. *Proceeding of the CHI EA '04 Extended Abstracts on Human Factors in Computing Systems*. Vienna, Austria, ACM, pp. 1509-1512.
- Federoff, M. A., 2002. Heuristics and usability guidelines for the creation and evaluation of fun in video games. University Graduate School (Master of Science Degree), Department of Telecommunications, Indiana University.
- Garris, R., Ahlers, R., and Driskell, J. E., 2002. Games, Motivation, and Learning: A Research and Practice Model. *Simulation & Gaming*, 33(4), 441–467
- Ibrahim, A. et al., 2014. Playability guidelines for educational video games. *International Journal of Game-Based Learning*, v. 2, i. 4, p. 18–40.
- Korhonen, H. and Koivisto, E., 2007. Playability heuristics for mobile multi-player games. *Proceedings of the 2nd international conference on Digital interactive media in entertainment and arts (DIMEA '07)*, Perth, Australia, p. 28–35.
- Malone, T. W., 1982. Heuristics for designing enjoyable user interfaces: Lessons from computer games. *Proceedings of the 1982 Conference on Human Factors in Computing Systems (CHI'82)*, Gaithersburg, Maryland, USA, pp. 63–68.

- Nielsen, J., 1994. Enhancing the explanatory power of usability heuristics. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'94)*, Boston, Massachusetts, USA, pp. 152–158.
- Nielsen, J., 1995. Severity ratings for usability problems. 1995. Available: <https://pdfs.semanticscholar.org/964e/328ae161d79cfe81b4dfe08d7a5021a18a15.pdf>.
- Pinelle et al., 2008. Heuristic evaluation for games: Usability principles for video game design. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI'08)*, pp. 1453-1462.
- Soomro, S., Ahmad, W. F. W. and Sulaiman, S., 2012. A preliminary study on heuristics for mobile games. *2012 International Conference on Computer & Information Science (ICCIS)*, Kuala Lumpur, Malaysia, pp. 1030–1035.