International Journal on Advanced Science Engineering Information Technology
Vol.8 (2018) No. 6
ISSN: 2088-5334

Evaluating Usability of Serious Games for Therapeutic Purpose: A Study of ASAH-*i*

Nadia Akma Ahmad Zaki[#], Tengku Siti Meriam Tengku Wook[*], Kartini Ahmad[+]

[#]Faculty of Art, Computing and Creative Industry, Universiti Pendidikan Sultan Idris,Tanjong Malim, Perak, Malaysia
E-mail: [1]nadiaakma@fskik.upsi.edu.my

[*]Center for Software Technology and Management, Faculty of Information Science and Technology,
Universiti Kebangsaan Malaysia, 43600, Bangi, Selangor, Malaysia
E-mail: [2]tsmeriam@ukm.edu.my

[+]Faculty of Health Sciences, Universiti Kebangsaan Malaysia, Jalan Raja Muda Abdul Aziz,Kuala Lumpur, Malaysia
E-mail: [3]kart@ukm.edu.my

***Abstract*— Usability has been determined to be connected to achieving predetermined practical goals by identifying usability problems during the development cycle of a product or system. In order for the next and future versions of the systems to be enhanced and improved, usability flaws and errors must first be identified. On the same stroke, identifying flaws encountered by the end users will indirectly steer other developers from committing the same mistakes. This paper reveals findings from a usability assessment conducted on the serious game, which was developed for therapeutic purpose known as ASAH-*i*, which was custom-made to develop and encourage the perception and cognitive skills for children with speech and language delay (CSLD). Two groups of users were formed, namely, test group and control group. Twelve participants were selected which include six CSLD as participants for test group and another six normal children for the control group. In a controlled environment, observation and think-aloud methods were applied to analyse ASAH-*i*. The data gathered were then evaluated accordingly to the task success, time of the user assignment accomplishment, error rates, and satisfaction in regards to the utilisation of ASAH-*i*. The result proved that there is no profound disputes between the two corresponding groups, where almost all participants for both groups can play ASAH-*i* prototype effectively since no major usability issues were found. With regards to competency or efficiency, the result reveals that the control group can play ASAH-*i* prototype more competently and efficiently than the test group. However, it is interesting to note that both group members also succeeded to play ASAH-*i* with very little mistakes. This factor yields a strong base that provides towards the group members' contentment and enjoyment, in which most group members in both teams were very much happy and satisfied with ASAH-i. With this fact in mind, we propose an endorsement or recommendation study which includes details and emphasises on future development to enhance the usability of ASAH-*i*.**

***Keywords*—therapeutic serious game; cognitive stimulation; human-computer interaction; user-centred design; usability.**

I. INTRODUCTION

Serious games have been used widely to encourage mental and physical exercise, military, education, health, training, and knowledge acquisition. Serious games had recently been made known to and got itself acquainted with the therapeutic purpose. Moreover, serious games blended well with the therapeutic context. It had caught the full attention of healthcare practitioners, and they concurred that therapeutic severe games to be a very efficient tool or form of medical therapy, by providing series of exercises that will improve individual skills of its users [1]. In a nutshell, therapeutic objectives are integrated with fun game elements and commonly used as additional tools to assist healthcare practitioners in providing treatment [2], [3]. Also, therapeutic severe games also increase user motivation and engagement during treatment, provide comfort and yield a positive impact to the therapists and patients as well [3], [4].

Recently, therapeutic severe games had emerged as a good option for cognition-focused interventions such as cognitive training, cognitive rehabilitation and cognitive stimulation which will benefit into gains on general cognitive ability [5]–[8]. Cognitive stimulation works in a way that it promotes increased user engagement in mentally stimulating activities where engagement levels are associated with the rates of cognitive decline. In a simpler term, it means that the more the user actively participates and gets involved in the activities, hence the lower the rate of the cognitive decline will be [9]. According to [10], cognitive stimulation intervention is suggested to be the most efficient remedy against all of the non-pharmacological therapies that give positive outcomes. These series of actions or exercises are individually custom-made to encourage and enhance the use of cognitive skills. It aims to indirectly help to curb and reduce cognitive weaknesses, while simultaneously enhancing it users' cognitive functions or skills. The basis of such an argument is that specific tasks or actions may enhance its user's cognitive skills, or at the very least keeps its users' cognitive skills to be functioning at a minimal level. This is achieved by performing the same, repetitious tasks or exercises in that domain and that any effects of practice could be generalized and also induce a general improvement in cognitive and social functioning.

Although there is an increasing numbers of researches being done recently on the utilisation of serious games as a therapeutic method to stimulate the cognitive skills of a diverse group of people [11], several hiccups were encountered in developing these games specifically those CSLD since they encounter big challenges in constructing their cognitive skills and abilities [12]. With that in mind, it is important that good usability should be considered and subsequently integrated into these games. Ultimately, this would greatly help to achieve the goal of enhancing and stimulating the cognitive abilities for CSLD users with therapeutic implications.

The human-computer interaction (HCI) greatly emphasizes on the usability factor and is a standard quality attribute for the design and evaluation of interactive systems. According to Jacob Neilson, "usability is a quality attribute that assesses the ease of employing user interfaces which also refers to methods for improving ease-of-use during the design process" [13]. He concludes that the usability of a design project should consist of some quality factors such as below [13]:

1) *Learnability* is the degree to which something can be learned efficiently and the simplicity of use by the users to complete elementary tasks when they encounter the new design.

2) *Efficiency* is a method to measure and appraise the number of works required to finish a certain task.

3) *Memorability* is an evaluation of how effortlessly or swiftly a system to remember after a substantial time-lapse between visits.

4) *Errors* are faulty moves that may lead to unsuccessful results or any action that does not achieve the pre-defined objectives.

5) *Satisfaction* refers to the users' comfort and positive attitudes in regards to the utilization of the system.

Meanwhile in ISO-9241-11:1998, usability is defined as "the extent to which a product can be used by identified users to attain specified goals effectively, efficiently, and satisfaction in a specified context of use" [14]. This definition elaborates on how to sift through and recognize the vital information that is needed to be considered when classifying or ranking usability regarding the evaluation of user performance and satisfaction. However, in the new version of ISO-9241-11, the original definition of usability has been widened to include systems and services and to also cover recent methods to usability and user experience [15].

According to usability definition in ISO-9241-11:1998, three main concerns can be used as guidelines to ensure a product or system to be usable; effectiveness, efficiency, and satisfaction. Effectiveness has been associated with completing a task completely and accurately. However, the new version of ISO-9241-11 has redefined this as "accuracy, completeness, and lack of negative consequences with which users achieved specified goals" [15]. Previously, efficiency was defined as "the ratio of effectiveness divided by the resources consumed," but this has been revised as "the resources (time, human effort, costs and material resources) that are expended when achieving a specific goal (e.g., the time to complete a specific task)" [16]. On another spectrum, in the current ISO 9241-11, satisfaction also considers the wider range of concerns that are now believed to be vital for user experience: "positive attitudes, emotions and/or comfort resulting from the use of a system, product or service" rather than "freedom from discomfort and positive attitudes toward the use of the product" [16]. These three factors, namely effectiveness, efficiency, and satisfaction are linked to the cognitive, affective, and psychomotor responses of an individual [15].

User-centered design (UCD) is the owner of a significant discipline, and one of its core responsibility is to make things more usable. UCD embodies and comprises the processes, methods, and procedures by positioning the user at the very core of the process for inventing a usable product. According to [17], there are three fundamental elements of UCD that are important to grasp the context for performing usability testing; one of them is an early focus on users and their tasks. As UCD aims to support and sustain how real users work, instead of bulldozing them to convert the way they are used to in doing things.

Usability testing can happen through all stages of development and known as one of significant activity for UCD [18]. Beginning from the early design phases or early prototypes, the usability researcher inaugurates usability trials to recognize usability flaws before launching the product in the real context, as well as to enhance the quality of an interface by identifying problematic-areas of the interface that need improvement [18], [19]. Usability testing can also mean a process that is executed and carried out by a representative user attempting representative tasks in representative environments as not all of them are the same thus they might face various problems [13], [17], [18].

Also, representative users can highlight any design issues that can be fixed at the early stage. Thus, in turn, this could help to eliminate frustration for users [17]. Another critical reason to conduct usability testing is to make sure game designers build the interfaces that match the needs of the users since most of them are inclined to depend solely on their understanding, intuition, familiarity, and interests [20]. Therefore, at any stage of development, the interface prototype needs to be tested by the users.

Interactive prototypes allow designers to test design concepts accordingly. The designers who utilized an interactive prototype are observed to detect more profound problems in the early design phase [21]. With this early detection, designers were able to rest assured that the prototype had been developed accordingly to match user requirements. On the same stroke, if there are any incompatibility issues, the developers would be able to correct any of the design issues before the target users use the product [22], [23].

The number of participants plays a leading role when devising a plan involving user testing with children to discover enough percentage of usability issues. Nielsen believes that "three to five test participants are sufficient to represent 80% of the usability problems" [24], [25]. In contrast, [26] argue that "eleven to thirteen children are needed as test participants to detect 80% of the problems". Moreover, Lazar et al. discovered that more than eight participants are sufficient to reveal much higher percentages of problems [18].

There were a significant amount of studies that involved the children in testing the product which was aimed for them [23], [27], [28]. When conducting usability testing with children, researchers are required to pay attention to the principles and issues of the participating children. The ethical highlights are regarding the safety of the children, such as to give guarantees to the parents and obtaining approvals from their parents to involve their offspring in the test. On top of that, researchers can choose the right children to involve in the testing [29].

During the design process, children can play interchangeable roles such as user, tester, informant, and design partner [19]. Lately, children with special needs have involved in the design process with diverse roles of participation [30]–[32]. Guha et al. state that "children with more severe disabilities such as autism are to be expected to have a more limited role such as tester whereas children for whom the nature of disability is less severe may have a more involved role such as designed partner" [30]. Besides that, when designing games for children with special needs and have wide range of disabilities; preferred culture, skills, and complexities and also their involvement in the design process must be considered by game designers [20], [30]. In the context of serious games that are related to health intervention, for children with the specific constraints on cognitive capabilities, their characteristic also needs to be considered since it might influence the method and result of the usability testing [33].

Numerous studies have engaged in associating the appropriate usability testing methods to children [29]. There are qualitative usability appraisal methods used for usability testing with children, such as introspection, direct observation, and interviews and questionnaires [29]. Simple observation and think-aloud are proper methods in direct observation to recognize the usability issues. Simple observation can be used to measure the usability components such as efficiency, effectiveness, and satisfaction and has been authenticated as the top choicest techniques in usability testing [34]. On another spectrum, think-aloud is a technique that required users to frequently and regularly vocalize their thoughts, instinctually declare their feel, excitement and the struggles they encounter during carrying out the user testing. Study of Donker and Markopoulos has highlighted that children who think-aloud tend to highlight problems more than children who answer specific questions during testing [35]. However, Barendregt & Bekker express their disagreement that young children are often not very good at thinking aloud and suggest that the evaluator has to interpret a lot of non-verbal behavior and incomplete sentences [26].

Two types of data can be obtained during usability testing, namely, qualitative and quantitative data. Qualitative data deals with a direct assessment of the usability of a system or product such as observation of participants and a verbal track to the talkative participants are provided. For participants who are not very talkative, moderators should try to engage them in conversation and urge them to express and articulate what they are feeling using the think-aloud method. Meanwhile, quantitative data deals with an indirect assessment of the usability of a design. Usually, it can be based on users' performance of a given task, time performance, or reflect participants' perception of usability[18].

Thus, this study aims to present the usability testing of ASAH-*i*. ASAH-*i* was developed and precisely to meet the needs of CSLD aged between 4 and 7 years old, with the aims to stimulate their cognitive abilities and skills. ASAH-*i* prototype will be evaluated and reviewed for usability by CSLD and normal children in this usability testing. The main usability metrics to be examined are effectiveness, efficiency, error, and user satisfaction with the ASAH-*i* prototype. Subsequent section then describes the research methodology employed. Sections III dissects the result, and Section IV sums up the paper and devises intentions for future work.

II. Material And Method

A. Participants

Although five participants are sufficient [24], increasing the number of tested participants will show a clearer picture of the harshness of the problems. Hence, more problems will be identified, and data confidence will increase [18], [26]. Therefore, twelve children (N=12) were selected for this usability testing. The children are then separated into two groups, individually, the test group and the control group. Control group was included in the study to enhance the validity of the measurements.

1) Test Group

Six participants in this group were diagnosed as CSLD and intentionally chosen by professionals (see Table I). They consist of four boys and two girls, aged between six and seven years old (Mean age = 6.3, SD = 0.5). All of them were from Audiology and Speech Clinic, which operated in the Faculty of Health Sciences, Kuala Lumpur. They met the criteria of CSLD (mild and moderate) and suffered from major weaknesses that influence the development of their cognitive abilities.

2) Control Group

For the control group, six children (two boys and four girls) between the age of four to six years (Mean age = 4.8, SD = 0.8) were normal children who were chosen by random (see Table I). The children selected for the control group attend the Emaan Kindy Islamic Kindergarten in Kajang, Selangor. The regular children also participated in this testing as they would be able to verbalize and justify their preferences actively, thus, can convey and sum up their experience into verbal statements [23], [35].

Parents' authorization was obtained to conduct the usability testing with the children. Parents were asked to fill in the permission forms for their children's participation and were also given the reassurance that this test will not harm their children. The study was performed in remote and isolated rooms in two different places. These rooms were equipped with tables, chairs, and a video recorder to monitor the participants as well as to capture and record the screen images. A usability study for CSLD was conducted at the Audiology and Speech Clinic after getting approval from the Ethics Review Board of Universiti Kebangsaan Malaysia. Meanwhile, another study was conducted at the Emaan Kindy Islamic Kindergarten.

TABLE I
DEMOGRAPHICS

Group	Participant	Gender	Age	Child's diagnosis
Test group	1	Male	6	CSLD, Learning disabilities
	2	Male	6	CSLD, Learning disabilities
	3	Male	6	CSLD, Autism
	4	Female	6	CSLD, Global Developmental Delay
	5	Female	7	CSLD, Autism, Learning disabilities
	6	Male	7	CSLD, Down Syndrome
Control group	7	Male	6	Normal child
	8	Female	5	Normal child
	9	Male	5	Normal child
	10	Female	5	Normal child
	11	Female	4	Normal child
	12	Female	4	Normal child

B. The ASAH-i Prototype

The interactive ASAH-*i* prototype was built following the proposed guidelines from [36]. ASAH-*i* is a 2D game and is running via the Android platform and has been installed in the tablet. ASAH-*i* prototype provides five different types of game (see Fig. 1). Games 1, 2 and 3 require children to give attention and identify the objects surround them. Game 4 trigger the children to focus and ability to make a decision. Game 5 was developed to recall the short term memory of the object that has been viewed.

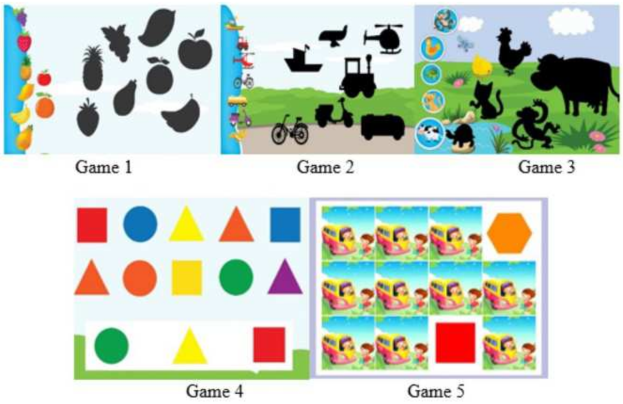

Fig. 1 ASAH-*i* games

C. Usability Test Procedure

Two researchers were involved in this test [20], where one of them acted as a moderator and sat facing the participants to guide them through the tasks and to stimulate for their views. Meanwhile, the other researcher observed the test and took the video recording of each participant as backup data to be used for data analysis.

The participants were tested individually in both groups which lasted approximately 20 to 30 minutes for a single test session for each participant (see Fig. 2). The stages of usability testing from [17] have been used as a guideline to execute this test. The session begins with a briefing by the moderator about the task the participants need to do. Each participant is required to play these five games twice in order to observe and record the differences during the first and second attempts.

Fig. 2 Participants during the usability testing

III. Results and Discussion

The ASAH-*i* prototype intends to explore and gauge primary usability components such as effectiveness, efficiency, errors and user satisfaction. These critical components were measured through a combination of observation and think-aloud method for both test and control groups and were presented as task success, time on task, error rates and satisfaction.

The participants from the test group, however, were found to be least responsive with this think-aloud methodology as they could only give simple answers when prompted for their opinions while playing the games. Moreover, they also responded mostly by making gestures or through body language and hardly produced any verbal expressions. So, in order to stimulate their thoughts and invoke meaningful verbal responses, a close engagement by normal children throughout the games was necessary.

The analysis performed based on the data collected from this testing employs 95% confidence interval level to show the inconsistency in the time data and help to visualize the difference between games [37], [38]. The summarised results are further explained below:

A. Task Success

Task success is a measure of how participants can complete a game effectively, with a score of 1 for success and 0 for failure [38]. Table II and III show the task success data for test and control group respectively with the mean score of two attempts calculated for each game.

TABLE II
TASK SUCCESS DATA FOR TEST GROUP

PARTICIPANT_ID	GAME _NO_1		GAME _NO_2		GAME _NO 3		GAME _NO 4		GAME _NO 5	
	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2
1	1	1	1	1	1	1	1	1	1	1
2	1	1	1	1	1	1	1	1	0	0
3	1	1	1	1	1	1	1	1	1	1
4	1	1	1	1	1	1	1	1	1	0
5	1	1	1	1	1	1	1	1	1	1
6	1	1	1	1	1	1	1	1	1	1
Average	**1.00**		**1.00**		**1.00**		**1.00**		**0.75**	
Confidence level (95 %)	**0%**		**0%**		**0%**		**0%**		**44%**	

TABLE III
TASK SUCCESS DATA FOR CONTROL GROUP

PARTICIPANT_ID	GAME _NO_1		GAME _NO_2		GAME _NO 3		GAME _NO 4		GAME _NO 5	
	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2
7	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	1	1	1	1	1
9	1	1	1	1	1	1	1	1	1	1
10	1	1	1	1	1	1	1	1	1	1
11	1	1	1	1	1	1	1	1	1	1
12	1	1	1	1	1	1	1	1	1	1
Average	**1.00**		**1.00**		**1.00**		**1.00**		**1.00**	
Confidence level (95 %)	**0%**		**0%**		**0%**		**0%**		**0%**	

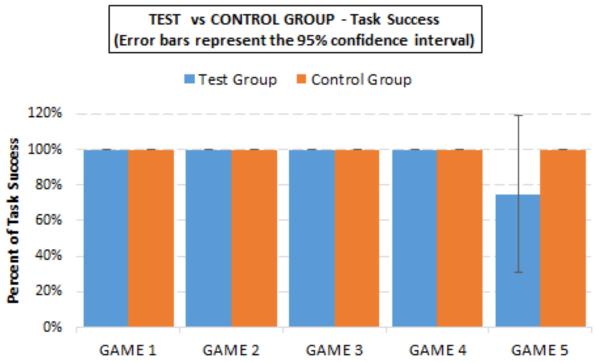

Fig. 3 Percentage of average task success rate for test and control group

Based on Fig. 3, it was evident that all participants from both groups managed to complete Game 1 to Game 4 successfully. However, only 17%, i.e., two out of twelve total participants failed Game 5, hence suggesting this game requires some modifications. Considering 83% of participants were able to finish all games successfully, it can be concluded that ASAH-*i* prototype did not display any significant usability issues.

B. Time on Task

The time taken for a participant to complete a given task was also being measured and known as time on task [38]. It began the moment a participant clicked on a start button and ended when the person completed the task. All participants from both groups would undergo two round of attempts for each game before the average score was calculated (see Table IV and V).

TABLE IV
TIME ON TASK DATA FOR TEST GROUP
(ALL DATA ARE EXPRESSED IN SECONDS)

PARTICIPANT_ID	GAME_ NO_1		GAME_ NO_2		GAME_ NO 3		GAME_ NO 4		GAME_ NO 5	
	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2
1	20	22	13	12	18	17	19	10	104	56
2	60	34	14	12	21	18	65	12	79	184
3	36	34	35	20	47	60	44	33	85	30
4	178	92	58	38	39	30	68	32	166	295
5	23	17	13	10	16	14	14	12	49	38
6	56	20	25	29	31	24	29	18	62	76
Average	**49.33**		**23.25**		**27.92**		**29.67**		**102.00**	
Confidence level (95 %)	**45.36**		**14.87**		**15.17**		**15.67**		**73.25**	

TABLE V
TIME ON TASK DATA FOR CONTROL GROUP
(ALL DATA ARE EXPRESSED IN SECONDS)

PARTICIPANT_ID	GAME_ NO_1		GAME_ NO_2		GAME_ NO 3		GAME_ NO 4		GAME_ NO 5	
	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2
7	47	18	15	11	31	17	22	13	55	43
8	30	18	14	9	14	17	17	10	94	48
9	69	29	25	14	31	19	24	10	101	94
10	29	24	17	11	33	30	32	18	82	52
11	50	26	20	14	34	25	32	17	153	101
12	23	31	12	10	31	28	23	18	47	90
Average	**32.42**		**13.67**		**19.83**		**19.25**		**84.83**	
Confidence level (95 %)	**13.43**		**3.05**		**4.03**		**10.03**		**29.16**	

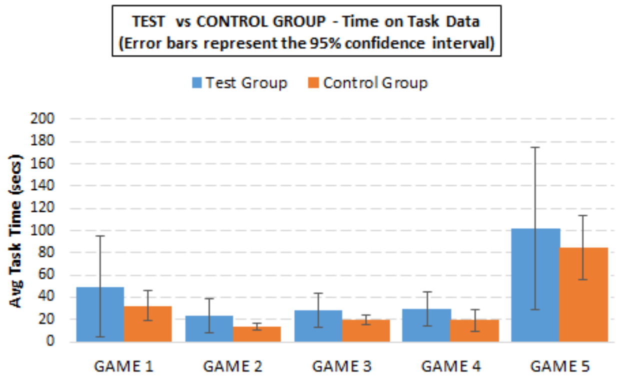

Fig. 4 Average task time for test and control group

Time on task is an appropriate way to measure the efficiency of a product [15], [38]. Fig. 4 compares the average task time for test and control group, and it was clear that the latter was able to complete the tasks much faster than the former – an indication that they can play the ASAH-*i* prototype more efficiently. There exists a correlation between task on time and learnability, in which a lower value of task on time (or lower common task on time), implies higher learnability. As both groups managed to complete all games within relatively a short period (under 1 minute for Game 1 to Game 4, and 2 minutes for Game 5), therefore it can be deduced that participants from both groups exhibited higher learning aptitude towards this prototype.

C. Error Rates

Errors committed by the participants while playing games were tracked and shown as error rates and they would receive a zero (0) score for no error and 1 for any mistakes that they made [38]. All participants from each group played twice for every game, and the error rates were recorded and presented in Table VI and VII.

TABLE VI
ERROR RATES DATA FOR TEST GROUP

PARTICIPANT_ID	GAME _NO_1		GAME _NO_2		GAME _NO 3		GAME _NO 4		GAME _NO 5	
	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2
1	0	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	0	1	1	0	0
3	0	0	1	1	1	0	1	1	0	0
4	0	0	0	0	0	0	1	1	1	0
5	0	0	0	0	0	0	0	0	0	0
6	0	0	0	0	0	0	1	0	0	1
Average	**0.00**		**0.20**		**0.10**		**0.60**		**0.20**	
Confidence level (95 %)	**0%**		**43%**		**21%**		**52%**		**27%**	

TABLE VII
ERROR RATES DATA FOR CONTROL GROUP

PARTICIPANT_ID	GAME _NO_1		GAME _NO_2		GAME _NO 3		GAME _NO 4		GAME _NO 5	
	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2	ATTEMPT_1	ATTEMPT_2
7	0	0	0	0	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0	0
10	0	0	0	0	0	0	1	1	0	0
11	0	0	0	0	0	0	1	0	0	0
12	0	0	0	0	0	0	1	1	0	0
Average	**0.00**		**0.00**		**0.00**		**0.40**		**0.00**	
Confidence level (95 %)	**0%**		**0%**		**0%**		**52%**		**0%**	

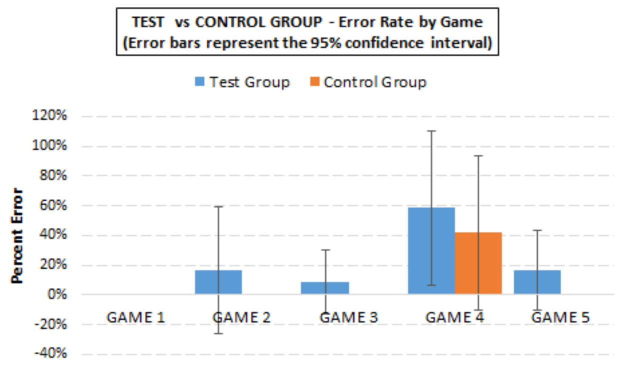

Fig. 5 Average error rates for test and control group

Further analysis by Fig. 5 shows that Game 4 in particular, has more than 50% error rates committed by participants from both groups. This was mainly attributed to the game's interface design which confused the players whereby they failed to drag the correct object. Nevertheless, they managed to complete other games successfully with very minimal errors and performed the right actions in order to accomplish the game's objectives.

D. Satisfaction

The participants' verbal and non-verbal expressions throughout the usability session are a suitable indicator of how satisfied they are with the game. Therefore, to gauge users' satisfaction, they were encouraged to express their thoughts about the functionalities or highlight any problems they encountered while playing the games. To achieve this, it was absolutely critical to make them at ease and as comfortable as possible by engaging them in casual conversations before the session [39].

CSLD users from the test group displayed their satisfaction mainly through facial expressions, body language, and gestures, such as nodding and shaking their heads. Some of them reacted positively by clapping their hands and laughing when they managed to complete the games. It was interesting to note that their body language and gesture suggest that most of them remained calm and composed throughout the session, except for a few who took slightly longer time to complete the game or encountered problems in recognizing specific objects or icons and their functions – they looked quite stressful. Furthermore, one of the participants expressed disappointment and anger by pushing away the tablet due to the difficulties in playing the games.

Unlike the test group, most participants from the control group expressed their mutual satisfaction with ASAH-*i* because they felt it was reasonably easy to use and learn. They were deeply engaged throughout the session, really passionate and excited while playing the games by asking for more to play. Also, their verbal expressions, including body language demonstrated a gratifying experience, indicating their pleasure and desire for the games. Alas, a few participants felt annoyed and frustrated, especially with Game 4 and 5 as they lamented that the games were slow in responding to their actions and that the level of these games was difficult to play.

In summary, the overall result of this study revealed that participants from both groups were generally satisfied with ASAH-*i*. However, for more satisfaction, positive attitudes, fun, and congeniality feel of real users resulting from the use of ASAH-*i*, improvements should be made so that the game can be used effectively and efficiently.

Following this usability testing, several recommendations are being proposed for an optimal design of ASAH-*i.* These recommendations will be helpful with some improvements:

- Alter the layout of Game 4 as it confused the participants who tend to make mistakes by pressing on the images instead of dragging them.
- Ensure the objects or icons can be recognized quickly and well-functioned.
- Limit the number of pictures that appear on each page to be within an acceptable limit by the size of the tablet.
- There should be an icon to clarify the appearance of other icons such as “Exit” or “Back” icon.
- Set a suitable and acceptable duration for the object to react so that participants do not get confused with the actions performed.

IV. CONCLUSION

This study managed to provide preliminary insights into the usability by recognizing the problems found by users and several recommendations have been made to enhance the usability of ASAH-*i* for CSLD. Hopefully, it would be helpful for the developers when improving the game's design and avoid them from making the same mistakes in designing serious games specifically for therapeutic purposes shortly. The outcomes of the testing will form the next version of the game to meet the remarks of the user. The focus of the future research should be directed towards improving and evaluating ASAH-*i* with the relevant expertise.

ACKNOWLEDGMENT

We thank all the parents and children for their participation in this study. We would also like to acknowledge the Audiology and Speech Clinic and Emaan Kindy Islamic Kindergarten. This research is funded by the scholarship of Ministry of Education, Malaysia and supported by Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia.

REFERENCES

[1] S. Mader, S. Natkin, and G. Levieux, “How to analyze therapeutic games: the player/game/therapy model,” Entertain. Comput. 2012, pp. 193–206, 2012.

[2] M. Wrzesien, M. A. Raya, C. Botella, J. Burkhardt, J. Breton-López, and A. R. Ortega, “A pilot evaluation of a therapeutic game applied to small animal phobia treatment,” in Serious Games Development and Applications, 2014, vol. 8778, pp. 10–20.

[3] N. Yusof, R. M. Rias, E. H. Yusoff, and M. Sciences, “Embedding Cognitive Behavior Therapy in Therapeutic Game Design Model for Depression Faculty of Medicine,” J. Converg. Inf. Technol., vol. 10, no. September, pp. 1–11, 2015.

[4] C. I. Springer, G. Colorado, and J. R. Misurell, “Structured therapeutic games for nonoffending caregivers of children who have experienced sexual abuse,” J. Child Sex. Abus., vol. 24, no. 4, pp. 412–428, 2015.

[5] H. M. Toh, S. E. Ghazali, and P. Subramaniam, “The acceptability and usefulness of Cognitive Stimulation Therapy for older adults with Dementia: A narrative review,” Int. J. Alzheimers. Dis., vol. 2016 (no p, no. 5131570, 2016.

[6] V. H. Carvalho, T. Martins, F. Soares, and M. Araújo, “Total Challenge : A Serious Game for Stimulating Cognitive Abilities,” Int. J. Adv. Corp. Learn., vol. 9, no. 1, pp. 4–11, 2016.

[7] A. Shapi'i, N. A. Mat Zin, and A. M. Elaklouk, “A game system for cognitive rehabilitation,” Biomed Res. Int., vol. 2015, 2015.

[8] J. Lumsden, E. A. Edwards, N. S. Lawrence, D. Coyle, and M. R. Munafò, “Gamification of Cognitive Assessment and Cognitive Training: A Systematic Review of Applications and Efficacy,” JMIR Serious Games, vol. 4, no. 2, p. e11, 2016.

[9] M. E. Kelly, D. Loughrey, B. A. Lawlor, I. H. Robertson, C. Walsh, and S. Brennan, “The impact of cognitive training and mental stimulation on cognitive and everyday functioning of healthy older adults: A systematic review and meta-analysis,” Ageing Res. Rev., vol. 15, no. 1, pp. 28–43, 2014.

[10] D. Mapelli, E. Di Rosa, R. Nocita, and D. Sava, “Cognitive stimulation in patients with Dementia: Randomized controlled trial,” Dement. Geriatr. Cogn. Dis. Extra, vol. 3, no. 1, pp. 263–271, 2013.

[11] N. A. A. Zaki, T. S. M. T. Wook, and K. Ahmad, “Analysis and classification of serious games for cognitive stimulation,” in International Conference on Electrical Engineering and Informatics (ICEEI), 2015, pp. 612–617.

[12] N. A. Ahmad Zaki, T. S. M. Tengku Wook, K. Ahmad, N. A. A. Zaki, T. S. M. T. Wook, and K. Ahmad, “Towards developing a therapeutic serious game design model for stimulating cognitive abilities: A case for children with speech and language delay,” in Computational Intelligence in Information Systems. CIIS 2016.

Advances in Intelligent Systems and Computing, vol. 532, S. Phon-Amnuaisuk, T.-W. Au, and S. Omar, Eds. Cham: Springer International Publishing, 2017, pp. 259–269.
[13] J. Nielsen, “Usability 101: Introduction to Usability,” Nielsen Norman Gr., p. Articles, 2012.
[14] International Organization for Standardization, “International Standard 9241-210, Ergonomics of human–system interaction - Human-centred design for interactive systems,” Int. Organ. Stand., 2010.
[15] N. Bevan, J. Carter, J. Earthy, T. Geis, and S. Harker, “New ISO standards for usability, usability reports and usability measures,” Lect. Notes Comput. Sci., vol. 9731, pp. 268–278, 2016.
[16] N. Bevan, J. Carter, and S. Harker, “ISO 9241-11 revised: What have we learnt about usability since 1998?,” Lect. Notes Comput. Sci., vol. 9169, pp. 143–151, 2015.
[17] J. Rubin and D. Chisnell, Handbook of usability testing, 2nd ed. Indianapolis, 2008.
[18] J. Lazar, J. H. Feng, and H. Hochheiser, “Usability testing,” in Research Methods in Human Computer Interaction, 2017, pp. 263–298.
[19] A. Druin, “The role of children in the design of new technology,” Behav. Inf. Technol., vol. 21, no. 1, pp. 1–25, 2002.
[20] B. Zaman, “Evaluating games with children,” Proc. Interact 2005 Work. Child Comput. Interact. Methodol. Res., pp. 9–11, 2005.
[21] J. Nielsen, “Paper versus computer implementations as mockup scenarios for heuristic evaluation,” in Proceedings of the IFIP TC13 Third Interational Conference on Human-Computer Interaction, 1990, pp. 315–320.
[22] C. Boletsis and S. McCallum, “Augmented reality cubes for cognitive gaming: Preliminary usability and game experience testing,” Int. J. Serious Games, vol. 3, no. 1, pp. 3–18, 2016.
[23] A.-W. Lyan, A.-G. Amjad, A.-Z. Shaden, and A.-N. Khalid, “A usability evaluation of Arabic mobile applications designed for children with special needs — Autism,” Lect. Notes Softw. Eng., vol. 3, no. 3, pp. 203–209, 2015.
[24] J. Nielsen, Usability engineering. California, 1993.
[25] N. Jacob, “Estimating the number of subjects needed for a thinking aloud test,” Int. J. Hum. Comput. Stud., vol. 41, no. 3, pp. 385–397, 1994.
[26] W. Barendregt and M. M. Bekker, “Guidelines for user testing with children,” Eindhoven, Netherlands, Tech. Rep, pp. 1–4, 2003.
[27] S. N. S. da Cunha, X. L. Travassos Junior, R. Guizzo, and C. de Sousa Pereira-Guizzo, “The digital memory game: An assistive technology resource evaluated by children with cerebral palsy,” Psicol. Reflex. e Crit., vol. 29, no. 1, 2016.
[28] N. Nasiri, S. Shirmohammadi, and A. Rashed, “A serious game for children with speech disorders and hearing problems,” 2017 IEEE 5th Int. Conf. Serious Games Appl. Heal. SeGAH 2017, 2017.
[29] M. A. Khanum and M. C. Trivedi, “Take Care: A study on usability evaluation methods for children,” Int. J. Adv. Res. Comput. Sci., vol. 3, no. 2, 2012.
[30] M. L. Guha, A. Druin, and J. A. Fails, “Designing with and for children with special needs: an inclusionary model,” Interact. Des. Child., pp. 61–64, 2008.
[31] T. Porcino, D. Trevisan, E. Clua, M. Rodrigues, and D. Barbosa, “A participatory approach for game design to support the learning and communication of Autistic children,” Int. Fed. Inf. Process. 2015, vol. 9353, pp. 238–251, 2015.
[32] L. Benton and H. Johnson, “Widening participation in technology design: A review of the involvement of children with special educational needs and disabilities,” Int. J. Child-Computer Interact., vol. 3–4, pp. 23–40, 2015.
[33] R. Yáñez-Gómez, D. Cascado-Caballero, and J. L. Sevillano, “Academic methods for usability evaluation of serious games: a systematic review,” Multimed. Tools Appl., vol. 76, no. 4, pp. 5755–5784, 2017.
[34] N. Mat Diah, M. Ismail, S. Ahmad, and M. K. M. Dahari, “Usability testing for educational computer game using observation method,” Int. Conf. Inf. Retr. Knowl. Manag. Explor. Invis. World, CAMP’10, no. April, pp. 157–161, 2010.
[35] A. Donker and P. Markopoulos, “A comparison of think-aloud, questionnaires and interviews for testing usability with children,” in People and Computers XVI - Memorable Yet Invisible: Proceedings of HCI 2002, F. X., F. J., and D. F., Eds. London: Springer London, 2002, pp. 305–316.
[36] N. A. A. Zaki, T. S. M. T. Wook, and K. Ahmad, “Therapeutic serious game design guidelines for stimulating cognitive abilities of children with speech and language delay,” J. ICT, vol. 2, no. 2, pp. 284–312, 2017.
[37] J. Sauro and J. R. Lewis, “Estimating completion rates from small samples using binomial confidence intervals: Comparisons and recommendations,” Proc. Hum. Factors Ergon. Soc. Annu. Meet., pp. 2100–2104, 2005.
[38] T. Tullis and B. Albert, Measuring the user experience: Collecting, analyzing, and presenting usability metrics, 2nd ed. 2013.
[39] T. S. M. T. Wook and S. S. Salim, “Visual interface for searching and browsing children’s WebOPAC,” Univers. Access Inf. Soc., vol. 13, no. 4, pp. 367–385, 2014.