CLUE: A Usability Evaluation Checklist for Multimodal Video Game Field Studies with Children Who Are Blind

Ticianne G. R. Darin	Rossana M. C. Andrade	Jaime Sánchez
Virtual University Institute	Dpt. Computer Science	Dpt. Computer Science,
Federal University of Ceara	Federal University of Ceara	University of Chile
Humberto Monte, S/N Bl 1430	Humberto Monte, S/N Bl 1430	Blanco Encalada, 2120
Fortaleza, CE Brazil	Fortaleza, CE Brazil	Santiago, Chile
ticiannedarin@great.ufc.br	rossana@great.ufc.br	jsanchez@dcc.uchile.cl

Abstract

Multimodal video games can enhance the cognitive skills of children who are blind by allowing interaction with scenarios that would be unfeasible in their everyday life. To assist the identification of relevant interface and interaction issues when children who are blind are playing multimodal video games, we propose a Checklist for Usability Evaluation of Multimodal Games for Children who are Blind (CLUE). CLUE was designed to assist researchers and practitioners in usability evaluation field studies, addressing multiple aspects of gameplay and multimodality, including audio, graphics, and haptics. Overall, initial evidence indicates that the use of CLUE during user observation helps to raise a greater number of relevant usability issues than other methods such as interview and questionnaire. CLUE makes the analysis of recorded user interactions a less time- and effort-consuming process by guiding the identification of interaction patterns and usability issues.

1. Introduction

The evolution of gaming technology has impacted the daily routine of children and adolescents worldwide, going beyond entertainment purposes [1,2]. In this context, serious video games play a valuable role in the development and enhancement of diverse types of cognitive skills [3], as well as for teaching and learning purposes [2], including people with multiple types of disabilities [2, 4, 5]. Particularly, learners who are blind have been using serious multimodal video games based on audio and haptics to foster mental skills, such as logical reasoning, navigation, mental mapping, and spatial cognition [6, 7, 8]. Such video games can also help people who are blind to transfer virtually acquired skills to real environments, and ultimately, to everyday life [6].

However, the development and enhancement of the intended cognitive skills will be possible through these video games only if they manage to combine the adequate modalities, while carefully coordinating interface and feedback to represent abstract information [4, 8, 9]. For learners who are blind, game modalities must afford a precise interpretation of the information conveyed, given the absence of vision [10]. Likewise, the multimodal gaming interface should support a comfortable and pleasant multimodal interaction, preventing learners from feeling confused, tired, or inattentive, which could negatively impact the learning of cognitive skills while playing [11].

Consequently, usability is fundamental in this context, especially considering that video games usually require constant interaction, and focusing on usability issues rather than on learning would be frustrating and undesirable [12, 13]. Administering an accurate usability evaluation is hence a necessary step towards assisting children and adolescents who are blind in the construction of cognitive skills while playing video games. Nevertheless, the usability evaluation of serious multimodal video games for learners who are blind lacks reasoning, regarding what game aspects to evaluate and how to proceed the assessment [14].

Evaluators conducting usability tests involving people who are blind should keep in mind that traditional Usability Evaluation Methods (UEM) [15] are usually designed for users without disabilities [16]. Studies comparing UEM have shown that the use of general UEM in different contexts is controversial [17, 18, 19]. Multimodality adds further complexity to this scenario since specific issues differentiate multimodal usability evaluation from the evaluation of traditional user interfaces, such as GUIs [10]. Besides, usability evaluation of multimodal games involving young learners who are blind can also be affected by the differences between children and adults with the same condition. Children who are blind cannot fully perceive anything at once; instead, everything has to be constructed [20]. In addition to that, they are still learning things and experiencing situations that they recognize and understand

HĭCSS

differently from adults, according to their linguistic, intellectual, and motor skills [21].

The facts mentioned above demonstrate that evaluating usability in this context requires UEM adaptation to assure that usability evaluation instruments and methods administered disclose most of the issues that might affect the game interaction of the target users.

The present study proposes an observational tool for usability field tests, the Checklist for Usability Evaluation of Multimodal Games for Children who are Blind (hereafter abbreviated as CLUE). CLUE is composed of 40 checkpoints to guide the observation of children who are blind, helping to identify issues on the multimodal interface and interaction of video games during field tests. The aim of this work is not to come up with a novel approach for evaluating the usability of serious video games or multimodal interfaces per se. The literature addresses the usability of both serious games [23, 24, 25] and multimodal interfaces [26, 27]. Instead, the main contribution brought by CLUE is focusing on the analysis of multimodal interface elements during the gaming interaction, disclosing issues that may affect the cognitive purposes of such video games for children who are blind. Our goal is that CLUE will be used not only by practitioners and researchers but also by specialized institutions and schools for learners who are blind, helping teachers and instructors to identify whether a game can be helpful to the children rather than create a barrier to their learning and cognition.

2. Related Work

To the best of our knowledge, few works are addressing usability evaluation of multimodal games for children who are blind. In a systematic literature review, Sánchez et al. (2015) analyzed 25 papers describing evaluation and design of multimodal video games and virtual environments for cognitive enhancement of people who are blind [14]. The authors discussed in details how studies with similar goals followed different procedures to conduct usability evaluation, identifying a lack of reasoning in this regard. They remarked that some studies made unconfirmed assumptions about ease of use, learnability, and interaction of these video games, as they do not perform usability evaluation involving potential users [28, 29, 30, 31].

In a later work [22], the authors proposed and discussed a 4-dimension classification to characterize multimodal video games for cognitive development of people who are blind: Interface, Interaction, Cognition, and Evaluation. They assembled such dimensions from the classification of features related to the development and evaluation of 17 multimodal video games and four virtual environments. Additionally, they identified the

instruments and methods applied during the evaluation of usability and cognitive impact of the considered games. Despite the helpful insights for the practical understanding of the issues involved in the design and assessment of such games, they did not offer guidance for carrying out usability evaluation in this field. As the authors did not consider usability in detail, adapting and analyzing the suitability of UEM and instruments is outside the scope of their work.

Darin et al. (2017) proposed a Standard List of Usability Problems (SLUP) based on the analysis of the usability evaluation reports of five target multimodal video games for cognitive improvement of children who are blind [33]. SLUP contains 61 issues related to the interface and interaction features that commonly impact the multimodal gaming interaction of learners who are blind. SLUP aims to help designers avoid recurrent usability issues regarding audio, adaptation, interaction mode, and feedback in the design of such games. Comparing data gathered from videotaped user observations applying Thinking Aloud Protocol (TAP), videotaped interviews, and answers to a questionnaire, the authors identified that SLUP could be used as a ground to develop specific usability evaluation instruments. In the present work, SLUP is one of the foundations for the proposal of CLUE.

Hereafter, we discuss some related work that analyzes the usability evaluation of multiple types of interfaces for people who are blind. They illustrate the need to adapt usability evaluation methods and instruments to fit better the context of individuals who are blind. None of them, however, focus on the usability evaluation in the context of children who are blind playing multimodal serious video games for cognitive enhancement.

Chandrashekar et al. (2006) analyzed usability testing sessions in which four users who were blind and six with visual disabilities used a screen reader and employed TAP during the evaluation of a website [16]. The analysis of recorded audio indicated that users with total blindness need alternative training strategies to apply TAP. Based on the number of comments raised by each group of users, the authors argued that TAP may not be effective in usability testing involving users with total blindness using a screen reader.

Raisamo et al. (2006) discussed a procedure for testing usability with children with visual disabilities based on the application of standard UEM adapted according to the knowledge and experience of the authors [21]. They tested a multimodal system using haptic feedback devices, stereo sound, and visual feedback, employing questionnaires, interviews, and observation methods in laboratory and field tests. They

analyzed the data gathered from both types of tests, including children's videotaped interviews, video recordings and log files of the children's use of the program, and questionnaires. As a result, the authors gave directions on how to consider the children's special testing requirements in different environments when conducting usability tests of multimodal applications for children with visual disabilities. For example, they stated that performing usability tests in a school where children attend for special education is a facilitating factor. Although the authors gave valuable advice on how to conduct field research, there is no discussion about practical ways to help researchers disclosing usability problems in this context.

Finally, Miao et al. (2016) investigated four usability methods involving people who are blind, partially sighted and sighted, comprising local test, synchronous remote test, tactile paper prototyping and computer-based prototyping [34]. The results showed that local tests were as efficient as synchronous remote tests, while tactile paper prototyping was comparable to computer-based prototyping, based on the number of usability problems uncovered by each approach in different categories.

3. Multimodal Video Games for Children Who Are Blind

In this section, we present a contextualization regarding interface and interaction features in multimodal video games for cognitive development and enhancement of children who are blind, as well as a brief discussion on usability evaluation in this field.

3.1 Interface and Interaction

Multimodal video games aiming to develop and enhance cognition of young learners who are blind can be described according to their motivating story [35] together with four dimensions: *Interaction*, *Interface*, *Cognition*, and *Evaluation* [22, 33]. These aspects indicate the key features of game interaction and interface characterization, along with the cognitive process meant to be developed and enhanced, and the type of evaluation implemented. According to this characterization, the interface and interaction features that most impact multimodal video games for learners who are blind are *Audio*, *Adaptation*, *Interaction Mode* and *Feedback* [22]. We enriched the description of such features by aggregating physical carriers, according to Bernsen's (2010) modalities taxonomy. As a result, Table 1 shows the modalities usually present in multimodal video games for people who are blind.

Multimodal applications, in general, orchestrate the fusion of multimodal inputs and the fission of multimodal outputs, resulting in a satisfactory outcome to the users, according to their context of use, and personal preferences and characteristics [32].

Table 1.   Usual modalities for children who are blind

MEDIUM	MODALITY	CHARACTERIZATION
GRAPHICS	Interface	Bidimensional images, maps or graphs Tridimensional images, maps or graphs Written text
	Feedback	Contextual visual cues (using graphic interface features)
	Adaptation	Size Contrast Color Scheme
ACOUSTICS	Interface	Spoken audio Speech synthesis Iconic sounds Spatialized sounds Stereo sounds Abstract earcons
	Feedback	Contextual audio cues (using sonorous interface features)
	Adaptation	Speed Intensity
HAPTICS	Interface	Tactile Kinesthetic
	Feedback	Force Vibration Pressure Motion
	Adaptation	Intensity Frequency

However, to stimulate cognitive processes, multimodal video games should additionally properly combine graphic, acoustic and haptic-related modalities in specific ways, according to the characteristics of the cognitive processes [4, 8]. The different combinations of modalities affect the users' behavior towards the game and determine how the learning takes place and how cognitive processes are stimulated. For instance, audio and visual cues coordinated with haptic elements distributed in a virtual navigational environment can serve as references for orientation and mobility, as well as help learners who are blind adopting and restructuring a mental model of spatial dimensions [31]. For that reason, the effectiveness of usability evaluation instruments and methods used in this regard depends on their capability to disclose issues related to the multimodal gaming interaction and interface.

3.2 Usability Evaluation

There is currently no consistent advice on which methods are appropriate in these circumstances, so the selection of methods relies on individual experience and

expertise [14]. In this context, the UEM most commonly applied when evaluating the usability of multimodal games for children who are blind are observation, interviews, and heuristic evaluation [22]. The use of observation is in accordance with the technique most used in the usability evaluation of serious video games, which is usually playtesting, often combined with ad-hoc questionnaires [25]. The instruments usually applied by researchers and practitioners in this context are ad-hoc questionnaires, Likert-based surveys, and specialized questionnaires [14, 22]. The first two types are created and used to evaluate only a specific game, comprising a set of opinion and attitude questions. Evaluators often use such instruments generated according to the overall goals of the evaluation, but not formally validated [37, 38, 39]. Specialized questionnaires are valid and reusable, and consist of a set of context-specific sentences, for which the users can define the degree of fulfillment on a scale, such as in the case of Software Usability for Blind Children Questionnaire (SUBC) [36].

However, traditional UEM may not disclose most of the issues that recurrently impinge on the game interaction of users who are blind [10, 21]. Some of which are discussed by [33] after an intensive analysis of the usability reports of multimodal video games, aiming the development of a Standard List of Usability Problems (SLUP). SLUP describes 61 common usability issues, addressing: 12 Overall Usability problems regarding learnability and satisfaction, errors, and efficiency; 15 problems related to the different types of audio previously listed in Table 1. There are five issues related to the adaptation of audio and graphic user interface; 19 problems related to interaction mode, including the use of different types of game controls; and eight items related to audible and haptic feedback.

Knowing the typical problems concerning the interaction of children who are blind with audio- and haptic-based multimodal video games is the first step towards planning a usability evaluation to identify and correct real problems. However, researchers and practitioners often choose to carry out informal usability evaluations due to either time or team issues, or even unfamiliarity with specific usability evaluation instruments and methods [28, 29, 31]. When usability evaluations do not consider the combination of multimodal inputs, the users' visual disabilities, and the addressed cognitive skills, an important part of the context of use is left out. In this scenario, a drawback is that applying ad-hoc questionnaires or interviews after a gameplay session is no guarantee of meeting the user's needs, neither the cognitive game requirements.

To help to fill this gap, we believe that a checklist based on the main issues that impact the interaction of children who are blind with such video games can be a valuable tool to provide practical guidance for properly evaluating usability. Such direction is especially valid considering that researchers, practitioners, and teachers with multiple backgrounds might be interested in this matter. Besides, evaluation checklists guide evaluators in gathering relevant evidence to determine the merit, worth, or significance of a subject, constituting a systematic tool highly significant and useful for evaluation purposes [40]. As an evaluation checklist, CLUE also decreases the possibility of forgetting to verify important aspects and are easier for the layperson to use and understand [41], which fits well our purposes.

4. CLUE Development and Characterization

As a research step towards guiding usability evaluation of video games for children who are blind, we propose a Checklist for Usability Evaluation of Multimodal Games for Children who are Blind (CLUE). CLUE was designed to be used by researchers, practitioners, and teachers during field tests involving children who are blind interacting with video games. In our exploratory research toward developing CLUE, we based our methodology on Stufflebeam's (2000) guidelines for developing evaluation checklists [42]. Our methodology consisted of three main steps, as summarized in Figure 1.
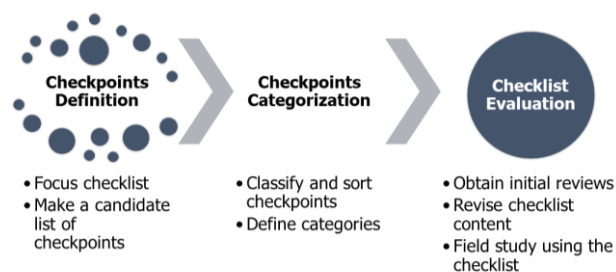
Figure 1. Steps followed for the development of CLUE

During the Checkpoints Definition, we first established the intended uses for CLUE, which are (i) help an evaluator to disclose usability issues in video games based on audio and haptics while watching a child who is blind playing. The video game purpose must include, but not necessarily be limited to, the development and enhancement of cognitive skills in children with total or partial blindness; and (ii) serve as an auxiliary resource for the analysis of recorded gaming sessions in the same context. To delimit our

understanding of what is a usability problem, we considered Manakhov and Ivanov's (2016) definition. According to the authors, a usability problem is any negative phenomenon in interaction induced by a combination of user interface design features and context-of-use factors, such as user's inability to reach a goal, or user's dissatisfaction [43]. Then, to assemble a candidate list of checkpoints, we searched the literature for studies that discussed usability for people who are blind. We selected SLUP [33] as the basis for our candidate checkpoints list, and classified them according to the modalities usually present in these games, as summarized in Table 1 [22,10].

After this, the Checkpoints Categorization step consisted in the listing, describing and defining of the checkpoints, before sorting them into categories based on the correspondence presented in Table 1. The first categories we used were *Audio, Adaptation, Interaction Mode,* and *Feedback*. Next, we designed a review version of the checklist, consisting of 42 checkpoints associated with three options, "Yes," "No" and "Not Applicable (N/A)." Each item addressed one or more issues in SLUP to simplify the experimenter's analysis while observing the user.

Finally, the Checklist Evaluation occurred in two main phases. First, five specialists (E1 to E5 in Table 2) analyzed the checkpoints and gave feedback via email and non-structured interviews, discussing the pertinence and adequacy of the checkpoints to the intended use. They also gave further details on the comprehensiveness of the checkpoints and suggested corrections. All the respondents were familiar with conducting usability evaluations of multimodal games with learners who are blind. According to the consolidation of the reviewers' answers, we rewrote and replaced a number of checkpoints; and excluded two of them, as all experts agreed they could not be observed.

The second phase towards validating CLUE was the conduction of preliminary usability testing sessions, which we will further describe in Section 5. This phase aimed to obtain reviews from intended users and experts while engaging them to field-test the checklist. Six children with different ophthalmic diagnoses (all legally blind), from 8 to 14 years old, attending from 2nd to 7th grades in schools for learners who are blind participated in the testing sessions. Four independent evaluators (E4, E5, E6, and E7) conducted the evaluation sessions using the updated review version of CLUE, which contained 40 checkpoints organized as Overall Usability (10 items), Interaction Mode (8 items), Feedback (6 items), Graphics/Adaptation (3 items), and Audio (13 items).

During user observation, evaluators filled the updated review version of CLUE to help them observe the children´s interactions, checking "Yes" or "No" for each checkpoint, indicating whether the event described in the checkpoint had occurred during the game session. After finishing each user observation, the evaluators would also answer "Yes," "No" or "N/A" to the question "Is it possible to verify the situation described in this item during a user observation?". They also offered input on the organization and use of checkpoints. All the evaluators agreed that a trained usability evaluator could disclose all the checkpoints using CLUE in a field test.

Table 2. Experts' Profiles

ID	Degree	Experience	Expertise
E1	Master of Computer Science	5-10 usability evaluations	Audio games usability
E2	Ph.D. Student in Computer Science	> 10 usability evaluations	Technology and applications for people who are blind
E3	Bachelor of Education	5-10 usability evaluations	Cognition of children who are blind
E4	Bachelor of Education	> 10 usability evaluations	Cognition of children who are blind
E5	Undergraduate degree in Computer Science	> 10 usability evaluations	Multimodal games design and usability
E6	Ph.D. Student in Computer Science	> 10 usability evaluations	Multimodal games evaluation and usability
E7	Undergraduate student in Education	5-10 usability evaluations	Evaluation for children who are blind

In a non-structured interview after the user observation, the evaluators described which specific usability issue led them to check each CLUE item marked as "Yes." For example, the checklist item number 7 instructed the experimenter to check whether the child found difficult to "accomplish the game tasks." If an evaluator checked this item as "Yes" during a user observation, it might have been caused by the SLUP issue Q19 "The user feels that the game does not allow him to be in control as much as he expected"; or by Q38 "The user has difficulties in understanding the game goals", or even by a combination of both. This procedure allowed us to verify whether each CLUE checkpoint was being correctly interpreted and used.

Finally, after finishing the Checkpoint Evaluation, we updated CLUE once again resulting in simplified checkpoints descriptions, and in a new organization. These changes aimed to facilitate the observation and to allow the modular use of the checklist. CLUE latest version contains 40 checkpoints grouped in 4 categories: Gameplay (14 items), Acoustics (11 items), Haptics (12 items) and Graphics (3 items).

2. Check if the child has *difficulties* to:	Did it happen?	How many times?
2.1 Hear the game sounds	☐YES ☐NO	☐ONCE ☐SEVERAL
2.2 Identify a specific sound	☐YES ☐NO	☐ONCE ☐SEVERAL
2.3 Recognize a specific sound	☐YES ☐NO	☐ONCE ☐SEVERAL
2.4 Understand the information conveyed by a sound	☐YES ☐NO	☐ONCE ☐SEVERAL
2.5 Realize that a specific sound is related to a specific action in the game	☐YES ☐NO	☐ONCE ☐SEVERAL
2.6 Associate the game sounds with his prior knowledge	☐YES ☐NO	☐ONCE ☐SEVERAL
2.7 Associate the game sounds with the right objects or actions in the game	☐YES ☐NO	☐ONCE ☐SEVERAL
2.8 Understand information about orientation and location in the game environment	☐YES ☐NO	☐ONCE ☐SEVERAL
2.9 Identify the purpose a specific audio feedback		
Check if the *audible feedback*:		
2.10 Are sufficient to the execution of the game activities	☐YES ☐NO	☐ONCE ☐SEVERAL
2.11 Are correctly applied to the game objects and execution of game activities	☐YES ☐NO	☐ONCE ☐SEVERAL
Check if the child demonstrates to *feel*:		
2.10 Uncomfortable with the speed of spoken audio or TTS	☐YES ☐NO	☐ONCE ☐SEVERAL
2.11 Annoyed by specific sounds or voices	☐YES ☐NO	☐ONCE ☐SEVERAL

Figure 2. Portion of current version of CLUE showing acoustics-related checkpoints

Gameplay contains items related to game overall usability and playability. *Acoustics* aggregates items related to the comprehensibility and adequacy of multiple types of sounds used in the user interface (Figure 2). *Haptics* contains checkpoints regarding the use of haptic interaction techniques and devices that may affect the user interaction with the game inputs and outputs. *Graphics* contains items related to visual aspects of the user interface. Each category also contains checkpoints related to the user feeling and satisfaction towards the specific game and modality aspects, and to the types of feedback provided by each modality.

The modular use of CLUE was an improvement designed to allow the independent use of the checkpoint groups, according to the context. For example, video games for children who are blind can be either based on audio-only or audio-haptic stimulus. In the first case, an experimenter could use the checkpoints related to *Gameplay* and *Acoustics*, while in the second the *Haptics* category would also be necessary. Whenever a child with low vision is participating in the tests, evaluators should also check the *Graphics* category. The current version of CLUE also contains a column where the experimenter can check approximately how often a situation occurred during observation. The full version of CLUE is available at https://goo.gl/pWuKLk.

5. Applying CLUE in a real scenario

In this section, we describe the testing sessions and discuss the results based on the number of usability issues uncovered by each approach in different categories.

5.1 Scenario Description

We conducted the user testing sessions in real environments, at two schools for children who are blind. Six legally blind children with distinct types of ophthalmic diagnosis interacted individually with the multimodal video games AbES and Audio Sims, which comprise diverse modalities and address multiple cognitive skills, as depicted in Table 3. The children participating in the field test were familiar with computer and mobile video games based on audio and haptics.

A team of two evaluators conducted each user test, to guarantee that the observation, the annotations, and the filling of CLUE were done properly. Each test session lasted about 40 minutes, in rooms designated by the schools. While a camera fixed to a tripod recorded the child's interactions and interview, one experimenter conducted the test activities including the mediation the children needed [21, 44], while the other experimenter observed, took notes and filled CLUE. The interaction data was gathered through user observation [21] together with a Think Aloud Protocol, followed by a semi-structured interview and the administration of the SUBC Questionnaire [36]. We chose that configuration because these are the most commonly used UEM and evaluation instruments in this context [14, 22].

5.2 Data Analysis

We performed a quali-quantitative analysis of the user's data gathered from the experimenter's notes from user observation, the information filled in the CLUE, the recorded user session, the recorded interview and the answers to the Software Usability for Blind Children Questionnaire (SUBC). The last one is a 10-question specialized instrument aimed at children who are blind, regarding traditional software usability.

The data analysis aimed to identify and compare sets of usability problems disclosed by the each UEM. We analyzed the data for each UEM independently.

Table 3. Characteristics of the multimodal video games used in the field study of with children

| Game	Interaction			Interface	Cognition
	Mode	*Feedback*	*Graphics*	*Audio*	
AbES (for PC)	Keyboard	Sonorous, visual	2D	Speech synthesis, iconic sounds, spatialized audio	Problem solving Spatial structure Orientation & Mobility
Audio Sims (for PC)	Joystick	Sonorous, visual, haptic	3D	Spoken audio, iconic sounds, spacialized audio	Mental mapping Orientation & Mobility Auditive and haptic sensory-perception

First, we systematized and sorted the raw data obtained. For CLUE and SUBC, we tabulated data and verified the presence or absence of the issues addressed, as well as their frequency. We transcribed the interview answers and video recordings and identified critical incidents. To organize the data, first, we listed user doubts and grouped similar answers and issues identified per session per UEM. Then, we examined data thoroughly for identification of interaction errors, difficulties and usability problems, including tasks in which users failed. We also identified causal explanations (considering the difficulty in understanding and mismatch with users' understanding) and described the problems. Then, we conducted a thematic analysis, categorizing problems into the following coding categories: Overall Usability, Audio, Adaptation, Interaction, and Feedback.

Given the sets of usability issues generated by each UEM, we ranked the problems by their severity rate (on a three-point scale [45]) and frequency. After this, we matched usability problems based on their description and severity, to identify repeated items. Frequency was analyzed using mean and standard deviation, according to each category of problems. Finally, for each UEM, we analyzed the percentage of problems belonging to each category.

5.3 Field Study Results

The field studies aimed to engage experts into testing CLUE in a real environment, to gather authentic feedback, which was useful to improve CLUE, as detailed in Section 4. However, comparing data collected using CLUE to data coming from video, interview, and SUBC also allowed us to identify preliminary evidence on which categories of usability issues each UEM could disclose in this context. We do not intend to establish the superiority of CLUE over other methods, as we acknowledge that each method has potential advantages and disadvantages. Instead, hereupon we

discuss the results obtained, considering the applicability of each UEM in this particular case. As expected, the analysis of the video-recorded user interactions could disclose the greater number of the usability issues listed in SLUP (Figure 4), in all the dimensions analyzed, especially those related to audio (Figure 5). From the total of 181 non-unique usability issues evaluators identified in all user sessions with both games, 112 came from video analysis, from which CLUE also disclosed 74. It means that, by using CLUE exclusively, evaluators could identify 66% of the usability issues disclosed later by video analysis and surpassed the problems revealed by using SUBC and interview.
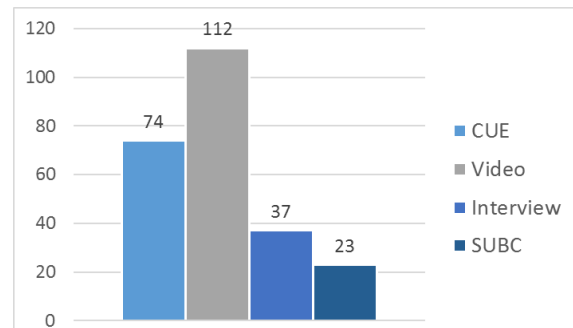
Figure. 4. Usability issues per UEM in six user tests

The strength of CLUE was in finding interaction, feedback, and overall usability issues. Compared to semi-structured interview and the specialized questionnaire SUBC, CLUE was better in the search for all dimensions of problems, except for adaptation. The identification of adaptation issues was more effective by receiving direct user feedback using SUBC. As summarized in Figure 5, the SUBC questionnaire could not disclose any problems related to feedback and interaction problems. However, it was more useful than interview in identifying adaptation and overall usability issues. The quantity of feedback, adaptation and audio issues found using interview indicate that the combined use of interview and SUBC would be beneficial for a rapid overall usability evaluation.
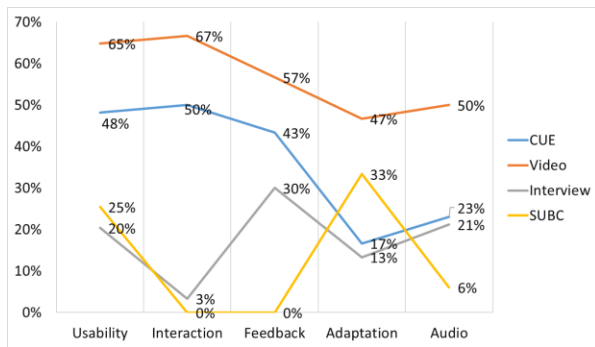
Figure 5. Comparison of types of usability issues disclosed per UEM in six user evaluations

In each category, the UEMs revealed usability issues at different levels. The problems related to audio features can be summarized as difficulties to recognize sounds, wrong association of sounds, misunderstanding information conveyed by a sound and somehow disliking a sound. Video analysis identified twice the number of problems related to misunderstanding and not recognizing sounds than CLUE. Interview and CLUE found the same quantity of issues regarding sounds that the user disliked, but CLUE was better than interview to help to identify the other types of audio issues. SUBC could reveal only a few indications of difficulty to recognize sounds.

The overall usability issues address problems of multimodal interaction, learnability, efficiency, and satisfaction. In this dimension, CLUE revealed a number of usability problems much superior to those obtained with interview and SUBC, in all subcategories, except satisfaction. CLUE was particularly good at identifying multimodal interaction and learnability problems, being comparable to the results obtained with video analysis. However, regarding efficiency, CLUE could identify only 25% of the problems disclosed by video, while results from interview and SUBC were unexpressive. Regarding user satisfaction, SUBC identified as many issues as video, followed by CLUE and interview respectively.

Regarding the interaction mode, which comprised problems related to the use of the game controls, only by using CLUE, and video, relevant results were revealed. CLUE found 67% of the problems identified in video analysis related to difficulties to learn and use the controls; and 59% of the problems associated with movement and rotation inside the game environment. The feedback problems can be related to difficulties in identifying a feedback and incorrect or insufficient feedback. CLUE found the same number of problems regarding feedback identification and use of incorrect feedback as video analysis. Interview identified the same quantity of problems reporting insufficient feedback as CLUE.

Overall, the results indicate that CLUE can provide a less time- and effort-consuming analysis when compared to video, yet revealing a substantial number of relevant usability issues. Alternatively, combining the use of CLUE with video analysis can be a powerful resource to assist researchers and practitioners in finding real problems while children who are blind play video games to improve cognition. The field study design targeted to engage experts into testing CLUE in a real environment. However, comparing the results obtained by each UEM was relevant to demonstrate that CLUE is a solid alternative to administering interview and questionnaire methods, which are broadly applied in this field, as discussed before. CLUE can additionally help to make the analysis of recorded user interactions an easier process, guiding the identification of interaction patterns and recurrent usability issues, even by evaluators with little experience.

We highlight that further tests are still needed to establish CLUE as a sound evaluation checklist. All the UEM applied in our tests have advantages and disadvantages. For example, while the combination of SUBC and interview can be easier to apply and good at revealing adaptation and feedback issues, CLUE help to raise a greater number of usability issues related to overall usability, audio, feedback, and interaction. Furthermore, CLUE analysis is faster and more straightforward than analyzing interview data. On the other hand, applying CLUE in usability testing demands at least two experienced evaluators to identify the issues, without jeopardizing the observation. It is up to the evaluators to decide, during the planning phase, which are the more suitable UEMs and tools to apply, given the available resources, and the evaluation goals.

6. Conclusion

In this work, we presented and discussed the development and evaluation of CLUE, a 40-item checklist designed to guide researchers, practitioners, and teachers of children who are blind in usability evaluation field studies, addressing multiple aspects of gameplay and multimodality, including acoustics, graphics, and haptics. CLUE aims to help in the identification of the real problems that affect the multimodal gaming interaction of these users, in a practical way. Furthermore, we considered primary evidence of what problems different UEM can reveal, indicating that using CLUE during user testing helped to raise a greater number of

relevant usability issues than other evaluation instruments.

The main purpose of CLUE is to clarify the basic aspects that should be considered when evaluating multimodal video games for cognitive development and enhancement of children who are blind. Thus, CLUE helps the evaluator not to forget important criteria and enhances the objectivity and reproducibility of evaluation. Our future work will expand the tests to cover the application of CLUE to a wider set of multimodal games and consider a larger user sample. Based on the results, we will further refine the checklist.

Usable and pleasant multimodal video games will impact the lives of children who are blind by helping them in developing skills that will allow them to be more independent in their everyday lives and better integrated and included into society. Our final goal in this research is to propose a sound instrument for the evaluation of usability of multimodal video games designed for developing and enhancing cognitive skills in children who are blind. Thus, we encourage the community to join our efforts, applying CLUE in their research and giving us feedback, as we continue to improve this instrument.

7. References

[1] Mayo, M.: Games for Science and Engineering Education. Communications of the ACM 30(35), 30–35 (2007)

[2] Cheng, M., Chen, J., Chu, S., Chen, S.: The use of serious games in science education: a review of selected empirical research from 2002 to 2013. Journal of Computers in Education. 2, 353-375 (2015)

[3] Rocha, R., Rego, P., Faria, B., Reis, L., Moreira, P.: A Web Platform of Serious Games for Cognitive Rehabilitation: Architecture and Usability Study. New Advances in Information Systems and Technologies. 1085-1095 (2016)

[4] Sánchez, J.Olivares, R.: Problem solving and collaboration using mobile serious games. Computers & Education. 57, 1943-1952 (2011)

[5] Durkin, K., Boyle, J., Hunter, S., Conti-Ramsden, G.: Video Games for Children and Adolescents With Special Educational Needs. Zeitschrift für Psychologie. 221, 79-89 (2013)

[6] Connors, E., Chrastil, E., Sanchez, J., Merabet, L.: Action video game play and transfer of navigation and spatial cognition skills in adolescents who are blind. Frontiers in Human Neuroscience. 8, (2014)

[7] Cheng, M.Annetta, L.: Students' learning outcomes and learning experiences through playing a Serious Educational Game. Journal of Biological Education. 46, 203-213 (2012)

[8] Lahav, O. and Mioduser, D.: Haptic-feedback support for cognitive mapping of unknown spaces by people who are blind. International Journal on Human-Computer Studies, 66, pp. 23–35 (2008)

[9] Sánchez, J., & Sáenz, M. (2010). Metro navigation for the blind. Computers & Education, 55(3), 970-981.

[10] Bernsen, N.Dybkjær, L.: Multimodal usability. Springer, Berlin (2009)

[11] Gonzalez, C., Best, B., Healy, A., Kole, J., Bourne, L.: A cognitive modeling account of simultaneous learning and fatigue effects. Cognitive Systems Research. 12, 19-32 (2011)

[12] Ardito, C., Costabile, M. F., De Marsico, M., Lanzilotti, R., Levialdi, S., Roselli, T., & Rossano, V.: An approach to usability evaluation of e-learning applications. Universal access in the information society, 4(3), 270-283. (2006)

[13] Bellotti, F.; Berta, R.; De Gloria, A. Designing Effective Serious Games: Opportunities and Challenges for Research. International Journal of Emerging Technologies in Learning (iJET), v. 5, n. SI3, 2010

[14] Sánchez, J., Darin, T., & Andrade, R.: Multimodal Videogames for the Cognition of People Who Are Blind: Trends and Issues. In International Conference on Universal Access in Human-Computer Interaction, pp. 535-546. Springer International Publishing (2015)

[15] Hartson, H., Andre, T., Williges, R.: Criteria For Evaluating Usability Evaluation Methods. International Journal of Human-Computer Interaction. 15, 145-181 (2003)

[16] Chandrashekar, S., Stockman, T., Fels, D., & Benedyk, R.: Using think aloud protocol with blind users: a case for inclusive usability evaluation methods. In: 8th international ACM SIGACCESS conference on Computers and accessibility. pp. 251-252. ACM (2006)

[17] Bowman, D. A., Gabbard, J. L., & Hix, D.: A survey of usability evaluation in virtual environments: classification and comparison of methods. Presence: Teleoperators and Virtual Environments, 11(4), 404-424 (2002)

[18] Chilana, P. K., Wobbrock, J. O., & Ko, A. J.: Understanding usability practices in complex domains. In SIGCHI Conference on Human Factors in Computing Systems, pp. 2337-2346. ACM (2010)

[19] Forbrig, P., Bernhaupt, R., Winckler, M., Wesson, J. 2011. 5th Workshop on Software and Usability Engineering Cross-Pollination: Patterns, Usability and User Experience. Proc. Interact 2011, pp 718-719. (2011)

[20] Spencer, C., Morsley, K., Ungar, S., Pike, E., & Blades, M. (1992). Developing the blind child's cognition of the environment: the role of direct and map-given experience. Geofo-rum, 23(2), 191-197

[21] Raisamo, R., Hippula, A., Patomaki, S., Tuominen, E., Pasto, V., Hasu, M.: Testing Usability of Multimodal Applications with Visually Impaired Children. IEEE Multimedia. 13, 70-76 (2006)

[22] Darin, T., Sánchez, J., Andrade, R.: Dimensions to Analyze the Design of Multimodal Video games for the Cognition of People Who Are Blind. In: 14th Brazilian Symposium on Human Factors in Computing Systems, pp 2-11. Brazilian Computer Society (2015)

[23] Mayer, I. et al. The research and evaluation of serious games: Toward a comprehensive methodology. British Journal of Educational Technology, v. 45, n. 3, p. 502-527, 2013

[24] Yáñez-Gómez, R., Cascado-Caballero, D. & Sevillano, JL. Multimed Tools Appl (2017) 76: 5755. doi:10.1007/s11042-016-3845-9

[25] Olsen T., Procci K., Bowers C. (2011) Serious Games Usability Testing: How to Ensure Proper Usability, Playability, and Effectiveness. In: Marcus A. (eds) Design, User Experience, and Usability. Theory, Methods, Tools, and Practice. DUXU 2011. Lecture Notes in Computer Science, vol 6770. Springer, Berlin, Heidelberg

[26] Coutaz, J., Nigay, L., Salber, D., Blandford, A., May, J., & Young, R. M. (1995). Four easy pieces for assessing the usability of multimodal interaction: the CARE properties. In Human—Computer Interaction (pp. 115-120). Springer US

[27] Dumas, B., Solórzano, M., & Signer, B. (2013, August). Design guidelines for adaptive multimodal mobile input solutions. In Proceedings of the 15th international conference on Human-computer interaction with mobile devices and services (pp. 285-294). ACM

[28] Torrente, J., Blanco, A., Moreno-Ger, P., Martinez-Ortiz, I., Fernandez-Manjon, B.: Implementing Accessibility in Educational Videogames with e-Adventure. In: 1st ACM International Workshop on Multimedia Technologies for Distance Learning. pp 57-66. ACM (2009)

[29] Guerrero, J., Lincon, J.: AINIDIU, CANDI, HELPMI: ICTs of a personal experience. Engineering Applications (WEA), 2012 Workshop on. pp 1-7. Colômbia (2012)

[30] Trewin, S., Hanson, V. L., Laff, M. R., Cavender, A.: PowerUp: An Accessible Virtual World. In Proceedings of the 10th international ACM SIGACCESS conference on Computers and accessibility, pages 177-184. Halifax, Canada (2008)

[31] Lahav, O. and Mioduser, D.: Construction of cognitive maps of unknown spaces using a multisensory virtual environment for people who are blind. Computers in Human Behavior 24(3), pp. 1139-1155 (2008)

[32] Dumas, B., Lalanne, D., & Oviatt, S. (2009). Multimodal interfaces: A survey of principles, models and frameworks. In Human machine interaction (pp. 3-26). Springer Berlin Heidelberg.

[33] Darin, T. G., Andrade, R., Merabet, L. B., & Sánchez, J. H. (2017, May). Investigating the Mode in Multimodal Video Games: Usability Issues for Learners who are Blind. In Proceedings of the 2017 CHI Extended Abstracts on Human Factors in Computing Systems (pp. 2487-2495). ACM

[34] Miao, M., Pham, H. A., Friebe, J., & Weber, G.: Contrasting usability evaluation methods with blind users. Universal Access in the Inform. Society, 15(1), 63-76. (2016)

[35] Allain, K., Dado, B., Van Gelderen, M., Hokke, O., Oliveira, M., Bidarra, R., & Kybartas, B.: An audio game for training navigation skills of blind children. In Sonic Interactions for Virtual Environments (SIVE), pp. 1-4, IEEE (2015)

[36] Sanchez, J.: Software Usability for Blind Children Questionnaire (SUBC). Usability Evaluation Test, University of Chile (2003)

[37] Yuan, B.: Towards Generalized Accessibility of Video Games for the Visually Impaired. Ph.D. Dissertation, Univ. of Nevada, Reno, Reno, NV, USA. Advisor (S Frederick C. Harris and Eelke Folmer. AAI3355610. (2009)

[38] Simões, D., & Cavaco, S.: An orientation game with 3D spatialized audio for visually impaired children. In: 11th Conference on Advances in Computer Entertainment Technology, pp. 37-41. ACM (2014)

[39] Westin, T.: Game accessibility case study: Terraformers – a real-time 3D graphic game. In: Proceedings of the 5th International Conference on Disability, Virtual Reality and Associated Technologies, ICDVRAT 2004, Oxford, UK, pp. 95–100 (2004)

[40] Martz, W. (2010). Validating an evaluation checklist using a mixed method design. Evaluation and program planning, 33(3), 215-222

[41] Scriven, M. (2005). Checklists. In S. Mathison (Ed.), Encyclopedia of evaluation (pp. 53– 59). Thousand Oaks, CA: Sage

[42] Stufflebeam, D. L. (2000). Guidelines for developing evaluation checklists: the checklists development checklist (CDC). Kalamazoo, MI: The Evaluation Center

[43] Manakhov, P., & Ivanov, V. D.: Defining Usability Problems. In: CHI Conference Extended Abstracts on Human Factors in Computing Systems, pp. 3144-3151, ACM (2016)

[44] Pagliano, P. (2013). Using a multisensory environment: A practical guide for teachers. Routledge

[45] Molich, R., & Dumas, J. S. (2008). Comparative usability evaluation (CLUE-4). Behaviour & Infor-mation Technology, 27(3), 263-281.