Usability Evaluation of a Tangible User Interface and Serious Game for Identification of Cognitive Deficiencies in Preschool Children

Sánchez-Morales, A.1, Durand-Rivera J.A.2, Martínez-González C.L.*3
Programa de Posgrado en Ingeniería de Sistemas, SEPI ESIIME-Z, Instituto Politécnico Nacional, México1,3
División de Neurociencias, Laboratorio de Neuroprotección, Instituto Nacional de Rehabilitación, México2

Abstract—Detecting deficits in reading and writing literacy skills has been of great interest in the scientific community to correlate executive functions with future academic skills. In the present study, a prototype of a serious multimedia runner-type game was developed, Play with SID, designed to detect deficiencies in cognitive abilities in preschool children (sustained attention, memory, working memory, visuospatial abilities, and reaction time), before learning to read and write. Usability tests are used in Human-Computer Interaction to determine the feasibility of a system; it is the proof of concepts before the development of real systems. The aim of this paper was to evaluate the usability of the interface of the serious game, as well as the tangible user interface, a teddy bear with motion sensors. A usability study using the Wizard of Oz technique was conducted with 18 neurotypical preschool participants, ages 4 to 6. Concepts related to interactivity (interaction, the fulfillment of the activity objective, reaction to stimuli, and game time without distraction) were observed, as well as eye-tracking to assess attention and the Usability Scale System (SUS) to measure usability. According to the usability evaluation (confidence interval between 74.74% and 90.47%), the prototype has good to excellent usability, with no statistically significant differences between the age groups. The observed concept with the highest score was the game time without distraction. This characteristic will allow evaluating the sustained attention. Also, we found out that the tangible interface use leads to the observation of laterality development, which will be added to the design of the serious game. The use of observation-based usability assessment techniques is useful for obtaining information from the participants when their communication skills are developing, and the expression of their perception in detail is limited.

Keywords—User interface; wizard of Oz; usability; HCI; input device

I. INTRODUCTION

Cognitive skills related to reading and writing (reading and writing literacy), such as working memory, verbal comprehension, processing speed, and perceptual reasoning [1] have been identified as determining factors for the personal and social development of an individual [2]. Detecting deficits in these skills has been of great interest in the scientific community to correlate executive functions, processed in the prefrontal cerebral cortex, with future academic skills [3].

The term used to refer to a child who acquires literacy skills is late-emerging poor reader and was proposed by Chall, who determined that the deficit increased as the student progressed in his academic life [4]. The identification of these deficiencies is commonly carried out at a stage when students have already faced problems related to poor school performance [5].

There are studies on the use of multimedia technology for therapeutic purposes to detect cognitive deficiencies and improve them [6][7], which have shown encouraging results, specifically with the use of serious games. These games are characterized by having implicit objectives, in addition to the explicit ones of the game, such as learning or developing skills [8].

On the other side, tangible user interfaces (TUI) are used to improve existing learning tasks and an alternative to graphical user interfaces (GUI) to allow the user to control or navigate in a system with physical objects [9].

A. Similar Works

Among similar works, Valladares-Rodriguez et al. [10] developed games to detect cognitive deficits in older adults, specifically to link them to early detection of Alzheimer's Disease. Jung et al. [11] reported a remote assessment of cognitive disability with a mobile game. Tong and Chignell [12] proposed interesting recommendations for the development of serious games for cognitive assessment. To date, no serious game was found aimed at detecting cognitive deficiencies for literacy in children.

Shamilov et al. [13] developed a computer game with a tangible interface that verifies that the sense of touch increases the user's attention and participation in an activity. Schneider et al. [14] proposed a learning system based on a tangible user interface and complemented it with learning with traditional materials, they observed that the participants who first used the TUI and then studied the texts, performed better than those who first read and then they used the system.

B. Usability of Serious Games

As well as for any technological development, in serious games, usability is one of the most relevant aspects to determine if it can be used by specific users to achieve certain goals, with effectiveness, efficiency, and satisfaction, in a certain context of use [15]. This feature is intrinsically related to user-centered design and human-centered design processes and its most relevant activities are summarized in 1) the understanding and specification of the context of use, 2) the specification of the user and the organizational requirements.
3) the production of design solutions and 4) the evaluation of the designs concerning the requirements [2].

Assessing usability early in technology development is important, although this process is iterative [16]. Usability tests are used in Human-Computer Interaction to determine the feasibility of a system; it is the proof of concepts before the development of real systems [17]. These tests are considered as user research, although their main objective is not aimed at the user itself, but on learning about the participants, the use of the interface and the possible technologies that can be used [18].

Some specific cases of usability evaluation, according to the profile of the user, are children, elderly adults, and people with disabilities. In these cases, user-centered design is more than fundamental to the utility and usability of the application. Some of the conventional usability assessment methods have been adapted for these types of user profiles. Cano et al. [19] applied an evaluation method for the user experience of serious games of children with a cochlear implant, where not only is the user pediatric, but also has a type of disability. In that same sense, but considering the context of use, Sun et al. [20] evaluated the usability of a mobile application for pain management in children, in a hospital environment.

Among the methods for the usability evaluation, there are different techniques to be used according to the purpose of the evaluation, the type of prototype to be evaluated, and the characteristics of the participants, among others. Analyzing the needs of the evaluation allows identifying the appropriate method to carry out the usability test [21].

The types of prototypes for these tests are paper prototypes, diagrams of the screens without any functionality or with partial functionality, prototypes that appear to be functional, but a human reply behind the computer. Also, the tests can be carried out with final software versions, before its launch or with systems already implemented [18]. Regarding serious games, Olsen, Procci y Bowers [22] emphasize that evaluation should be carried out from the paper version [23] since the implicit objective of the game must be revised considerably.

In the present study, a prototype of a serious multimedia runner-type game was developed, Play with SID (SID for the acronym in Spanish for deficiency identification system), designed to detect deficiencies in cognitive functions in preschool children, before learning to read and write. The aim of this paper was to evaluate the usability of the interface of the serious game, as well as the tangible user interface, a teddy bear with motion sensors.

II. MATERIALS AND METHODS

A. Design of the Evaluated Prototype

The design of the serious game for the identification of cognitive deficiencies in preschool children and its control interface was obtained from a study based on rapid contextual design and participatory design techniques, reported in [24]. The design of the exercises to evaluate different cognitive skills, such as sustained attention, memory, working memory, visuospatial abilities, and reaction time involved in literacy, is described in a report to be published.
B. Study Design

A usability study was conducted with 18 participants whose inclusion and exclusion criteria were as follows: preschool students between 4 and 6 years old, who had not been previously diagnosed with any cognitive deficiency or motor disability and who did not suffer from allergies due to textiles of the tangible user interface (even so, different t-shirts for the teddy bear were used, made with hypoallergenic fabric).

The recruitment was carried out in a public school of preschool level in the metropolitan area of Mexico City. Table I depicts the demographic information of the participants. For a usability test, only 5 participants are needed to find approximately 80% of the problems of using an interface [26], this number is recurrent since it is stated that the number of failures found may depend more on the type of tasks and the design, than on the number of users [27].

C. Data Collection

The Human-Computer Interaction technique called Wizard of Oz was used; this technique mainly consists of simulating functionality that has not yet been developed. [28]. The user perceives that he is interacting with the system when, in reality, he is interacting with a human being (magician or evaluator) who is the one who provides the answers [29]. This technique allows evaluating a prototype before the development stages.

The test consisted of playing the game using the teddy bear or the tangible user interface as an input device. With the movement of the bear, the participants controlled the virtual character within the game.

The movements were made from the observation of the children, by the human wizard, who simulated the control movements. The context of the use was a preschool classroom, in an environment without distractors, as the serious game system would be used. Regardless of this non-threatening environment [30], a non-participatory observation was made by a teacher and some parents. A computer, a monitor, and a web camera were used. Also an additional camera also recorded the test. The layout of the installation is shown in Fig. 4(a)-(d).

<table>
<thead>
<tr>
<th>Age distribution</th>
<th>Gender</th>
<th>Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 year-old</td>
<td>5 year-old</td>
<td>6 year-old</td>
</tr>
<tr>
<td>Number</td>
<td>4</td>
<td>10</td>
</tr>
<tr>
<td>%</td>
<td>22.22</td>
<td>55.55</td>
</tr>
<tr>
<td>Male</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>Female</td>
<td>38.88</td>
<td>61.11</td>
</tr>
<tr>
<td>2nd</td>
<td>16</td>
<td>88.88</td>
</tr>
</tbody>
</table>
After explaining the test procedure to the parent or guardian, the child was instructed to play a game on the computer, using the bear. Subsequently, the informed consent of both was requested, also for the recording of the test. The teddy bear was given to the child as the input device, and he/she was given a T-shirt in the color of his/her choice. During the test, the instructor (bee) shows how to use the control (teddy bear) with animations, no more detailed indications of the exercise were given. The average test time with each participant was 10 minutes.

D. Observed Concepts

Four concepts were observed during the usability test to review the interactivity of the serious game interface. These are described in Table II.

The observation of each participant was carried out during the test, and each session was videotaped, to be evaluated later. An evaluation scale of 0 to 2 was proposed to quantify the observation data, based on [31], where 0 is equivalent to the fact that the participant failed to achieve the observed concept, 1 is equivalent to the fact that the participant managed to achieve difficulties and 2 means that the participant managed to achieve without any problem.

E. Usability Measure Instrument

The System Usability Scale (SUS) questionnaire was used to measure usability [32]. It consists of 10 statements (Table III) in which users rate the level of agreement or disagreement; the scores are on a scale of 1 to 5, where 1 corresponds to totally disagree and 5 to totally agree. In the present work, the SUS statements were adapted according to the age range from 4 to 6 years and the evaluated technological development, a serious game with a physical control interface. It was used in a Spanish version.

As it is stated for the evaluation of the SUS, results were carried out on with a scale of 0 to 4, obtained by subtracting a point from the odd statements and for even questions, the number given by the user in the answer must be subtracted from five. The sum of these results must be multiplied by 2.5 to obtain an evaluation percentage. This percentage is interpreted as not acceptable (<50%), marginal (50-70%) or acceptable (> 70%). Among these items, 4 and 10 are usually identified to refer to learnability and the rest to usability [27].

<table>
<thead>
<tr>
<th>Observed concept</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interaction with the game</td>
<td>Identify that the movements of the bear control the character within the game</td>
</tr>
<tr>
<td>Fulfillment of the objective of the activity</td>
<td>Understand the purpose of the activity presented in the game</td>
</tr>
<tr>
<td>Reaction to stimuli</td>
<td>Identify that the necessary action is carried out for each stimulus, either pick apples or dodge obstacles</td>
</tr>
<tr>
<td>Game time without distraction</td>
<td>Maintaining attention in the game throughout the duration of the test</td>
</tr>
</tbody>
</table>

F. Eye-Tracking

Eye-tracking is a technique that allows evaluating eye movements and their sequence to understand the processing of the information received from the screen and the behavior during a usability test [34]. It has also been linked to the point of interest of attention in an interface and has previously been used in the study of serious games [35]. This technique was used to obtain additional information about the interactivity with the prototype. The eye-tracking software used was Gaze Recorder, with the webcam placed on the monitor.

G. Analysis of Data

Kruskall-Wallis tests were run for independent samples to determine statistically significant differences between the age groups for the results of the concepts observed during the test (Table II), as well as for the results of the usability test with the SUS (Table III). Statistical analysis was performed with SPSS Statistics software.

III. RESULTS

The results of each user for the concepts observed during the test to evaluate interactivity are depicted in Table IV.

The mean of the evaluation of the observed concepts for all the participants was 1.5 for the interaction with the game (SD = 0.57), 1.5 for the fulfillment of the activity objective (SD = 0.51), 1.61 for the reaction to the stimuli (0.51) and 1.66 (SD = 0.5) for the game time without distraction (SD = 0.59). The scores obtained with the System Usability Scale SUS are shown in Table V.

<table>
<thead>
<tr>
<th>Age group</th>
<th>Game interaction</th>
<th>Activity objective</th>
<th>Stimulus reaction</th>
<th>Game time without distraction</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 year-old</td>
<td>Mean: 1.5 SD: 0.57</td>
<td>Mean: 1.75 SD: 0.5</td>
<td>Mean: 1.75 SD: 0.5</td>
<td>Mean: 1.5 SD: 0.57</td>
</tr>
<tr>
<td>5 year-old</td>
<td>Mean: 1.3 SD: 0.48</td>
<td>Mean: 1.4 SD: 0.51</td>
<td>Mean: 1.5 SD: 0.52</td>
<td>Mean: 1.7 SD: 0.67</td>
</tr>
<tr>
<td>6 year-old</td>
<td>Mean: 2 SD: 0</td>
<td>Mean: 1.5 SD: 0.57</td>
<td>Mean: 1.75 SD: 0.5</td>
<td>Mean: 1.75 SD: 0.5</td>
</tr>
</tbody>
</table>
TABLE V. SYSTEM USABILITY SCALE SCORE

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Usability 95%-CI</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>82.61%</td>
<td>15.82%</td>
<td>53</td>
<td>100</td>
<td>[74.74%, 90.47%]</td>
</tr>
<tr>
<td>Female</td>
<td>79.09%</td>
<td>18.78%</td>
<td>55</td>
<td>100</td>
<td>[66.47%, 91.71%]</td>
</tr>
<tr>
<td>Male</td>
<td>89.14%</td>
<td>8.33%</td>
<td>73</td>
<td>100</td>
<td>[81.43%, 96.85%]</td>
</tr>
</tbody>
</table>

CI = Confidence interval

The mean of the SUS test results was 82.61% (SD = 15.82) with the participants in the age range. Fig. 5 shows the results of the evaluation of the prototype by age range. In the group of 4 years, the results were found from 65% to 100% with a mean of 86.5% (SD = 15.15); in the 5-year group, 53% to 100%, with a mean of 82% (SD = 17.34) and finally, the 6-year-old group had a minimum evaluation of 58% and reached a maximum of 95% with a mean of 80.25% (SD = 16.04). One of the volunteers was outside the age range; however, his participation was considered to contrast their answers illustratively, but it was not counted within the sample.

Table VI and Table VII show the results of the Kristall-Wallis test for independent samples between age groups for evaluation with SUS.

TABLE VI. KRISKALL-WALLIS TEST RANKS RESULTS FOR SUS EVALUATION AMONG AGE GROUPS, RANKS

<table>
<thead>
<tr>
<th>Age group</th>
<th>N</th>
<th>Mean rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>4 year-old</td>
<td>4</td>
<td>10.75</td>
</tr>
<tr>
<td>5 year-old</td>
<td>10</td>
<td>9.50</td>
</tr>
<tr>
<td>6 year-old</td>
<td>4</td>
<td>8.25</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
<td></td>
</tr>
</tbody>
</table>

TABLE VII. KRISKALL-WALLIS TEST STATISTICS FOR SUS EVALUATION AMONG AGE GROUPS, TEST STATISTICS A, B

<table>
<thead>
<tr>
<th>SUS Evaluation</th>
<th>Chi-square</th>
<th>df</th>
<th>Asymp. Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.443</td>
<td>2</td>
<td>0.801</td>
</tr>
</tbody>
</table>

a. Kruskal-Wallis Test
b. Grouping variable: Age group

Fig. 6 depicts color maps of the zones where the sight of the participants was focused.
IV. DISCUSSION

The SUS confidence interval, obtained for the evaluation of the sample, between 74.74% and 90.47%, according to the evaluation scale, implies that the usability of the system is in the “acceptable” range, defined in [33]. Therefore, users evaluate the prototype favorably which meets the usability criteria. Regarding the ranking of adjectives, it can be classified between “good” and “the best imaginable”, according to [36].

The results of the Kristall-Wallis test for independent samples indicated that there are no statistically significant differences between the age groups in the evaluation of the observed concepts, as well as in the evaluation with the SUS. Therefore, the result obtained for the sample used in this study can be generalized.

It is important to highlight that the only participant outside the age range obtained an evaluation of 0 in the concepts observed for interactivity, presented in Table II (interaction with the game, fulfillment of the activity objective, reaction to stimuli, and time of game without distraction). In contrast, its percentage evaluation of the game with the SUS was 20%, indicating a high score in the item that evaluated the difficulty of the game. Despite being a single participant under the age of 4, it is notorious that the serious game design is not aimed for this age range, under four years.

Regarding eye monitoring, bias was found in the experimental procedure, since not all participants were adequately captured due to the webcam used for this purpose and placement. In the case of child volunteers, with different heights, it is necessary to fix the face to homogenize the calibration of eye-tracking. However, this monitoring allowed studying the area of the game interface in which the attention of the participants was focused, in such a way it could be confirmed that there are distractors not considered in the design. The results obtained showed that there is a relationship of attention in the areas where the stimuli were presented. In the welcome screen where the control of the main character is explained, the attention was focused on the animation of the instructor character (the bee).

Among the limitations of this study, it was focused on evaluating the usability of a serious game using the proposed tangible user interface. Hence, not all the exercises that involve the evaluation of cognitive deficiencies for the detection of cognitive deficits were evaluated.

In this sense, although the SUS usability and learnability subscales have been commonly used according to the items identified for such, Lewis [33] recommended reporting it as a one-dimensional metric.

Regarding the use of SUS in Spanish, it has been successfully used in this language, although there is no validated Spanish version [37], [38]. On the other side, considering the use of the SUS with children, in a major part of these usability studies, children are accompanied by their parent or guardian. In certain applications, adult intervention is necessary, for example, with therapeutic education for children or adolescents and their caregivers [20], [39]–[41]. In our study, the specific user profile was in an age range between 4 and 6 years, and no adult intervened to explain the detail of the activity; there were no difficulties in completing it. Subsequently, no problems were applying the SUS.

V. CONCLUSIONS

With the completion of the Wizard of Oz test, the importance of evaluating the usability before the development of the systems can be identified, obtaining information on how to use it, the opinions of the participants, and identify characteristics that require improvement. The observation of the environment where the activity took place within the school and not in a controlled environment, allowed to know in greater detail the technical requirements of the system.

According to the usability evaluation with the SUS questionnaire, the prototype has good to excellent usability, with no statistically significant differences between the age groups.

Regarding observations during and after the usability test, show that participants were favorably evaluated in the observed concepts on the interaction with the game, causing the most conflict to understand each other. We assume that the function of the tangible user interface was not explained intentionally before starting the test, and the volunteers had to learn how to use it on the fly.

This study is part of the iterative design of the prototype. Therefore, the observations made will improve the high-fidelity prototype.

The observed concept with the highest score was game time without distraction, which means that the prototype design allowed participants to maintain their attention for the required time. This characteristic will allow evaluating sustained attention as a cognitive ability. In this regard, one of the participants presented a deficit in the development of laterality, which was observed, since the movement of the bear made it to the opposite side of the required one.

The use of observation-based usability evaluation techniques is useful to obtain information from the participants when they are preschool children because their communication skills are still developing, the expression of their perception in detailed form is limited.

On the other hand, carrying out the usability evaluation in low and medium fidelity prototypes, before development, provides a significant complement, which allows identifying user behavior according to the methodological process and the context of use.

In this study, the characteristics of a serious game and a tactile user interface for children were successfully evaluated. The Wizard of Oz test in such young children is outstanding to test prototypes, allowing them to collect information about users through observation since it is challenging to achieve extensive descriptions at such an early age.

We also successfully applied the SUS to evaluate the usability of a tangible user interface, all along with a serious game in preschool children, without the intervention of the parents or other adults.
As the aim of the study was to evaluate the user interface with the serious game, not all the exercises that involve the evaluation of cognitive deficiencies for the detection of cognitive deficits were evaluated. Though, useful improvements were achieved for the final design, including additional executive functions to be evaluated and the validation of the design for children between four and six-years-old.

Future studies in this subject would include the comparison of different usability measurement tools specifically for tangible interfaces and serious games for children, also for the case of special needs software.

ACKNOWLEDGEMENT

We acknowledge the support of Instituto Politécnico Nacional and Consejo Nacional de Ciencia y Tecnología (CONACyT). We also acknowledge the participants, their parents, teachers, and administrative staff involved in this study.

REFERENCES


